UNIVERSITY OF CALIFORNIA
SANTA CRUZ

MASK DESIGN FOR SINGLE AND DOUBLE EXPOSURE OPTICAL MICROLITHOGRAPHY: AN INVERSE IMAGING APPROACH

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

COMPUTER ENGINEERING

by

Amyn Poonawala

September 2007

The Dissertation of Amyn Poonawala
is approved:

Professor Peyman Milanfar, Chair

Yan Borodovsky, Senior Intel Fellow

Professor Matthew Guthaus

Lisa C. Sloan
Vice Provost and Dean of Graduate Studies

Contents

List of Figures vi

Abstract xi

Acknowledgements xiv

Dedication xvi

1 Introduction 1
- 1.1 ITRS and Rayleigh Criterion . . . 3
- 1.2 Resolution Enhancement Techniques . . . 7
  - 1.2.1 Optical and Process Correction (OPC) . . . 7
  - 1.2.2 Phase Shift Masks . . . 11
  - 1.2.3 Off-axis Illumination . . . 14
  - 1.2.4 Double Exposure and Double Patterning . . . 15
- 1.3 Thesis Organization . . . 16

2 The Inverse Lithography Problem 19
- 2.1 Lithography Modeling . . . 19
- 2.2 The Inverse Problem . . . 22
  - 2.2.1 Introduction . . . 22
  - 2.2.2 Past Work on ILT . . . 24
- 2.3 Approximate Forward Model . . . 26
  - 2.3.1 Sigmoid Approximation of the Resist . . . 27
  - 2.3.2 Aerial Image Models . . . 28
- 2.4 ILT Optimization Problem . . . 33

3 Single Exposure OPC Mask Design 37
- 3.1 OPC Mask Design Algorithm . . . 38
- 3.2 Preliminary Results for Incoherent Imaging with Gaussian PSF . . . 42
- 3.3 Regularization Framework . . . 49

3.3.1 Discretization Penalty Term . . . 50
3.3.2 Complexity Penalty Term . . . 53
3.3.3 Aerial Image Penalty Term . . . 57
3.3.4 An Alternative Two Step Strategy for Optimization . . . 59
3.4 More Results . . . 61
3.4.1 Assist Bars for Coherent Imaging . . . 61
3.4.2 Incoherent Imaging with Circular Pupil . . . 63
3.4.3 Comparison with Combinatorial Optimization (Sherif, et al.) . . . 65
3.5 Weight Mask and Selected Parameter Optimization . . . 67
3.5.1 Weight Mask . . . 67
3.5.2 Selected Parameter Optimization . . . 69

4 Single Exposure PSM Design **72**
4.1 PSM Optimization . . . 72
4.1.1 Attenuated PSM . . . 73
4.1.2 100% transmission PSM . . . 75
4.1.3 Strong PSM with Chrome . . . 77
4.2 Results . . . 78
4.2.1 2-Bar Separation . . . 78
4.2.2 Random Logic Contact Holes . . . 81
4.2.3 Periodic Contact Holes . . . 83
4.2.4 2-D Periodic Dense Pattern . . . 84
4.2.5 A Complicated Tight-Pitch Pattern . . . 85

5 ILT for Double Exposure Lithography with $k_1 \geq 0.25$ **89**
5.1 Double Exposure and Double Patterning . . . 90
5.1.1 Double Exposure Lithography (DEL) . . . 90
5.1.2 Dual Patterning (DPL) . . . 91
5.2 Double Exposure Inverse Lithography Technology (DEL-ILT) Framework 94
5.2.1 Motivation . . . 94
5.2.2 The Double Exposure Inverse Lithography Problem . . . 96
5.3 DEL-ILT Optimization . . . 99
5.3.1 Method 1: Simultaneous Descent . . . 101
5.3.2 Method 2: Cyclic Coordinate Descent . . . 103
5.4 Results . . . 104
5.4.1 T Joint . . . 104
5.4.2 U-Joint . . . 106
5.4.3 Pattern Decomposition for More Complicated Patterns . . . 108
5.4.4 Verification using Prolith . . . 110

6 DEL with Novel Materials for $k_1 < 0.25$ 114
6.1 Motivation . . . 114
6.2 DEL Using Contrast Enhancement Layer (CEL) . . . 116
6.2.1 Modeling . . . 117
6.2.2 Working of CEL . . . 119
6.3 ILS Improvement of the CEL-DEL Image . . . 122
6.3.1 CEL Film Thickness . . . 122
6.3.2 Incoming Dose Magnification . . . 123
6.3.3 Quenchers . . . 125
6.3.4 Matching with Arsel^{TM} . . . 125
6.4 ILT using CEL for 2-D Patterns . . . 126
6.5 Two-Photon Absorption Resist . . . 129

7 Contributions, Impact, and Future Work 135
7.1 Contributions . . . 135
7.2 Recent Work . . . 138
7.2.1 Luminescent Technologies . . . 138
7.2.2 Mentor Graphics . . . 139
7.2.3 GILT - University of Delaware . . . 140
7.2.4 Invarium Inc. . . . 140
7.2.5 Synopsys Inc. . . . 141
7.3 Future Work and Extensions . . . 142
7.3.1 Robustness of DEL to Alignment Error . . . 142
7.3.2 Thick-Mask Effects . . . 147
7.3.3 Mask Design for 3-D Resist and Etch Profiles . . . 149
7.3.4 Applications to E-beam Lithography . . . 150
7.3.5 RET for Optical Maskless Lithography and DUV Laser Writing 153
7.3.6 Pattern Splitting and Stitching for DPL . . . 155
7.3.7 Practical Extensions . . . 156

A Derivation of the Gradient for Incoherent Imaging Systems 158

B Double Exposure Theorem 160

C Relationship Between ILS and Dose Magnification. 162

D Intel Fab Results 164

Bibliography 167

List of Figures

1.1 The optical microlithography process (from [82]) 1
1.2 Steps involved in optical microlithography 2
1.3 International Technology Roadmap for Semiconductors 4
1.4 Optical and Process Correction (from [82]) 8
1.5 Feedback optimization loop for model-based edge-based OPC (from [17]) 10
1.6 Pattern transfer using binary mask 11
1.7 Pattern transfer using Attenuated PSM 12
1.8 Pattern transfer using Alternating PSM 13
1.9 Pattern transfer using Chromeless PSM 14
1.10 Commonly employed source configurations; left to right - conventional, annular, fourfold source, separated source, quadrapole, CQUEST, QUASAR (from [83]) 15

2.1 A simplified view of the lithography process 20
2.2 Forward Model and ILT 22
2.3 Approximated forward process model 27
2.4 The effect of the steepness parameter a on the sigmoid function $\text{sig}(u) = 1/(1+e^{-a(u-0.5)})$. 29
2.5 The jinc function for NA=0.85 and $\lambda = 193nm$. 30
2.6 The jinc squared function for NA=0.85 and $\lambda = 193nm$. 31
2.7 Aerial image computation of partially coherent imaging systems using SOCS. 33

3.1 Flow-chart showing the sequence of operations required for calculating the gradient $\nabla F_1(\underline{\boldsymbol{\theta}})$. Note that *multiply* indicates element-by-element multiplication of 2-D arrays. 40

3.2 Top row (input masks), left to right: original pattern, optimized gray pattern, and binary pattern obtained using an optimum global threshold t_m. The middle and bottom rows indicate the corresponding aerial images and binary output patterns. The parameters are $a = 90,\ t_r = 0.5,\ 15 \times 15$ filter with $\sigma = 5$ pixels, $s = 0.4$, and number of iterations $= 200$. 43
3.3 Cost function (calculated using (2.15)) versus steepest-descent iterations for the pattern in Fig. 3.2 44
3.4 Top row, left to right: original pattern, optimized gray pattern using the constrained optimization routine from [41], and the binary pattern obtained using an optimum global threshold t_m. The bottom row indicates the corresponding binary output patterns. The parameters are $a = 90,\ t_r = 0.5,\ 15 \times 15$ filter with $\sigma = 5$ pixels. 46
3.5 Top row, left to right: original pattern, optimized gray pattern, and binary pattern obtained using an optimum global threshold t_m. The bottom row indicates the corresponding binary output patterns. The parameters are $a = 80,\ t_r = 0.5,\ 11 \times 11$ filter with $\sigma = 14$ pixels, $s = 0.5$, and number of iterations $= 200$. 47
3.6 Cost function versus steepest-descent iterations for Fig. 3.5. 48
3.7 Quadratic penalty cost function $R_{dis}(\underline{\mathbf{m}})$ 51
3.8 Left to right: original pattern, optimized gray pattern, and binary pattern obtained using global thresholding with $t_m = 0.5$. The cost function minimized was $F(\underline{\mathbf{m}})+\gamma_{dis}R_{dis}(\underline{\mathbf{m}})$ and the parameters are $a = 80,\ t_r = 0.5,\ 11 \times 11$ filter with $\sigma = 14$ pixels, $s = 1.5$, number of iterations $= 200$, and $\gamma_{dis} = 0.015$ 52
3.9 Left to right: original pattern, optimized gray pattern, and binary pattern obtained using global thresholding with $t_m = 0.5$. The cost function minimized was $F(\underline{\mathbf{m}})+\gamma_{dis}R_{dis}(\underline{\mathbf{m}})+\gamma_{TV}R_{TV}(\underline{\mathbf{m}})$ and the parameters are $a = 90,\ t_r = 0.5,\ 15 \times 15$ filter with $\sigma = 5$ pixels, $s = 1$, number of iterations $= 200$, $\gamma_{dis} = 0.025$ and $\gamma_{TV} = 0.045$ 55
3.10 Left to right: original pattern, optimized gray pattern, and binary pattern obtained using global thresholding with $t_m = 0.5$. The cost function minimized was $F(\underline{\mathbf{m}})+\gamma_{dis}R_{dis}(\underline{\mathbf{m}})+\gamma_{TV}R_{TV}(\underline{\mathbf{m}})$ and the parameters are $a = 80,\ t_r = 0.5,\ 11 \times 11$ filter with $\sigma = 14$ pixels, $s = 1$, number of iterations $= 400$, $\gamma_{dis} = 0.01$ and $\gamma_{TV} = 0.02$ 56
3.11 The synthesized masks (top row) and the output binary patterns (bottom row) for 90nm random logic contact patterns. $\widehat{\underline{\mathbf{m}}}_b$ (left) is the estimated mask at the end of Step 1 ($\gamma_{aerial} = 1$), $\widehat{\underline{\mathbf{m}}}_b$ (center) is estimated at the end of Step 2 ($\gamma_{fid} = 1,\ \gamma_{aerial} = 0.25,\ \gamma_{dis} = 0.005,\ \gamma_{TV} = 0$) and $\widehat{\underline{\mathbf{m}}}_b$ (right) is estimated using the above parameters with $\gamma_{TV} = 0.005$ 62

3.12 Convergence curves for Fig. 3.11. The left curve indicates the cost function behavior for Step 1 and the right curve indicates the regularization cost ($\gamma_{dis}R_{dis}(\underline{\mathbf{m}})+\gamma_{TV}R_{TV}(\underline{\mathbf{m}})$) versus iteration number for Step 2 of the optimization. 63
3.13 Horizontal slice at row number 35 for the aerial image obtained using the synthesized mask in Fig. 3.11. Note that the contrast has improved and the side-lobes will not print. 64
3.14 The top row consists of the original pattern and the synthesized masks before and after regularization. The bottom row indicates the corresponding output wafer patterns. $\widehat{\underline{\mathbf{m}}}_b$ (center) is estimated using $\gamma_{fid} = 1$ and $\widehat{\underline{\mathbf{m}}}_b$(right) is obtained using $\gamma_{fid} = 1$, $\gamma_{dis} = 0.01$, and $\gamma_{TV} = 0.1$. . 65
3.15 The top row indicates the synthesized masks obtained using Sherif, et al. (from [84]) and our algorithm for a square pupil with $d = 4.0$ and $d = 4.5$ pixels. The bottom row indicates the corresponding output wafer patterns assuming a hard-limiter with $t_r = 0.2$. 66
3.16 The employed weight mask (left), synthesized mask pattern (center), and the final output pattern (right). Note that 5x5 region around the corners have zero weight whereas the weight is one everywhere else. . . 68
3.17 Left to Right: User-selected parameters (marked in white), the corresponding optimized mask pattern, and the final output pattern. The parameters are a=90, $t_r = 0.5$, 15x15 filter with $\sigma = 5$, $\lambda_{dis} = 0.01$, and s=1. 70

4.1 Discretization penalty term for AttPSM (maximum penalty is at $m_j = 0.3775$) 75
4.2 Discretization term for 100% transmissive PSM (maximum penalty is at $m_j = 0$) 76
4.3 Discretization penalty cost for strong PSM (with chrome). The minima are at -1, 0, and 1 77
4.4 The estimated 6% AttPSM mask (left), the corresponding aerial image (center), and the final binary output pattern (right) for coherent imaging system with $\lambda = 193$nm and $NA = 0.85$. The black and white regions in $\widehat{\underline{\mathbf{m}}}_b$ correspond to -0.245 and 1 respectively. 79
4.5 The estimated 100% transmission PSM (left), the corresponding aerial image (center), and the final binary output pattern (right) for coherent imaging system with $\lambda = 193$nm and $NA = 0.85$. The black and white regions in $\widehat{\underline{\mathbf{m}}}_b$ correspond to -1 and 1 respectively. 79
4.6 The estimated strong PSM (left), the corresponding aerial image (center), and the final binary output pattern (right) for a coherent imaging system with $\lambda = 193$nm and $NA = 0.85$. The black, gray, and white regions in $\widehat{\underline{\mathbf{m}}}_b$ correspond to -1, 0, and 1 respectively. 80

4.7 Horizontal slices along the center of the aerial images obtained using the desired pattern $\underline{\mathbf{z}}^*$ and the various synthesized PSMs in Fig. 4.4, 4.5 and 4.6 as inputs. 81
4.8 The top row consists of the desired pattern, the estimated strong PSM with chrome, and the 100% transmission PSM. The black, gray, and white regions correspond to transmission values of -1, 0, and 1 respectively. The bottom row consists of the aerial images corresponding to the masks in the top row. Here $\lambda = 193$nm, $NA = 0.7$, and $\sigma = 0$. . . 82
4.9 Estimated CPL masks for a 100nm contact through varying pitch for $NA = 0.85$. Left to right the pitch value is 200nm, 260nm, 340nm, and 400nm. 84
4.10 Two commonly arising cases of phase conflict (from [54]). The pattern on the left is a T-Joint which results in lack of phase-shift along the critical dimension. The one on right is a line-end which suffers from the problem of unwanted phase-shift. 85
4.11 The top row indicates the periodic target pattern and the estimated 100% transmission PSM. The bottom row indicates the aerial image and its contour at $t_r = 0.3$. Here $NA = 0.85$, $\sigma = 0$, and $k_1 = 0.35$. . . 86
4.12 The top row indicates the complicated target pattern with 120nm pitch and the estimated 100% transmission PSM. The bottom row indicates the aerial image and its contour at $t_r = 0.3$. Here $NA = 0.93$, $\sigma = 0$, and $k_1 = 0.29$. 87

5.1 Flowchart of the steps involved in double exposure (single development) lithography. 91
5.2 The original (1:1) grating and two decomposed (1:3) gratings for dual patterning. The two patterns are shifted versions of each other and the pitch is doubled enabling them to be individually printed using single exposure. 92
5.3 Steps involved in double patterning for frequency doubling of grating (figure courtesy Intel Corporation). 93
5.4 The top row indicates the vertical and horizontal (decomposed) target patterns and the original target. The bottom row indicates the corresponding aerial images obtained using single exposure ILT masks, and the combined aerial image. 95
5.5 Forward Model and ILT for Double Exposure Lithography Systems . . 97
5.6 The T-junction target pattern (left) and the two estimated chromeless PSM (center and right). 104
5.7 The final aerial image (left) and the aerial images corresponding to the two individual exposures (center and right) for the mask patterns in Fig. 5.6 106

5.8 The graph on the left is the aerial image obtained by solving the single exposure ILT problem for the T-junction. The center and right graphs are the contours at $t_r = 0.55$ (sampled at 2.5nm) for single and double exposure ILT. The corresponding pattern errors are 1047 and 389. . . 107
5.9 The U-Joint target pattern (left) and the two estimated chromeless PSM (center and right) obtained using Method-2. 107
5.10 The final aerial image (left) and the aerial images corresponding to the two individual exposures (center and right) for the mask patterns in Fig. 5.9. 108
5.11 The final aerial image (left) and the individual aerial images (center and right) for the U-joint target pattern. The corresponding masks were synthesized using Method-1. 109
5.12 Cost function behavior for simultaneous and cyclic descent algorithms when solving for the target pattern in Fig. 5.9. 110
5.13 Top row illustrates the estimated CPL masks (black=-1 and white=+1) and the (binary) target pattern consisting of 60nm features. The pitch p=120nm is indicated by dotted lines in the figure. Bottom row indicates the aerial images for the individual exposures and the combined aerial image. Here $NA = 0.93$ and $k_1 = 0.289$. 111
5.14 Top row illustrates the estimated CPL masks and the target pattern consisting of 80:320 vertical grating interlaced with 80nm wide horizontal features. Bottom row indicates the aerial images for the individual exposures and the combined aerial image. Here $NA = 0.85$ and $k_1 = 0.35$.112
5.15 The complex target pattern (left) and the two estimated chromeless PSM (center and right). 113
5.16 The final aerial image (left) and the aerial images corresponding to the two individual exposures (center and right) for the mask patterns in Fig. 5.15 113
5.17 The final aerial image (left) and the aerial images for the two individual exposures (center and right) obtained by feeding the mask patterns from Fig. 5.15 to Prolith. 113

6.1 The resist stack with CEL. 117
6.2 CEL film transmittance for different film thickness. 118
6.3 Transmitted dose (D_t) versus input dose D_i for different film thickness. 119
6.4 The two figures in the top row indicate the aerial images (dotted blue) and the CEL images (solid red) for the two individual exposures. The bottom row indicates the combined aerial image (showing no modulation) and the combined CEL image (showing good modulation). In this case $a = 30$, $t_c = 0.15$, and $c = 0.45$. 121
6.5 Effect of CEL film thickness assuming fixed D_i. 123
6.6 Effect of dose magnification p on the CEL image using fixed resist with threshold ($t_r = 0.33$). Here, $a = 60$ and $t_c = 0.15$. 124

6.7 OD versus D_i for the proposed model (left) and the measured data using Arsel^{TM}(right). 126
6.8 The target pattern consisting of 40nm features on 80 nm pitch and the (manually) decomposed patterns used for individual exposures. 127
6.9 The top row indicates the CEL-ILT results obtained by solving (6.5). The bottom row indicates the DEL results where the second exposure was performed by shifting the mask by 240nm. 130
6.10 The intensity and square of intensity modulation curves corresponding to the conventional and 2-photon absorption resists respectively. . . . 131

7.1 Left to right: The resist contours for no misalignment, $\underline{\mathbf{b}}$ shifted right by 20nm, $\underline{\mathbf{b}}$ shifted down by 20nm, and $\underline{\mathbf{b}}$ shifted both right and down by 20nm, where the ILT problem is solved for the contour fidelity with no misalignment. 145
7.2 The top row indicates the aerial images from the first exposure (left), second exposure (center), and the combined aerial image (right). The bottom row indicates the resist contours with no misalignment, and the second mask is shifted right or down or both by 20nm. The DEL-ILT problem is solved only for the last case ($\underline{\mathbf{b}}$ misaligned in both directions). 146
7.3 Kirchhoff (thin) versus rigorous (thick) mask modeling (from [25]). . . 148
7.4 Example of 3-D resist profiles (from [39]). 150
7.5 E-Beam Lithography (from [18]) 151
7.6 Mask manufacturing to circuit printing process 153
7.7 Traditional mask-based versus maskless flow for optical lithography (from [78]). 154

D.1 SEM image of the fabricated wafer pattern obtained using the single exposure ILT mask pattern from Fig. 4.12 (courtesy Intel Corporation). 164
D.2 SEM images of the fabricated patterns using the DEL-ILT mask patterns from Fig. 5.13 (left) and Fig. 5.17 (right). The figure is courtesy Intel Corporation. 165

Abstract

Mask Design for Single and Double Exposure Optical Microlithography: An Inverse Imaging Approach

by

Amyn Poonawala

Optical microlithography, a technique similar to photographic printing, is used for transferring circuit patterns onto silicon wafers. The above process introduces distortions arising from optical limits and non-linear resist effects, leading to unacceptable pattern fidelity and yield loss. The input to the above system is a photomask (or reticle), which can be controlled (and engineered) such that it cancels out (or pre-compensates) for the process losses to come. This forms the basis of optical proximity correction (OPC) and phase shift masks (PSM), two commonly employed resolution enhancement techniques (RETs) for patterning small feature sizes close to the optical limit (low k_1 values).

In this thesis, we propose a novel inverse lithography technology (ILT) framework to synthesize OPC and PSM for high-fidelity patterning of random logic 2-D features. ILT attempts to synthesize the input mask which leads to the desired output wafer pattern by inverting the mathematical forward model from mask to wafer. Our framework employs a pixel-based mask parametrization, continuous function formulation, and analytic gradient-based optimization techniques to synthesize the masks.

We also introduce a regularization framework to control the tone and complexity of the synthesized masks, and inculcate other user-defined properties. Assist bars are automatically generated and optimally placed as part of the solution.

Another contribution of this thesis is an ILT framework for mask design of double exposure (single development) lithography (DEL) systems. The proposed DEL-ILT algorithm introduces a new paradigm of decomposing the *gray-level* target aerial image into two complementary parts, automatically resolves the phase conflicts, gives superior contrast aerial images, and is also robust to mask-to-mask alignment errors. Finally, we investigate two novel materials which allow violation of the linear superposition of the intensities from the two exposures, thereby enabling imaging below the conventional theoretical limit ($k_1 < 0.25$).

Acknowledgements

I embarked onto this journey, fresh out of college from Mumbai, and I am indebted to all the people who converted this journey into a very memorable experience.

First and foremost I would like to express my deep gratitude to my advisor, Dr. Peyman Milanfar. I have learned a lot from him. He transformed me from a raw graduate student to a researcher. His vast experience and expertise in the area of inverse imaging helped me to envision, formulate, and solve the mask design problem as an inverse problem. He gave creative suggestions, wonderful intuitions and insights, constant encouragement, and also inculcated good presentation and technical writing skills in me. His mentoring while deciding my career options was invaluable. Its been an honor to work under him.

I would also like to thank our sponsor, Intel Corporation, for their support and interest. A special thanks to Yan Borodovsky for providing an industry perspective of things, and helping us close the gap between academia research and industry applications. His input in convincing the reviewers (lithography experts) about the significance and impact of our work was very beneficial. I would like to thank my qualifying and dissertation committee members Dr. Matthew Guthaus, Dr. Michael Isaacson, Dr. Pat Mantey, and Dr. Ali Shakouri. Thanks to Gary Allen, Paul Davids, and Richard Schneker from Intel, for carrying out experiments in the fab and providing us with data. I also acknowledge my lab-mates Morteza, Dirk, Sina, Hiro, Priyam, and Hae-Jong for making the workplace interesting.

I am lucky to have had some great friendships and bonds over the past few years. My friends were my support system at UCSC. A special thanks to Morteza and Srikumar, the members of the original DCH chaai club. The extended (sometimes all night long) tea sessions, the incoherent ramblings, the discussions on random topics in life, etc are sweet memories which will stay with me forever. Amin and Noureen, you guys gave me a sense of home away from home. Kalu, your yahoo chat company was always a refreshing break from work during the night-outs. I would also like to thank Karen (my kutti sister), Chandu and Kaushik (gundaas from my P.A.R.T.Y. gang), Suratna (my dinner/Chanakya company), and many many other friends. As years passed by, people moved out of Santa Cruz, but I was fortunate enough to find new friendships to cherish. Special thanks to DJ, Rajjo, Sangy, and Ambu for their support during the crucial final stages of my PhD. And to the entire baskin-tea gang (Pri)yam, (Pri)tam, Su(month), Ani, Suresh Anna, Sudha, Mundu, Maddy, Anindya, Reemavik, for all the beautiful west-cliff memories and happy times together.

I also recall the sacrifices my parents made in their time and quality of life, to ensure that I have the knowledge, education, and all the tools required to pursue my desired career options. Their love, patience, enthusiasm, and encouragement enabled me to keep going and reach my destination. I would like to dedicate this to my parents and my brother (Azim). And Dad, I am one step closer to being addressed as Prof. Poonawala, like you!

Santa Cruz, CA Aug 31, 2007 Amyn Poonawala

To my parents, Daulat and Alaudin Poonawala

Chapter 1

Introduction

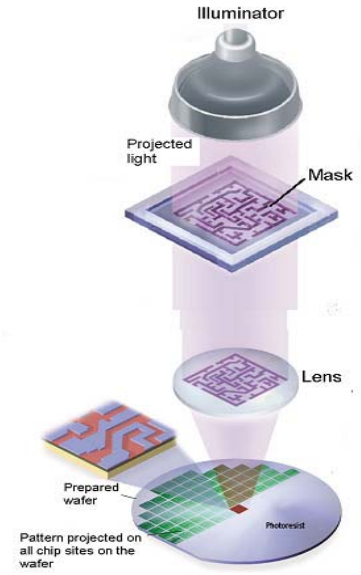

Figure 1.1: The optical microlithography process (from [82])

Optical microlithography, a process similar to photographic printing, is used for transferring circuit patterns onto silicon wafers (see Fig. 1.1) and forms a very critical step in the IC fabrication flowchart [13]. Fig. 1.2 illustrates the steps involved

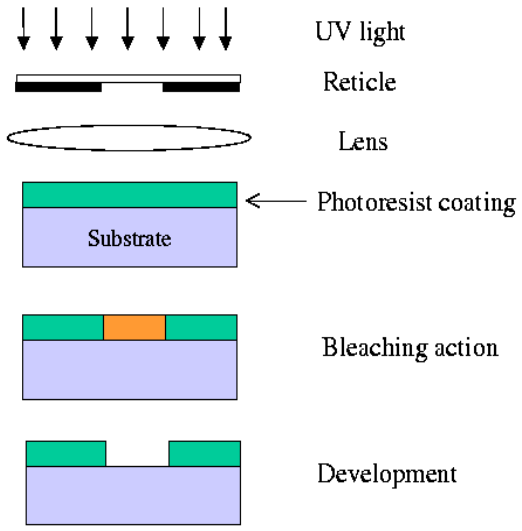

Figure 1.2: Steps involved in optical microlithography

in a typical lithography process. The pattern to be replicated on the wafer is first carved on a reticle (mask) composed of quartz and chrome features. Light passes through the clear quartz areas and is blocked by the chrome region. An illuminator (UV source) is shone through this mask producing an image of the pattern through the lens system, which is eventually projected down onto a photoresist-coated silicon wafer using a projection system (typical aspect ratio 4:1). The photoresist is chemically sensitive to light and acts as the recording medium for the above imaging system. It responds non-linearly to the incoming dose energy (commonly referred to as the *aerial image*) and undergoes bleaching action. New chemical bonds are formed or existing bonds are broken in the exposed regions (depending on whether it is a negative or positive photoresist respectively), thereby changing their dissolution properties. The photoresist is finally developed (washed using an appropriate chemical solution) leaving

behind a likeliness of the mask pattern on the substrate (wafer). In an ideal world, the output circuit pattern on the wafer will be exactly similar to the mask pattern. Unfortunately, the band-limited imaging system introduces distortions caused by the diffraction effects of the lens-projection system, particularly at higher resolutions. The loss in higher order frequency components results in blurring and poor quality aerial image. This coupled with the resist effect results in loss of pattern fidelity, and the output pattern obtained on the wafer is generally a warped version of the input [82].

1.1 ITRS and Rayleigh Criterion

The semiconductor industry strives to adhere to the International Technology Roadmap for Semiconductors (ITRS) [1] road-map which is driven by Moore's law for the past three decades. Moore's law with regard to lithography requires the critical dimensions (CD) [1] to shrink by thirty percent every two years (currently CD = 65nm). This puts very stringent requirements on lithography thereby making it one of the tightest bottlenecks faced by the semiconductor industry.

The resolution($w_{\min}$) of the lithography system in Fig. 1.2 is related to the Rayleigh's criterion [2] as follows,

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

The resolution of the imaging system can be improved by increasing the numerical

[1] Critical dimension is defined as the smallest dimension, typically the gate length

[2] The Rayleigh limit for the resolution of two sources is when the first diffraction pattern minimum of the image of one source point falls on the central maximum of the other. This corresponds to a separation of $0.61\lambda/NA$.

Year	Node (CD)	Wavelength used
1990	800nm	435 nm (g-line)
1993	500nm	365 nm (i-line)
1995	350nm	200-300nm (DUV)
1997	250nm	DUV
1999	180nm	DUV
2001	130nm	DUV
2003	90nm	193nm (ArF laser)
2005	65nm	193nm
2007	45nm	193nm
2009	32nm	193nm
2011	22nm	193nm or EUV
2013	16nm	193nm or EUV??

Figure 1.3: International Technology Roadmap for Semiconductors

aperture of the imaging system (NA) or decreasing the wavelength (λ). Fig. 1.3 is part of the ITRS and it is interesting to note that the CD size dived below the wavelength of the source around 1997. There were many attempts to reduce the wavelength by using a 157nm laser. However they were all futile due to the cost, risk, and immaturity of the technology (see [4] for a review). EUV, a next generation lithography technology employs 13nm wavelength thereby having an immense potential to be useful for many more generations [10]. Unfortunately, it faces serious technical issues like lack of adequate power source, lack of resists, and mask-inspection difficulties. EUV was initially planned for high volume manufacturing in 2007, but it has now been pushed beyond 2011. Thus, we have no choice but to use 193nm lithography until the 22nm technology node or maybe even beyond.

The second option to increase the resolution is by increasing the NA of the imaging system. This has an adverse effect of decreasing the depth of focus making the process extremely sensitive to the deviations from the optimal process conditions. The currently employed NA values are 0.93 and the upper-bound for dry lithography systems (using air as the medium) is 1. Note that $NA = n\sin(\theta)$ where n is the refractive index of the medium, and θ is the half angle of the maximum cone of light that can enter the lens. Dry lithography uses air which has a refractive index of 1. Immersion lithography uses water ($n = 1.44$ at 193nm) or other transparent fluid with high refractive index values ($n \approx 1.6$) in order to increase the NA of the system beyond 1. It was inserted into the ITRS road-map and is being currently employed for mass-production of 45nm node by IBM [57] and many other foundries.

The third and most commonly employed way to decrease the minimum feature-size is by employing very low values of k_1. At this point, we would like to highlight that the resolution limits for pattern periodicity and pattern dimensions are different [93, 53]. A pitch is defined as the sum of a line and space width pair and determines the packing density of the transistors. The process constant in this context is referred to as k_{pitch} and is lower bounded by 0.5 [11] corresponding to the Sparrows limit. On the other hand, k_1 corresponds to the smallest printable feature size (also known as the critical dimension or CD), which dictates the gate-length, speed, and the power consumed by the individual transistors. There are no theoretical limits on k_1 because the feature sizes can also be manipulated using over-exposure, etching, and other tricks like chemical shrinking [87]. Thus the limits are bound by our ability to

faithfully and repeatedly produce a feature with small CD [93]. Unfortunately, the lithography literature often refers to k_1 as being lower bounded by 0.25 (but this is only true if the aspect ratio is 1:1).

A typical imaging system is characterized by its modulation transfer function (MTF). However, an optical lithography system has a high contrast recording device (the photoresist). The latter is capable of taking even a weak modulation and converting it into a high contrast (binary) pattern. Therefore, the metric of evaluation of lithography systems are the developed images and their robustness to process variations. The lithography difficulty is inversely proportional to the k_1 factor. Value of $k_1 > 0.7$ indicate very easy process conditions, whereas $0.7 > k_1 > 0.25$ becomes increasingly difficult as k_1 becomes smaller.

As the k_1 values falls below 0.7, the image quality degrades noticeably. The photoresist non-linearity is alone insufficient for production-worthy low k_1 lithography because of the poor CD control. There are two insertion points where process changes can be made to overcome the above issue; namely, manipulating the optical wavefront and manipulating the photoresist stack. The latter forms the domain of chemists who are actively working on tuning the resist response system. However, the domain of this thesis lies in the wavefront manipulation point known as resolution enhancement techniques (RETs).

Resolution Enhancement is defined by Frank Schellenberg as "manipulating the wavefront at the mask to make structures in a high contrast medium that are finer than those possible under conventional exposure conditions" [83]. RETs have

become an integral part of the IC manufacturing flow for the past several generations due to the lack of flexibility of any other wavelength besides 193nm. The k_1 values are continuously falling (current values in the range of 0.29-0.35) requiring very aggressive lithography strategies. Indeed RET have and will continue to play a huge part in our effort to stay abreast with Moores's law.

1.2 Resolution Enhancement Techniques

RETs are based on exploring three properties of the wavefront; namely, its amplitude, phase, and direction. These are respectively referred to as optical and process correction, phase-shift method and off-axis illumination [94]. In what follows, we provide a brief introduction to these methods.

1.2.1 Optical and Process Correction (OPC)

In low k_1 imaging, the high spatial frequency components carry sizable fraction of the light energy. The lens pupil cannot capture higher order components, thereby introducing distortions. The three types of distortions typically arising are 1) feature with the same nominal CD, printing differently under environmental differences (like dense, semi-dense, or isolated), 2) line-end pull back (or line shortening), and 3) corner rounding. The above are also referred to as optical proximity effects.

OPC was initially an acronym for optical proximity correction. But nowadays it implies optical and *process* correction since it also accounts for resist and etch effects in addition to the optical ones. OPC works by manipulating the first property,

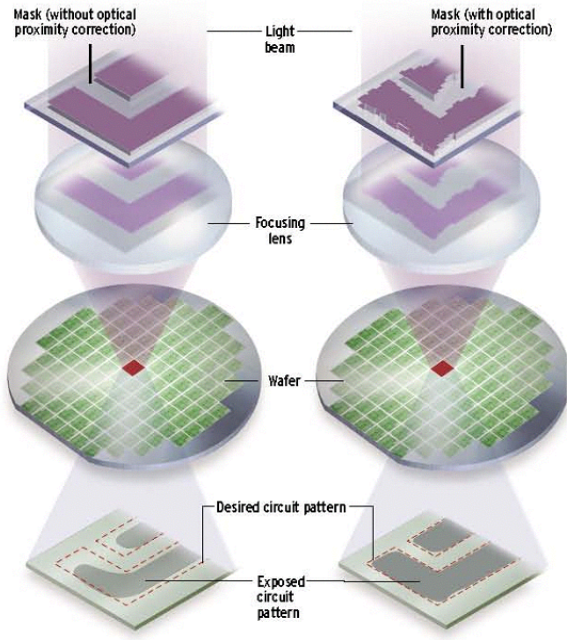

Figure 1.4: Optical and Process Correction (from [82])

namely the amplitude of the wavefront to counteract the imaging effects that distort the reproduced patterns on the wafer. The amplitude of the light-wave can be adjusted by changing the sizes of the openings thereby allowing more or less amount of light through as shown in Fig. 1.4. More specifically, OPC consists of adding sub-resolution features (very small features whose dimensions are below the resolution of the imaging system) to the original layout, a process also known as micromanipulation. Typical modification involves lengthening of a feature, displacing the edge of a pattern (edge-biasing), or creating serifs to reduce corner rounding. In addition to layout modifications, new features called assist-bars can also be added (more on this in Chapter 2). Thus the idea is to pre-compensate for the process losses by modifying the original layout, which eventually leads to a better pattern fidelity and improved

resolution [81].

OPC has been traditionally carried out using two approaches; namely rule-based, or model-based. As the name suggests, in the rule-based OPC scheme, empirical rules are developed to counteract the commonly occurring problems around corners, edges, local interactions, etc. These are then applied throughout the pattern to provide a general improvement in pattern fidelity. Rule-based OPC is very simple to implement but it only compensates for the local features and does not optimize the global performance depending on the overall layout. Hence it has limited scope, effectiveness and applications [94].

Model-based methods on the other hand use a mathematical description to represent the pattern transfer process (forward model). As such, they are more universal and represent a more aggressive OPC strategy. The success of these methods relies heavily on accurate modeling of the distortion process which has been studied carefully by physicists, chemists, and other researchers in the lithography community [24] (more on this in Chapter 2).

Model-based OPC itself has two flavors. The forward model-based OPC techniques employed by Cobb and Zakhor [15] parameterize (or fragment) the pattern into edges and corners. The algorithm proceeds by *locally* moving the edges, using a forward model simulator to determine if the new locations are better, moving them somewhere else, and iteratively repeating this process until certain error criteria are satisfied. Fig. 1.5 illustrates the model simulation feedback optimization loop proposed by them. The above technique is suboptimal because the edge movements are

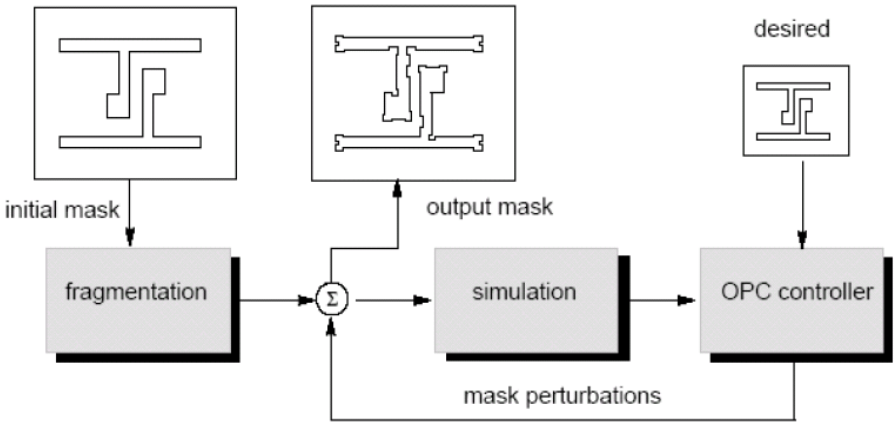

Figure 1.5: Feedback optimization loop for model-based edge-based OPC (from [17])

local and trial-and-error based. Furthermore, the number of fragments are becoming unmanageable for 45nm node making it computationally very demanding and there is also no built-in provision for sub-resolution assist features (see Chapter 3). Edge-based OPC is also referred to as sparse OPC.

Backward (or inverse) model-based techniques on the other hand attempt to synthesize the optimized OPC pattern by directly inverting the mathematical model from mask to wafer [94]. They are also referred to as inverse lithography technology which forms the focus of this thesis. The pioneering work in this area was done by Sherif, [84], et al. in the 1980s at the University of Wisconsin, who employed branch and bound algorithms to synthesize pixel-based OPC masks. A more formal introduction of inverse lithography is presented in Chapter 2.

1.2.2 Phase Shift Masks

Phase Shift Masks (PSM) exploit the second property of the wavefront, namely its phase, to achieve higher resolution and were originally introduced by Levenson, et al. [42] at IBM in 1982. Fig. 1.6 illustrates the pattern transfer process for a binary mask comprising of only quartz and chrome features. Note that there is some unwanted energy in the opaque (chrome) regions due to the close proximity of the neighboring transmitting (quartz) features. This unwanted energy affects the resist profile and the achievable resolution. In this section, we introduce three widely used flavors of PSM and provide some intuitive understanding on how phase manipulation can help us improve the resist profile and pattern fidelity.

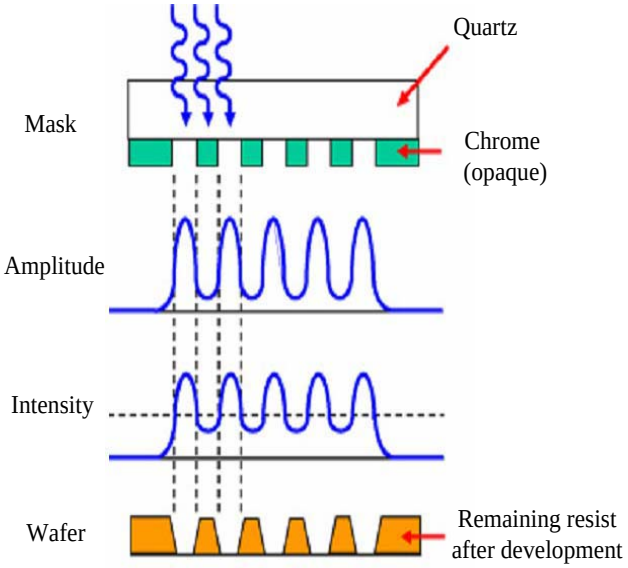

Figure 1.6: Pattern transfer using binary mask

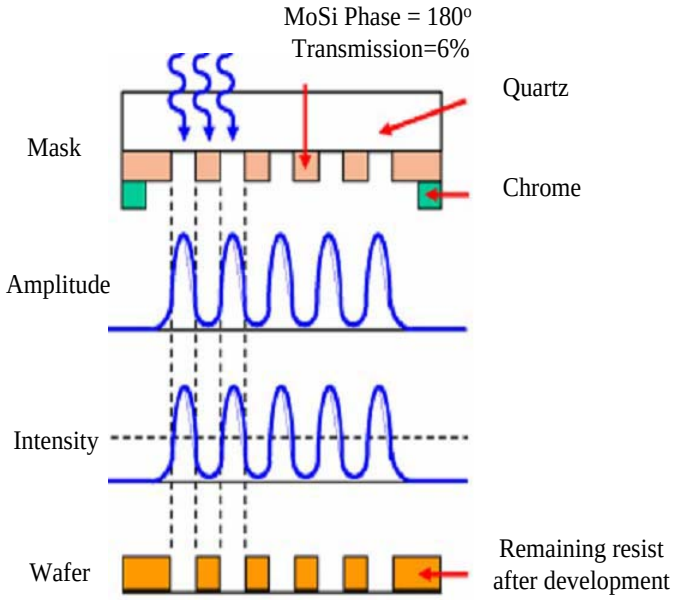

Figure 1.7: Pattern transfer using Attenuated PSM

Attenuated Phase-Shift Mask: The features in Attenuated Phase Shift Masks consists of quartz and molybdenum silicide (MoSi) instead of chrome. Unlike chrome, MoSi allows a small percentage of light (typically 6% or 18%) to pass through. The thickness of MoSi is chosen such that light which passes through is 180° out of phase compared to the transmitting quartz regions. As seen from Fig. 1.7, the light passing through the MoSi areas is very weak to expose the resist. However, the phase difference causes destructive interference thereby pushing the intensity further down in the opaque regions. This cancels out some of the unwanted energy resulting in a sharpened intensity profile.

Alternating Phase-Shift Mask: In this case, the opaque chrome features are surrounded on one side by quartz of phase zero degree, and on the other side by quartz

of phase 180 degrees. The phase shift is achieved by carefully etching the quartz region. As seen from Fig. 1.8, the transmitted electric field amplitude is positive and negative in the non-etched and etched quartz regions, respectively. The electric field amplitude is forced to pass through zero as it changes from positive to negative. This transition occurs underneath the opaque (chrome) area which now has zero energy deposition. This leads to very dark and sharp lines on the wafer and an improved resist profile. It is important to note that Alt-PSM, although powerful, is a very complex technology and the resulting masks are very hard and costly to manufacture.

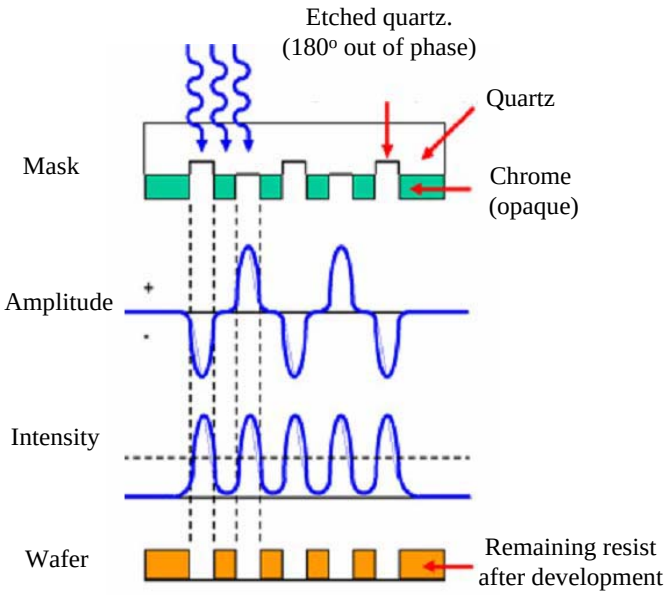

Figure 1.8: Pattern transfer using Alternating PSM

Chromeless Phase Lithography: Chromeless Phase Lithography is a 100% transmission version of the Attenuated PSM since they have no opaque (chrome) features. Dark lines are produced on the wafer due to the interaction of the two adjacent (180

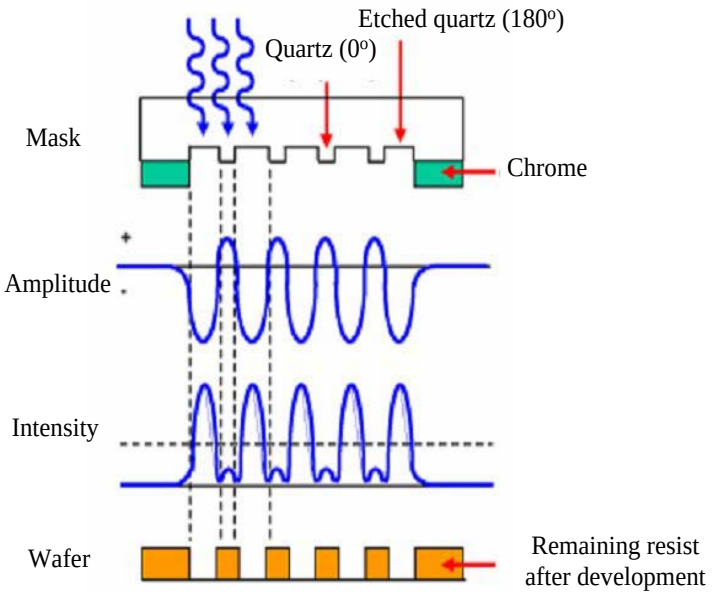

Figure 1.9: Pattern transfer using Chromeless PSM

degree) phase-shifted quartz features. Once again by forcing the amplitude to pass through zero, the intensity which is proportional to the square of the amplitude also hits zero, resulting in sharp lines on the wafer (see Fig. 1.9).

In Chapter 4, we propose an ILT framework for automatically synthesizing attenuated and strong phase shift masks for single exposure lithography. Strong PSM also suffer from phase conflict issue which can be overcome using double exposure lithography (see Chapter 5).

1.2.3 Off-axis Illumination

Off-axis illumination (OAI) manipulates the direction or the angle at which the light falls on the mask thereby affecting the diffraction orders captured by the lens. It consists of modifying the source configurations into specific geometric patterns like

annulus, quadrapole, etc as shown in Fig. 1.10. This is achieved by placing a light blocking hard stop aperture or the more recently invented diffractive optical elements (DOE) [35], between the laser and the mask. DOE have the capability of forming very complex source shapes and smooth distribution of light across the aperture. Each configuration can emphasize certain spatial frequencies while de-emphasizing other ones and has its own advantages or disadvantages. DOE source design can also be treated as an inverse problem (as observed by Granik in [30]), but is beyond the scope of this thesis.

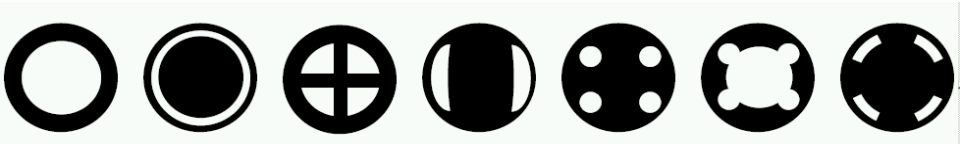

Figure 1.10: Commonly employed source configurations; left to right - conventional, annular, fourfold source, separated source, quadrapole, CQUEST, QUASAR (from [83])

We would like to highlight that all the resolution enhancement techniques increase the complexity, storage requirements, and cost of the masks (currently running into millions). Therefore an important challenge is to implement RETs while keeping the above factors under control.

1.2.4 Double Exposure and Double Patterning

We now provide a brief introduction to two additional lithography techniques which are becoming increasingly popular for 45nm and smaller nodes.

Double exposure lithography (DEL) is a technique commonly employed to improve the aerial image contrast. It consists of exposing the same photo-resist layer se-

quentially using two separate exposures, typically using two different reticles. Double-exposure can be done with existing production lithography technology, and is often combined with other RETs discussed above. DEL enables patterning very close to the lower bound ($k_1 = 0.25$) by providing a very high contrast aerial image, and is currently employed with AltPSM [28], double dipole lithography (DDL) [12], etc.

Double patterning involves splitting a dense pattern into two separate less dense circuit patterns (also know as pitch splitting). The two relatively simple patterns are then printed separately on a target wafer. The wafer is removed from the exposure chuck in between the two exposures for development and etch processing. Double patterning can theoretically double the resolution of the imaging systems and the lower bound on k_{pitch} becomes 0.25. Thus, it has the potential to enable 32nm and 22nm nodes using 193nm lithography. However, it requires extremely accurate overlay between the two exposures, and has a lower throughput because of the additional exposure and wafer processing step. The above techniques are discussed in more detail in Chapter 5 and 6.

1.3 Thesis Organization

The thesis is organized as follows.

In Chapter 2, we start by discussing lithography modeling, formally introduce the inverse lithography problem, and do a literature survey of the past work on ILT. We then introduce the *sigmoidal* approximation of the resist behavior, and discuss

our approximate forward lithography model for the cases of coherent, incoherent, and partially coherent imaging systems. This is followed by the continuous function optimization formulation of the ILT problem, a key difference in our approach compared to others.

In Chapter 3, we discuss the single exposure OPC mask design framework. We propose a cost function to minimize, and derive the analytic gradient of the cost function. We then discuss the parametric transformation technique to reduce the bound-constrained optimization problem to an unconstrained one. This is followed by the motivation and introduction of the regularization framework for ILT, which enables the user to promote certain desirable properties in the solution. The above framework is used to restrict the tone of the OPC masks, reduce the manufacturing complexity, and improving the aerial image contrast. The results demonstrate automatic generation and placement of assist-bars for coherent imaging, and un-intuitive feature breaking behavior for incoherent imaging systems. Finally, we also introduce two useful and practical extensions of the mask design problem, namely, weighted mask and selected parameter optimization.

In Chapter 4, the ILT framework is extended to single exposure phase shift mask synthesis. In particular, we deal with the automated design of attenuated PSM, 100% transmission PSM, and strong PSM with chrome. Appropriate regularization terms are introduced to control the tone of the estimated masks. The phase assignments are automatically carried as part of the optimization, and the results indicate a much superior contrast for strong PSM.

Chapter 5 is dedicated to double exposure inverse lithography technology. We start by introducing double exposure and double patterning lithography systems. We then exclusively focus on the former technique, where the resist is developed only once. We introduce the DEL-ILT mask design problem, and discuss the continuous function optimization formulation employed to solve it. We propose two gradient-based optimization algorithms to minimize the cost function. The results indicate a new paradigm of splitting the gray-level aerial image between the two exposures and a very superior contrast.

In Chapter 6, we propose a novel paradigm of using DEL for $k_{pitch} < 0.5$ by employing novel materials. We study reversible contrast enhancement layer (CEL) and 2-photon absorption resists as two potential candidates. We propose a mathematical model for CEL, discuss possible ways of improving the image log-slope (ILS), and also derive lower bounds on ILS improvement. We extend the single exposure ILT framework to also incorporate for the CEL effects, and synthesize CPL masks to enable patterning 2-D features with $k_1 = 0.185$.

In Chapter 7, we review five recent bodies of work on ILT (published over the past 1.5-2 years), and show connections of these approaches with our proposed framework. We also present preliminary DEL-ILT results for increasing the robustness of the lithography system to mask-to-mask alignment errors. We then discuss the extension of our approach to e-beam proximity effect correction as well as other interesting future directions emanating from this research project. Finally, we provide conclusive remarks outlining the contribution of this thesis.

Chapter 2

The Inverse Lithography Problem

We now present a detailed discussion of the inverse lithography mask design problem. The discussion consists of the lithography forward model description, literature survey of ILT, our proposed approximate forward model, and the optimization problem setup.

2.1 Lithography Modeling

Before we move on to the inverse lithography problem, it is desirable to discuss the forward (process) model. Accurate modeling of the optical lithography system is critical because it allows prediction of the wafer pattern without going through the cumbersome fabrication process. Fast and accurate lithography simulation has become an integral part of the development of new technologies.

Fig. 2.1 illustrates a block diagram view of the lithography process. It consists

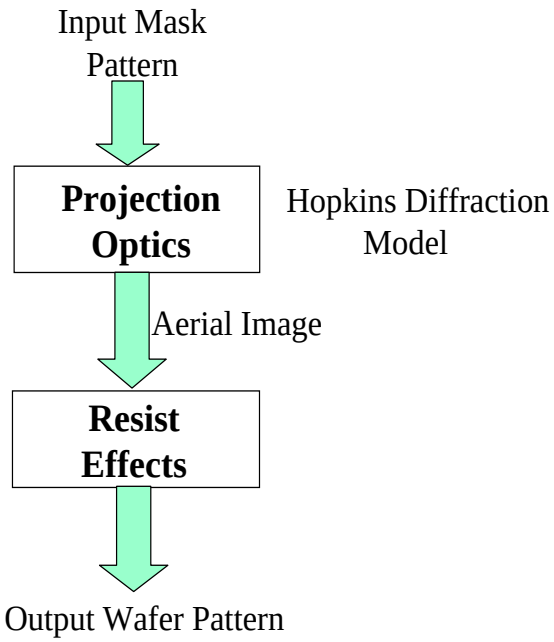

Figure 2.1: A simplified view of the lithography process

of two functional blocks; namely the projection optics effects (aerial image formation), and the resist effects. The aerial image calculations are based on the underlying optical system being coherent, incoherent, or partially coherent [11, 16]. Practical lithography systems are partially coherent imaging systems which can be simulated using the Hopkins (scalar or vector) imaging model [11]. Pati and Kailath in [64] proposed the sum of coherent systems (SOCS) decomposition of the Hopkins imaging equation, which was later used by Cobb and Zakhor [16] to speed up the forward model calculations.

The resist exposure model consists of bulk image formation, photo-active compound (PAC) generation, post-exposure bake, and resist development [13]. Bulk image is the intensity within the thin resist film. It is formed by interference of

counter-propagating plane waves within the resist thin film stack and also takes into account the reflections from the bottom of the substrate. The resist bleaches under the influence of the exposure energy and undergoes changes in photochemical properties. Dills model [13] is used to model the conversion of the bulk intensity in the resist film into the concentration of the photochemical compound. It is this concentration that determines the resist dissolution properties and allows for the resist to be patterned. The post exposure bake step that follows the exposure causes diffusion of the photo-active compound and smooths the standing waves thereby giving good 3-D resist profiles [13]. Finally the resist is developed or etched away using a developer solution, a process which can be modeled (for example) using the Mack model [52, 77].

Note that resist exposure and development involves a 3-D visualization with all of the above steps affecting the resist profile. However, for the sake of speed and simplicity, lithographers often consider the 2-D contour view of the (final) developed resist profile (where the contours are measured at the resist/substrate interface). The developed 2-D contours are then directly compared to the (desired) 2-D target layouts to determine the pattern fidelity. There are numerous 2-D resist models like Variable Threshold Resist (VTR) model [76], Constant Threshold Resist (CTR) model, or empirical models [2] which predict the 2-D resist contours from the incoming aerial image. We use the CTR model in our analysis (see Section 2.3 for more details on our imaging model).

Finally, there are numerous lithography simulation softwares available from various commercial and academic organizations like iPHOTO-II from Intel [74], SPLAT/SAMPLE

from University of California Berkeley [6], [40], SOLID from the Fraunhofer Institute [24], and Prolith from KLA-Tencor to name a few, which employ the aforementioned (and many other proprietary) models.

2.2 The Inverse Problem

As discussed in Chapter 1, there has been a revival of interest in "inverse lithography" or "layout inversion" techniques in recent times, which are introduced more formally in this section.

2.2.1 Introduction

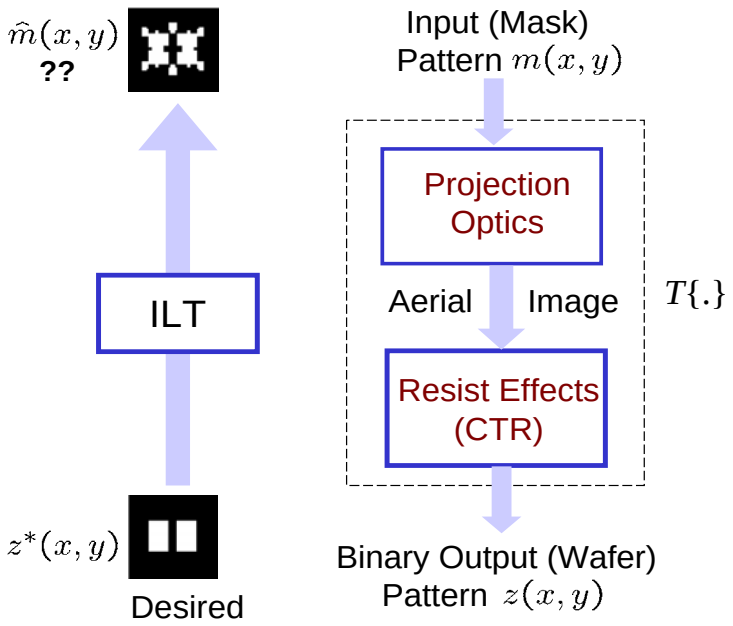

Figure 2.2: Forward Model and ILT

The first step towards solving an *inverse* problem is to define a *forward* (or

process) model which is a (possibly approximate) mathematical description of the given imaging system. The lithography imaging system from the mask to the wafer consists of two steps (as seen earlier in Section 2.1, Fig. 2.1, and Fig. 2.2).

The image formation process can be mathematically expressed as,

$$z(x, y) = T\{m(x, y)\}, \tag{2.1}$$

where $T\{.\}$ is the forward model which maps the input intensity function $m(x, y)$ to the output intensity function $z(x, y)$. Let $z^*(x, y)$ be the desired output intensity function. The goal of inverse lithography technique (ILT) is to estimate the input intensity function which will give us a close approximation to the desired output $z^*(x, y)$ (see Fig. 2.2). This is achieved by searching the space of all inputs and choosing $\widehat{m}(x, y)$ which minimizes a distance $d(z(x, y), z^*(x, y))$, where $d(\cdot, \cdot)$ is some appropriate distance metric to be defined later. Thus,

$$\widehat{m}(x, y) = \arg \min_{m(x,y)} d[z^*(x, y), T\{m(x, y)\}]. \tag{2.2}$$

In our case, $T\{.\}$ is the lithography forward (process) model, $z^*(x, y)$ is the desired output wafer pattern, and $\widehat{m}(x, y)$ is the estimated optical proximity correction or phase shift mask pattern.

Inverse lithography is an *image synthesis* or *image design* problem. The problems of *image restoration* (reconstruction) and *image synthesis* (design) are related but not the same as pointed out by Sayegh, et al. [79, 80]. In both cases, the output and the imaging system are known, but the input is unknown [79, 80]. In the restoration case, the output image results from an actual but unknown input image [29], and

therefore at least one solution must exist in the absence of measurement error or noise. However, in the synthesis case, there may be no input capable of producing the prescribed output, so that we have to target the output as close as possible to the desired ideal output image. Furthermore, the solution need not be unique.

2.2.2 Past Work on ILT

The pioneering work in ILT was by B.E.A. Saleh and his students at the University of Wisconsin in the early 1980's. In 1985 Sayegh, et al. [80] used linear-programming technique to do OPC for a system modeled as a band-limited linear system followed by hard threshold operation. They suggested using "pixel flipping" method to synthesize the masks. Pixel values of 0 or 1 are tried. If the fidelity error decreases, then the pixel is accepted, else it is rejected and the next pixel is tried. Nashold and Saleh [56] employed iterative alternating projections, and Sherif, et al. [84] used mixed linear integer programming to synthesize binary masks for the above imaging system. Note that the above work (including their proposed techniques) are restricted to only the linear aerial imaging model.

Pati and Kailath [64] considered the more prevalent partially coherent approximation to a coherent imaging system, using optimal coherent approximations. The above system becomes linear in the electric field complex amplitude, which therefore enabled them to use POCS (Projection on Convex Sets) to synthesize phase shift masks. Liu and Zakhor [46] formulated the mask design problem as the minimization of the L_2 norm of the difference between the ideal and the actual wafer images, thereby

obtaining an integer constrained optimization problem. Later in [47], they optimized the image log slope in the vicinity of the target. For incoherent imaging, they observed that the above problem is a linear programming one which was solved using branch and bound algorithms. However, it becomes quadratic for partially coherent imaging systems, and hence they proposed simulated annealing algorithms to design phase shift masks (PSM). Peckerar, et al. [65, 68, 67] employed L_2 norm based cost function and proposed gradient-based optimization technique to solve the proximity effects arising in the related e-beam lithography problem.

More recently, Oh, et al. [59] used random-pixel flipping and Erdmann, et al. [23] proposed genetic algorithms to solve the joint mask and source optimization problem. Socha, et al. [86] proposed interference map lithography (IML) technique for automatic generation and placement of assist bars for contact holes optimization. In 2005, Granik [31] did a comprehensive review of the past ILT work and discussed the reduction of the mask design problem to linear, quadratic, and general non-linear programming problems. In 2006, Liu, et al. [63] demonstrated the commercial viability of inverse lithography techniques. In Chapter 8, we provide a more detailed discussion of some the very recently published work [31, 63, 96, 50] on ILT, and relate them to our proposed framework.

We observe that while ILT has been extensively researched for the past 20 years, it has found commercial following only recently. Cobb and Dudau [14] compare the computational complexity of sparse (edge-based) and dense (pixel-based) OPC, and conclude that the latter might become more efficient at the 45nm scale. Another

reason for the recent success of ILT is the availability of more computing power and the development of new optimization algorithms. As discussed above, most of the previous work on ILT employed genetic algorithm, POCS, integer programming, branch and bound, etc. which have prohibitively high computational complexity.

However, our proposed ILT framework employs continuous function optimization and (analytically calculated) gradient information to systematically explore the solution space. As seen from Fig. 2.2, the input mask is binary (or trinary for Alternating PSM), whereas the output pattern is binary. The above two factors are detrimental towards having a continuous function formulation. In what follows next, we propose an approximated forward model for ILT which overcomes the binary output restriction. The discrete mask tone considerations are explored in Chapter 3.

2.3 Approximate Forward Model

We now discuss the forward model used for our inverse problem formulation. This section is divided into two parts. We start by proposing an approximation to the resist model, and then explore the details of the aerial imaging models.

Since we employ the pixel-based approach, the first step is to represent the input, output, and the desired patterns using 2-D discrete images. We define vectors $\underline{\mathbf{z}}^*,\ \underline{\mathbf{z}},\ \underline{\mathbf{m}} \in \Re^{MN\times 1}$ which are obtained by sampling and lexicographic ordering of $z^*(x,y)$, $z(x,y)$, and $m(x,y)$ respectively. The number of samples along the horizontal and vertical directions are given by M and N respectively. Note that the edge place-

ment error is related to the sampling interval $\delta_s$nm and is upper-bounded by $(\delta_s/2)$nm. Smaller pixels enable better edge placement but also increase the number of samples M and N which are related to the algorithmic complexity by $O(MN\log(MN))$. Thus, there is a trade-off between speed and accuracy. Throughout our discussion, $\underline{\mathbf{z}}^*$ represents the prescribed *binary* pattern, $\underline{\mathbf{z}}$ represents the *gray-level* output pattern, and $\underline{\mathbf{m}}$ represents the input pattern fed to the imaging system (which can be *binary* or *gray-level*).

2.3.1 Sigmoid Approximation of the Resist

The simplest way to simulate the resist effect is using the constant threshold resist (CTR) model [37]. Thus, for positive resists, the areas having aerial image intensity higher than threshold t_r are completely removed leaving behind a *space* in the wafer. The above operation can be described using a Heaviside operator (hard threshold) defined as

$$\Gamma(u) = \begin{cases} 0, & u \leq t_r \\ 1, & u > t_r \end{cases} \tag{2.3}$$

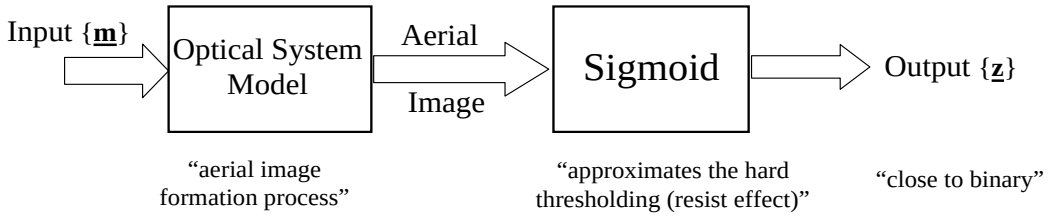

Figure 2.3: Approximated forward process model

The Heaviside operator brings us into the discrete domain and necessitates the use of branch and bound or other integer optimization algorithms (like our predecessors [79, 47]) to solve the mask design problem. However, the focus of our work is to solve the inverse problem in the continuous domain. With this in mind, we approximate the heaviside operator using a sigmoid: a smooth, continuous function [22]. The approximated forward process model is illustrated in Fig. 2.3. We employ the logarithmic sigmoid function,

$$\mathrm{sig}(u) = \frac{1}{1 + e^{-a(u-t_r)}}, \tag{2.4}$$

where the parameter a dictates the steepness of the sigmoid. The parameter t_r is the threshold parameter of the sigmoid and is set equal to the threshold level of the resist in accordance with the constant threshold resist model. Fig. 2.4 illustrates the behavior of the sigmoid for different values of a with $t_r = 0.5$. A large value of a leads to a very steep sigmoid which closely resembles the hard thresholding operation. Owing to the above approximation, the output pattern $\underline{\mathbf{z}}$ will not be binary, but a (continuous tone) " close-to-binary" pattern. The above approximation enables us to use gradient-based continuous function optimization techniques like steepest-descent to solve the mask design problem.

2.3.2 Aerial Image Models

We now discuss aerial image calculations and the (cascaded) forward models for coherent, incoherent, and partially coherent imaging systems.

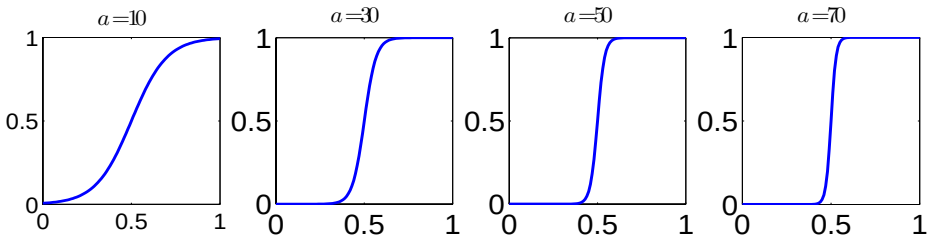

Figure 2.4: The effect of the steepness parameter a on the sigmoid function $\text{sig}(u) = 1/(1 + e^{-a(u-0.5)})$.

Coherent Imaging System

In the case of coherent imaging systems, the spatial distribution of the output electric field amplitude $e(x, y)$ is linearly related to the input electric field amplitude generated by the mask $m(x, y)$. This can be mathematically described as,

$$e(x, y) = m(x, y) * h(x, y) \tag{2.5}$$

where $h(x, y)$ is referred to as the amplitude spread function (ASF) of the given imaging system [92].

Typical lithography systems employ a circular lens aperture, where the coherent imaging system now acts as an ideal low pass filter with cutoff frequency NA/λ. The higher frequency components of the diffracted mask image are lost by the finite lens aperture stop thereby causing a blurry version of the mask image at the imaging (wafer) plane. The convolution kernel $h(x, y)$ is defined as the Fourier transform of the circular lens aperture with cutoff frequency NA/λ [11, 92]. Therefore,

$$h(x, y) = jinc(r\frac{NA}{\lambda}) = \frac{J_1(2\pi r\ NA/\lambda)}{2\pi r\ NA/\lambda} \tag{2.6}$$

where $r = \sqrt{x^2 + y^2}$ and $J_1(\cdot)$ is the first order Bessel function of first kind. Fig. 2.5

illustrates the jinc function $h(x, y)$ for the case of $NA = 0.85$ and $\lambda = 193nm$ sampled at 10nm. Note that it has very prominent side-lobes which lead to the generation of assist-bars (more on this in Chapter 3).

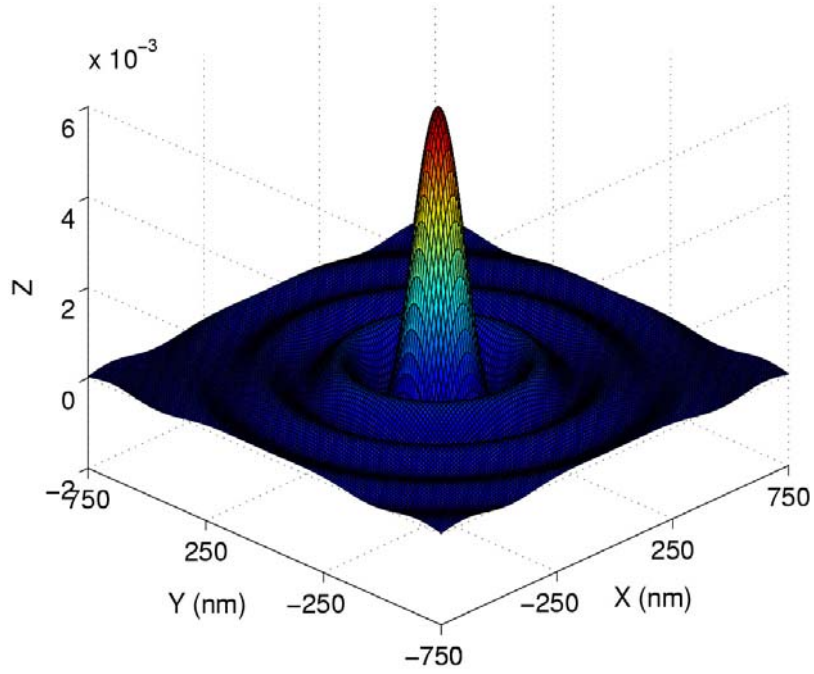

Figure 2.5: The jinc function for NA=0.85 and $\lambda = 193nm$.

The photo-resist responds to the intensity of the electric field, where intensity is defined as the square of the complex amplitude $e(x, y)$. Therefore, the forward model is defined as,

$$\underline{\mathbf{z}}^{\mathbf{c}} = \text{sig}(|\underline{\mathbf{e}}|^{\mathbf{2}}) = \text{sig}(|\mathbf{H}\underline{\mathbf{m}}|^2), \tag{2.7}$$

where $\underline{\mathbf{z}}^{\mathbf{c}}$ represents the output pattern for coherent imaging systems. The sigmoid function simulates the resist behavior and acts on the aerial image $|\mathbf{H}\underline{\mathbf{m}}|^2$ (square of the amplitude), giving the output pattern $\underline{\mathbf{z}}^{\mathbf{c}}$. Two important points to note: the kernel $\mathbf{H} \in \Re^{MN \times MN}$ is the jinc function $h(x, y)$ sampled using the same sampling rate as

$\underline{\mathbf{z}}^{\mathbf{c}}$, and $|\cdot|^2$ operator here implies element-by-element absolute square of the individual vector entries. Finally, for partially coherent imaging systems, the optical kernel $h(x, y)$ can be instead substituted by the optimal coherent approximation proposed by Pati and Kailath [64].

Incoherent Imaging System

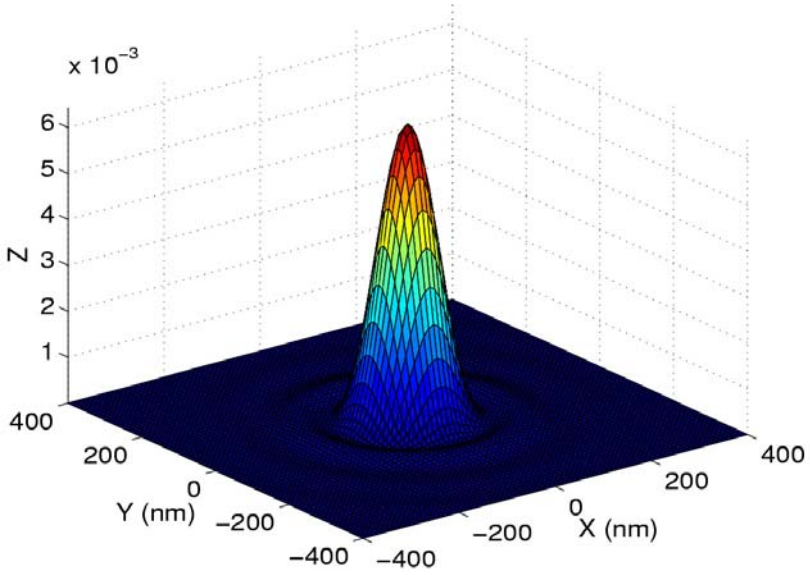

Figure 2.6: The jinc squared function for NA=0.85 and $\lambda = 193nm$.

An incoherent imaging system is linear in intensity (or irradiance) and can be mathematically described as [11],

$$|e(x, y)|^2 = |m(x, y)|^2 * |h(x, y)|^2. \tag{2.8}$$

The kernel $h(x, y)$ is defined in (2.6), and $|\cdot|^2$ once again implies absolute square of the individual elements. Note that the phase of the input electric field does not contribute towards the output. Hence, for incoherent imaging systems we restrict our attention

to only binary masks. The photo-resist directly responds to the above electric field intensity and the forward model for a binary mask is defined as,

$$\underline{\mathbf{z}}^{\mathbf{i}} = \text{sig}(|\mathbf{H}|^2|\underline{\mathbf{m}}|^2) = \text{sig}(\tilde{\mathbf{H}}\underline{\mathbf{m}}), \tag{2.9}$$

where $\underline{\mathbf{z}}^{\mathbf{i}}$ represents the output pattern for incoherent imaging systems. The filter $\tilde{\mathbf{H}}$ in (2.9) is known as the point spread function of the imaging system and is a jinc-squared function (see Fig. 2.6). It is defined as the square of the PSF shown in (2.6). Also, since the mask is binary, $|\underline{\mathbf{m}}|^2 = \underline{\mathbf{m}}$.

Partially Coherent Imaging System

Real-world lithography systems are partially coherent and can be modelled using the Hopkins diffraction model [11]. Pati and Kailath [64] proposed an approximation to the above model called the sum-of-coherent-system (SOCS) by using the singular value decomposition of the transmission cross-coefficient matrix. In their approach, the P^{th} order approximation to the aerial image formulation can be calculated using the weighted sum of P coherent systems. The forward model now becomes,

$$\underline{\mathbf{z}}^{\mathbf{p}} = \text{sig}\left(\sum_{j=1}^{P} \sigma_j |\mathbf{H}_j \underline{\mathbf{m}}|^2\right) \tag{2.10}$$

where $\mathbf{H}_j$ for $j = 1, \ldots, P$ are the amplitude spread functions (also referred to as optical system kernels) of the coherent systems, and $\sigma_1, \ldots, \sigma_P$ are the corresponding singular values. The singular values quickly decay to zero, thereby facilitating an accurate reduced order approximation. Fig. 2.7 is a block diagram view of the aerial image

formation model for partially coherent imaging system computed as the weighted sum of coherent imaging systems.

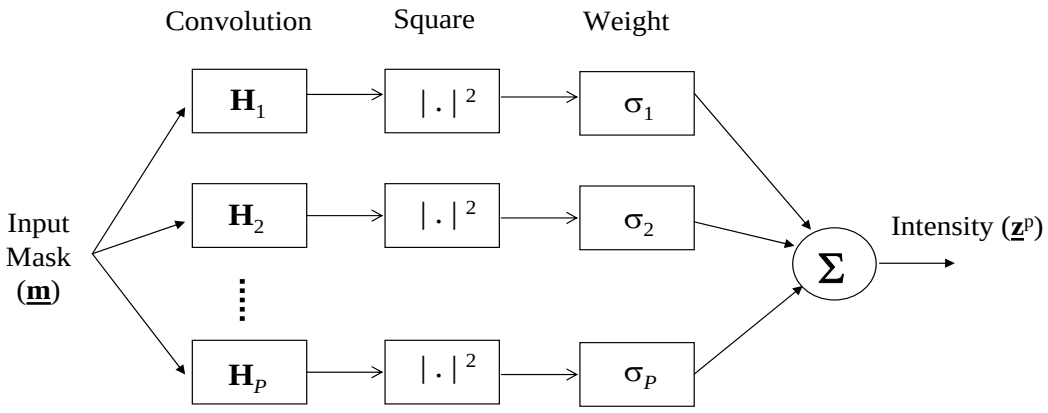

Figure 2.7: Aerial image computation of partially coherent imaging systems using SOCS.

In our work, we have focused only on fully coherent and incoherent imaging systems. However, our framework is also readily extendible to partially coherent imaging systems via the SOCS model discussed above.

2.4 ILT Optimization Problem

We now formulate the mask design optimization problem for coherent and incoherent imaging systems. Note that every pixel m_j represents the transmission values of the mask which can be represented as a complex term $m_j = p_j + iq_j$, where $i = \sqrt{-1}$. For the present discussion, we restrict ourselves to only 180 degree phase shifting[1]. Thus, angle$(m_j) = 0$ (no phase-shifting) or angle$(m_j) = \pi$. The above conditions imply that q_j is always zero and the sign of p_j dictates the phase (positive

[1] Later in Chapter 5, we discuss double exposure lithography which is equivalent to allowing multiple phases.

values corresponding to no phase-shift and negative values corresponding to 180 degree phase-shift). Furthermore, $|p_j|$ gives the mask transmission magnitude. Therefore, we are required to estimate only the real part (p_j) of the mask (which we henceforth refer to as m_j for notational convenience). We formulate the mask design problem as finding the optimized mask layout $\widehat{\underline{\mathbf{m}}}$ that minimizes the cost function $F(\underline{\mathbf{m}})$, defined as the L_2 norm of the difference between the desired pattern $\underline{\mathbf{z}}^*$ and the output pattern $\underline{\mathbf{z}}$ [2]. That is,

$$\widehat{\underline{\mathbf{m}}} = \arg\min_{\underline{\mathbf{m}}}\{F(\underline{\mathbf{m}})\} = \arg\min_{\underline{\mathbf{m}}} ||\underline{\mathbf{z}}^* - \underline{\mathbf{z}}||_2^2 = \arg\min_{\underline{\mathbf{m}}} \sum_{k=1}^{MN} (z_k^* - z_k)^2 \tag{2.11}$$

Later in Chapter 3, we refine this approach by introducing the regularization terms and augmenting the cost function.

Coherent Imaging System:

For coherent imaging system $\underline{\mathbf{z}} = \underline{\mathbf{z}}^{\mathbf{c}}$ and the cost function can be formulated using (6.5) as follows,

$$F(\underline{\mathbf{m}}) = F^c(\underline{\mathbf{m}}) = ||\underline{\mathbf{z}}^* - \underline{\mathbf{z}}^{\mathbf{c}}||_2^2 = \sum_{k=1}^{MN} (z_k^* - z_k^c)^2 \tag{2.12}$$

From (2.7) we observe that every pixel in a coherent imaging system undergoes a cascade of convolution, squaring, and sigmoidal operation. Therefore, the output pixel z_k^c in (2.12) can be represented as,

$$z_k^c = \frac{1}{1 + \exp\left[-a\left(\sum_{j=1}^{MN} h_{kj} m_j\right)^2 + at_r\right]}\ , \tag{2.13}$$

[2] Alternately, it is also possible to employ other vector norms.

for $k = 1, \ldots, MN$ where h_{kj} are the elements of the k^{th} row of $\mathbf{H}$, which is the jinc function defined in (2.6).

The estimated mask is either two or three tone depending on the employed RET. Therefore, the transmission values m_j for $j = 1, \ldots, MN$ should be allowed to take only specific values as summarized in the table below.

RET	Allowable transmission values
OPC	0 or +1
6% Attenuated PSM	-0.2449 or +1
18% Attenuated PSM	-0.4243 or +1
Strong PSM (100% transmission)	-1 or +1
Strong PSM (With chrome)	-1 or 0 or +1

The optimization problem (2.12) is therefore subject to the constraints given by the allowable transmission values of m_j. This unfortunately makes the search space discrete thereby again reducing our problem into an integer optimization one. To overcome this issue and move back into the continuous domain, we relax the parameter values to lie within a *range* $[\underline{m}, \overline{m}]$. Thus the discrete equality constraints are substituted by the inequality (bound) constraints and the optimization problem in (2.12) is made subject to

$$\underline{m} \leq m_j \leq \overline{m} \quad \text{for } j = 1, \ldots, MN \ . \tag{2.14}$$

The above procedure reduces the mask design problem into a bound constrained optimization problem which will be henceforth employed in our discussion.

Incoherent Imaging System:

For an incoherent imaging system $\underline{\mathbf{z}} = \underline{\mathbf{z}}^{\mathbf{i}}$ and the cost function can be formulated

using (6.5) as follows,

$$F(\underline{\mathbf{m}}) = F^i(\underline{\mathbf{m}}) = ||\underline{\mathbf{z}}^* - \underline{\mathbf{z}}^{\mathbf{i}}||_2^2 = \sum_{k=1}^{MN} \left(z_k^* - z_k^i\right)^2 \tag{2.15}$$

From (2.9) we observe that every pixel in an incoherent imaging system undergoes a cascade of convolution followed by a sigmoidal operation. Therefore, the z_k^i in (2.15) is given as,

$$z_k^i = \frac{1}{1 + \exp\left(-a \sum_{j=1}^{MN} \tilde{h}_{kj} m_j + a t_r\right)} , \tag{2.16}$$

for $k = 1, \ldots, MN$ where $\tilde{h}_{kj}$ are the elements of the k^{th} row of $\tilde{\mathbf{H}}$, which is the jinc-squared function.

As discussed earlier, we are only interested in binary masks, and the optimization problem in (2.15) is subject to the bound constraints

$$0 \leq m_j \leq 1 \quad \text{for } j = 1, \ldots, MN \ . \tag{2.17}$$

To summarize, we have been successful in formulating the mask design problem as a continuous function optimization problem. Our next step is to efficiently solve this large-scale optimization problem, which forms the topic of discussion of the next chapter.

Chapter 3

Single Exposure OPC Mask Design

In this chapter, we discuss the optimization algorithm for synthesizing OPC masks for coherent and incoherent imaging systems. We then introduce the regularization framework to guarantee near binary results, ease of manufacturing, and good quality aerial images. The discussion is facilitated by numerous examples illustrating the working of the proposed framework.

Before proceeding, we take a quick look at the notations used in the manuscript. Matrices are represented using bold upper-case letters, and vectors are represented as bold, underlined, lower-case letters. Also scalar values are denoted using simple lower-case letters. More specifically, $\underline{\mathbf{H}}$ and $\underline{\tilde{\mathbf{H}}}$ correspond to the convolution matrix for coherent and incoherent imaging systems, respectively. For a circular lens pupil, $\underline{\mathbf{H}}$ and $\underline{\tilde{\mathbf{H}}}$ now represent the jinc and jinc-squared functions. Furthermore, $\underline{\mathbf{z}}^{\mathbf{c}}$ and $\underline{\mathbf{z}}^{\mathbf{i}}$ are

vectors representing the output *gray-level* patterns for coherent and incoherent imaging systems respectively. Finally, scalars with sub-scripts are used to index corresponding entries in a vector or a matrix, unless otherwise stated. Thus, z_k^i indicates the k^{th} element of the vector $\underline{\mathbf{z}}^{\mathbf{i}}$.

3.1 OPC Mask Design Algorithm

As discussed in Section 1.2.1, an OPC mask consists of only chrome and quartz features. Therefore, the mask transmission values are restricted to be either 0 or 1. We begin by considering the case of an incoherent imaging system. Recall that our goal is to solve the optimization problem defined in (2.15) subject to the constraints in (2.17).

The above problem can be directly solved using constrained optimization algorithms like gradient-projection [41], which consists of a descent along the direction of gradient, followed by a projection onto the constrained space. There is also a possibility of employing the projected BFGS (Broyden, Fletcher, Goldfarb, Shanno) [41, 73] algorithm. However, the bound-constrained optimization problem can be further reduced to an unconstrained optimization problem using the following parametric transformation,

$$m_j = \frac{1 + \cos(\theta_j)}{2} \quad \text{for} \quad j = 1, \ldots, MN, \tag{3.1}$$

where $\underline{\boldsymbol{\theta}} = [\theta_1, \ldots, \theta_{MN}]^T$ is the unconstrained parameter vector. The re-parameterized cost function for the incoherent imaging case can be formulated in terms of the pa-

rameter vector $\underline{\boldsymbol{\theta}}$ as follows,

$$F_1^i(\underline{\boldsymbol{\theta}}) = \sum_{k=1}^{MN} \left(z_k^* - \frac{1}{1 + \exp\left[-a \sum_{j=1}^{MN} \tilde{h}_{kj} \frac{1+\cos(\theta_j)}{2} + at_r \right]} \right)^2 . \tag{3.2}$$

We can now employ steepest-descent search to minimize the above cost function. This requires the first order derivatives of (3.2), and the gradient vector $\underline{\mathbf{d}}^{\mathbf{i}} = \nabla F_1^i(\underline{\boldsymbol{\theta}}) \in \Re^{MN \times 1}$ can be analytically calculated (see Appendix A) using the following expression:

$$\nabla F_1^i(\underline{\boldsymbol{\theta}}) = \underline{\mathbf{d}}^{\mathbf{i}} = a(\tilde{\mathbf{H}}^{\mathbf{T}} \ [(\underline{\mathbf{z}}^* - \underline{\mathbf{z}}^{\mathbf{i}}) \odot \underline{\mathbf{z}}^{\mathbf{i}} \odot (\underline{\mathbf{1}} - \underline{\mathbf{z}}^{\mathbf{i}})] \ ) \odot \ \sin(\underline{\boldsymbol{\theta}}) \tag{3.3}$$

where $\odot$ is the Hadamard product (element-by-element multiplication) of the two vectors, $\underline{\mathbf{1}} = [1, \ldots, 1]^T$, and $\underline{\mathbf{z}}^{\mathbf{i}}$ is defined in (2.9).

The flow-chart in Fig. 3.1 illustrates the steps involved in calculating the gradient given by (3.3). We would like to highlight the useful fact that due to the structure of (3.3), the steepest descent iterations can be quickly and directly carried out on the 2-D image array (matrices) with no need for the (algebraically convenient) raster scanning operation (see Fig. 3.1). Note that the gradient calculation involves two convolution operations which dictate the complexity of our algorithm. Thus, the algorithmic complexity is $O(MN \log(MN))$.

Having evaluated the gradient, the n^{th} iteration of steepest descent algorithm is given as,

$$\underline{\boldsymbol{\theta}}^{n+1} = \underline{\boldsymbol{\theta}}^n - s\underline{\mathbf{d}}^n \tag{3.4}$$

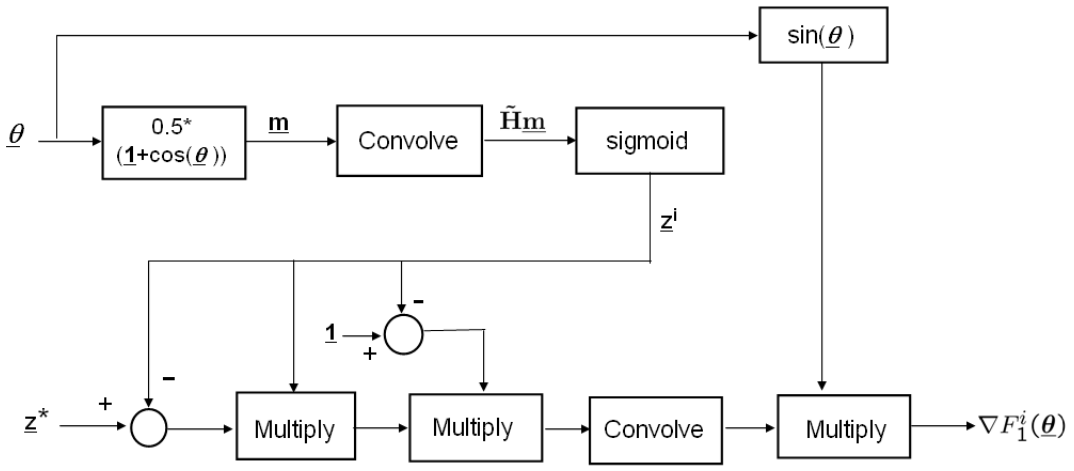

Figure 3.1: Flow-chart showing the sequence of operations required for calculating the gradient $\nabla F_1(\underline{\boldsymbol{\theta}})$. Note that *multiply* indicates element-by-element multiplication of 2-D arrays.

where s is the step-size and $\underline{\mathbf{d}}^n = [\underline{\mathbf{d}}^{\mathbf{i}}]^n$ (the gradient vector $\underline{\mathbf{d}}^{\mathbf{i}}$ calculated at the n^{th} iteration). We are also aware that the OPC solution is typically a perturbation of the prescribed pattern. Therefore, the steepest descent algorithm can be initialized using the prescribed pattern, thereby leading to quick convergence. Thus, we can let $\underline{\boldsymbol{\theta}}^0 = \cos^{-1}(\mathbf{2}\underline{\mathbf{z}}^* - \underline{\mathbf{1}})$. Finally, after completing the iterations, the optimized pattern $\underline{\widehat{\mathbf{m}}}$ can be obtained from $\underline{\widehat{\boldsymbol{\theta}}}$ using (3.1).

For a coherent imaging system, recall that our goal is to solve the optimization problem defined in (2.12) subject to the constraints in (2.17). We follow the same parametric transformation approach discussed earlier to reduce it to an unconstrained optimization problem. The re-parameterized cost function for the case of coherent

imaging systems is given as,

$$F_1^c(\underline{\boldsymbol{\theta}}) = \sum_{k=1}^{MN} \left(z_k^* - \frac{1}{1+\exp\left[-a\left(\sum_{j=1}^{MN} h_{kj}\frac{1+\cos(\theta_j)}{2}\right)^2 + at_r\right]} \right)^2 \tag{3.5}$$

and the gradient vector in this case is given as,

$$\nabla F_1^c(\underline{\boldsymbol{\theta}}) = \underline{\mathbf{d}}^{\mathbf{c}} = a(\mathbf{H}^{\mathbf{T}}\ [(\underline{\mathbf{z}}^* - \underline{\mathbf{z}}^{\mathbf{c}}) \odot \underline{\mathbf{z}}^{\mathbf{c}} \odot (\underline{\mathbf{1}} - \underline{\mathbf{z}}^{\mathbf{c}}) \odot (\mathbf{Hm})]\ ) \odot\ \sin(\underline{\boldsymbol{\theta}}). \tag{3.6}$$

Once again we observe that the gradient calculation involves two convolution operations leading to a computational complexity of $O(MN\log(MN))$. However, the cost-function defined in (3.5) is a quartic function and is non-convex with multiple local-minima (as also noted by Granik in [31]). Since we are using a local gradient-based search technique; there is no guarantee of reaching the global minimum. However, ILT is an ill-posed problem and it is often not necessary to arrive at the global optima (see [63]). Any *good local minimum* (where the *goodness* is defined using data-fidelity and user-desired properties), can suffice as an acceptable solution. In Section 3.3, we introduce the regularization framework to incorporate the above requirement.

We note that the pattern $\underline{\widehat{\mathbf{m}}}$ obtained using the above method is not binary. Instead each pixel can have gray values anywhere in [0,1]. This makes the resulting mask practically unrealizable and hence we need a post-processing step to obtain the synthesized *binary* OPC mask $\underline{\widehat{\mathbf{m}}}_b$. The simplest way to obtain $\underline{\widehat{\mathbf{m}}}_b$ from $\underline{\widehat{\mathbf{m}}}$ is using a global threshold parameter t_m, such that the error between $\underline{\mathbf{z}}^*$ and the output binary pattern obtained using $\underline{\widehat{\mathbf{m}}}_b$ (as the input) is minimized. The optimum value of t_m can

be obtained using a simple line search operation [5]. However, the above approach for mask binarization is sub-optimal and in Section 3.3 we propose an alternate approach.

3.2 Preliminary Results for Incoherent Imaging with Gaussian PSF

We now demonstrate the working of our algorithm using the simple case of an incoherent imaging system with a Gaussian PSF. We also compare our results with those obtained by directly solving the constrained optimization problem.

Unconstrained Optimization:

The bottom row in Fig. 3.2 illustrates the binary output patterns (after resist thresholding) obtained using the (left) original pattern (i.e. $\underline{\mathbf{m}} = \underline{\mathbf{z}}^*$), (center) the optimized gray pattern ($\underline{\mathbf{m}} = \underline{\widehat{\mathbf{m}}}$) and (right) the binary mask pattern ($\underline{\mathbf{m}} = \underline{\widehat{\mathbf{m}}}_b$) as inputs. The estimated mask $\underline{\widehat{\mathbf{m}}}$ was calculated using the unconstrained gradient descent iterations described in (3.3) and (3.4).

The corresponding aerial images (which form an intermediate step in the process flow) are illustrated in the middle row. The PSF effect was simulated using a Gaussian filter of size 15×15 with $\sigma = 5$ pixels, with the sigmoid parameters $a = 90$, $t_r = 0.5$ and the gradient descent iterations were carried out with step size $s = 0.4$. The final output binary patterns were obtained by thresholding the aerial images using the heaviside operator given in (2.3) with $t_r = 0.5$. Fig. 3.3 illustrates the cost function behavior for the first 200 iterations of steepest descent, indicating quick convergence.

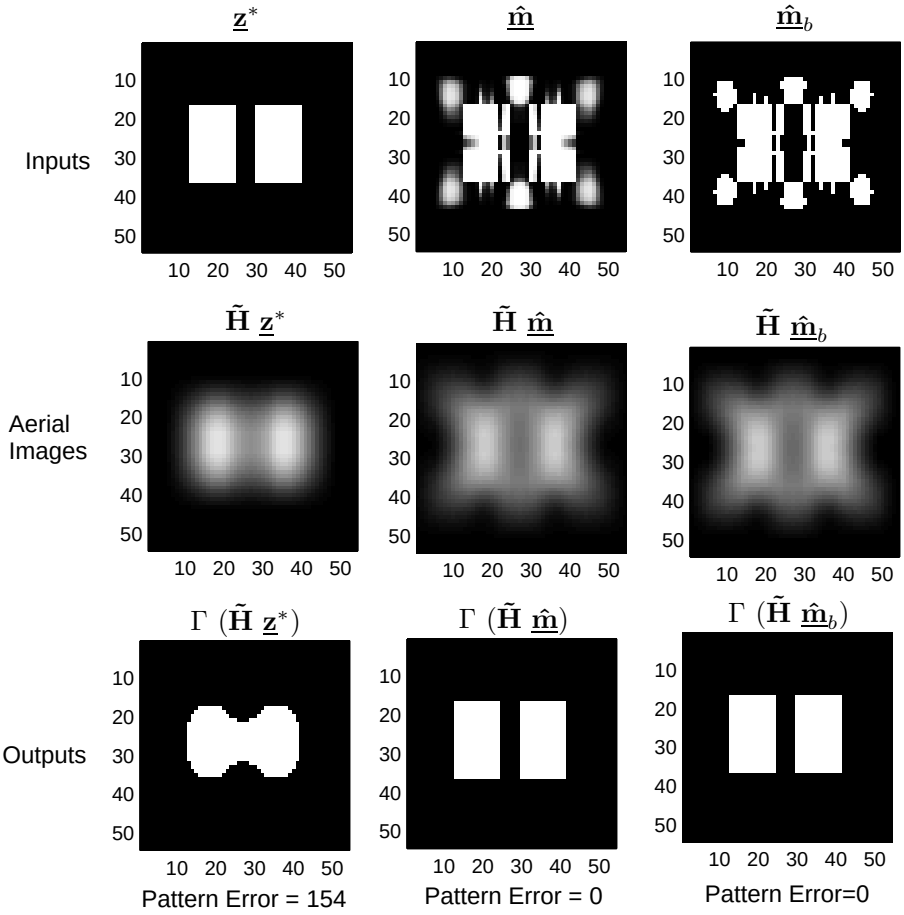

Figure 3.2: Top row (input masks), left to right: original pattern, optimized gray pattern, and binary pattern obtained using an optimum global threshold t_m. The middle and bottom rows indicate the corresponding aerial images and binary output patterns. The parameters are $a = 90,\ t_r = 0.5$, 15×15 filter with $\sigma = 5$ pixels, $s = 0.4$, and number of iterations = 200.

The algorithm was coded using Matlab and the execution time was 7 seconds on a 1.4GHz Pentium-M processor.

The cost $F^i(\underline{\mathbf{m}})$ associated with the above three input patterns can be calculated using (2.15); where for this example, $F^i(\underline{\mathbf{z}}^*) = 142.76$, $F^i(\widehat{\underline{\mathbf{m}}}) = 4.09$, and $F^i(\widehat{\underline{\mathbf{m}}}_b) = 8.46$. However, for the given imaging system (see Fig. 2.2), the aerial image is actually subjected to the Heaviside operation to obtain the binary output image.

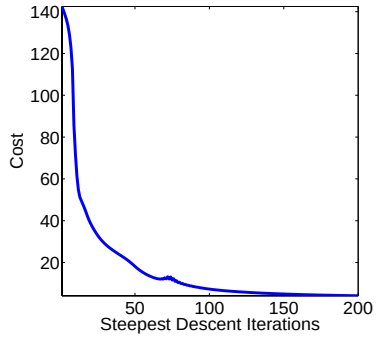

Figure 3.3: Cost function (calculated using (2.15)) versus steepest-descent iterations for the pattern in Fig. 3.2

Hence, it is more appropriate to express the error using the binary output image instead of the sigmoidal transformed (gray level) output image. We employ a metric we call *pattern error*, defined as the total number of pixels which are not faithfully reproduced in the binary output pattern, [1]

$$\text{error}(\underline{\mathbf{m}}) = ||\underline{\mathbf{z}}^* - \Gamma(\tilde{\mathbf{H}}\underline{\mathbf{m}})||_1, \tag{3.7}$$

where $\Gamma(\cdot)$ is the heaviside operator defined in (2.3). From Fig. 3.2, we observe that $\text{error}(\underline{\mathbf{m}}) = 154$, $\text{error}(\widehat{\underline{\mathbf{m}}}) = 0$, and $\text{error}(\widehat{\underline{\mathbf{m}}}_b) = 0$.

Note that the imaging system was incapable of rendering the two bars distinguishable if the desired image is itself used as the input mask (i.e. $\underline{\mathbf{m}} = \underline{\mathbf{z}}^*$). However, the patterns are reproduced very faithfully using the synthesized OPC mask. The results also indicate a perfect corner reproduction which is physically unrealizable. The perfect resist reproduction appears because of the coarse sampling of the underlying pattern coupled with the hard thresholding operation. The current pixel size is too

[1] Although we define (3.7) using L_1 norm, note that for binary patterns, L_0, L_1 and L_2 norms are all the same.

big to capture the rounding at the corners. However, if we choose a smaller sampling interval, the corners will not be *perfectly* reproduced.

Constrained Optimization:

Fig. 3.4 illustrates the input and output patterns obtained by directly solving the bound-constrained optimization problem in (2.15) instead of performing the parametric transformation. We employed the *gradient projection algorithm* described in [41] and utilized the Matlab code provided by the author for this experiment. The experimental parameters are exactly the same as Fig. 3.2, and 200 iterations were performed. We observe that the optimized mask $\underline{\widehat{\mathbf{m}}}$ is structurally very similar albeit much more gray compared to one obtained in Fig. 3.2. As a result, the *pattern error* increases from 4 to 40 after the mask binarization step rendering it ineffective. A similar behavior was observed even for other test patterns. The above comparison illustrates that the parametric transformation technique also has an added indirect advantage; the unconstrained optimization routine tends to produce near-binary results, thereby easing the post-processing binarization step. We attribute this to the cosine term in (3.1), which results in a non-linear mapping onto $[0, 1]$ while favoring values closer to the extremities. Furthermore, the execution time was 14 seconds on a 1.4GHz Pentium-M processor (which is double the amount of time required for the unconstrained case). Note that the above example is only for illustration purposes and all the remaining simulations in this article are carried out using unconstrained optimization.

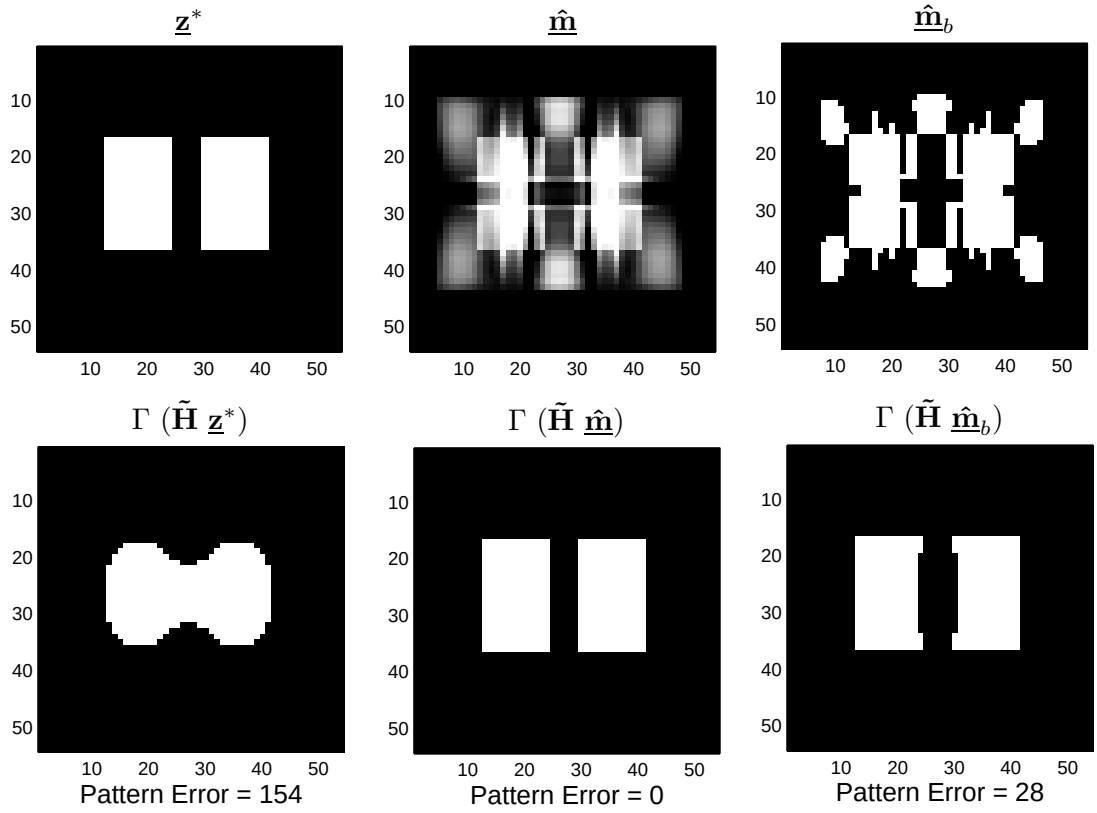

Figure 3.4: Top row, left to right: original pattern, optimized gray pattern using the constrained optimization routine from [41], and the binary pattern obtained using an optimum global threshold t_m. The bottom row indicates the corresponding binary output patterns. The parameters are $a = 90,\ t_r = 0.5$, 15×15 filter with $\sigma = 5$ pixels.

Shortcoming of Mask Binarization:

The mask binarization step in Fig. 3.2 did not cause any increase in the pattern error. However, this may not be necessarily true in general. Fig. 3.5 illustrates the input and output patterns for a more complicated circuit pattern. The Gaussian filter employed has size 11×11 with $\sigma = 14$ pixels (the high value of σ gives the effect of a severe averaging filter), while the sigmoid parameters are $a = 80$, $t_r = 0.5$, and $s = 0.5$. We observe that our algorithm performs an excellent job of pre-distortion

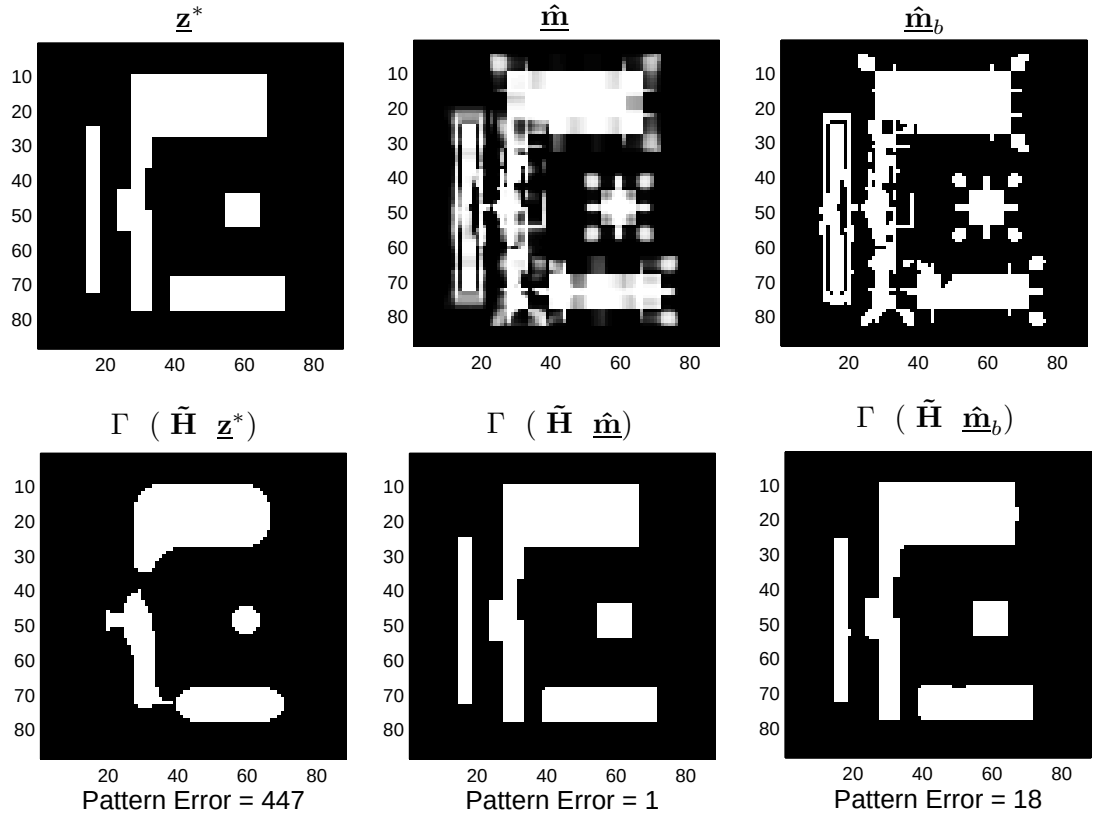

Figure 3.5: Top row, left to right: original pattern, optimized gray pattern, and binary pattern obtained using an optimum global threshold t_m. The bottom row indicates the corresponding binary output patterns. The parameters are $a = 80,\ t_r = 0.5$, 11×11 filter with $\sigma = 14$ pixels, $s = 0.5$, and number of iterations = 200.

and the output is extremely faithful to the desired pattern (error($\underline{\mathbf{m}}$) = 447 versus error($\underline{\widehat{\mathbf{m}}}$) = 1). Fig. 3.6 illustrates the cost function behavior for the first 200 iterations of the steepest-descent procedure. Note that the long vertical bar (on the extreme left) in the prescribed pattern in Fig. 3.5 is completely missing when reproduced using $\underline{\mathbf{m}} = \underline{\mathbf{z}}^*$. However, our optimization procedure starts adding prewarping elements to correct it at around the 85^{th} iteration, which results in a steep decrease in the cost function observed in Fig. 3.6. The kinks in the curve are due to the large step size

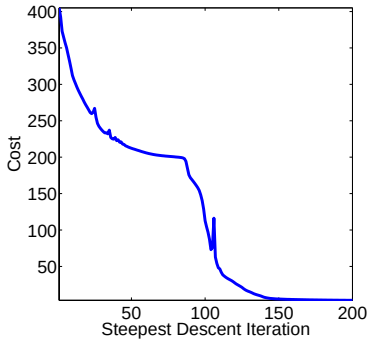

Figure 3.6: Cost function versus steepest-descent iterations for Fig. 3.5.

and can be overcome by choosing a smaller s.

Fig. 3.5 also demonstrates the performance of the estimated binary mask obtained using the mask binarization process. We observe that in this case the pattern fidelity of the binary mask is poorer than the gray level mask (error($\widehat{\underline{\mathbf{m}}}$) = 1 versus error($\widehat{\underline{\mathbf{m}}}_b$) = 18). This is not surprising because the imaging system outlined in Fig. 2.2 is a binary-in-binary-out (BIBO) system. The binary output part is modeled by using the sigmoid based function (see Fig. 2.3) which guarantees that the output pattern is always close to binary. However, our current setup does not incorporate the fact that the estimated input pattern should also be binary. The disadvantage is the need for an extra post-processing step (mask binarization) which is sub-optimal with no guarantee that the *pattern error* will be under control (as observed in Fig. 3.5). There is also a possibility to jointly optimize the binarizing threshold and the gray pattern. In the next section, we use the regularization framework to overcome these problems and complete the binary-to-binary loop.

3.3 Regularization Framework

As we noted earlier, the inverse lithography problem is an ill-posed problem [90]. The continuous function formulation implies that there can be infinitely many input (gray level) patterns all giving rise to the same binary pattern at the output. Similarly, there may be many discrete-tone masks all capable of providing good contour fidelity for a given pattern. Since multiple solutions exist, our goal is to choose a solution which is more favorable to us compared to others. For example, continuous tone masks are physically not realizable and we are only interested in solutions comprising of two or three tones. Furthermore, we want the synthesized mask patterns to have low-complexity in order to control the mask manufacturing costs. The user may also have other requirements like low MEEF (mask error enhancement factor), large process-window, minimum feature spacing, etc [63]. In general, we may want to inculcate (or promote) certain desirable properties in the solution. The regularization framework [90] incorporates these requirements as *prior* information about the solution, and helps us arrive at the *preferred* solution. The "regularized" problem formulation can be described as follows,

$$\widehat{\underline{\mathbf{m}}} = \arg\min_{\underline{\mathbf{m}}} \left[\gamma_{fid} F(\underline{\mathbf{m}}) + \gamma_{reg}\ R(\underline{\mathbf{m}})\right] \tag{3.8}$$

where $F(\underline{\mathbf{m}})$ is the data-fidelity term, and $R(\underline{\mathbf{m}})$ is the regularization function (or the penalty term) used to direct the unknown parameter $\underline{\mathbf{m}}$ towards the desired solution space. γ_{fid} and γ_{reg} are user-defined scalars for adequately weighing the first (data fidelity) term against the second (regularization) term. The prior knowledge is con-

tained in the penalty term $R(\underline{\mathbf{m}})$ and solutions in closer agreement with the prior are penalized less compared to others.

The regularization framework was first employed in the context of lithography by Peckerar, et al. [68, 67] to solve the proximity effect problem arising in e-beam lithography. The estimated dose was obtained by solving an unconstrained continuous function optimization problem using gradient-descent. This led to impractical negative dose values which was overcome by employing a regularization framework similar to (3.8). In what follows next, we discuss three regularization terms targeted to achieve three different goals.

3.3.1 Discretization Penalty Term

The first penalty term employed to obtain near binary patterns is as follows,

$$R_{dis}(\underline{\mathbf{m}}) = \sum_{j=1}^{MN} \left[1 - (2m_j - 1)^2\right] = 4\underline{\mathbf{m}}^T(\underline{\mathbf{1}} - \underline{\mathbf{m}}), \tag{3.9}$$

where $\underline{\mathbf{1}} = [1, \ldots, 1]^T \in \Re^{MN}$. Thus every pixel m_j has an associated penalty given by the quadratic function (see Fig. 3.7),

$$r(m_j) = 1 - (2m_j - 1)^2.$$

As described earlier, the mask transmission values will be constrained to lie in $[0, 1]$ and hence we are only interested in the behavior of the cost function in that range. The penalty incurred is zero for transmission values 0 or 1 and increases as we move away from the binary region in either direction (maximum at $m_j = 0.5$).

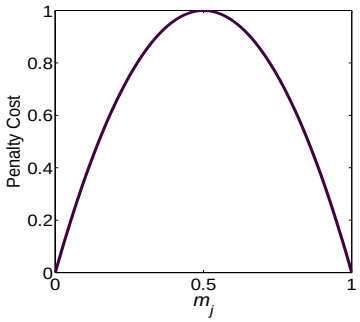

Figure 3.7: Quadratic penalty cost function $R_{dis}(\underline{\mathbf{m}})$

Therefore, the algorithm now favors pixel values closer to 0 and 1 and penalizes pixel values around 0.5, which eases the mask binarization process.

The gradient $\nabla R_{dis}(\underline{\mathbf{m}}) \in \Re^{MN\times 1}$ of the quadratic penalty term is given by,

$$\nabla R_{dis}(\underline{\mathbf{m}}) = (-8\underline{\mathbf{m}} + 4), \tag{3.10}$$

which can be used in conjunction with (3.3) and (3.4) while carrying out the steepest-descent iterations as before.

Fig. 3.8 illustrates the results with the same prescribed pattern, PSF and sigmoidal parameters used in Fig. 3.5 (Section 3.2). However, here we employ the quadratic regularization term and minimize the overall cost function

$$J(\underline{\mathbf{m}}) = F(\underline{\mathbf{m}}) + \gamma_{dis} R_{dis}(\underline{\mathbf{m}}),$$

where $F(\underline{\mathbf{m}})$ is defined in (2.15) and $R_{dis}(\underline{\mathbf{m}})$ is defined in (3.9), with $\gamma_{dis} = 0.015$. The masks in Fig. 3.8 were obtained using 200 iterations with $s = 1.5$. Note that unlike Fig. 3.5, the optimized gray-pattern is very close to binary. Hence, there is no need for the line search operation discussed in Section 3.1 to find the optimal global threshold

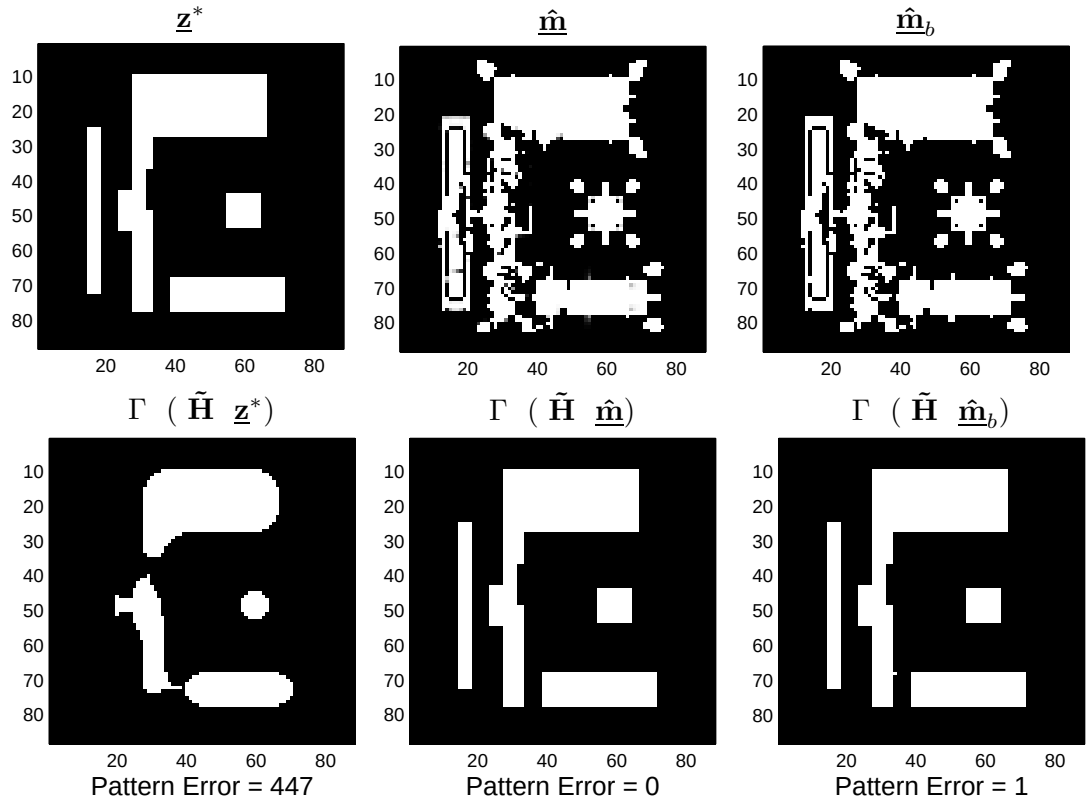

Figure 3.8: Left to right: original pattern, optimized gray pattern, and binary pattern obtained using global thresholding with $t_m = 0.5$. The cost function minimized was $F(\underline{\mathbf{m}})+\gamma_{dis}R_{dis}(\underline{\mathbf{m}})$ and the parameters are $a = 80,\ t_r = 0.5$, 11×11 filter with $\sigma = 14$ pixels, $s = 1.5$, number of iterations $= 200$, and $\gamma_{dis} = 0.015$

parameter t_m. In fact, we can simply obtain the synthesized binary pattern $\widehat{\underline{\mathbf{m}}}_b$ by thresholding $\widehat{\underline{\mathbf{m}}}$ with $t_m = 0.5$. Note that the *pattern error* only increases from 0 to 1 pixel due to the above step. We also mention that the inclusion of the regularization penalty that promotes binary results did not introduce any new error when applied to the pattern in Fig. 3.2.

3.3.2 Complexity Penalty Term

The pixel-based approach allows tremendous flexibility in representing the mask patterns but also suffers the inherent disadvantage that the masks are rather complex and hence difficult to manufacture and inspect. Liu and Zakhor addressed this issue in the past using a cell-based approach [48]. The cells are selected and moved around either randomly or using the knowledge from previous moves. Researchers in the past have also reverted to post-processing operations to simplify the output [59], but this approach is sub-optimal. We follow the regularization framework and employ a penalty function to direct our algorithm towards generating low-complexity masks. Isolated perturbations, protrusions, etc are not preferred because they increase the storage and manufacturing cost. Hence we seek a penalty term which suppresses these effects. Thus we integrate a mask simplicity criterion into the optimization objective and inherently favor low-complexity masks while exploring the search space.

To achieve this, let us first define an auxiliary variable called *the activation pattern* $\underline{\mathbf{f}}$ where,

$$f_j = |m_j - z_j^*| \quad \text{for} \quad j = 1, \ldots, MN.$$

The *on* pixels in $\underline{\mathbf{f}}$ indicate the positions where pre-warping occurred; so the pre-warped pattern can be obtained by simply flipping the corresponding pixels in $\underline{\mathbf{z}}^*$ from 1 to 0 or 0 to 1.

There are a variety of penalty terms that one can employ depending upon how one defines *mask complexity*. Akin to the idea of Total Variation (TV) [90] penalty,

we choose to penalize the mask complexity using the local variation of the activation pattern as follows [27]:

$$||\nabla \underline{\mathbf{f}}||_1 = ||\mathbf{Q_x}\underline{\mathbf{f}}||_1 + ||\mathbf{Q_y}\underline{\mathbf{f}}||_1, \tag{3.11}$$

where $\mathbf{Q_x}, \mathbf{Q_y} \in \Re^{MN}$ represent the first (directional) derivatives and are defined as $\mathbf{Q_x} = \mathbf{I} - \mathbf{S_x}$ and $\mathbf{Q_y} = \mathbf{I} - \mathbf{S_y}$ where $\mathbf{S_x}$ and $\mathbf{S_y}$ shift the 2-D mask represented by $\underline{\mathbf{m}}$ along horizontal (right) and vertical (up) direction by one pixel, respectively.

This approach, while relatively simple, enables us to decouple the features of the underlying prescribed pattern $\underline{\mathbf{z}}^*$ from $\underline{\mathbf{m}}$ thereby capturing only the *changes* occurring due to prewarping. Isolated holes, protrusions, and jagged edges have higher associated penalty. The regularization term in (3.11) suppresses these effects and forces the *changes* to be spatially smoother and less abrupt. This leads to simple, and easy to manufacture OPC masks.

The gradient $\nabla R_{TV}(\underline{\mathbf{m}}) \in \Re^{MN\times 1}$ of the TV penalty term is given as,

$$\begin{aligned}\nabla R_{TV}(\underline{\mathbf{m}}) = [&\mathbf{Q_x}^T \text{sign}(\mathbf{Q_x}\underline{\mathbf{f}}) \\ &+\mathbf{Q_y}^T \text{sign}(\mathbf{Q_y}\underline{\mathbf{f}})] \ \odot \ \text{sign}(\underline{\mathbf{m}} - \underline{\mathbf{z}}^*),\end{aligned} \tag{3.12}$$

which can be used in conjunction with (3.3) and (3.4) while carrying out the steepest-descent iterations as before.

Fig. 3.9 illustrates the results with the same prescribed pattern, PSF, and sigmoidal parameters used in Fig. 3.2. However, here we employ both the quadratic

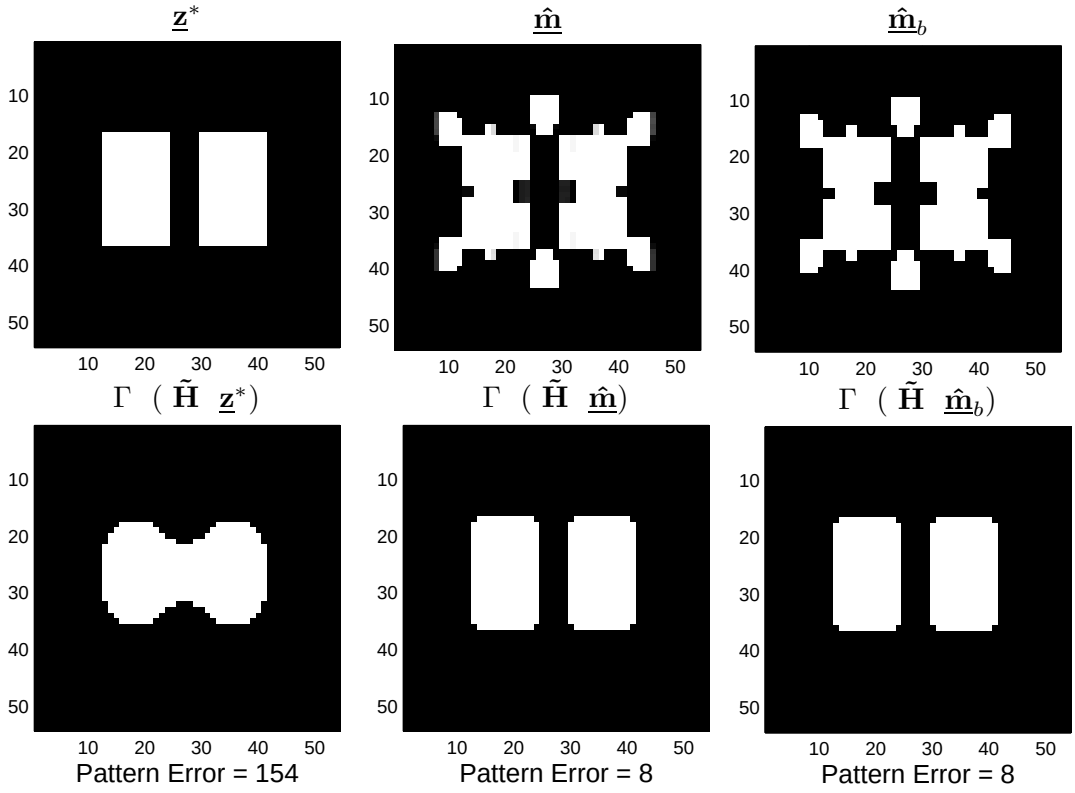

Figure 3.9: Left to right: original pattern, optimized gray pattern, and binary pattern obtained using global thresholding with $t_m = 0.5$. The cost function minimized was $F(\underline{\mathbf{m}})+\gamma_{dis}R_{dis}(\underline{\mathbf{m}})+\gamma_{TV}R_{TV}(\underline{\mathbf{m}})$ and the parameters are $a = 90$, $t_r = 0.5$, 15×15 filter with $\sigma = 5$ pixels, $s = 1$, number of iterations $= 200$, $\gamma_{dis} = 0.025$ and $\gamma_{TV} = 0.045$

and TV regularization terms and minimize the overall cost function

$$J(\underline{\mathbf{m}}) = F(\underline{\mathbf{m}}) + \gamma_{dis}R_{dis}(\underline{\mathbf{m}}) + \gamma_{TV}R_{TV}(\underline{\mathbf{m}}),$$

where $F(\underline{\mathbf{m}})$ is defined in (6.5), $R_{dis}(\underline{\mathbf{m}})$ is defined in (3.9), and $R_{TV}(\underline{\mathbf{m}})$ is defined in (3.11) with $\gamma_{dis} = 0.025$ and $\gamma_{TV} = 0.045$. The masks were obtained using 200 iterations with $s = 1.0$. We observe that the synthesized OPC pattern is much simpler than the one in Fig. 3.2, yet the *pattern error* is still quite small.

Similarly, Fig. 3.10 illustrates the results obtained using both regulariza-

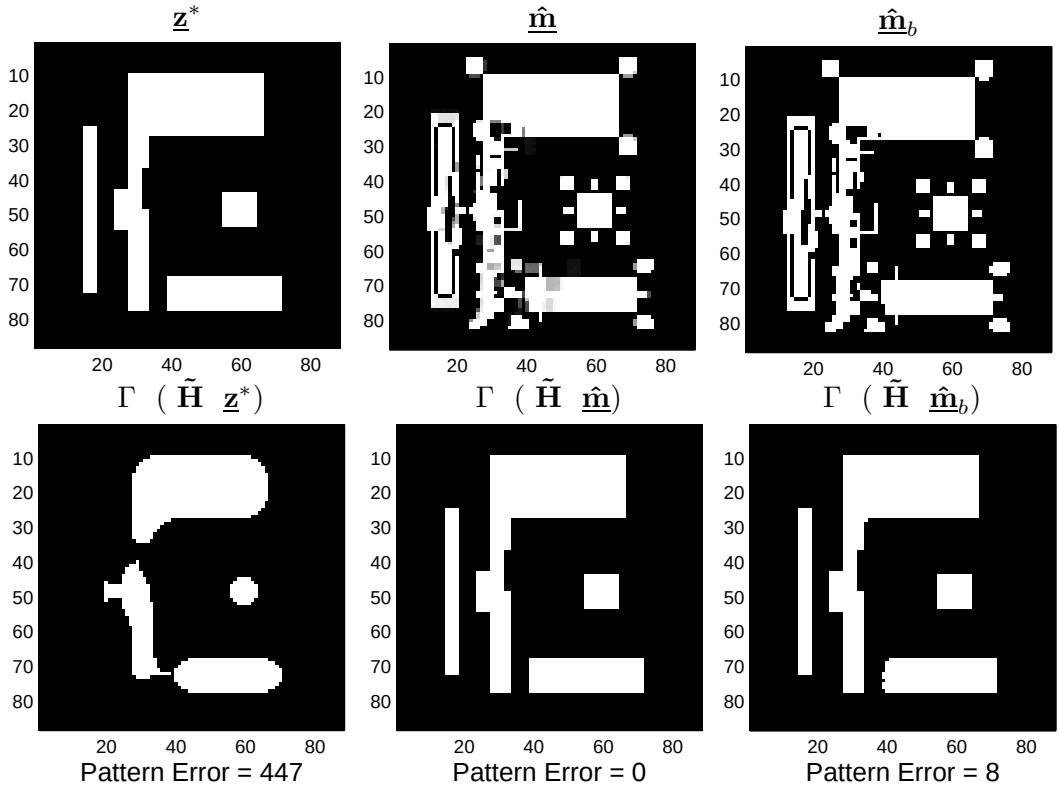

Figure 3.10: Left to right: original pattern, optimized gray pattern, and binary pattern obtained using global thresholding with $t_m = 0.5$. The cost function minimized was $F(\underline{\mathbf{m}})+\gamma_{dis}R_{dis}(\underline{\mathbf{m}})+\gamma_{TV}R_{TV}(\underline{\mathbf{m}})$ and the parameters are $a = 80,\ t_r = 0.5$, 11×11 filter with $\sigma = 14$ pixels, $s = 1$, number of iterations $= 400$, $\gamma_{dis} = 0.01$ and $\gamma_{TV} = 0.02$

tion terms with the same prescribed pattern, PSF, and sigmoidal parameters used in Fig. 3.8 with $\gamma_{dis} = 0.01$ and $\gamma_{TV} = 0.02$. The above result was obtained in 400 iterations with $s = 1$. In both the above cases, we observe that the isolated holes, protrusions, and jagged edges are suppressed and hence the resulting mask is simpler, and therefore easier, and cheaper to fabricate and inspect.

3.3.3 Aerial Image Penalty Term

The lithography process needs to be robust to the process errors introduced by undesirable focus and exposure variations. This can be achieved by availing a good quality aerial image; one with sharp contrast or steep transitions along the desired edge locations. Sub-resolution assist features (SRAF) are small features below the resolution limit of the imaging system, but influence the lithography behavior of the larger features near them. They help improve the contrast of the main feature, but do not themselves print in the final wafer pattern [69].

Interference Map Lithography (IML) is a commonly employed technique to generate SRAF for contact holes [86]. It is based on the principle of Fresnel lens and aims at increasing the peak intensity at the wafer. In the simplest case of binary mask for an isolated contact, this can be achieved by placing clear areas on the mask where the PSF is greater than zero, and by placing opaque areas on the mask where the PSF is less than zero. Similarly for strong phase shift mask, there are 100% transmittance 0 degree areas (+1) where the PSF is greater than zero, and 100% transmittance 180 degree areas (-1) where the PSF is less than zero. The above idea is extended to random

logic contacts by treating each of them as a Dirac delta function, and calculating the electric field on the wafer plane (the interference map) as a linear superposition of the individual contributions of contact holes. The above technique is restricted however, to coherent imaging systems only. SRAF also give rise to side-lobes, which should be below the resist threshold to avoid their printability. IML cannot automatically take the SRAF printability into account. Therefore, there has recently been much interest in ILT-based techniques for SRAF generation geared towards the 65, 45, and 32nm nodes [38, 55].

There can be multiple ways of promoting assist-bars in the solution (as also noted by Granik in [32]). In our study, we employ a penalty term defined as the L_2 norm of the difference between the desired (binary) pattern and the aerial image obtained using the input mask. For a coherent imaging system, the penalty term is given as,

$$R_{aerial}(\underline{\mathbf{m}}) = ||\underline{\mathbf{z}}^* - |\mathbf{H}\underline{\mathbf{m}}|^2||_2^2, \tag{3.13}$$

and the gradient $\nabla R_{aerial}(\underline{\mathbf{m}})$ can be calculated as,

$$\nabla R_{aerial}(\underline{\mathbf{m}}) = -\mathbf{H^T}\Big[(\underline{\mathbf{z}}^* - |\mathbf{H}\underline{\mathbf{m}}|^{\mathbf{2}}) \odot (\mathbf{H}\underline{\mathbf{m}})\Big]. \tag{3.14}$$

It is important to ensure that slight variations in the lithography process conditions do not result in the printing of the SRAF (a common check performed during RET verification). However, the penalty term in (3.13) automatically ensures that the energy in the side-lobes is minimal, thereby also accounting for SRAF printability.

Note that the optimization problem (2.12) can also be formulated using (3.13)

instead of the contours fidelity term. That is, $F(\underline{\mathbf{m}}) = F^c(\underline{\mathbf{m}}) = R_{aerial}(\underline{\mathbf{m}})$. This will guarantee a high quality aerial image. However, the contour fidelity will be very poor. The above approach is useful in case of *isolated* contacts where the dose can be varied to bring the CD on target and hence contour fidelity is not an issue of concern.

The augmented cost function to be minimized is defined as the sum of the data fidelity term and the regularization terms:

$$J(\underline{\mathbf{m}}) = \gamma_{fid} F(\underline{\mathbf{m}}) + \gamma_{aerial} R_{aerial}(\underline{\mathbf{m}}) + \gamma_{dis} R_{dis}(\underline{\mathbf{m}}) + \gamma_{TV} R_{TV}(\underline{\mathbf{m}}) \tag{3.15}$$

where γ_{aerial}, γ_{dis}, and γ_{TV} are the regularization weights corresponding to their respective penalty terms.

3.3.4 An Alternative Two Step Strategy for Optimization

A closer look at the above regularization terms indicates that on some occasions they tend to conflict and suppress certain type of features in the estimated mask. For example, the gradient descent optimization is usually initialized at $\underline{\mathbf{m}}^0 = \underline{\mathbf{z}}^*$. In such cases, the assist bar generation implies a switch from 0 to 1 in regions away from the main feature in the estimated mask. However, in the continuous domain, any movement away from 0 is penalized by the quadratic penalty term $R_{dis}(\underline{\mathbf{m}})$. This tends to counteract the generation of assist-bars. We also observe from our experiments (see Fig. 3.14) that often the best contour fidelity and aerial image contrast is obtained by *breaking* the feature into two disjoint ones. This effect is again suppressed by both the complexity and discretization penalty terms.

Hence, we propose an alternative optimization strategy which involves distributing the optimization objectives into two steps. In what follows next, we discuss a possible strategy for assist-bar generation and strong phase shift masks (PSM) synthesis.

Step 1 - In the first step, data fidelity is defined as the aerial image fidelity and we aim to minimize the cost function,

$$J(\underline{\mathbf{m}}) = R_{aerial}(\underline{\mathbf{m}})$$

Since the only term employed is the aerial image penalty, our goal is to estimate a mask which improves the overall contrast of the aerial image. The steepest descent iterations are initialized as $\underline{\mathbf{m}}^0 = \underline{\mathbf{z}}^*$. Note that since the data fidelity does not account for the contour fidelity, the resulting contours will not be on target yet.

Step 2 - We next minimize the augmented cost function defined in (3.15) which consists of the contour fidelity, complexity, discretization, and aerial image penalty terms. The key point is that the steepest descent algorithm for step two is initialized using the estimated mask pattern obtained from step one. In this case, the phase assignments for strong PSM and generation of assist-bars for binary and AttPSM masks occurs during the first step. In step two, we start exploring the search space from the above solution point and incorporate the other objectives.

There is also a possibility of solving for only the contour fidelity in the first step. The resulting (continuous-tone) mask solution is then subject to the discretiza-

tion and complexity penalty terms in the second step to obtain the final regularized solution.

3.4 More Results

Typical lithography systems employ a circular pupil as discussion in Section 2.2. In this section, we demonstrate results for coherent and incoherent imaging systems using circular and square lens pupils.

3.4.1 Assist Bars for Coherent Imaging

Fig. 3.11 illustrates the estimated masks and the binary output patterns obtained using a coherent imaging system ($\sigma = 0$) with a circular pupil, $\lambda = 193$nm, $NA = 0.85$, and the resist threshold $t_r = 0.3$. The desired pattern consists of 90nm random logic contact holes sampled at 10nm ($k_1 = 0.39$). We follow the two step optimization approach and Fig. 3.11 illustrates the estimated mask obtained at the end of step one. The result was obtained in 60 iterations with s=20 and the run-time was 22 seconds on a 1.4Ghz Pentium-M machine using Matlab. The cost function behavior is illustrated in Fig. 3.12 and indicates quick convergence. We have observed that the algorithm scales the center feature during the initial iterations and adds assist bars during the remaining ones.[2] Note that the assist features are also shared between adjacent contacts and the mask has continuous tone. Furthermore, the output wafer pattern (bottom row) obtained using the above mask does not have the contours on

[2]The above effect of scaling of the main feature was also recently reported in the ILT-based assist bar generation work in [55].

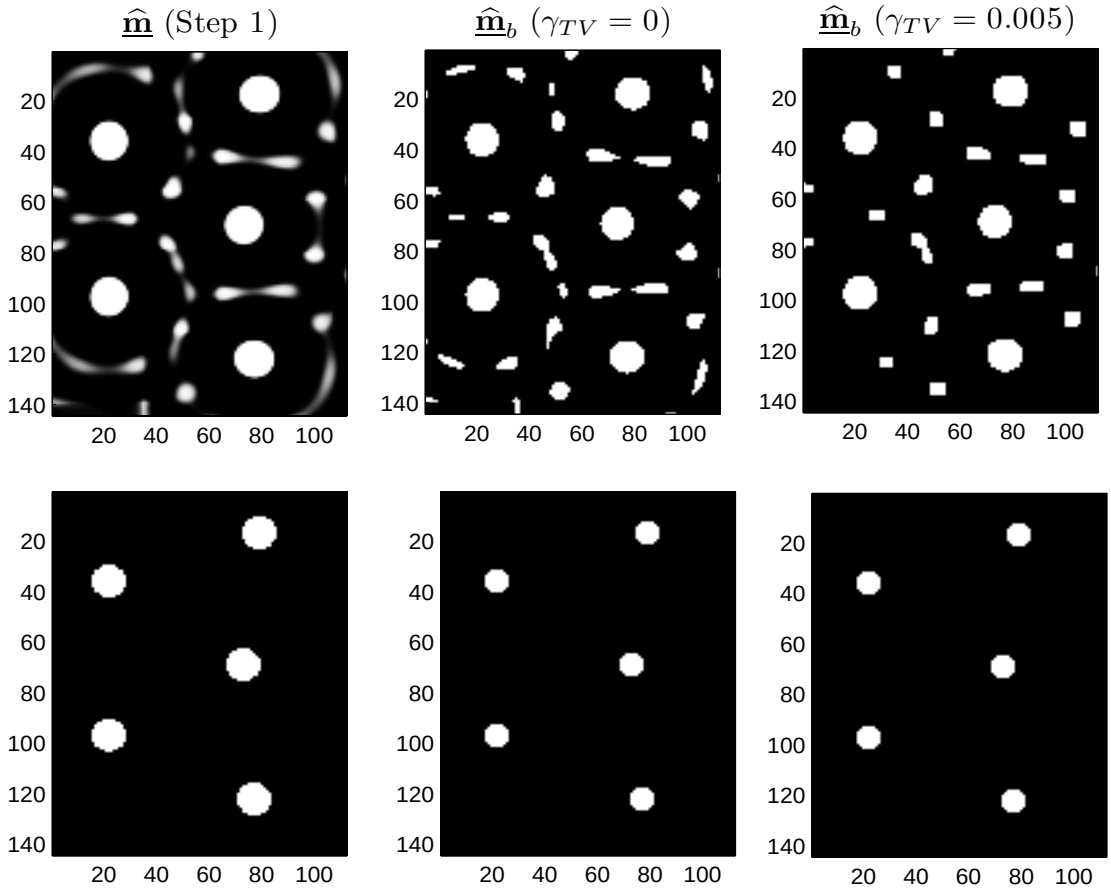

Figure 3.11: The synthesized masks (top row) and the output binary patterns (bottom row) for 90nm random logic contact patterns. $\widehat{\underline{\mathbf{m}}}_b$ (left) is the estimated mask at the end of Step 1 ($\gamma_{aerial} = 1$), $\widehat{\underline{\mathbf{m}}}_b$ (center) is estimated at the end of Step 2 ($\gamma_{fid} = 1$, $\gamma_{aerial} = 0.25$, $\gamma_{dis} = 0.005$, $\gamma_{TV} = 0$) and $\widehat{\underline{\mathbf{m}}}_b$ (right) is estimated using the above parameters with $\gamma_{TV} = 0.005$

target. This is not surprising since the contour fidelity was not considered as an optimization objective.

The estimated mask in the center of Fig. 3.11 illustrates the result at the end of the second step. The parameters are: $\gamma_{fid} = 1$, $\gamma_{aerial} = 0.25$, and $\gamma_{dis} = 0.005$. We observe that the mask is having discrete tone and the output contours for all the contacts are on target. Fig. 3.13 compares the aerial image slices and indicates a

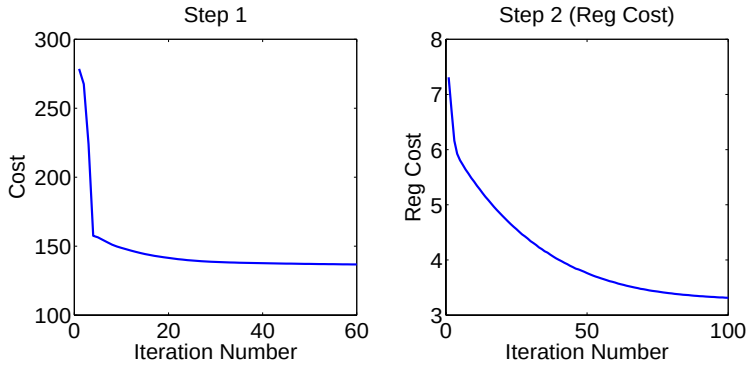

Figure 3.12: Convergence curves for Fig. 3.11. The left curve indicates the cost function behavior for Step 1 and the right curve indicates the regularization cost ($\gamma_{dis} R_{dis}(\underline{\mathbf{m}}) + \gamma_{TV} R_{TV}(\underline{\mathbf{m}})$) versus iteration number for Step 2 of the optimization.

tremendous improvement in the peak intensity. We also observe that the side-lobes are under control and the assist bars do not print. However, the assist features are broken, irregular, curvy, and complex, making the mask very difficult to manufacture. Hence, we repeat the second step by employing the complexity penalty term ($\gamma_{TV} = 0.005$). This leads to a comparatively simple pattern (see $\widehat{\underline{\mathbf{m}}}_b$ on the right in Fig. 3.11) and the assist bars tend to become square or rectangular in shape, which are more preferable. The above result was obtained by performing 100 iterations (see Fig. 3.12) with $s = 8$ and the run-time was 60 seconds.

3.4.2 Incoherent Imaging with Circular Pupil

Fig. 3.14 illustrates the estimated masks and the binary output patterns obtained using an incoherent imaging system with $\sigma = 1$, $\lambda = 193$nm, $NA = 0.95$, and the resist threshold $t_r = 0.5$. The desired pattern is a more complicated pattern sampled at 5nm with features as small as 50nm ($k_1 = 0.25$). The top row indicates

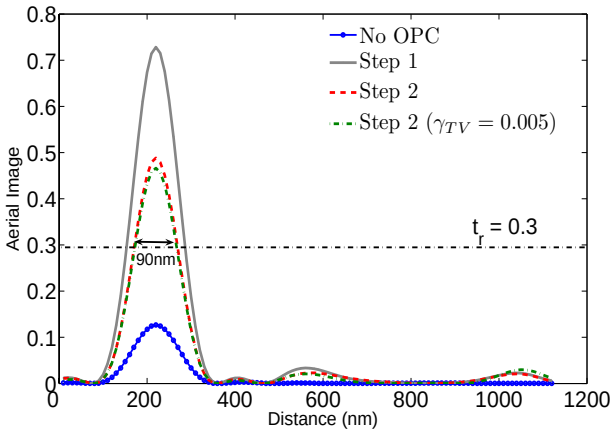

Figure 3.13: Horizontal slice at row number 35 for the aerial image obtained using the synthesized mask in Fig. 3.11. Note that the contrast has improved and the side-lobes will not print.

the input patterns and the bottom row indicates the corresponding output wafer patterns. The center mask was obtained without any regularization with only $\gamma_{fid} = 1$. We observe that although the contours are on target, the mask itself is very choppy and irregular, making it very hard to manufacture. Hence we employ the two step procedure outlined earlier and optimize using the augmented cost function (3.15) with $\gamma_{fid} = 1$, $\gamma_{TV} = 0.1$ and $\gamma_{dis} = 0.01$. This results in an estimated mask which is comparatively much smoother with little loss in performance. It is also interesting to note that the algorithm automatically decided to *break* some features into two disjoint sections.[3] For example, the elbow on the top-left region in Fig. 3.14 is split into two parts in $\widehat{\underline{\mathbf{m}}}_b$, but is reproduced accurately at the output. Such counter-intuitive results are hard to obtain using edge-based parametrization and would require extensive and tedious segmentation scripts. The results were obtained using 100 iterations (runtime

[3]The above effect of *breaking* of features was also observed for partially coherent imaging systems in [45].

55 seconds) for each step.

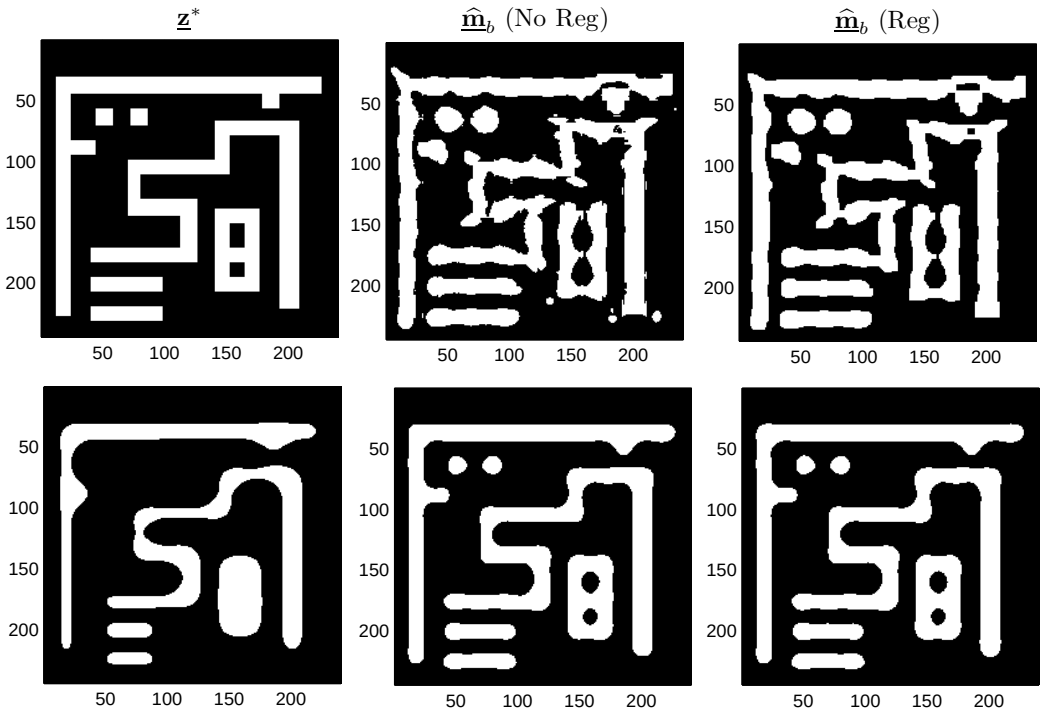

Figure 3.14: The top row consists of the original pattern and the synthesized masks before and after regularization. The bottom row indicates the corresponding output wafer patterns. $\underline{\widehat{\mathbf{m}}}_b$ (center) is estimated using $\gamma_{fid} = 1$ and $\underline{\widehat{\mathbf{m}}}_b$(right) is obtained using $\gamma_{fid} = 1$, $\gamma_{dis} = 0.01$, and $\gamma_{TV} = 0.1$.

3.4.3 Comparison with Combinatorial Optimization (Sherif, et al.)

Here we compare the results obtained using our continuous function optimization method against those obtained by integer optimization. Sherif, et al. [84] employed mixed-linear integer programming to estimate the resist threshold and the OPC mask for an idealized lithography system with a square aperture. Their cost function is very similar to the *pattern error* defined in (3.7) (note that we have used the latter as a metric to evaluate the quality of our results). The authors in [84]

reduce the non-linear cost function to a linear objective function with unconditional constraints by tripling the number of estimated variables. The problem can then be solved using branch and bound method.

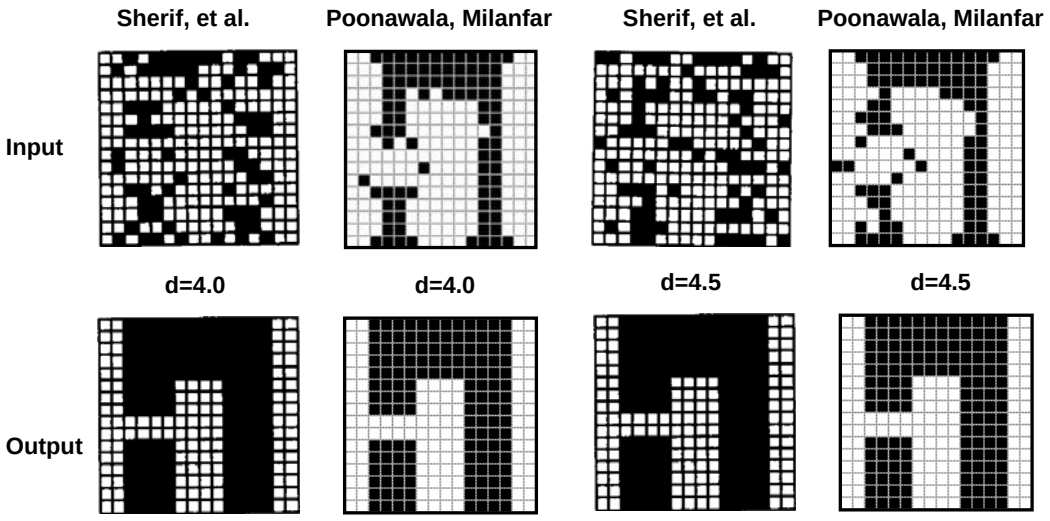

Figure 3.15: The top row indicates the synthesized masks obtained using Sherif, et al. (from [84]) and our algorithm for a square pupil with $d = 4.0$ and $d = 4.5$ pixels. The bottom row indicates the corresponding output wafer patterns assuming a hard-limiter with $t_r = 0.2$.

In order to make a fair comparison, we choose the same imaging system and target pattern ($\underline{\mathbf{z}}^*$) as employed by the authors in [84]. The imaging system considered in [84] is an incoherent imaging system with a square aperture. In this case, the PSF is defined as,

$$h(x,y) = \frac{1}{d^2}\text{sinc}^2(\frac{x}{d})\text{sinc}^2(\frac{y}{d}),$$

where d is the width of the impulse response function in units of pixels. The results for the cases of $d = 4.0$ and $d = 4.5$ are illustrated in Fig. 3.15. The experimental parameters for our algorithm are as follows: $a = 60$, $t_r = 0.2$, and $\gamma_{dis} = 0.03$. We observe that the synthesized masks obtained using our algorithm are different

from those obtained by solving the combinatorial optimization problem. However, the output wafer pattern is reproduced with 100% accuracy proving that our results are correct. It is also worth noting that our masks were synthesized in 5 seconds using a 1.4GHz Pentium-M machine compared to 15-25 minutes reported in [84] (using the computational resources available in 1995).

3.5 Weight Mask and Selected Parameter Optimization

In this section we discuss two additional useful features which can be incorporated in the above framework for solving the OPC mask design problem.

3.5.1 Weight Mask

In many practical scenarios, it is of interest to reproduce certain regions or areas of the pattern more faithfully as compared to others. Consider two parallel wires; the key requirement is that they never fuse on the reproduced pattern, whereas we are much less concerned about their actual reproduced shape. Similarly the mask designer can mark other critical regions depending on the shape, size and vicinity of the neighboring features. The above information can be incorporated into a *weight pattern* which can then be used for selective region correction. The goal is that only the critically marked (high priority) regions on the desired pattern should be faithfully replicated using suitable prewarping. The unimportant (low weight) regions of the underlying pattern may be reproduced with relatively less fidelity. The weight pattern is assigned before running the OPC routine and can be generated heuristically or by

the mask designer.

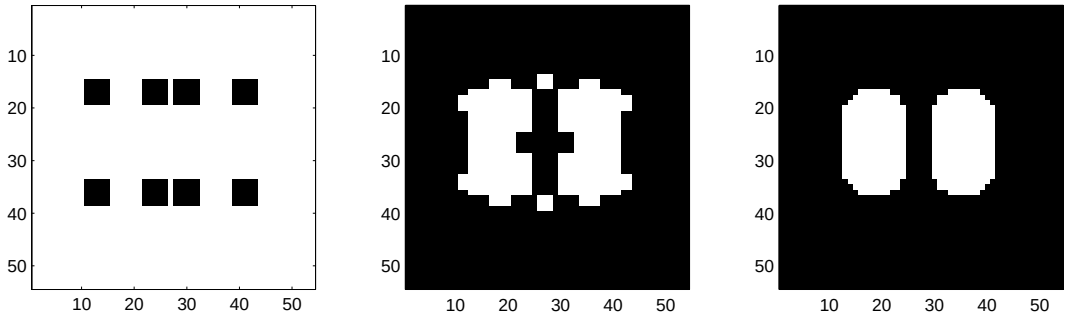

Figure 3.16: The employed weight mask (left), synthesized mask pattern (center), and the final output pattern (right). Note that 5x5 region around the corners have zero weight whereas the weight is one everywhere else.

Let $\underline{\mathbf{w}} \in \Re^{MN\times 1}$ represent the weight mask where each element $w_k \in [0,1]$ signifies the importance (or weight) associated with reproducing that particular pixel at the output. For an incoherent imaging system, we reformulate the optimization problem in (2.15) using the weighted cost function as follows,

$$\widehat{\underline{\mathbf{m}}} = \arg\min_{\underline{\mathbf{m}}} \sum_{k=1}^{MN} w_k \left(z_k^* - z_k^i\right)^2 \tag{3.16}$$

The gradient calculations and steepest-descent iterations can be easily extended using (3.3) and (3.4) as follows,

$$\underline{\boldsymbol{\theta}}^{n+1} = \underline{\boldsymbol{\theta}}^n - s\{a\tilde{\mathbf{H}}^{\mathbf{T}}[\underline{\mathbf{w}} \odot (\underline{\mathbf{z}}^* - \underline{\mathbf{z}}^{\mathbf{i}}) \odot \underline{\mathbf{z}}^{\mathbf{i}} \odot (\underline{\mathbf{1}} - \underline{\mathbf{z}}^{\mathbf{i}})] \odot \ \sin(\underline{\boldsymbol{\theta}}^n)\} \tag{3.17}$$

Fig. 3.16 illustrates the weight mask (left), optimized OPC pattern (center), and the final output pattern (right) using the same parameters as in Fig. 3.2 with $\gamma_{dis} = 0.02$ and $\gamma_{TV} = 0.04$. In this case, we are not concerned about reproducing the corners accurately at the output. However, we want the other features of the pattern

(length, width and separation of the bars) to be accurately reproduced. The above objective can be achieved by incorporating the information from the weight mask in Fig. 3.16. Note that a 5×5 region around every corner has been assigned zero weight, whereas the remaining regions have been assigned a higher weight equal to one. The resulting prewarping is different from that occurring in Fig. 3.2 and all the regions, barring corners, are perfectly reproduced.

3.5.2 Selected Parameter Optimization

All the algorithms discussed up to this point assume the entire synthesized pattern to be completely unknown. The user (mask designer) has no control over which regions of the underlying pattern are allowed to be predistorted. The entire layout space is not always at our disposal and some design rules may require us to force our variations to fall only within a specified area or region of the synthesized mask. In this section we outline an approach which provides the above flexibility. This may also be beneficial in reducing the computational time required to reconfigure OPC solutions in the event of design re-spins, errors, etc. Note that the focus here is on controlling the areas of the input that undergo changes, whereas in Section 3.5.1, the focus was on controlling the areas of the output to be accurately reproduced.

The procedure starts with the mask designer (or the OPC engineer) marking (or selecting) regions (or pixels) in the original pattern where predistortion is permissible. Alternatively, the above procedure can also be automated based on feature type, density, vicinity, etc and incorporated into the CAD software. The above pixels are

stacked together to form the unknown parameter $\underline{\dot{\mathbf{m}}} \in \Re^{P\times 1}$. The remaining pixels will stay unchanged in the synthesized mask and are stacked together to obtain the known vector $\underline{\ddot{\mathbf{m}}} \in \Re^{(MN-P)\times 1}$.

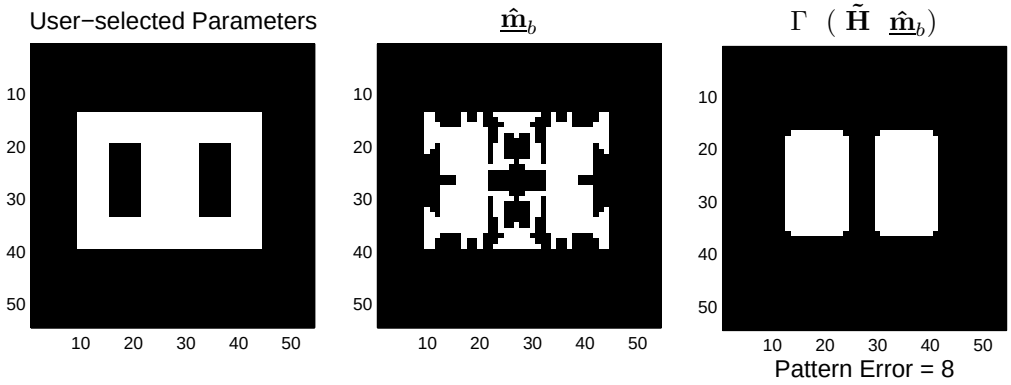

Figure 3.17: Left to Right: User-selected parameters (marked in white), the corresponding optimized mask pattern, and the final output pattern. The parameters are a=90, $t_r = 0.5$, 15x15 filter with $\sigma = 5$, $\lambda_{dis} = 0.01$, and s=1.

Note that an element $\tilde{h}_{kj}$ in matrix $\tilde{\mathbf{H}}$ denotes the contribution of the j^{th} pixel towards the formation of the output z_k. The above matrix can be rewritten as $\tilde{\mathbf{H}} = [\tilde{\mathbf{h}}_\mathbf{1}, \tilde{\mathbf{h}}_\mathbf{2}, \ldots, \tilde{\mathbf{h}}_\mathbf{MN}]$ where $\tilde{\mathbf{h}}_\mathbf{p} \in \Re^{MN\times 1}$ represents the contribution of the p^{th} pixel toward the formation of the output vector $\underline{\mathbf{z}}$. The $\tilde{\mathbf{h}}_\mathbf{p}$'s corresponding to the pixels in $\underline{\dot{\mathbf{m}}}$ are stacked together to form $\dot{\mathbf{H}} \in \Re^{MN\times P}$ and the remaining $\tilde{\mathbf{h}}_\mathbf{p}$'s are stacked to form $\ddot{\mathbf{H}} \in \Re^{MN\times(MN-P)}$. Thus $\tilde{\mathbf{H}}$ is split into two components where $\dot{\mathbf{H}}$ determines the contribution from pixels which are unknown (allowed to be modified) and $\ddot{\mathbf{H}}$ determines the contribution from pixels which are fixed. The forward model in (2.9) can now be rewritten as,

$$\underline{\mathbf{z}}^\mathbf{i} = \text{sig}(\dot{\mathbf{H}}\underline{\dot{\mathbf{m}}} + \ddot{\mathbf{H}}\underline{\ddot{\mathbf{m}}}). \tag{3.18}$$

Note that the second summation term inside the sigmoid is a constant because $\underline{\ddot{\mathbf{m}}}$ remains unchanged. Finally, the steepest descent iterations can be directly used to update the modifiable pixels in $\underline{\dot{\mathbf{m}}}$ as follows,

$$\underline{\dot{\boldsymbol{\theta}}}^{n+1} = \underline{\dot{\boldsymbol{\theta}}}^{n} - s\{a\ (\dot{\mathbf{H}}^{\mathbf{T}}\ [(\underline{\mathbf{z}}^{*} - \underline{\mathbf{z}}^{\mathbf{i}}) \odot \underline{\mathbf{z}}^{\mathbf{i}} \odot (\underline{\mathbf{1}} - \underline{\mathbf{z}}^{\mathbf{i}})]\ ) \odot\ \sin(\underline{\dot{\boldsymbol{\theta}}}^{n})\} \tag{3.19}$$

Fig. 3.17 illustrates a user-defined band (three pixels thick on both side of the edges) indicating the region allowed to undergo predistortion. The user-selected region is marked in white, and the remaining region where predistortion is not permissible is marked in black. The chosen parameters were used to solve for the same experiment and pattern as in Fig. 3.2 albeit with $\gamma_{dis} = 0.01$ and $s = 1$. Fig. 3.17 also illustrates the optimized binary OPC pattern and the final binary output pattern using the above method. As expected, the changes are confined to lie within the band and are very different (far less scattered) compared to the synthesized OPC mask in Fig. 3.2.

In conclusion, we were able to design OPC masks to reduce line-end shortening, corner-rounding, and improve pattern fidelity for both coherent and incoherent imaging systems. Our ILT routine is also capable of automatic generation and placement of assist-features to improve the aerial image contrast. In the next chapter, we extend our ILT framework to synthesize strong and weak phase shift masks.

Chapter 4

Single Exposure PSM Design

We now move our discussion to the design of phase shift masks (PSM) and demonstrate the extension of our framework to the cases of attenuated PSM (AttPSM), 100% transmission PSM, and strong PSM with chrome. The results will indicate that ILT leads to very interesting and un-intuitive mask designs, especially for strong PSM.

4.1 PSM Optimization

An incoherent imaging system does not account for the phase information of the mask. Therefore, in this chapter, we focus exclusively on coherent imaging systems. Recall from (2.12) and (2.13), that the ILT problem for a coherent imaging system consists of minimizing the cost function:

$$F^c(\underline{\mathbf{m}}) = \sum_{k=1}^{MN} \left(z_k^* - \frac{1}{1 + \exp\left[-a \left(\sum_{j=1}^{MN} h_{kj} m_j \right)^2 + a t_r \right]} \right)^2 \tag{4.1}$$

As seen from Section 2.4, the estimated mask pixels are now allowed to have negative values (corresponding to the 180 degree phase shifted regions). In what follows next, we discuss the mask design algorithms for the three commonly employed variants of PSM.

4.1.1 Attenuated PSM

Attenuated phase shift masks (AttPSM) consist of quartz and molybdenum silicide (MoSi) instead of chrome (see Section 1.2.2). MoSi (unlike chrome) allows a small percentage of light intensity (typically 6% or 18%) to pass through it. AttPSM are also referred to as weak phase shift masks in the lithography literature. The thickness of MoSi is chosen such that light which passes through is 180° out of phase compared to the transmitting quartz regions. In our discussion, we focus on the 6% intensity transmission AttPSM.

Note that every pixel m_j can now have only two amplitude transmission values equal to $-\sqrt{0.06} = -0.245$ (the 180 degree phase shift with weak transmission) or $+1$ (100% transmission with no phase shift). Therefore, the optimization problem in (4.1) is subject to the constraints that $m_j = -0.245$ or 1. The above integer optimization problem can be reduced to an unconstrained optimization problem using

the same procedure outlined in Section 3.1 (for OPC masks). Therefore, the first step is to translate the integer constraints into inequality constraints $-0.245 \leq m_j \leq 1$, thereby reducing it to a bound constrained optimization problem. The parametric transformation should now map the unconstrained variable θ_j to the range [-0.245, 1] and is given as,

$$m_j = 0.6225(1 + \cos(\theta_j)) - 0.245 \tag{4.2}$$

The (re-parameterized) unconstrained optimization objective is now given as

$$F_1^c(\underline{\boldsymbol{\theta}}) = \sum_{k=1}^{MN} \left(z_k^* - \frac{1}{1 + \exp\left[-a \left(\sum_{j=1}^{MN} h_{kj} \left\{ 0.6225(1 + \cos(\theta_j)) - 0.245 \right\} \right)^2 + at_r \right]} \right)^2 \tag{4.3}$$

The above objective can be minimized using steepest descent iterations similar to (3.4) to obtain the optimal mask solution. The above solution will of course have continuous transmission values between -0.245 and 1, and hence a post-processing operation will be required to obtain the two-tone 6% AttPSM masks.

Alternatively, we can also employ the regularization framework introduced in Section 3.3, with the goal of obtaining easy to discretize AttPSM mask solutions. This can be achieved by employing a quadratic penalty term, where now the pixels having values -0.245 and 1 have zero penalties, and the cost increases as we move away from them in either direction (also see Fig. 4.1). Therefore, we can choose

$$r(m_j) = -m_j^2 + 0.755m_j + 0.245, \tag{4.4}$$

and the penalty is highest at the center of the range ($m_j = 0.3775$). Finally, the two-tone AttPSM $\widehat{\underline{\mathbf{m}}}_b$ is simply obtained by thresholding the estimated mask $\widehat{\underline{\mathbf{m}}}$ with $t_m = 0.3775$.

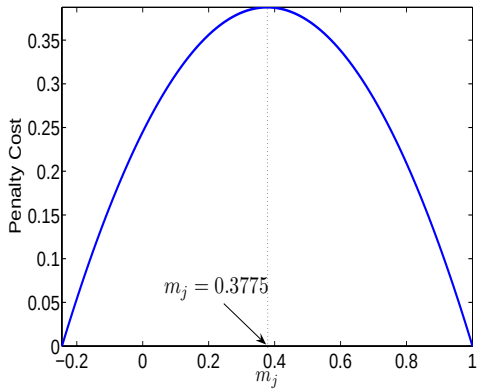

Figure 4.1: Discretization penalty term for AttPSM (maximum penalty is at $m_j = 0.3775$)

There is also the possibility of using the complexity and aerial image penalty terms, and the two step optimization procedure outlined in Section 3.3. Finally, the steepest descent iterations are initialized using the target pattern $\underline{\mathbf{z}}^*$, such that the zero regions are substituted by the phase-shifted MoSi regions. Thus, $(z_i^* = 0) \rightarrow (m_j^0 = -0.245)$ and $(z_i^* = 1) \rightarrow (m_j^0 = 1)$ for $j = 1, \ldots, MN$.

4.1.2 100% transmission PSM

As the name suggests, 100% transmission PSM does not use the opaque chrome features at all, and is an extreme case of AttPSM. It is an all transmissive mask consisting of only zero and 180 degree phase shift features. It is also referred to as strong phase shift mask or chromeless phase lithography (CPL) in the lithography literature (see Fig. 1.9).

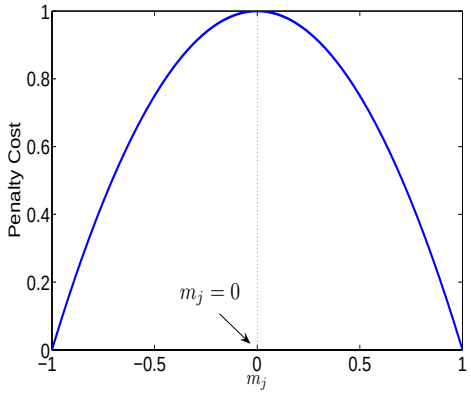

Figure 4.2: Discretization term for 100% transmissive PSM (maximum penalty is at $m_j = 0$)

Here every pixel m_j can only have values -1 (fully transmissive with 180 degree phase shift) or +1 (fully transmissive with 0 degree phase shift). The optimization problem in (4.1) is therefore subject to the integer constraints that $m_j = -1$ or 1. Once again we relax the condition and impose the bound constraints $-1 \leq m_j \leq 1$, and perform the parametric transformation $m_j = \cos(\theta_j)$ to reduce the problem to an unconstrained optimization problem. The quadratic penalty term (used for regularization) now has zero penalty at -1 and 1, and maximum penalty for $m_j = 0$ (see Fig. 4.2),

$$r(m_j) = -m_j^2 + 1. \tag{4.5}$$

Hence, the transmission values are pushed towards -1 and 1 thereby easing the discretization step. The two tone 100% transmission mask $\widehat{\underline{\mathbf{m}}}_b$ is obtained by thresholding the estimated mask $\widehat{\underline{\mathbf{m}}}$ with $t_m = 0$.

It is important to note that in case of strong PSM, the estimated mask seldom resembles the target (see Fig. 4.5, 4.8, and 4.11). Hence, the steepest descent algorithm

was initialized to all zeros, thereby allowing the algorithm to automatically perform the phase assignments for all the regions.

4.1.3 Strong PSM with Chrome

The final case we consider is that of strong PSM where the mask features can have values 0 (chrome) or 1 (quartz with no phase shift) or -1 (quartz etched to provide 180 degree phase shift). The bound constraints and parametric transformation are similar to those employed for 100% transmission PSM. The only difference occurs in the discretization regularization term since we now want a three-tone mask. There

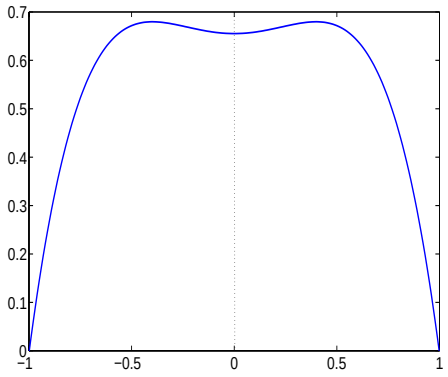

Figure 4.3: Discretization penalty cost for strong PSM (with chrome). The minima are at -1, 0, and 1

can be different ways of formulating the above regularization term. One possible approach is to divide the interval [-1 1] into three equal parts and employ a fourth order quartic penalty term where each pixel has an associated cost,

$$r(m_j) = -0.967m_j^4 + 0.307m_j^2 + 0.655 \tag{4.6}$$

Fig. 4.3 is the plot of the above function. The above curve was obtained by fitting a fourth order polynomial to an over-determined linear system. The latter was formulated to obtain stationary points at $m_j = -0.33,\ 0,$ and 0.33, zero penalty at $m_j = -1,\ 0,$ and 1, and high penalty at $m_j = -0.33$ and 0.33. The required three-tone mask $\widehat{\underline{\mathbf{m}}}_b$ can be finally obtained by quantizing $\widehat{m}_j \in [-1, -0.33)$, $\widehat{m}_j \in [-0.33, +0.33)$, and $\widehat{m}_j \in [+0.33, 1]$ to -1, 0, and 1 respectively.

4.2 Results

We now demonstrate some results obtained using the above framework for both weak and strong PSM.

4.2.1 2-Bar Separation

Experiment 1:

In the first experiment, our goal is to print two 120nm thick bars separated by 50nm ($k_1 = 0.22$) with high contour fidelity. Fig. 4.4 illustrates the synthesized 6% attenuated PSM for a coherent imaging system with $\sigma = 0$, $\lambda = 193$nm and $NA = 0.85$. We employ the two-step optimization strategy outlined in Section 3.3.4 and the experimental parameters are as follows: $a = 25$, $t_r = 0.3$, $\gamma_{fid} = 1$, $\gamma_{aerial} = 0.25$, $\gamma_{TV} = 0.01$, $\gamma_{dis} = 0.01$, and $s = 5$. The number of iterations for the two steps were 80 and 160, and the run-times were 30 and 60 seconds respectively. We observe that the optimization algorithm automatically adds assist bars in all four directions. The width of the assist bars and their placements from the center feature are calculated

as part of the optimization procedure. Fig. 4.7 illustrates the central horizontal slice of the aerial image. If the desired binary pattern is itself fed as the input to the imaging system, the aerial image barely has any modulation and the two bars are not distinguishable. However, the synthesized AttPSM mask causes good modulation and the assist features do not print.

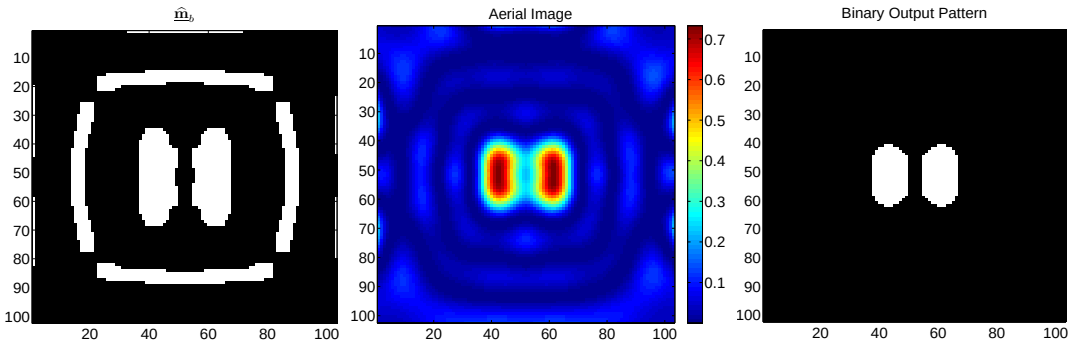

Figure 4.4: The estimated 6% AttPSM mask (left), the corresponding aerial image (center), and the final binary output pattern (right) for coherent imaging system with $\lambda = 193$nm and $NA = 0.85$. The black and white regions in $\underline{\widehat{\mathbf{m}}}_b$ correspond to -0.245 and 1 respectively.

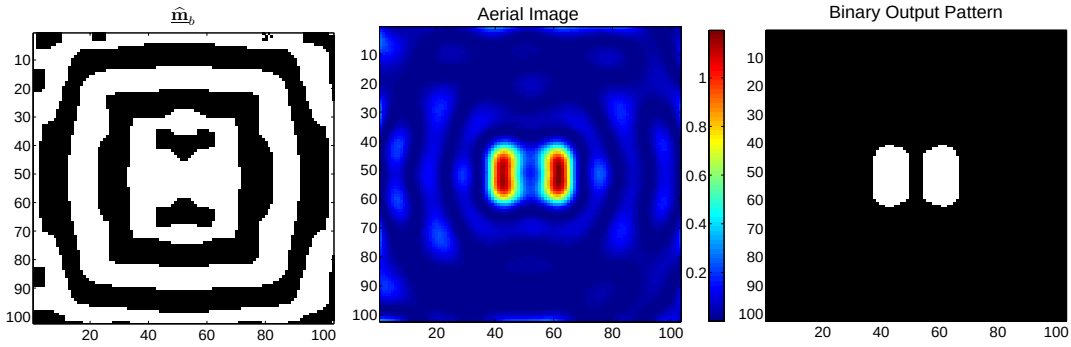

Figure 4.5: The estimated 100% transmission PSM (left), the corresponding aerial image (center), and the final binary output pattern (right) for coherent imaging system with $\lambda = 193$nm and $NA = 0.85$. The black and white regions in $\underline{\widehat{\mathbf{m}}}_b$ correspond to -1 and 1 respectively.

Fig. 4.5 illustrates the synthesized 100% transmission PSM for the same

imaging system. The experimental parameters are as follows: $a = 25$, $t_r = 0.3$, $\gamma_{fid} = 0.75$, $\gamma_{aerial} = 0.25$, $\gamma_{TV} = 0.01$, $\gamma_{dis} = 0.01$, and $s = 5$. The number of iterations for the two steps were 120 and 240, and the run-times were 45 and 90 seconds respectively. Note that our goal is to create a horizontal separation between the two bars. However, the corresponding region in the synthesized mask is a zero phase shift fully transmissive (white) contiguous region. The destructive interference is actually created by the two vertically separated 180 degree phase shift (black) features giving a sharp contrast aerial image as observed in Fig. 4.7. The result also demonstrates that our algorithm can synthesize masks which can be very different from the desired patterns.

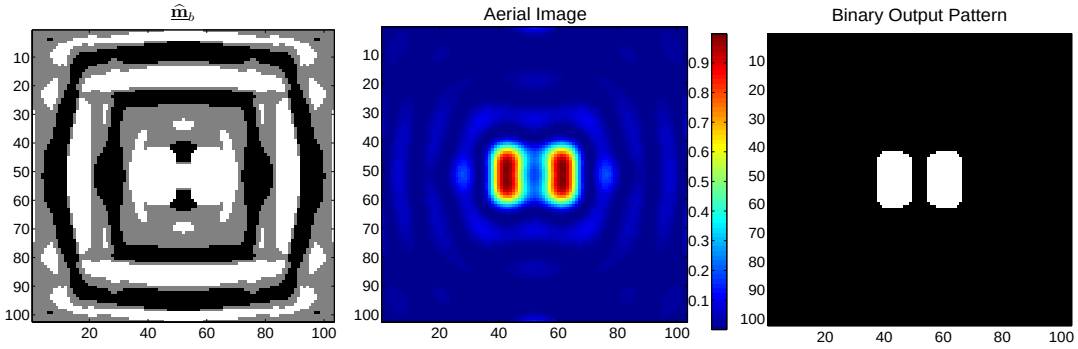

Figure 4.6: The estimated strong PSM (left), the corresponding aerial image (center), and the final binary output pattern (right) for a coherent imaging system with $\lambda = 193$nm and $NA = 0.85$. The black, gray, and white regions in $\underline{\widehat{\mathbf{m}}}_b$ correspond to -1, 0, and 1 respectively.

Finally, Fig. 4.6 illustrates the result using a strong PSM with chrome. The experimental parameters are as follows: $a = 25$, $t_r = 0.3$, $\gamma_{fid} = 1$, $\gamma_{aerial} = 0.25$, $\gamma_{TV} = 0$, $\gamma_{dis} = 0.01$, and $s = 1.5$. The number of iterations for the two steps were 80 and 160, and the run-times were 30 and 60 seconds respectively. Once again the

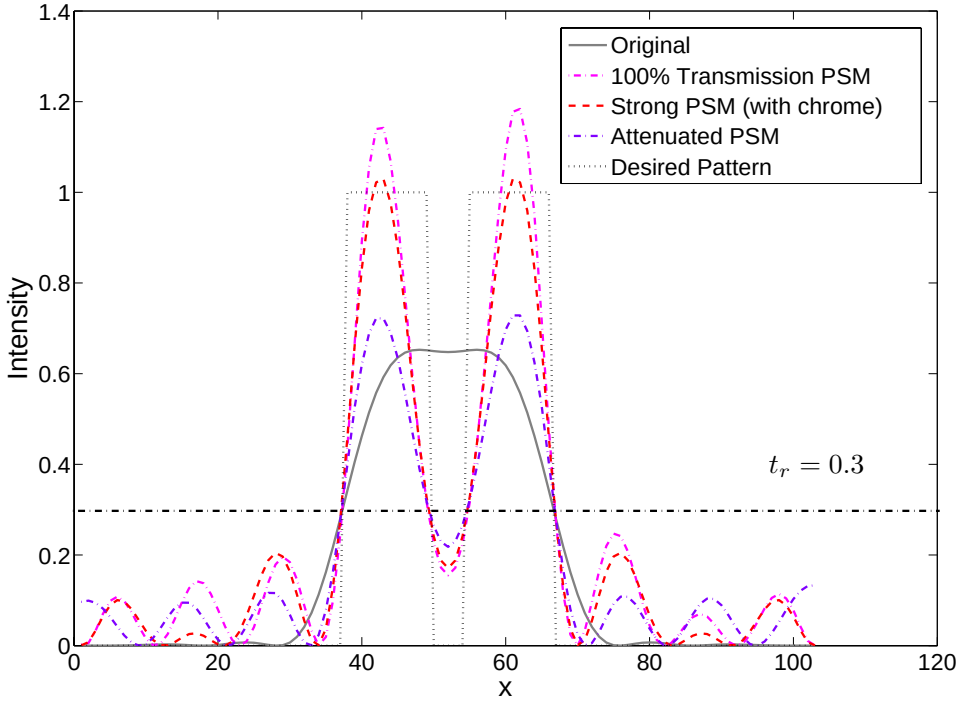

Figure 4.7: Horizontal slices along the center of the aerial images obtained using the desired pattern $\underline{\mathbf{z}}^*$ and the various synthesized PSMs in Fig. 4.4, 4.5 and 4.6 as inputs.

synthesized mask is quite different from the desired pattern. We also see three assist features around the main pattern which improve the contrast of the aerial image. The aerial image slices in Fig. 4.7 demonstrate an improvement in the contrast. Furthermore, the side-lobes are below the resist threshold t_r, and hence they will not print.

4.2.2 Random Logic Contact Holes

In the second experiment, our goal is to improve the contrast of the aerial image and distinguish the 100nm random logic contacts (see Fig. 4.8) using strong PSM. The lithography system parameters are: $\sigma = 0$, $\lambda = 193$nm, and $NA = 0.7$ ($k_1 =$

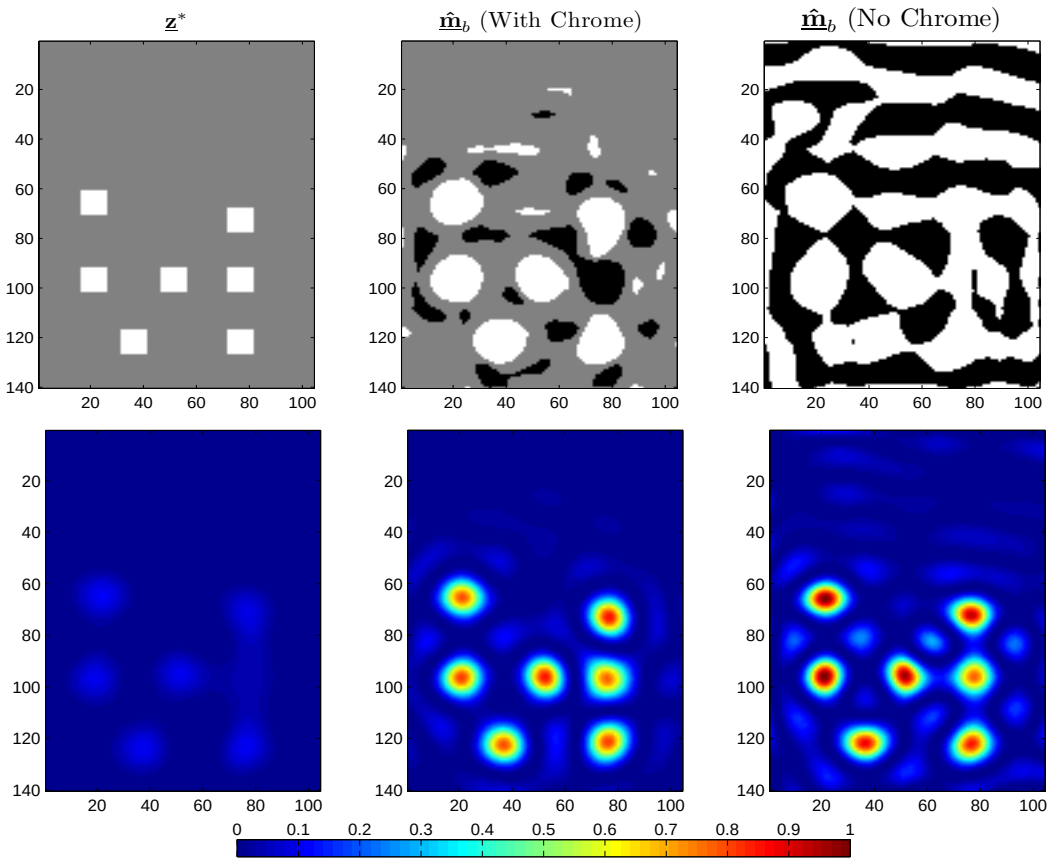

Figure 4.8: The top row consists of the desired pattern, the estimated strong PSM with chrome, and the 100% transmission PSM. The black, gray, and white regions correspond to transmission values of -1, 0, and 1 respectively. The bottom row consists of the aerial images corresponding to the masks in the top row. Here $\lambda = 193$nm, $NA = 0.7$, and $\sigma = 0$.

0.36). The first row in Fig. 4.8 indicates the desired pattern ($\underline{\mathbf{z}}^*$) and the estimated discrete tone masks ($\widehat{\underline{\mathbf{m}}}_b$) for strong PSM with chrome and 100% transmission PSM. The colors black, gray, and, white correspond to transmission values -1, 0, and +1 respectively. Note that in this case we are only concerned about matching the aerial image without worrying about the resist effects. Hence we set the contour fidelity term $\gamma_{fid} = 0$. The remaining experimental parameters for strong PSM with chrome are $\gamma_{aerial} = 1$, $\gamma_{dis} = 0.03$, $\gamma_{TV} = 0$, $s = 7$, and number of iterations = 100. The

parameters for 100% transmission PSM are $\gamma_{aerial} = 1$, $\gamma_{dis} = 0.01$, $\gamma_{TV} = 0$, $s = 7$, and number of iterations = 100. The optimization was carried out in a single step for both cases and each had a runtime of 50 seconds.

The aerial images corresponding to the input masks are illustrated in the bottom row of Fig. 4.8. We observe that strong phase shift masks bring tremendous improvement in the aerial image contrast thereby making the contacts distinguishable. Note that the upper half of the desired pattern is a contiguous zero region. The 100% transmission PSM deals with such regions by placing the $+1$ and -1 assist bars in a manner which destructively interfere leading to very little energy deposition. The PSM with chrome blocks the energy deposition by simply using chrome features.

4.2.3 Periodic Contact Holes

We now illustrate results for printing 100nm contact hole grating for different pitches. Since a contact train periodically repeats in the X and Y directions, the ILT simulations are performed for only one period. The boundary conditions are simulated as periodic (while performing the convolution operations) to account for the pattern periodicity. Fig. 4.9 indicates one period of the estimated CPL masks for 100nm contact hole with pitch increasing from 200nm to 400nm. The simulations were performed for a coherent imaging system with $NA = 0.85$, $\lambda = 193$nm, and pixel-size of 20nm. The experimental parameters are $\gamma_{aerial} = 1$, $\gamma_{fid} = 0$, and $\gamma_{dis} = 0.01$. We observe that even though the contact hole size is fixed, the estimated masks look drastically different over the pitch. The tight pitch of 200nm leaves no room to place

assist bars on the mask, but the main feature appears enlarged. As pitch increases, the size, position, and shape of the assist bars significantly vary.

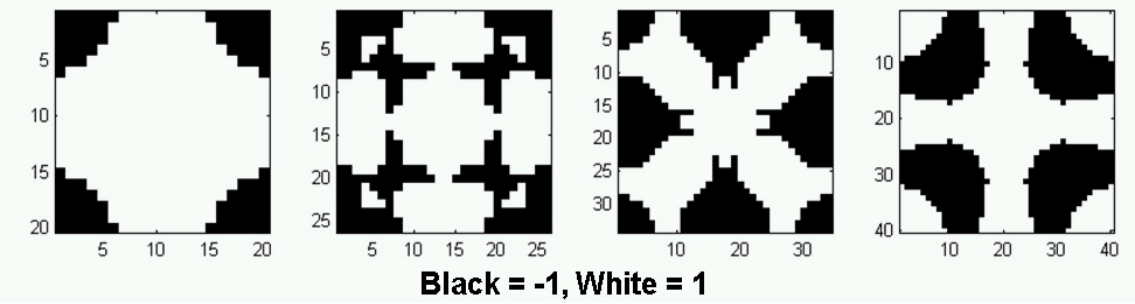

Figure 4.9: Estimated CPL masks for a 100nm contact through varying pitch for $NA = 0.85$. Left to right the pitch value is 200nm, 260nm, 340nm, and 400nm.

4.2.4 2-D Periodic Dense Pattern

The next experiment is for a periodic dense pattern consisting of *phase-conflicts* [54], a critical issue with strong PSM. Fig. 4.10 illustrates the two basic types of phase-conflicts: no phase-shift where we want it, and phase-shift where we do not. For the T-junction, we observe that there is lack of phase-shift along the horizontal critical feature due to the odd wrapping of phase assignments. The second type of phase assignment problem is called the termination problem, since it usually occurs near a line-end. When two opposite phases meet, they create a dark interference region (see Fig. 1.9 in Section 1.2.2), thereby resulting in an unwanted resist line. In the past, researchers have proposed using alternating phase shift mask with trim mask (double exposure) [28] or layout modification using graph-cut methods [7] to resolve the above problem.

The target pattern in Fig. 4.11 consists of both T-joints and line-ends. The

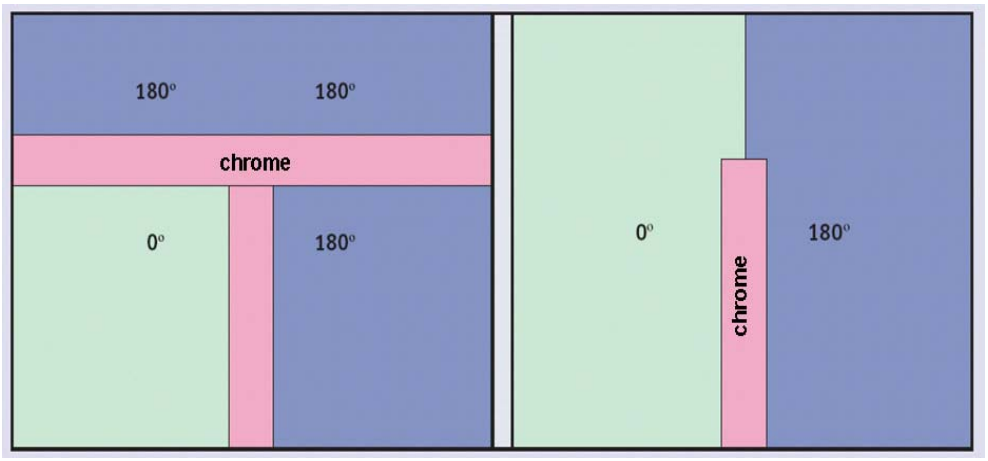

Figure 4.10: Two commonly arising cases of phase conflict (from [54]). The pattern on the left is a T-Joint which results in lack of phase-shift along the critical dimension. The one on right is a line-end which suffers from the problem of unwanted phase-shift.

grating pitch is 160nm, NA=0.85, $k_1 = 0.35$, and the pattern is sampled using 20nm resolution. We employ our ILT approach and solve for one period of the pattern (marked in dotted red box in Fig. 4.11). The boundary conditions are simulated using circularly symmetric padding for the convolution operations. The experimental parameters are $\gamma_{fid} = 1$, $\gamma_{dis} = 0.01$, $s = 0.5$, number of iterations = 1500, and the runtime was 50 seconds. The estimated 100% transmission mask corresponding to one period of the pattern is indicated in Fig. 4.11. The results indicate that the aerial image has low peak-intensity value in the phase-conflict regions, but the overall contour fidelity (at $t_r = 0.3$) is good.

4.2.5 A Complicated Tight-Pitch Pattern

The target pattern in Fig. 4.12 consists of a mixture of line, space, and contacts on a 120nm pitch and was provided to us by Intel Corporation. The imaging

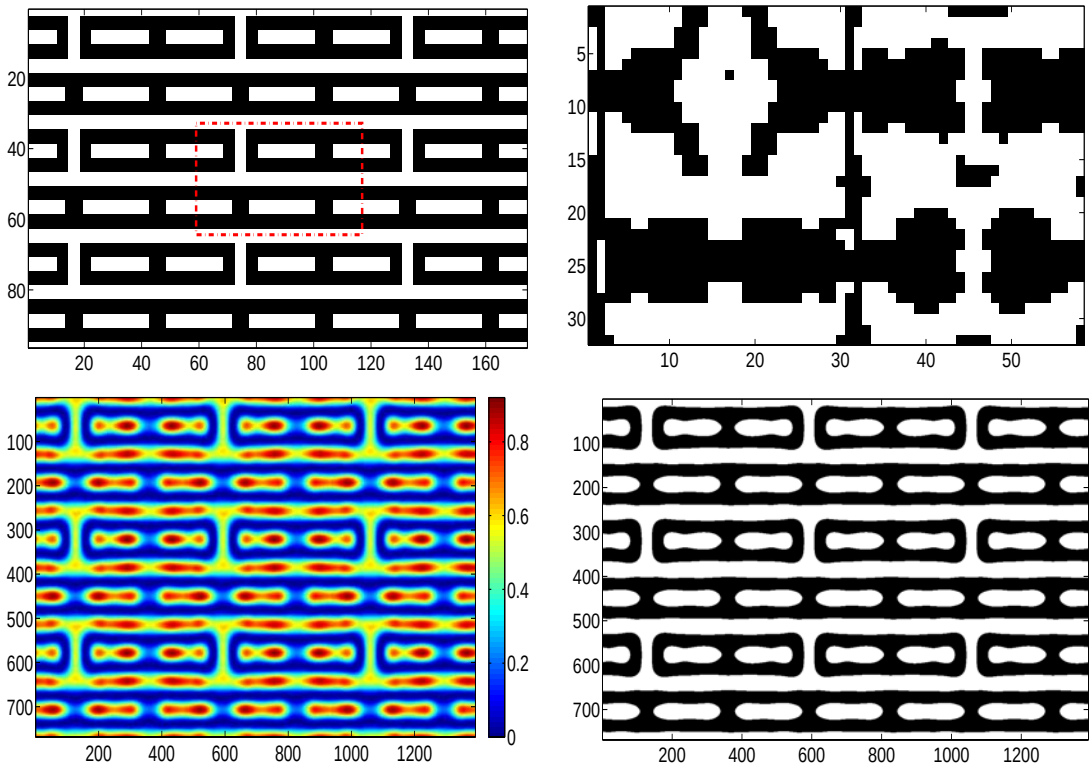

Figure 4.11: The top row indicates the periodic target pattern and the estimated 100% transmission PSM. The bottom row indicates the aerial image and its contour at $t_r = 0.3$. Here $NA = 0.85$, $\sigma = 0$, and $k_1 = 0.35$.

system has NA=0.93 giving a very low k_1 value of 0.29, and the mask pixel size is 20nm. We employ our ILT approach to solve for a CPL mask, and the experimental parameters are $\gamma_{fid} = 1$, $\gamma_{aerial} = 0.1$, $\gamma_{dis} = 0.03$, and $s = 1$. The estimated 100% transmission mask is indicated in Fig. 4.12. The very low k_1 value and the phase conflict regions lead to a relatively poor contrast aerial image, but the overall contour fidelity (at $t_r = 0.3$) is still good.

As discussed earlier, the cost-function for coherent and partially coherent imaging systems is a quartic cost function, and is non-convex with multiple local-

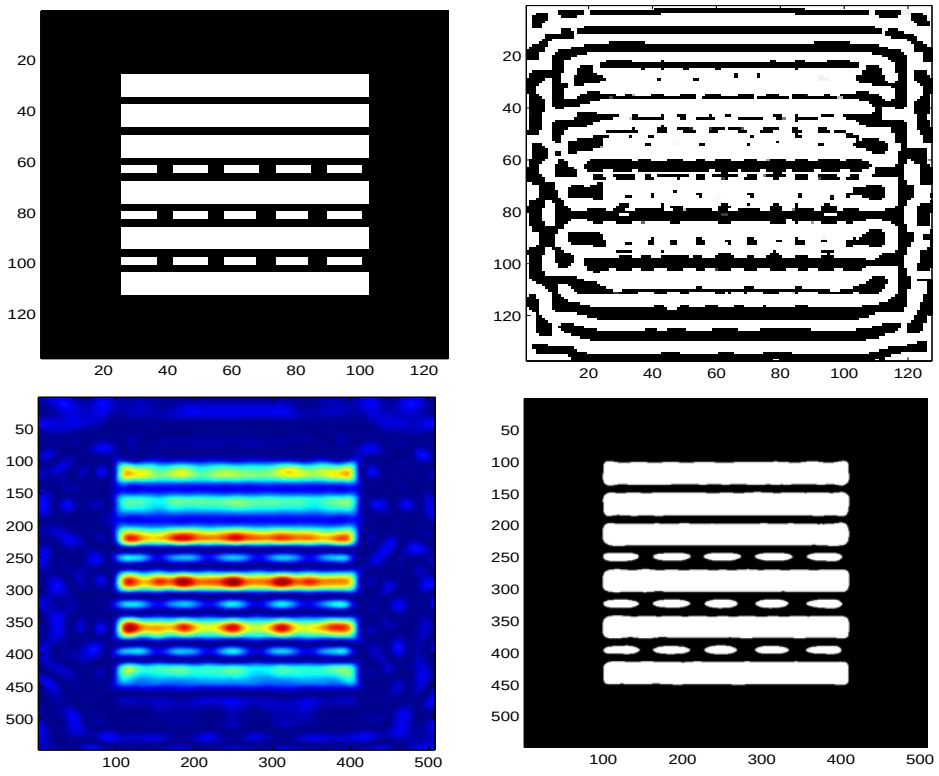

Figure 4.12: The top row indicates the complicated target pattern with 120nm pitch and the estimated 100% transmission PSM. The bottom row indicates the aerial image and its contour at $t_r = 0.3$. Here $NA = 0.93$, $\sigma = 0$, and $k_1 = 0.29$.

minima (as also noted by Granik in [31]). Note that the solutions obtained using OPC and AttPSM masks look structurally similar to the target pattern (see Fig. 3.11, Fig. 3.14, and Fig. 4.4). Our observation is that in such cases, the target itself acts as a good initial guess, and the (local) gradient-based search algorithm converges to a "good local minimum".

In the case of strong PSM, the estimated masks do not structurally resemble the target (see Fig. 4.5, Fig. 4.6, Fig. 4.8, and Fig. 4.11). PSM mask design often runs into "phase entanglement issues" as observed in [31]. In our case, the gradient

descent algorithm was initialized to all zeros thereby forcing the algorithm to choose the phase (0 or 180 degree) for all the regions. However, in the presence of dense features with numerous "phase-conflicts", the algorithm sometimes gets stuck in a "bad local minimum".

One possible approach (suggested in [31]) is to inspect the local minima of a simplified quadratic programming (QP) problem, and choose a good minimum as the initialization guess to the quartic cost function optimization routine. Another option is to employ four-phase masks, which are adequate for any Manhattan type geometry, as proved in [64]. The above approach was also employed under the alias of generalized inverse lithography in [50]. Yet another approach is to initialize with several different starting conditions, and reach a "better' local minima. But this is computationally more intensive. A more practical alternative is to use double exposure ILT, which automatically splits the target pattern into two parts and resolves the phase-conflicts [70]. The latter is more thoroughly explored in the next chapter of the thesis.

Chapter 5

ILT for Double Exposure Lithography with $k_1 \geq 0.25$

As we approach the 45nm and 32nm technology nodes, the operating k_1 value is rapidly decreasing (currently $k_1 \approx 0.3$). For such low k_1 values, the aerial image contrast obtained using single exposure lithography is poor, and the process is very sensitive to focus and exposure errors. Furthermore, single exposure lithography also suffers from the phase-conflict issue (discussed in Section 4.2.4).

The above issues and the non-readiness of EUV lithography has forced the industry to look into alternate advanced lithography solutions. Double exposure techniques have gained much importance recently, and are tipped as important technology enablers for 45nm and smaller nodes [34, 36]. They employ two (same or different) masks and two (same or different) illumination settings in order to print the desired circuit pattern. There are two fundamentally different flavors of double exposure de-

pending on whether the resist is developed once or twice. These are referred to as double exposure lithography (DEL) or double patterning lithography (DPL), and are described in more detail in Section 5.1.

In this chapter, we extend our inverse lithography framework to the case of DEL. The above combination allows patterning very close to the Rayleigh limit ($k_1 = 0.25$) with high contrast aerial images, and also overcomes the phase-conflict problem.

5.1 Double Exposure and Double Patterning

In this section, we compare and contrast the two different implementations of double exposure.

5.1.1 Double Exposure Lithography (DEL)

A flowchart of the steps involved in DEL is outlined in Fig. 5.1. We observe that the two exposures are carried out sequentially on the same resist layer, and the resist is developed only once after the second exposure. If a conventional resist is used, the final aerial image is the linear super-position of the two individual exposures. Hence, the minimum resolvable pitch is the same as single-exposure, i.e., minimum $k_{pitch} = 0.5$ and,

$$\text{Min Pitch} = 0.5\frac{\lambda}{NA}. \tag{5.1}$$

The above technique is widely adopted because it gives superior contrast for very low k_1 values (close to 0.25), not achievable using single exposure lithography. It

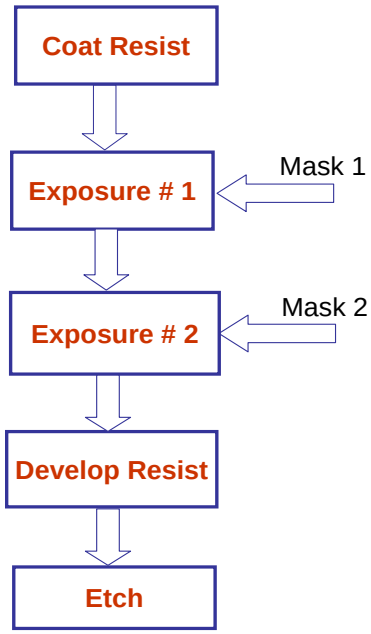

Figure 5.1: Flowchart of the steps involved in double exposure (single development) lithography.

finds applications such as trim mask with alternating phase shift masks [28] (to overcome the phase-conflict issue discussed in Section 4.2.4), double-dipole lithography (H/V decomposition), IDEAL (Innovative Double Exposure by Advanced Lithography) [34], etc. Double exposure with high-index immersion fluids (NA>1.5) has the potential to enable 32nm technology node. DEL also forms the focus of this thesis, and is explored in much more detail in the later sections.

5.1.2 Dual Patterning (DPL)

The double development case (also referred to as dual patterning) requires additional resist development, etching, and coating steps in between the two exposures

[36, 26]. The concept behind multiple exposure/develop techniques is that the fundamental single-exposure limit given in (5.1) is not on the individual feature size (CD), but on the pattern pitch [9]. Using non-linearities in the pattern transfer process, CDs can be much smaller than half of the smallest allowed pitch. Dual patterning uses the above concept by patterning features at dimensions close to 1/4 of the minimal pitch allowed by constraints of the imaging system (see Fig. 5.2), and stores the information from the first exposure in a hard mask (or other sacrificial layer). The second exposure is then carried out with similar feature size target and appropriate shift and again preserved in a sacrificial layer, thereby doubling the frequency of each exposed feature present on each of the masks (also see Fig. 5.3). The minimum pitch resolution for dual patterning is $0.25\frac{\lambda}{NA}$. It was demonstrated that dual patterning allows one to reach patterning with effective k_1 as low as 0.16 [58].

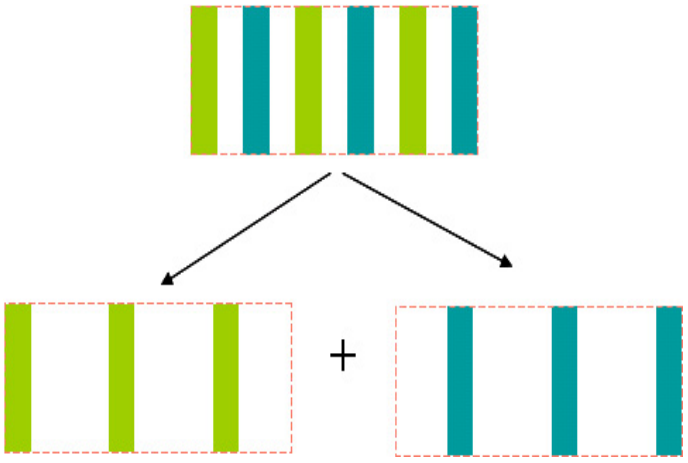

Figure 5.2: The original (1:1) grating and two decomposed (1:3) gratings for dual patterning. The two patterns are shifted versions of each other and the pitch is doubled enabling them to be individually printed using single exposure.

However, additional processing steps related to preserving the image of the

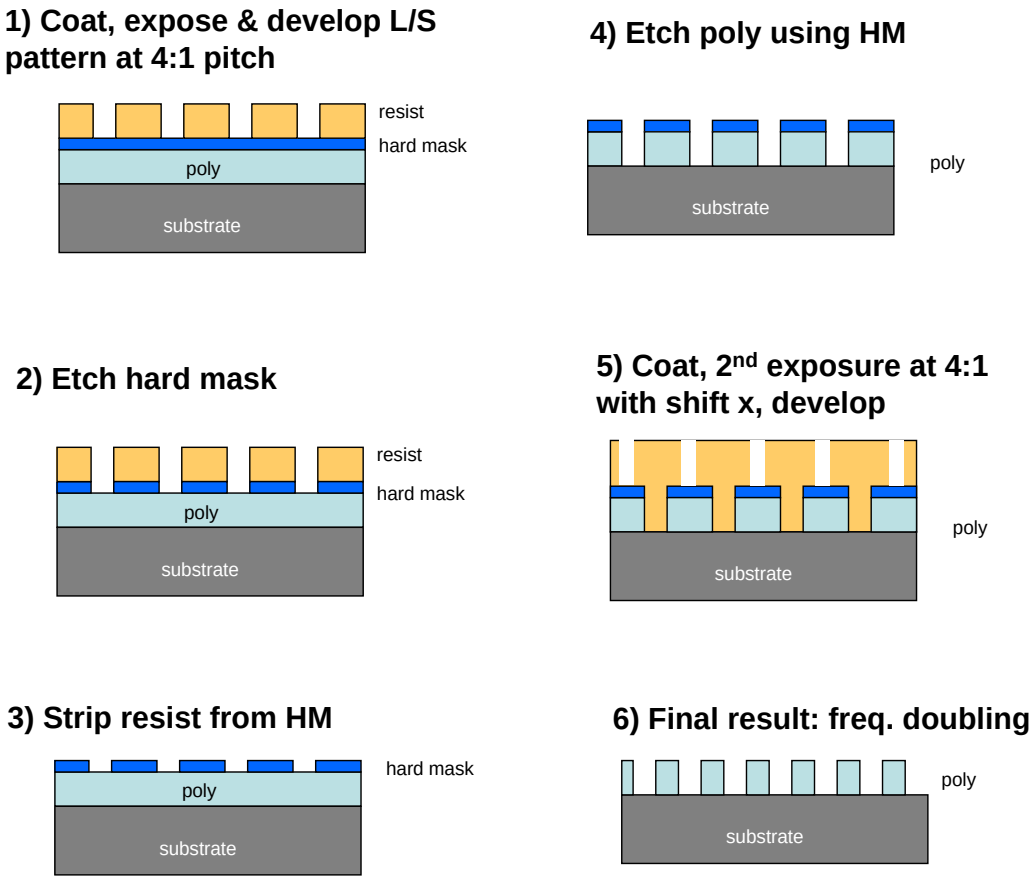

Figure 5.3: Steps involved in double patterning for frequency doubling of grating (figure courtesy Intel Corporation).

first exposure in sacrificial layer (such as need to load and unload wafer twice from the exposure tool), carry additional overlay and dimensional control (CD) penalty. Furthermore, the extra processing steps such as a second develop, hard-mask deposition, and etch steps might result in additional yield liability as well as increased wafer cost.[1]

In this thesis, we focus exclusively on DEL, where the resist is developed only once at the end of the two exposures. With regard to the above, we explore two

[1]Despite of these misgivings, the unavailability of EUV and high-index immersion fluids have made dual patterning a strong contender for the 32nm technology node.

cases of $k_1 \geq 0.25$ (above Rayleigh resolution limit), and $k_1 < 0.25$ (below Rayleigh resolution limit). In this chapter we extend our ILT framework for $k_1 \geq 0.25$ using conventional resist, and present a new paradigm of decomposing the gray-level aerial image for the individual exposures. Later (in Chapter 6), we explore an ambitious paradigm of using novel materials which may allow frequency doubling ($k_1 < 0.25$) similar to dual patterning, but without the need for a sacrificial layer.

5.2 Double Exposure Inverse Lithography Technology (DEL-ILT) Framework

Here we demonstrate the shortcoming of DEL and motivate the need for an inverse lithography based approach to mask design for double exposure lithography systems. This is followed by a discussion of the forward model for DEL, and a formal definition of the double exposure inverse lithography problem.

5.2.1 Motivation

As discussed earlier, in DEL systems the resist is developed only once at the end of the two exposures. All DEL systems involve some sort of pattern decomposition targeted towards the individual exposures. A typical double dipole lithography scheme proposed by Hsu, et al. [36] starts by separating the target into horizontal and vertical geometries for use with D-Y and D-X dipoles respectively. The final image with the two exposures is the sum of the energies at the wafer plane. Therefore, a critical shielding/protecting step is essential in order to guarantee a high aerial image contrast.

Thus the vertical features are protected when the horizontal features are imaged and vice-versa. The above steps are followed by scattering bar insertion and model-based OPC in order to improve robustness and bring the contours on target.

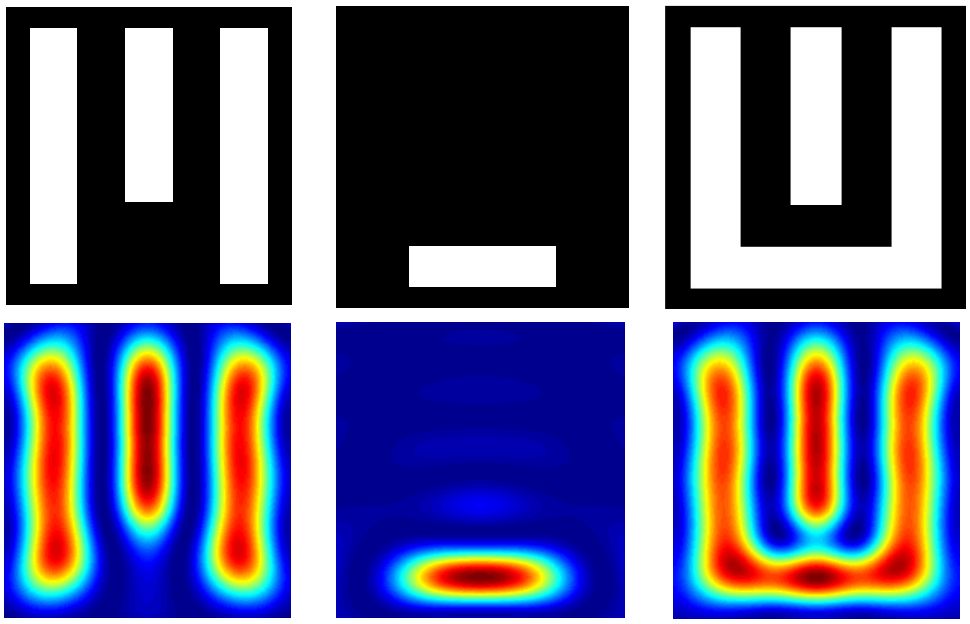

Figure 5.4: The top row indicates the vertical and horizontal (decomposed) target patterns and the original target. The bottom row indicates the corresponding aerial images obtained using single exposure ILT masks, and the combined aerial image.

Fig. 5.4 is a simple example which demonstrates the effect of residual aerial image energy in DEL. Here, the target pattern was first decomposed into horizontal (H) and vertical (V) structures (see the left and center patterns on the top row), and the corresponding masks were synthesized. The individual aerial images are illustrated in Fig. 5.4 and they have a good contrast. However, we observe that the combined aerial image has an unacceptable quality at the critical junction (the gap between the horizontal and vertical features) due to the unwanted residual energy from the

previous exposure. Thus some sort of model-based pattern decomposition or shielding is essential for the success of DEL. The double exposure inverse lithography (DEL-ILT) framework proposed in this thesis is an "*aerial image aware*" pattern decomposition. Most commonly employed techniques aim to split the (binary) target into two parts followed by some post-processing steps to obtain a high contrast aerial image. However, the DEL-ILT framework is a fresh paradigm which acts by decomposing the (gray-level) aerial image into two overlapping parts (see Fig. 5.10 for the DEL-ILT decomposition of the target pattern in Fig. 5.4). The estimated masks are simultaneously synthesized to guarantee a superior contrast of the combined aerial image without worrying about the residual energies from the previous exposure.

5.2.2 The Double Exposure Inverse Lithography Problem

The first step towards solving an *inverse* problem is to define a *forward* (or process) model which is a mathematical description of the given imaging system. Consistent with our earlier notation, the image formation process for a DEL system can be mathematically expressed as,

$$z(x,y) = T\{a(x,y), b(x,y)\}, \tag{5.2}$$

where $T\{.\}$ is the forward model which maps the input intensity functions $a(x,y)$ and $b(x,y)$ to the output intensity function $z(x,y)$. Let $z^*(x,y)$ be the desired output intensity function. The goal of double exposure inverse lithography mask design is to estimate the input intensity functions $\hat{a}(x,y)$ and $\hat{b}(x,y)$ which will give us a close approximation to the desired output $z^*(x,y)$ (see Fig. 5.5). This is achieved by searching

the space of both the inputs and choosing a pair $\widehat{a}(x,y)$ and $\widehat{b}(x,y)$ which minimizes a distance $d(z(x,y), z^*(x,y))$, where $d(\cdot,\cdot)$ is a distance metric. Therefore,

$$\{\widehat{a}(x,y), \widehat{b}(x,y)\} = \arg \min_{a(x,y),b(x,y)} d[z^*(x,y), T\{a(x,y), b(x,y)\}]. \tag{5.3}$$

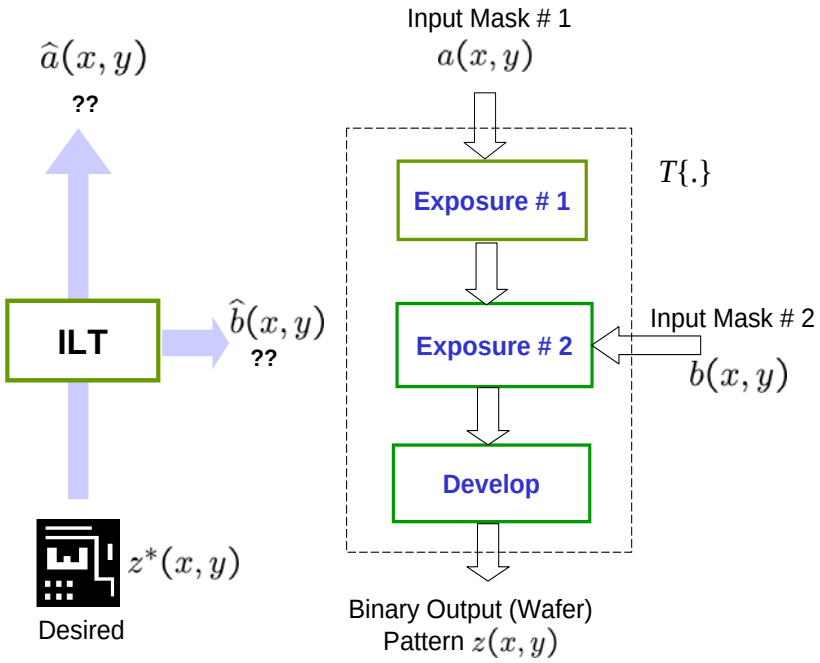

Figure 5.5: Forward Model and ILT for Double Exposure Lithography Systems

A typical lithography process consists of aerial image formation and the resist process. If the forward model $T\{.\}$ accounts for the resist effects, the output pattern $z(x,y)$ represents the resist pattern in the wafer. If $T\{.\}$ only models the imaging optics, $z(x,y)$ represents the aerial image in the wafer plane. The objective of the inverse problem in the first case is to improve the contour fidelity, and in the second case is to improve the aerial image contrast. The present discussion focuses on the

latter case of aerial image improvement.[2]

The final aerial image for the given DEL system is equal to the sum of the aerial images obtained from the two individual exposures. Using the SOCS approximation for partially coherent imaging systems, the aerial image $\underline{\mathbf{z}}$ can be modeled as,

$$\underline{\mathbf{z}} = \left(\sum_{j=1}^{P_1} \alpha_j |\mathbf{H}_j \underline{\mathbf{a}}|^2 \right) + \left(\sum_{j=1}^{P_2} \beta_j |\mathbf{G}_j \underline{\mathbf{b}}|^2 \right), \tag{5.4}$$

where $\underline{\mathbf{a}},\ \underline{\mathbf{b}},\ \underline{\mathbf{z}} \in \Re^{MN \times 1}$ are obtained by sampling and lexicographic ordering of $a(x,y),\ b(x,y)$, and $z(x,y)$ respectively, $\mathbf{H}_1, \ldots, \mathbf{H}_{P_1}$ and $\mathbf{G}_1, \ldots, \mathbf{G}_{P_2}$ are the SOCS kernels corresponding to the two exposures, and $\alpha_1, \ldots, \alpha_{P_1}$ and $\beta_1, \ldots, \beta_{P_2}$ respectively correspond to their singular values. Note that the above model can account for different illumination settings for the first and second exposures (commonly seen in double dipole lithography).

As a proof of principle, we focus on the simpler case of a DEL system using a coherent imaging system ($\sigma = 0$) with the same exposure settings for both the exposures. Under the above assumptions, the forward model in (5.4) reduces to,

$$\underline{\mathbf{z}} = |\mathbf{H}\underline{\mathbf{a}}|^2 + |\mathbf{H}\underline{\mathbf{b}}|^2, \tag{5.5}$$

where $\mathbf{H}$ is a *jinc* function with cutoff frequency NA/λ. Alternatively, $\mathbf{H}$ can also be substituted by the optimal coherent approximation (OCA) kernel reported in [64].

The only related work (in the open literature) on double exposure mask design using inverse lithography is by Wang, et al. [91] and Pati, et al. [64]. Their original goal

[2]The contour fidelity model will be discussed in Section 7.3 in context of alignment error.

was to actually solve the inverse problem for single exposure phase shift mask design. The mask was synthesized in the frequency domain using POCS (projection onto convex sets) iterations. The resulting mask was complex-valued having continuous phase and amplitude. Double exposure was suggested only as a *phase quantization* technique by employing the double-exposure theorem [91]. The latter states that any image attainable using a complex-valued mask can be obtained using a suitable pair of two strong continuously varying phase shift masks. Finally, the amplitude was quantized using half-toning. Note that they did not originally set out to design masks for double exposure systems, and hence their framework is more restricted in that it can inherently handle only coherent imaging systems with the *same* illumination settings for both exposures. Furthermore, the half toning technique [91] tremendously increases the complexity of the masks. Finally, it is hard and not intuitive to impose simplicity and discretization constraints on the masks in the frequency domain.

In the next section we discuss our inverse imaging approach for pixel-based DEL mask design.

5.3 DEL-ILT Optimization

The mask synthesis problem can be formulated as minimizing the L_2 norm of the difference between the desired pattern $\underline{\mathbf{z}}^*$ and the aerial image $\underline{\mathbf{z}}$ defined in (5.5),

$$[\widehat{\underline{\mathbf{a}}}, \widehat{\underline{\mathbf{b}}}] = \arg\min_{\underline{\mathbf{a}},\underline{\mathbf{b}}} F(\underline{\mathbf{a}}, \underline{\mathbf{b}}) = \arg\min_{\underline{\mathbf{a}},\underline{\mathbf{b}}} ||\underline{\mathbf{z}}^* - |\mathbf{H}\underline{\mathbf{a}}|^2 - |\mathbf{H}\underline{\mathbf{b}}|^2||_2^2. \tag{5.6}$$

The above estimated masks are subject to integer constraints depending by the employed RET (as seen earlier in Chapter 2). Throughout the current discussion, we restrict ourselves to the case where both the estimated masks are 100% transmission strong PSM. Thus, the transmission values of the estimated mask pixels should be either -1 or +1 leading to an integer optimization problem, which can be reduced to a continuous variable optimization problem by imposing the following $2MN$ bound constraints on the estimated variables,

$$-1 \leq a_j \leq 1 \quad \text{and} \quad -1 \leq b_j \leq 1 \text{ for } j = 1, \ldots, MN \ . \tag{5.7}$$

Having reduced the problem to the continuous domain, the analytically calculated gradient can once again be used to systematically explore the solution space.

Note that double exposure lithography under the bound constraints in (5.7) is equivalent to employing single exposure masks with continuously varying phase and amplitude as per the double exposure theorem (also see Appendix B). The estimated masks are finally thresholded using optimum thresholds t_a and t_b to obtain the (physically realizable) two-toned masks.

It is important to realize that the number of unknown parameters in double exposure ILT are twice as many as those in a single exposure ILT. Indeed the inverse problem for DEL is severely ill-posed and there may be multiple pairs of eligible mask solutions. We employ the regularization framework (introduced in Section 3.3) to overcome the above issue and arrive at preferred (good) local minima. For example, the tone of the estimated masks can be controlled by minimizing the augmented cost

function,

$$J(\underline{\mathbf{a}}, \underline{\mathbf{b}}) = F(\underline{\mathbf{a}}, \underline{\mathbf{b}}) + \gamma_{dis_a} R_{dis}(\underline{\mathbf{a}}) + \gamma_{dis_b} R_{dis}(\underline{\mathbf{b}}),$$

where $F(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ is the data fidelity term defined in (5.6), $R_{dis}(\cdot)$ is the quadratic penalty term for 100% transmission strong PSM (defined in (4.5)), and γ_{dis_a}, γ_{dis_b} are the regularization weights. The discrete tone CPL mask can then be obtained by simply thresholding the estimated masks with $t_a = t_b = 0$. The regularization framework can also possibly be used to promote other *desirable* properties like tolerance to alignment errors, low-complexity and ease of manufacturing, low MEEF, and good process-window.

We now discuss two gradient-based algorithms for solving the bound-constrained optimization problem defined by (5.6) and (5.7). The above algorithms will also help to demonstrate the existence of multiple solutions.

5.3.1 Method 1: Simultaneous Descent

The first method consists of finding the gradient of the cost function defined in (5.6) with respect to the $2MN$ unknown parameters, and searching through the *entire* $2MN$ dimensional space using steepest descent iterations. The gradient of the cost function (5.6) with respect to the parameter vectors $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ are denoted as $\nabla_{\underline{\mathbf{a}}} F(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ and $\nabla_{\underline{\mathbf{b}}} F(\underline{\mathbf{a}}, \underline{\mathbf{b}})$ where,

$$\nabla_{\underline{\mathbf{a}}} F(\underline{\mathbf{a}}, \underline{\mathbf{b}}) = -\mathbf{H}^{\mathbf{T}} \left[(\underline{\mathbf{z}}^* - |\mathbf{H}\underline{\mathbf{a}}|^{\mathbf{2}} - |\mathbf{H}\underline{\mathbf{b}}|^{\mathbf{2}}) \odot (\mathbf{H}\underline{\mathbf{a}}) \right] \in \Re^{MN \times 1} \tag{5.8}$$

$$\nabla_{\underline{\mathbf{b}}} F(\underline{\mathbf{a}}, \underline{\mathbf{b}}) = -\mathbf{H}^{\mathbf{T}} \left[(\underline{\mathbf{z}}^* - |\mathbf{H}\underline{\mathbf{a}}|^{\mathbf{2}} - |\mathbf{H}\underline{\mathbf{b}}|^{\mathbf{2}}) \odot (\mathbf{H}\underline{\mathbf{b}}) \right] \in \Re^{MN \times 1} \tag{5.9}$$

The steepest descent algorithm requires a user-defined initialization to start its iterations. However, it is interesting to note from (7.4) and (7.5) that the gradients have the same functional form. Hence if $\underline{\mathbf{a}}^0 = \underline{\mathbf{b}}^0$ (i.e. the two masks are initialized to the same value), then for all the iterations we will have $\underline{\mathbf{a}}^k = \underline{\mathbf{b}}^k$. This is not helpful and against the spirit of double exposure, since the above effect can be easily simulated with a single mask by merely doubling the dose. Hence, it is compulsory to initialize $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ at separate points for the success of this algorithm.

Furthermore, note that the optimization problem in (5.6) is subject to bound constraints given in (5.7). The above can be reduced to an unconstrained optimization problem by using the parametric transformation $a_j = \cos(\omega_j)$ and $b_j = \cos(\psi_j)$ for $j = 1, \ldots, MN$. The new parameter vector is $\underline{\boldsymbol{\theta}} = [\underline{\boldsymbol{\omega}}^T,\ \underline{\boldsymbol{\psi}}^T]^T \in \Re^{2MN \times 1}$, where $\underline{\boldsymbol{\omega}}$ and $\underline{\boldsymbol{\psi}}$ are the unconstrained vectors corresponding to $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$. The cost function defined in (5.6) is reformulated to obtain $F_1(\underline{\boldsymbol{\omega}}, \underline{\boldsymbol{\psi}})$ and the gradient vectors $\underline{\mathbf{d}}_1 = \nabla_{\underline{\boldsymbol{\omega}}} F_1(\underline{\boldsymbol{\omega}}, \underline{\boldsymbol{\psi}})$ and $\underline{\mathbf{d}}_2 = \nabla_{\underline{\boldsymbol{\psi}}} F_1(\underline{\boldsymbol{\omega}}, \underline{\boldsymbol{\psi}}) \in \Re^{MN \times 1}$ are evaluated. The steepest descent iterations are used to update both masks together as follows,

$$\underline{\boldsymbol{\theta}}^{k+1} = \underline{\boldsymbol{\theta}}^k - s\underline{\mathbf{d}}^k \tag{5.10}$$

where,

$$\underline{\mathbf{d}}^k = \begin{bmatrix} \underline{\mathbf{d}}_1^k \\ \\ \underline{\mathbf{d}}_2^k \end{bmatrix}.$$

Thus, for every iteration we move along the direction defined by the $2MN$ dimensional gradient vector $\underline{\mathbf{d}}$, and *simultaneously* update both $\underline{\boldsymbol{\theta}}_1$ and $\underline{\boldsymbol{\theta}}_2$. The step size s is fixed

by the user. The above method works in practice, but its convergence is slow. In the next section, we explore a possible way of speeding up the convergence.

5.3.2 Method 2: Cyclic Coordinate Descent

We now introduce an alternative optimization strategy called cyclic coordinate descent. The latter involves optimizing the cost function with respect to only one parameter at a time, and sequentially covering up the whole parameter vector. Thus each coordinate axis (corresponding to a given parameter) is searched, and descent is made only along that particular dimension. The above procedure is then repeated (cycling) until convergence is achieved. Thus at any point, we only make one dimensional moves along $a_1, \ldots, a_{MN}, b_1, \ldots, b_{MN}$ and back to a_1. However, we propose a slightly modified approach, which is in spirt of the above scheme. Instead of searching the $2MN$-dimensional space one dimension at a time, we split the search into *two* MN dimensional sub-spaces, corresponding to the two masks.

The procedure is as follows. We start by initializing the variables arbitrarily (usually $\underline{\mathbf{a}}^0 = \underline{\mathbf{b}}^0 = \underline{\mathbf{0}}$) and do the parametric transformations to obtain the unconstrained vectors $\underline{\boldsymbol{\omega}}^0$ and $\underline{\boldsymbol{\psi}}^0$. However, we update $\underline{\boldsymbol{\omega}}$ and $\underline{\boldsymbol{\psi}}$ separately on every iteration. Each iteration now consists of two steps. The first step involves updating only $\underline{\boldsymbol{\omega}}$ by calculating the gradient of (5.6) with respect to $\underline{\boldsymbol{\omega}}$. The second step involves updating only $\underline{\boldsymbol{\psi}}$, but the gradient calculations now utilize the updated value of $\underline{\boldsymbol{\omega}}$ (from the previous step). To summarize, for the k^{th} iteration,

$$\underline{\boldsymbol{\omega}}^{k+1} = \underline{\boldsymbol{\omega}}^k - s_1 \nabla_{\underline{\boldsymbol{\omega}}} F_1(\underline{\boldsymbol{\omega}}^k, \underline{\boldsymbol{\psi}}^k) \tag{5.11}$$

$$\underline{\boldsymbol{\psi}}^{k+1} = \underline{\boldsymbol{\psi}}^{k} - s_2 \nabla_{\underline{\boldsymbol{\psi}}} F_1(\underline{\boldsymbol{\omega}}^{k+1}, \underline{\boldsymbol{\psi}}^{k}) \tag{5.12}$$

At any point, we restrict our search only to the subspace defined by one mask. Simulation results (see Section 5.4) indicate that Method 2 has a faster convergence compared to Method 1. It is also worth noting that the solutions obtained using Method 1 and 2 are not always equal implying that there are multiple ways of splitting the patterns for double exposure (see Section 5.4.2 for more details).

5.4 Results

We now discuss some results obtained using the above framework. All the experiments were carried out for fully coherent imaging systems with 193nm radiation wavelength.

5.4.1 T Joint

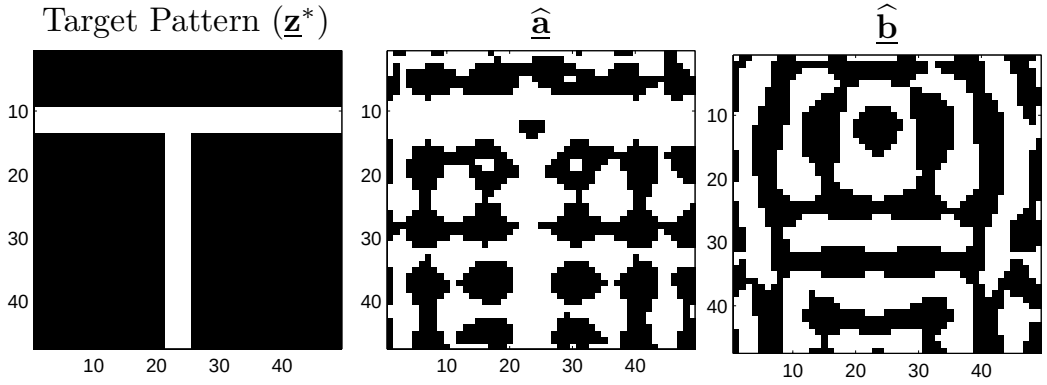

Figure 5.6: The T-junction target pattern (left) and the two estimated chromeless PSM (center and right).

A T-Joint suffers from the phase conflict problem (as seen in Fig. 4.10 in

Chapter 4). Hence direct rule-based phase assignments cannot be made, and special solutions have been proposed in the past to address this issue [36, 8]. Fig. 5.6 illustrates the desired pattern and the two synthesized chromeless PSM obtained using our double exposure inverse lithography approach with Method-2. The horizontal and vertical bars are 80nm wide ($k_1 = 0.35$), and the mask pixel-size is 20nm in wafer dimensions (actual size 80nm assuming 4X reduction factor). The experimental parameters are $s_1 = s_2 = 8$, $\gamma_{dis} = 0.002$, and number of iterations = 300. The aerial images in Fig. 5.7 illustrate that the first exposure prints the horizontal and vertical bars but leaves a gap at their intersection[3]. This is the phase-conflict region. The second exposure deposits the missing energy in the above region to resolve it (see Appendix B). The final aerial image has good contrast and is uniform along all portions (including the junction). Fig 5.8 illustrates the aerial image obtained using a chromeless PSM, which was synthesized by solving the *single exposure* ILT problem. We observe that the aerial image in this case is weak (tending to break) near the junction indicating a hot-spot. The L_2 norm cost function for single and double exposure cases was 39.81 and 35.12 respectively, indicating that the latter achieves a much better contrast. Fig. 5.8 also indicates that the aerial image contours at $t_r = 0.55$ have higher fidelity and much smaller *pattern error* for DEL-ILT[4]. Thus, DEL was able to resolve the phase-conflict. Finally, the aerial image decomposition obtained by employing Method-1 is similar to the one seen in Fig. 5.7.

[3]The aerial images are up-sampled by 8X using spline interpolation for display purpose.

[4]Note that the optimization was performed only for the aerial image and not the pattern fidelity. The pattern error can be further decreased using contour fidelity as the objective goal (see Section 7.3).

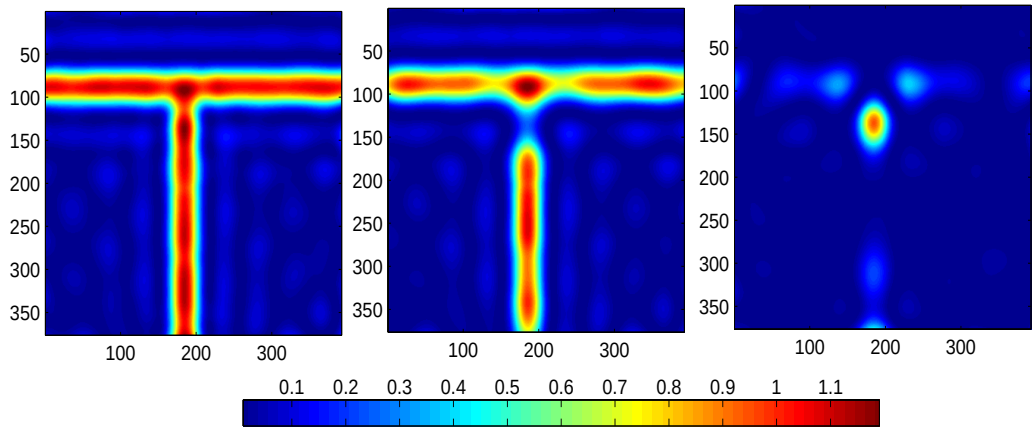

Figure 5.7: The final aerial image (left) and the aerial images corresponding to the two individual exposures (center and right) for the mask patterns in Fig. 5.6

5.4.2 U-Joint

Fig. 5.9 illustrates the target pattern and the synthesized PSM for a vertical bar sandwiched between a U-shaped feature. The width of the above features and their horizontal and vertical separations are 80nm ($k_1 = 0.35$). The masks were synthesized using Method-2 and 20nm pixels similar to the previous case. The aerial images in Fig. 5.10 indicates very interesting pattern decomposition. It appears that the proposed algorithm is trying to split the original pattern into two less dense patterns, thereby making them easier to print using individual exposures. The combined aerial image has a superior contrast.

Fig. 5.11 illustrates the aerial images obtained using the masks synthesized by Method-1. We observe that the decomposition is different from that in Fig. 5.10. We believe that in general there may be multiple ways of splitting the target pattern for double exposure. The above can be attributed to the ill-posedness of the inverse DEL

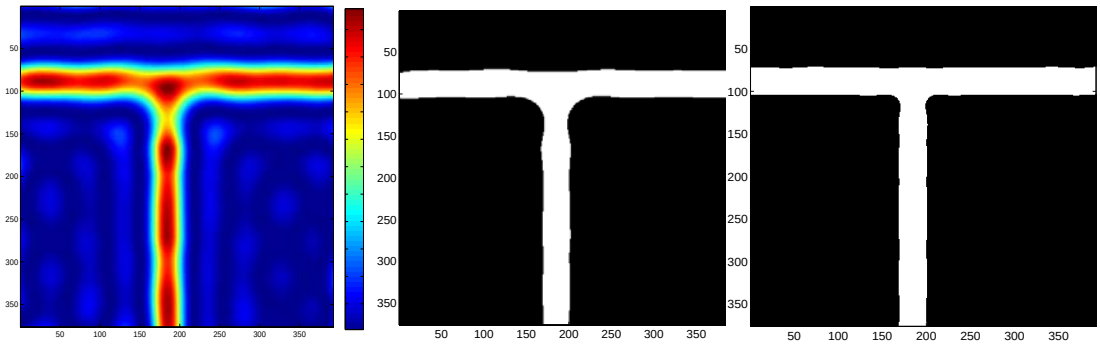

Figure 5.8: The graph on the left is the aerial image obtained by solving the single exposure ILT problem for the T-junction. The center and right graphs are the contours at $t_r = 0.55$ (sampled at 2.5nm) for single and double exposure ILT. The corresponding pattern errors are 1047 and 389.

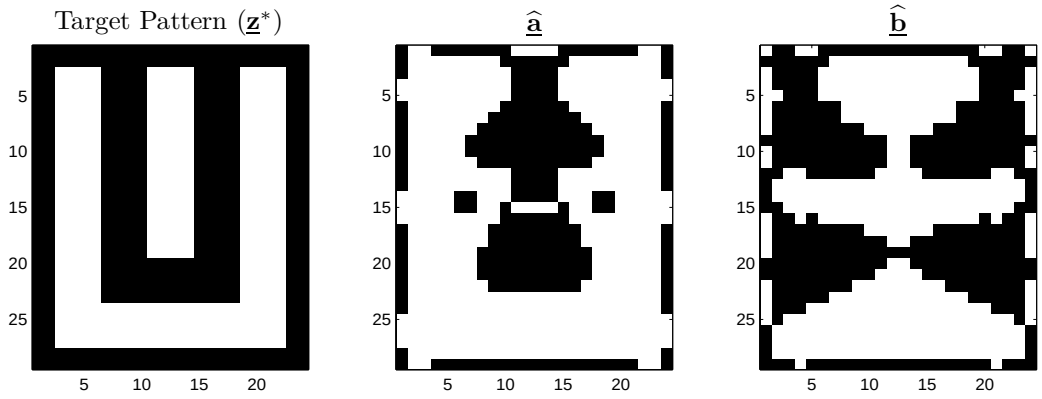

Figure 5.9: The U-Joint target pattern (left) and the two estimated chromeless PSM (center and right) obtained using Method-2.

problem. In Section 7.3, the optimization objective is reformulated to find a (*preferred*) pattern split which is robust to alignment errors. Similarly, the pattern decomposition can be motivated by other user-defined objectives (via novel regularization terms).

The graph in Fig. 5.12 indicates the cost function behavior of the two algorithms (discussed in Section 5.3) for the first 50 iterations of the steepest descent algorithm. The rate of convergence depends on the chosen step-size. In order to make a fair comparison, optimal step sizes were calculated at every iteration for both the

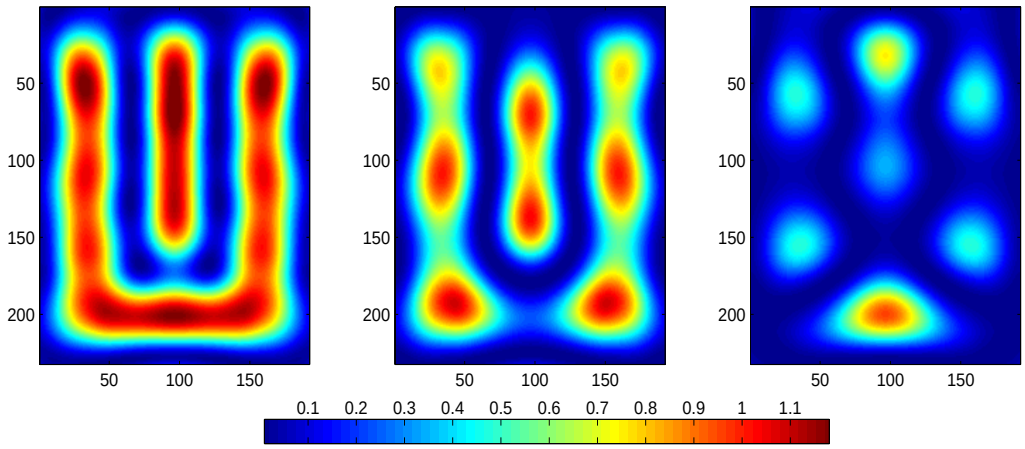

Figure 5.10: The final aerial image (left) and the aerial images corresponding to the two individual exposures (center and right) for the mask patterns in Fig. 5.9.

algorithms. The step-sizes were calculated using the electrical caching technique discussed in [31]. The double exposure ILT problem was solved for the target pattern in Fig. 5.9. We observe that the cyclic descent method converges faster than the simultaneous descent method. This is true for all the target patterns discussed in this chapter. The computational complexity and number of convolutions required by the two algorithms is the same. Therefore, we conclude that the cyclic coordinate descent technique is a superior way of searching the solution space.

5.4.3 Pattern Decomposition for More Complicated Patterns

We now demonstrate automatic pattern decomposition results for some very complicated patterns (provided to us by Intel Corporation), with numerous phase-conflicts, and very low k_1 values. Our first assignment was to print the complex pattern consisting of 60nm features (as indicated in Fig. 5.13) using 193/0.93NA dry lithography. The mask pixel size is 20nm in wafer dimensions (actual size 80nm for

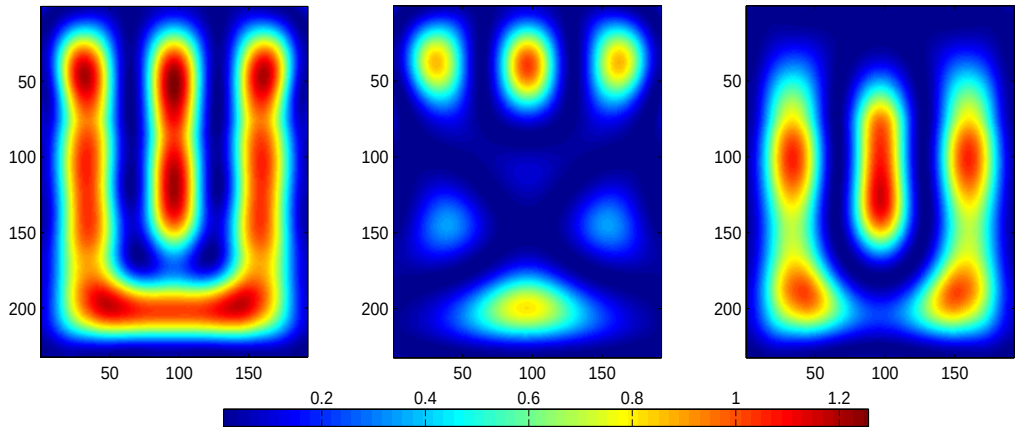

Figure 5.11: The final aerial image (left) and the individual aerial images (center and right) for the U-joint target pattern. The corresponding masks were synthesized using Method-1.

a 4X reduction system). The estimated CPL masks and their corresponding aerial images are illustrated in Fig. 5.13, and are un-intuitive, but very successful. The DEL-ILT algorithm splits the aerial image into two (overlapping) parts, and in the process automatically resolves all the possible phase-conflicts. The combined aerial image has very good quality and closely resembles the target pattern.

Our next assignment was to print the target illustrated in Fig. 5.14 consisting of 80:320 vertical gratings (thin lines in between large spaces). Here the pixel-size is 20nm, $NA = 0.85$, radiation wavelength = 193nm with corresponding $k_1 = 0.35$ for smallest pattern pitch. The target also consists of an 80nm horizontally sandwiched space thereby making the whole layout a challenging patterning problem. The estimated CPL masks obtained using DEL-ILT along with their corresponding aerial images are illustrated in Fig. 5.14. We again observe an unconventional pattern decomposition, but the combined aerial image has very good contrast. In particular,

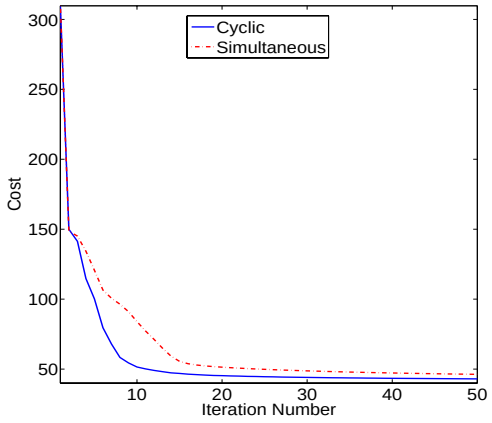

Figure 5.12: Cost function behavior for simultaneous and cyclic descent algorithms when solving for the target pattern in Fig. 5.9.

note that there is a very well-defined separation (dark-area) between the horizontal space and grating line-ends; something very challenging to obtain using conventional H-V decomposition.

5.4.4 Verification using Prolith

We now want to verify our DEL-ILT aerial images using Prolith (a commercially available lithography simulation software). Fig. 5.15 illustrates a (complicated) target pattern and the synthesized CPL masks obtained using Method-2. The target pattern was provided by Intel and consists of 60nm features on a 120nm pitch, the mask pixel size is 20nm, $NA = 0.93$, and $k_1 = 0.29$. The corresponding aerial images obtained using our forward model (5.5) are given in Fig. 5.16. We observe that the algorithm attempts to alternately distribute the column of contacts between the two exposures. Thus it decomposes the original pattern into two comparatively sparser patterns which can then be tackled separately during the individual exposures. The

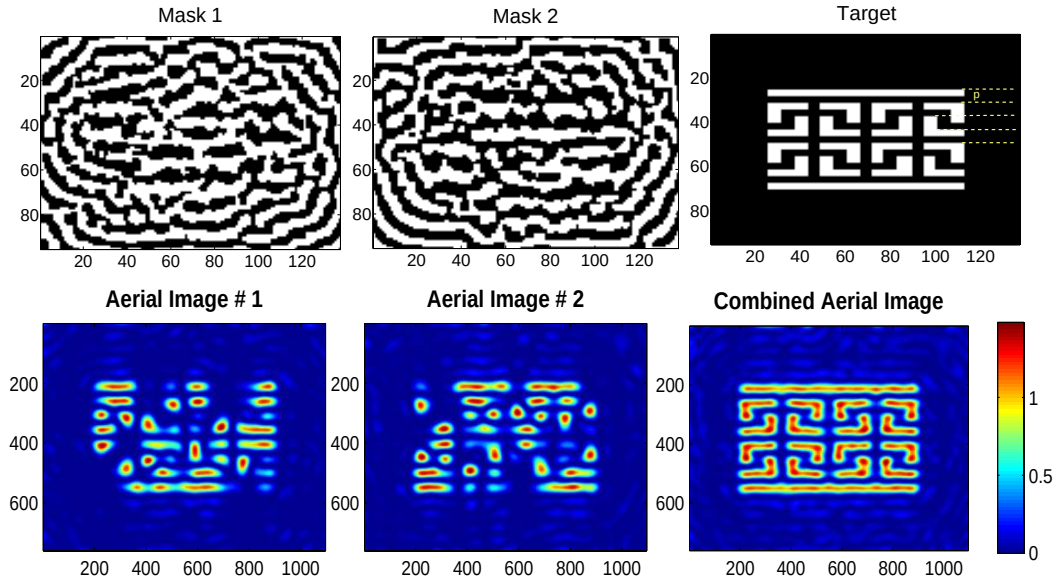

Figure 5.13: Top row illustrates the estimated CPL masks (black=-1 and white=+1) and the (binary) target pattern consisting of 60nm features. The pitch p=120nm is indicated by dotted lines in the figure. Bottom row indicates the aerial images for the individual exposures and the combined aerial image. Here $NA = 0.93$ and $k_1 = 0.289$.

combined aerial image has a superior contrast as observed in Fig. 5.16.

Fig. 5.17 illustrates the aerial images obtained by feeding the estimated ILT masks (in Fig. 5.15) to Prolith (provided by KLA-Tencor). The simulation was performed assuming Kirchhoff (thin-mask) approximation and using the same experimental settings as earlier. We observe that the results are in good overall agreement with those obtained in Fig. 5.16. The slight discrepancy in the results may be due to the numerical approximations used by Prolith to simulate the sharp cutoff of a coherent imaging system (an *ideal* low-pass filter). We believe that the differences can be overcome if we employ the *exact* same forward model as Prolith.

The DEL-ILT framework proposed in this chapter was successful in patterning solutions close to the $k_1 = 0.25$ physical limit. In the next chapter, we discuss the

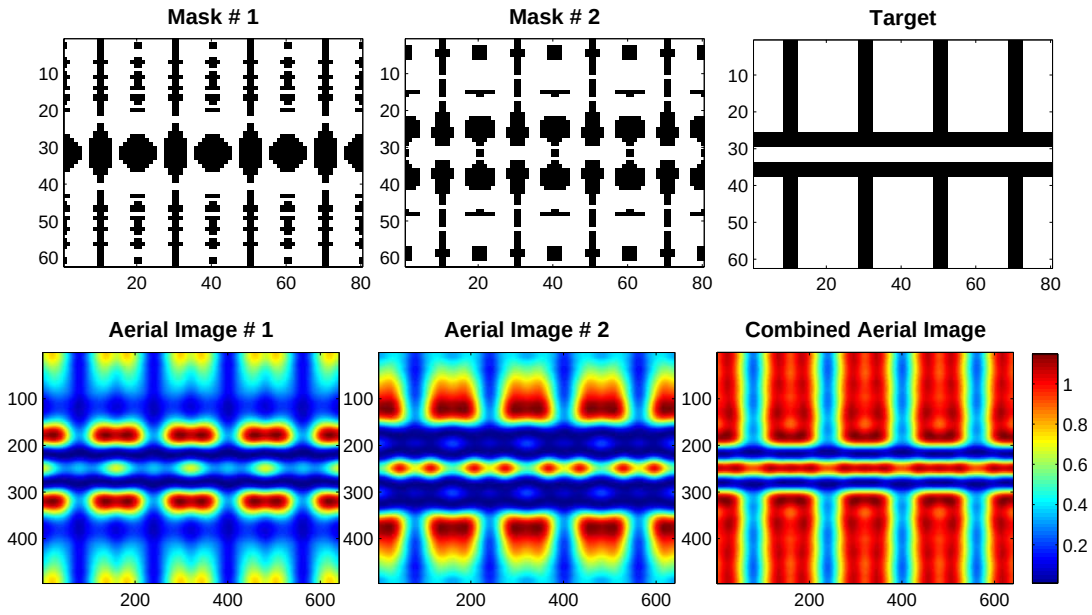

Figure 5.14: Top row illustrates the estimated CPL masks and the target pattern consisting of 80:320 vertical grating interlaced with 80nm wide horizontal features. Bottom row indicates the aerial images for the individual exposures and the combined aerial image. Here $NA = 0.85$ and $k_1 = 0.35$.

extension of ILT to double exposure lithography systems with physical limit $k_1 = 0.125$ enabled by the use of novel materials.

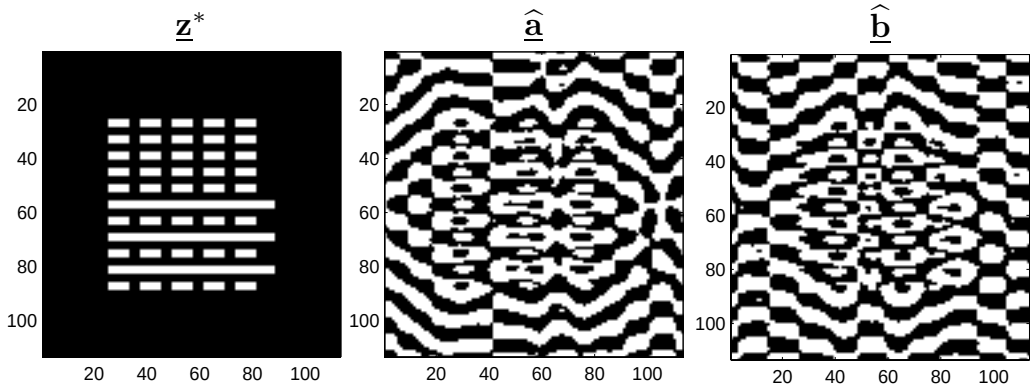

Figure 5.15: The complex target pattern (left) and the two estimated chromeless PSM (center and right).

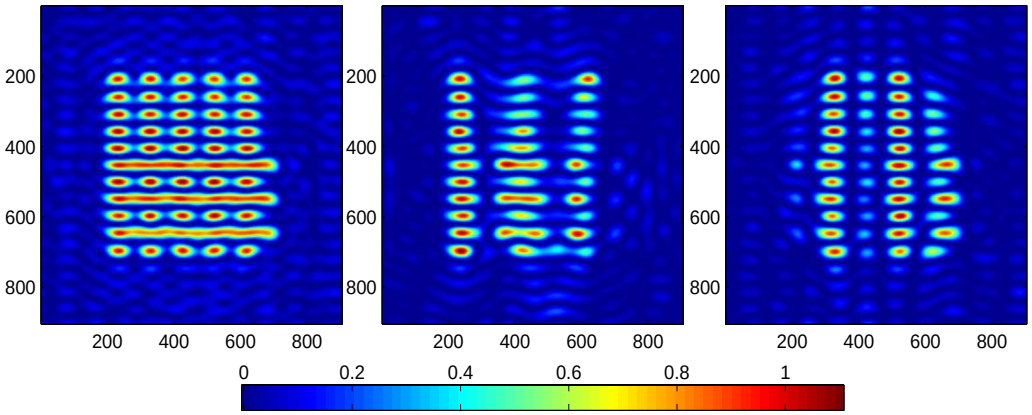

Figure 5.16: The final aerial image (left) and the aerial images corresponding to the two individual exposures (center and right) for the mask patterns in Fig. 5.15

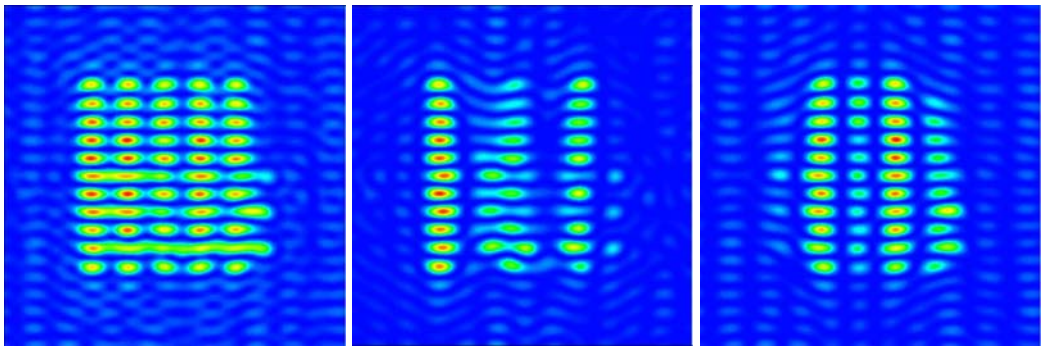

Figure 5.17: The final aerial image (left) and the aerial images for the two individual exposures (center and right) obtained by feeding the mask patterns from Fig. 5.15 to Prolith.

Chapter 6

DEL with Novel Materials for $k_1 < 0.25$

In the previous chapter, we discussed double exposure single development (DEL) lithography using conventional resists. Although DEL gives superior contrast, the resolution limit is the same as that of single exposure. In this chapter, we explore a novel paradigm of DEL which enables patterning below the Rayleigh limit ($k_1 < 0.25$).

6.1 Motivation

In case of DEL with conventional resists (discussed in Chapter 5), the aerial image intensities from the two exposures are linearly superimposed. Therefore, the lower bound on k_{pitch} (which dictates the minimum resolvable pitch) is the same as that of single exposure (equal to 0.5). Using the currently available 193nm/1.1NA immersion scanner systems, the minimum resolvable pitch is equal to $0.5 \times 193/1.1 \approx$

87nm. As we approach the 32nm and 22nm nodes, there is an increasing need for lithography techniques which can pattern even tighter pitches. Double patterning (see Section 5.1.2) with a minimum $k_{pitch} = 0.25$ has emerged as a very strong contender to fulfil the above requirement. Dual patterning consists of preserving the image of the first exposure in a sacrificial layer. The above step requires loading and unloading the wafer twice from the exposure tool, which induces overlay and dimensional control (CD) penalty. Furthermore, the extra processing steps such as second develop, hard-mask deposition, and etch steps might result in additional yield liability as well as increased wafer cost

There is strong incentive to identify techniques able to support patterning with two masks containing split design information, yet capable of producing combined final latent image in resist without unloading the wafer off the exposure tool chuck. We thus need only one single exposure to produce an image ready for *pitch doubling*, thereby avoiding the need for a sacrificial layer, the follow-up processing steps, and their corresponding yield liabilities.

In this chapter, we consider the paradigm of double exposure single development (DEL) for $k_{pitch} < 0.5$. Considering the single development scenario, the only way this is possible is by using new out-of-sight out-of-mind (OSOM) materials with properties that allow non-linear combination of the two aerial images obtained from the individual exposures. We have identified two potential candidates with regard to above which are discussed in this chapter. The first one is called contrast enhancement layer (CEL) which was originally proposed in the 1980s in an effort to improve the

resist profiles for single exposure lithography [49, 51]. In this chapter, we propose a mathematical model for CEL, illustrate its working for the 1-D case, and also extend our ILT framework to enable DEL-CEL with 2-D patterns. The second OSOM material is two-photon absorption resist which responds to the square of the intensity of the aerial image. We discuss its suitability for DEL (1-D case) and point to a need for novel materials with much enhanced properties to make such an approach feasible for manufacturing.

6.2 DEL Using Contrast Enhancement Layer (CEL)

We now discuss the CEL as a potential technology enabler for the $k_1 < 0.25$ case. The contrast enhancement layer (CEL) is a photo-bleachable film which is coated on top of the conventional photo-resist (see Fig. 6.1). It was studied in the 1980s in order to improve the sidewall angle of single exposure lithography process [49, 51, 33, 61]. The CEL film is initially unbleached with (ideally) zero or very low transmittance for small dose values. As the incident dose increases, the film starts undergoing bleaching action and the transmittance increases. The bleaching rate is related to the incident energy. The transmittance of an ideal fully bleached CEL film is equal to one. Thus, the film which was initially opaque to the exposure wavelength, becomes transparent upon exposure. When the aerial image is incident on the CEL, higher intensity regions bleach faster compared to lower intensity ones. This non-linear transformation of the incident aerial image leads to a higher contrast of the

dose received by the resist. During the exposure, the unbleached portion of the CEL will act as a portable conformable mask. Finally, we wish to point out that our current model takes a simplified view and ignores the reflections from the substrate.

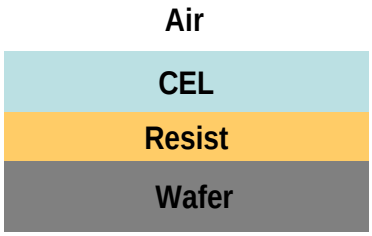

Figure 6.1: The resist stack with CEL.

6.2.1 Modeling

The (non-ideal) bleaching behavior of the CEL is similar to a soft-threshold operation. Therefore, the CEL film transmittance $T(D_i)$ can be modeled as a sigmoid function as follows,

$$T(D_i) = \text{sig}(D_i, a, t_c, c) = \frac{1}{1 + e^{-acD_i + at_c}}, \tag{6.1}$$

where D_i is the incident dose, the parameter c is inversely related to the thickness, a is related to the slope of the transition region, and t_c is the threshold dose parameter of the sigmoid. Fig. 6.2 indicates the behavior of a typical sigmoid function with $a = 60$, $t_c = 0.15$, and varying values of c. As discussed earlier, the transmittance is low for small values of incident dose. As the input dose builds up, the film starts bleaching and its transmittance increases ($T(D_i) = 0.5$ for $D_i = t_c/c$). Fig. 6.2 also indicates the effect of the film thickness c on the film transmittance. A thin CEL

film bleaches quickly compared to a thicker one, and hence has a steeper transition region. Furthermore, a thick CEL film also requires higher dose values to become fully transmissive.

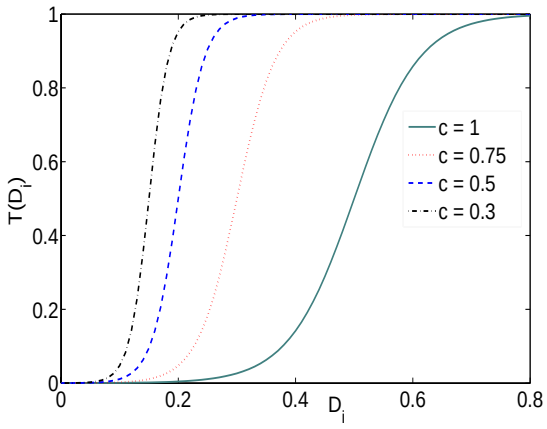

Figure 6.2: CEL film transmittance for different film thickness.

The final transmitted dose $D_t\{D_i\}$ (henceforth referred to as simply D_t), is obtained by integrating the transmittance as the incident dose builds up from 0 to D_i.

$$D_t = \int_0^{D_i} T(D_i) dD_i = \int_0^{D_i} \frac{1}{1 + e^{-acD_i + at_c}} dD_i$$

Therefore,

$$D_t = D_i - \frac{t_c}{c} + \frac{1}{ac} \log \left[1 + e^{(-acD_i + at_c)}\right] \tag{6.2}$$

Fig. 6.3 indicates that the graph of transmitted dose versus incident dose consists of three regions. For low doses (unbleached film), the incident dose is absorbed by the CEL film and D_t is very low. When D_i increases, the film starts bleaching and there is a non-linear increase in the transmitted energy. Finally, the CEL film gets fully bleached after which the transmitted dose increases linearly. As one would expect, the

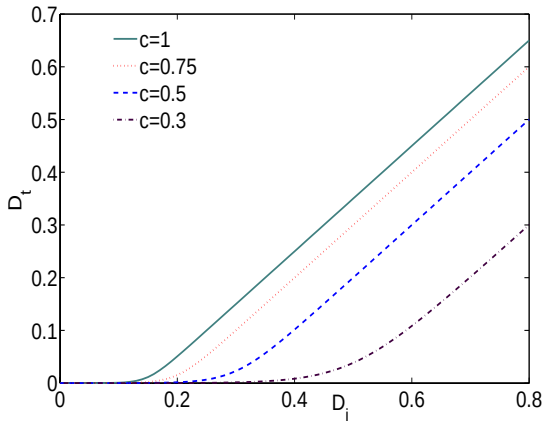

Figure 6.3: Transmitted dose (D_t) versus input dose D_i for different film thickness.

above regions also depend on the thickness of the CEL film. Thinner CEL films enter the linear region earlier compared to the thicker ones.

6.2.2 Working of CEL

We now discuss the working of CEL for the simple example of 1-D gratings. The target consists of 80nm (1:1) grating which needs to be printed using 193/0.85NA ($k_{pitch} = 0.35$). The first step is to divide the target into two 160nm pitch gratings (also see Fig. 5.2) which can be printed using single exposure CPL masks (with pitch = 320nm). Note that the second exposure consists of simply shifting the mask by 80nm. The left plot in the top row in Fig. 6.4 indicates the aerial image and the transmitted CEL image (calculated using (6.2)), both for the first exposure. The CEL transmitted dose is absorbed by the photo-resist and stored in the form of PAC (photo-active compound) concentration. The right plot indicates the aerial image and the CEL dose for the second exposure. It is very important to highlight here that in

order to support DEL, the CEL considered should have *reversible bleaching* property. In other words after the first exposure is completed, the entire CEL film returns back to its original (opaque) state. If the CEL were not reversible, the regions previously exposed would remain bleached and as such have different transmittance values than those indicated in Fig. 6.2. Development of the above material is central to the use of CEL for double exposure. Such materials are not just hypothetical as is evidenced by class of reversible CEL (Arsel$^{\text{TM}}$) under development by Pixelligent Technologies LLC.

The total dose absorbed by the resist is a superposition of the two CEL transmitted doses. The graph in the bottom row of Fig. 6.4 indicates that the combined transmitted CEL dose (assuming the reversibility property) shows good modulation. An intuitive explanation is that the CEL suppresses the aerial image energy in the unwanted regions owing to the soft-thresholding action. This has the effect of truncating the tail of the optical PSF (making it narrower), and resulting in less energy being deposited in the areas targeted by the second exposure. Fig. 6.4 also indicates the aerial image in the absence of CEL layer (just for demonstration purpose). Since $k_{pitch} < 0.5$, no useful modulation is observed.

Image log-slope (ILS) is a commonly employed metric to evaluate the quality of the aerial image. The ILS at the desired edge location is evaluated as,

$$\text{ILS}(D_t) = \frac{1}{D_t}\frac{dD_t}{dx}$$

Higher ILS value implies higher image slope, which in turn gives good contrast. The

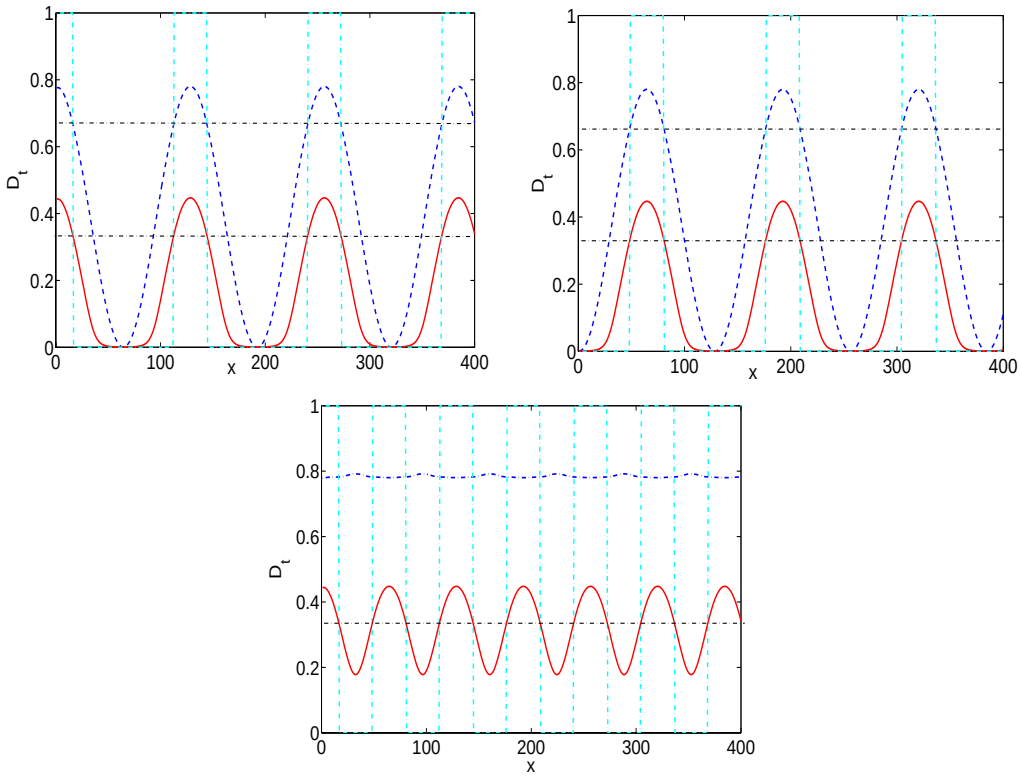

Figure 6.4: The two figures in the top row indicate the aerial images (dotted blue) and the CEL images (solid red) for the two individual exposures. The bottom row indicates the combined aerial image (showing no modulation) and the combined CEL image (showing good modulation). In this case $a = 30$, $t_c = 0.15$, and $c = 0.45$.

calculated ILS at the desired edge location (CD = 40nm) for the incoming aerial image (D_i) and the CEL image (D_t) are 0.016 and 0.032 respectively. Thus the CEL also improves the aerial image contrast which was indeed one of the earliest applications of CEL [60]. Finally, we also observe that the transmitted CEL doses have lower magnitude compared to the incoming (aerial image) dose. We discuss this in more detail in the next section.

6.3 ILS Improvement of the CEL-DEL Image

Our goal now is to improve the robustness of the double exposure CEL-based lithography system. This will also enable us to understand the fundamental trade-offs involved in designing a CEL-DEL system. We have identified three potential ways to improve the operating ILS at the desired edge location which are discussed below. The experimental setup employed here is the same as in Section 6.2.

6.3.1 CEL Film Thickness

The effect of the CEL film thickness parameter c on the film transmittance and the transmitted dose were illustrated earlier in Fig. 6.2 and Fig. 6.3 respectively. Fig. 6.5 illustrates the effect of c on the transmitted dose modulation assuming a fixed value of the incoming (aerial image) dose modulation D_i. The remaining simulation parameters are fixed ($a = 30$, $t_c = 0.15$), and c was varied from 0.8 to 0.2. We observe that as the CEL film becomes thicker, the magnitude of the transmitted dose modulation decreases. Therefore, in order to obtain 40nm CD, the photo-resist threshold t_r should be equal to 0.48, 0.36, and 0.076 for $c = 0.8, 0.5$, and 0.2 respectively. The ILS values for the corresponding t_r values are 0.022, 0.03, and 0.054. To summarize, the ILS can be improved by using a thicker CEL film. However, patterning the above image also requires very sensitive (high-acid generating) resists, having very low t_r value.

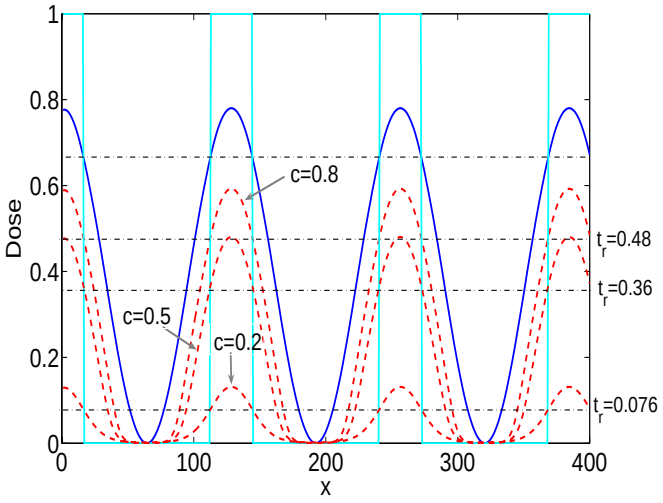

Figure 6.5: Effect of CEL film thickness assuming fixed D_i.

6.3.2 Incoming Dose Magnification

The second mode of operation is to use the same photoresist (fixed value of t_r), and magnify the incoming dose by a factor p. The CEL film thickness c also needs to be adjusted in order to match D_t with the target CD at the chosen threshold value t_r. Fig. 6.6 illustrates the original aerial image (D_i) and the transmitted CEL doses for four different values of p assuming a resist with operating threshold $t_r = 0.33$. The corresponding CEL thickness parameters are also indicated in the legend. We observe that as the dose magnification increases, the transmitted CEL image quality improves. The ILS values corresponding to magnification factor $p = 2, 3, 5$, and 7 are $0.032, 0.047, 0.072$, and 0.087 respectively. Thus we conclude that for a given resist, the ILS can be improved by magnifying the incoming dose and using a thicker CEL film. Dose magnification can be realized in practice using a more powerful light source

or longer exposure times. A very high intensity light source endangers the operating life of the (very expensive) lenses, and is a less preferable option. On the other hand, longer exposure times lead to reduction in the overall yield. Hence, we are interested in finding out the maximum "mileage" (in terms of ILS improvement) possible for a given magnification factor.

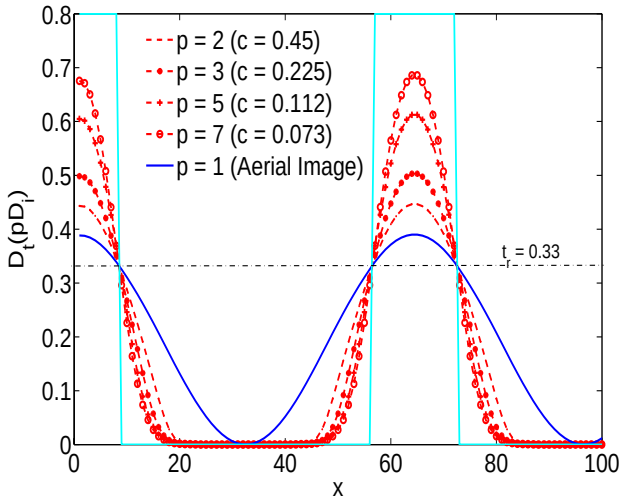

Figure 6.6: Effect of dose magnification p on the CEL image using fixed resist with threshold ($t_r = 0.33$). Here, $a = 60$ and $t_c = 0.15$.

It can be proved (see Appendix C) that for a given value of p, the ILS of the transmitted dose is upper-bounded by p times the ILS of the original incoming dose D_i. In other words,

$$\text{ILS}(D_t(pD_i)) <= p \times \text{ILS}(D_i) \tag{6.3}$$

The above equation is a fundamental relationship which will also be beneficial to evaluate the quality of the CEL material. A good CEL material should be capable of improving the ILS by a full-factor p (particularly for large values of p).

6.3.3 Quenchers

The third and final way to improve the ILS is motivated by the fact that the combined CEL image sits on top of a background energy, as observed in Fig. 6.4. As observed earlier, ILS at the desired edge location is evaluated as

$$\text{ILS}(D_t) = \frac{1}{D_t}\frac{dD_t}{dx}$$

Therefore a possible way to improve the ILS is by adding (base) quenchers to the photoresist. Quenchers are alkaline-based chemicals which can be used to neutralize the extra acid generated by the (undesired) background energy present in D_t. The above process is equivalent to subtracting a constant q ($q \geq 0$) from D_t thereby effectively shifting the entire modulation curve (indicated in Fig. 6.4) down. The effect on ILS is as follows:

$$\text{ILS}(D_t - q) = \frac{1}{D_t - q}\frac{d(D_t - q)}{dx} = \frac{1}{D_t - q}\frac{d(D_t)}{dx} > \text{ILS}(D_t)$$

Thus the effective ILS is higher than before.

6.3.4 Matching with Arsel$^{\text{TM}}$

We now compare the behavior of the proposed CEL model with the measured data of Arsel$^{\text{TM}}$kindly provided by Pixelligent Technologies LLC. Fig. 6.7 indicates the measured optical density (calculated as $\text{OD} = \log(D_i/D_t)$) as a function of input dose D_i. The blue and red curves correspond to the measured OD when D_i is increased from 0 to maximum and decreased back to 0, respectively. The two curves will coincide if the CEL is completely reversible. The graph on the left represents the OD calculated

using the proposed model with parameters $a_c = 1.6$, $t_c = 2.5$, and $c = 0.3$. Both graphs indicate a structurally similar behavior; a linear drop in the OD (attributed to bleaching) followed by flattening at higher dose values (attributed to very high $T(D_i)$).

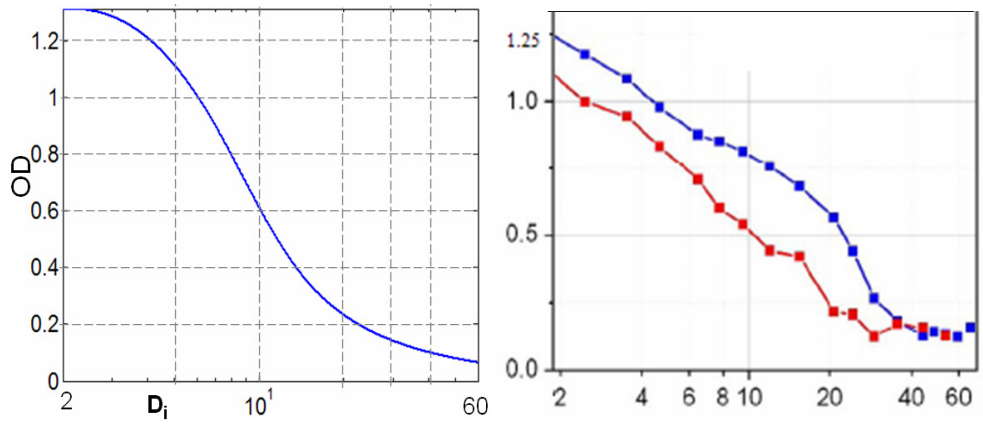

Figure 6.7: OD versus D_i for the proposed model (left) and the measured data using Arsel^{TM}(right).

6.4 ILT using CEL for 2-D Patterns

In the last section, we demonstrated the working of CEL for 1-D gratings. We now demonstrate the applicability of CEL for printing 2-D patterns using DEL for $k_1 < 0.25$. The goal is to use ILT to synthesize CPL masks for printing the target pattern in Fig. 6.8 which consists of 40nm features on an 80nm pitch using 193/0.9NA ($k_1 = 0.187$). The target is sampled at 20nm implying that the features are merely two 20nm pixels wide.

Firstly, our DEL-ILT algorithm (with the integrated CEL model) did not converge to an acceptable local minima for $k_1 < 0.25$. Therefore, we were unable to perform the automatic pattern decomposition similar to that illustrated in Section 5.4.

We suspect that this may be due to the local gradient-based search technique used for optimization. The algorithm is unable to explore the more complicated search space resulting from $k_1 < 0.25$, and better heuristic based initialization points may help solve the problem. However, this remains a direction of future work.

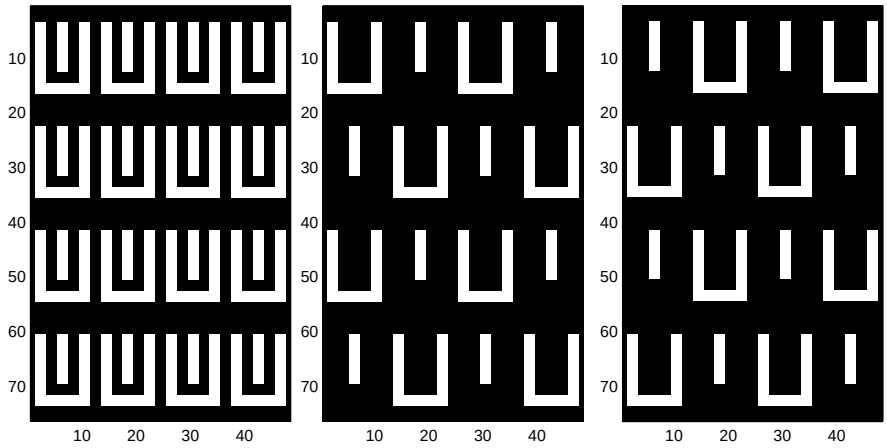

Figure 6.8: The target pattern consisting of 40nm features on 80 nm pitch and the (manually) decomposed patterns used for individual exposures.

We therefore perform a rule-based pattern decomposition as illustrated in Fig. 6.8 and employ the single exposure based ILT algorithm to synthesize the CPL masks. The augmented forward model also accounting for the CEL is as follows:

$$\underline{\mathbf{z}} = F(G(p|\mathbf{H}\underline{\mathbf{m}}|^2)), \tag{6.4}$$

where the function $G(.)$ represents the CEL transmittance function defined in (6.2). $F(.)$ is a sigmoid function which models the thresholding operation performed by the resist (see Section 2.3.1),

$$F(y) = \text{sig}(y) = \frac{1}{1 + \exp(-a_r y + a_r t_r)}.$$

Here a_r and t_r are the slope and threshold parameters of the resist respectively. Thus the forward model in (6.4) accounts for the magnified incoming aerial image dose, the CEL transferred dose, as well as the resist contour formation. The estimated mask $\underline{\widehat{\mathbf{m}}}$ is defined as one which minimizes the L_2 norm of the difference between the target contour $\underline{\mathbf{z}}^*$ and output contour $\underline{\mathbf{z}}$ modeled using (6.4),

$$\underline{\widehat{\mathbf{m}}} = \arg\min_{\underline{\mathbf{m}}} C_1(\underline{\mathbf{m}}) = \arg\min_{\underline{\mathbf{m}}} ||\underline{\mathbf{z}}^* - \underline{\mathbf{z}}||_2^2. \tag{6.5}$$

The cost function in (6.5) is still continuous allowing the use of non-linear programming techniques for optimization. We follow our earlier proposed parametric transformation (Section 3.1) and regularization (Section 4.1.2) techniques, and employ the gradient-descent optimization algorithm to minimize $C_1(\underline{\mathbf{m}})$. The gradient can be calculated as follows:

$$\nabla C_1(\underline{\mathbf{m}}) = -4a_r p \mathbf{H}^{\mathbf{T}}[(\underline{\mathbf{z}}^* - \underline{\mathbf{z}}) \odot \mathbf{z} \odot (\underline{\mathbf{1}} - \underline{\mathbf{z}}) \odot (\mathbf{Hm}) \odot \underline{\mathbf{q}}], \tag{6.6}$$

where

$$\underline{\mathbf{q}} = \frac{\underline{1}}{\underline{1} + \exp(-a_c c p |\mathbf{H}\underline{\mathbf{m}}|^2 + a_c t_c)}.$$

Fig. 6.9 illustrates the results obtained using the CEL-ILT algorithm discussed above. The top row indicates the estimated CPL mask (white = 1 and black = -1), the aerial image, and the CEL image for a single exposure. The simulations parameters were: $a_c = 60$, $t_c = 0.15$, $c = 0.2$, $p = 4$, $a_r = 60$, $t_r = 0.3$. We observe that although the aerial image is very poor, the CEL transmitted image (D_t) has much better contrast, and closely resembles the target. This is because our ILT algorithm is

aware of the presence of CEL, and estimates the masks taking its action into account. By the end of the first exposure, the CEL transmitted dose is completely absorbed by the photoresist, where it is stored in the form of photo-active-compound (PAC) concentration. The second exposure is performed by simply shifting the mask horizontally by 320nm (see Fig. 6.8). The bottom row in Fig. 6.9 indicates that the combined aerial image (the dose received by the resist in the absence of CEL) has no useful modulation. This is expected because $k_{pitch} < 0.5$. However, the linear superposition of the two CEL images shows very good modulation, which will enable patterning of the desired features. The contours at $t_r = 0.3$ indeed indicate good pattern fidelity.

6.5 Two-Photon Absorption Resist

The second OSOM material candidate for $k_1 < 0.25$ DEL is two-photon absorption (2-PA) resist. They have been well-known to double the resolution of optical lithography, and were actively explored by Yablonovitch, et al. [95].

The left graph in Fig. 6.10 illustrates the aerial image *intensity* obtained using an interferometric arrangement with 2-coherent beams incident at $\pm\theta$. Therefore, $I(x) = 1 + \cos(Kx)$ where $K = 2k \sin\theta$ and $k = 2\pi/\lambda$. For the case of two-photon absorption, the resist responds to the *square* of the intensity. Thus $I^2(x) = 3/2 + 2\cos(Kx) + 1/2\cos 2Kx$, which consists of superposition of both the single and double-frequency components. As observed in Fig. 6.10, the above mixture merely enables patterning at the same resolution as single-photon absorption (although with a better

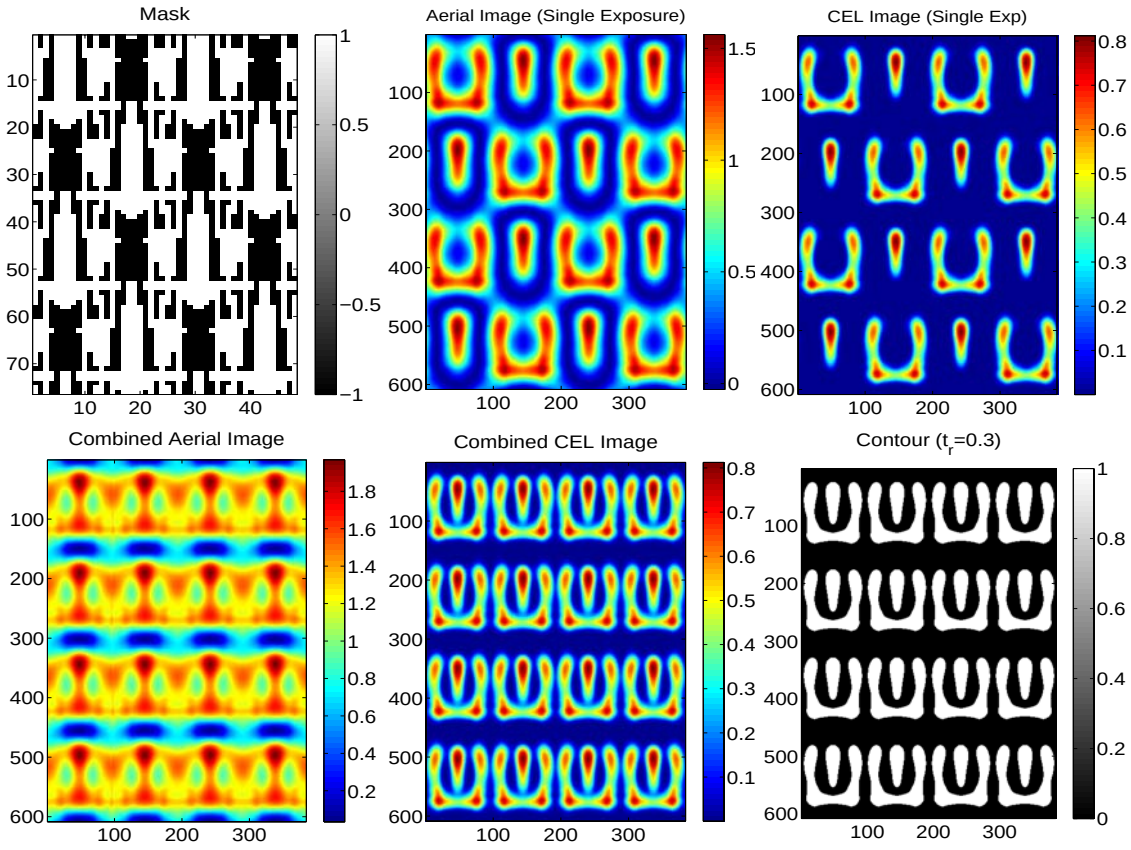

Figure 6.9: The top row indicates the CEL-ILT results obtained by solving (6.5). The bottom row indicates the DEL results where the second exposure was performed by shifting the mask by 240nm.

contrast). Therefore, we need to suppress the linear frequency component in order to enable frequency doubling, and allow $k_{pitch} < 0.5$. Yablonovitch, et al. [95] proposed a new (single) exposure arrangement which destroys the stationary interference patterns corresponding to the linear frequency component, leaving behind the double-frequency interference pattern on a background.

An alternative way to enable frequency doubling using 2-photon absorption resists is by employing DEL. In the case of gratings, the second exposure should have

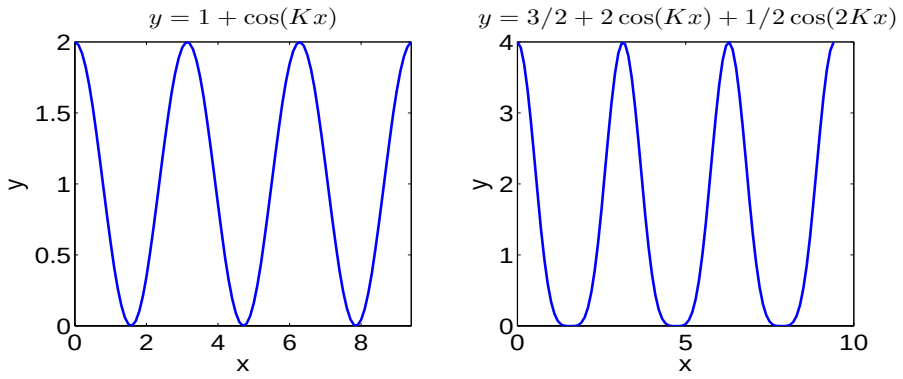

Figure 6.10: The intensity and square of intensity modulation curves corresponding to the conventional and 2-photon absorption resists respectively.

an offset of half a period compared to the first. Therefore the combined contribution from the two exposures is given as follows:

$$\begin{aligned} I_1^2(x) + I_2^2(x) &= [1 + \cos(Kx)]^2 + [1 + \cos(Kx + \pi)]^2 \\ &= [1 + 2\cos(Kx) + \cos^2(Kx)] + [1 - \cos(Kx)]^2 \\ &= [1 + 2\cos(Kx) + \cos^2(Kx)] + [1 - 2\cos(Kx) + \cos^2(Kx)] \\ &= 2 + 2\cos^2(Kx) \\ &= 3 + \cos(2Kx) \end{aligned}$$

We observe that by merely phase shifting and adding $I^2(x)$ to itself, the linear frequency component cancels out, leaving behind only the double-frequency component with a DC background. Thus 2-photon absorption resists are naturally suitable to DEL. However, we are not yet aware of materials exhibiting suitably large 2-photon absorption cross section (δ) values at the wavelength of interest with existing radiation sources. In order to get an estimate of the required value for δ, lets assume

a lithography system with a radiation source consisting of a 193nm laser with 10ns pulse length (l), 60W output power (P), and 6KHz repetition rate (f_p). Therefore, its corresponding peak intensity (I) within 1cm^2 exposure slit area A will be,

$$I = P/(Af_pl) = 60/(6 \times 10^3 \times 10^{-8}) = 10^6 \mathrm{W/cm}^2$$

Conventional photo-resists are usually engineered with one-photon absorption coefficient of $\alpha \approx 10^4 \mathrm{cm}^{-1}$. For "two-photon resist" one would need the two-photon absorption to be at least equal to the one-photon absorption. Therefore, the two-photon absorption coefficient can be calculated as,

$$\beta = \alpha/I = 10^4/10^6 = 10^{-2}\mathrm{cm/W} \tag{6.7}$$

We can then estimate a lower bound on the required 2-photon absorption cross-section for such materials (for use with existing 193nm radiation sources) by employing the following equation:

$$\delta = \beta \frac{h\nu}{N}$$

Substituting β from the calculations in (6.7), $N = 10^{20}$ (number of 2-photon absorption sensitive molecules per cm^3), and $h\nu = 10^{-18}$J (excitation photon energy for 193nm photons) gives $\delta = 10^{-2} \times 10^{-18}/10^{20} = 10^{-40}\mathrm{cm}^4 \times \mathrm{s} \times \mathrm{photon}^{-1}$.

The (above calculated) required minimal value for 2-photon absorption cross section is many orders of magnitude higher than $\delta = 1 \times 10^{-46}$ cm$^4 \times s \times \mathrm{photon}^{-1}$ (in the near-IR), which is amongst the highest δ values ever reported for any organic molecule [21, 3, 20]. This makes the prospects of synthesizing such material for use with

existing 193nm exposure systems highly unlikely. For example, an attempt to drastically reduce the exposure slit area (in order to raise the peak intensity), will also result in correspondingly drastic loss of system productivity, thereby making it unacceptable for volume manufacturing. Another possible option to find imaging solutions worthy of volume IC manufacturing is by employing femto-second laser sources (such as traditionally used in 2-photon microscopy or MEMS manufacturing). Although this enables the use of materials with more realistic, yet still gigantic $\delta = 10^{-46}\text{cm}^4 \times s \times \text{photon}^{-1}$, it also has its own forbidding set of problems. High volume manufacturing lithography that require 1 or 2 cm^2 working (output) area in support of needed productivity levels, will typically require sub-200nm wavelength femto-second lasers with power output exceeding the existing ones by several orders of magnitude. Unfortunately, imaging systems with such source will also face equally hard challenges to deliver diffraction-limited exposure system solutions, that can correct the large chromatic aberrations arising due to the use of a femto-second laser source. Furthermore, it is also challenging to find optical materials that will not degrade rapidly under the unusually high power densities present in such imaging system. Overall in our opinion, the efforts in exploring two-photon absorptive materials in support of pitch doubling with double patterning for 32nm half pitch and below, do not appear to be a promising direction to pursue at this time.

Thus, we explored two potential candidates which enable frequency doubling without the need of the second development step. Based on the current material availability, CEL appears to be a more promising candidate. Future work in this direction

consists of modeling the non-ideal transparent and opaque behavior of the CEL film for the fully bleached and unbleached states, and accounting for the reflections from the CEL-resist interface. It will also be very interesting to verify the effect of CEL thickness using Arsel^{TM}.

Chapter 7

Contributions, Impact, and Future Work

In this chapter, we first discuss the contributions of this thesis to the area of mask design using inverse lithography. Since the inception of this project, inverse lithography based OPC, PSM, and SRAF solutions have generated much of interest in the industry and academia alike. Therefore, we next discuss some of the very contemporary work (published in the past 2 years), and connect them to our proposed framework. Finally, we identify some key areas of future work, and some interesting problems in the lithography industry, which may benefit from an inverse imaging approach.

7.1 Contributions

In this thesis, we proposed a novel inverse lithography framework for pixel-based OPC and PSM mask design. The key contribution of this work is to reduce

the mask design problem to a continuous function optimization problem, and use analytically calculated gradient to systematically explore the solution space (via the steepest-descent algorithm). We also introduced the regularization framework to inculcate user-defined properties in the synthesized mask solutions. The inverse lithography approach was successfully extended to double exposure lithography, which led to a new paradigm of aerial-image based pattern decomposition. We sincerely hope that our ILT approach will help the lithography industry to design new patterning solutions in pursuit of Moore's law.

- In Chapter 2, we posed the single exposure inverse lithography mask design problem. We noted that the binary input and output nature of the problem had prohibited the past researchers from employing non-linear programming techniques. To overcome this, we introduced the sigmoid function approximation to the hard threshold (resist) operation, and reduced the output to the continuous domain. We then discussed the forward process model for coherent, incoherent, and partially coherent imaging systems, and formulated the mask design problem as minimizing the L_2 norm of the difference between the (binary) target and the (near-binary) output pattern. Finally, we substituted the integer RET constraints by bound constraints, thereby reducing the mask design problem to a bound-constrained continuous function optimization problem.

- In Chapter 3, we discussed the single exposure OPC mask design algorithm. The ILT mask design problem was first reduced to an unconstrained optimiza-

tion problem. We then calculated the analytic gradient of the cost function, and employed a steepest-descent algorithm to search for the optimal mask solution. The analytic gradient had a computational complexity of $O(MN \log(MN))$ making it practically feasible. We then introduced the regularization framework to control the tone and complexity of the estimated mask patterns, thereby easing the mask binarization and mask manufacturing process. We also introduced an aerial image penalty term for promoting SRAF generation. The results demonstrated automatic generation and placement of assist bars, serifs, anti-serifs, and un-intuitive feature-splitting. The pattern fidelity was superior and side-lobes were not printed. Finally, we extended the ILT framework to weighted-region and selected-parameter optimization.

- In Chapter 4, we addressed the single exposure PSM mask design problem. We introduced the tone control penalty terms for AttPSM, 100% transmission PSM, and strong PSM with chrome. The estimated masks obtained using ILT were very un-intuitive and did not resemble the target. However, the results demonstrated good aerial image contrast and pattern fidelity for low k_1 values.

- In Chapter 5, we motivated and posed the double exposure inverse lithography problem. Our approach involved estimating two real-valued masks, which was shown to be equivalent to estimating one complex-valued mask. We formulated the mask design problem as minimizing the L_2 norm of the difference between the aerial image and the target, and proposed two optimization algorithms to search

the $2MN$ dimensional space. The results indicated a very interesting paradigm, whereby the proposed algorithms decomposed the *gray-level* aerial image into two parts, and automatically resolved the phase-conflicts. The combined aerial image demonstrated very good contrast, and enabled patterning of very complex target patterns close to $k_1 = 0.25$.

- In Chapter 6, we explored a novel paradigm of pitch doubling using double exposure single development lithography with novel materials allowing non-linear superposition of the two aerial images. We proposed a mathematical model for reversible CEL, and three different ways of improving the image log slope. The single exposure ILT framework was successfully extended to synthesize CPL masks using the new CEL-resist stack. We also concluded that 2-photon absorption resists were not suitable for dual patterning applications due to lack of materials with desired absorption cross-section.

7.2 Recent Work

We now discuss some recently published literature on the inverse lithography mask design technique.

7.2.1 Luminescent Technologies

The inverse lithography problem was first formulated by Saleh, et al. [80] in the 1980s. However, ILT never found commercial success owing to lack to efficient algorithms, and the satisfactory performance provided by edge-based techniques. The

shortcomings of edge-based OPC solvers became prominent at the 65nm nodes, which led to the debut of Luminescent Technologies (in 2005), who launched the first "commercial" software (Luminizer) for mask design based on ILT [63, 44, 45].

The Luminescent approach employs a level-set function $\psi(x, y)$, and the contours of the photomask pattern are represented using the "level-set" $\psi(x, y) = 0$. The merit function referred to as the Hamiltonian measures the similarity between desired target pattern and the printed wafer pattern. The mask design problem is formulated as estimating the level-set function $\hat{\psi}(x, y)$ which minimizes the Hamiltonian. The above problem was solved using (analytically calculated) gradient-descent algorithm, and the regularization framework was also employed to impose mask-manufacturing constraints. Their results illustrated sizing of the main feature while placing assist-bars [55] and feature-splitting [45], both of which were also observed by us in Section 3.4.

7.2.2 Mentor Graphics

Granik [31, 32] (2005/2006) used pixel-based mask representation and sum of coherent imaging system (SOCS) formulation to solve the inverse lithography problem. The cost function was defined as the L_2 norm of the difference between the target and the aerial image (as defined in (3.13)). Contour fidelity was accounted by using an additional term, which penalized the deviation of the intensity from the given threshold t_r along the boundary C of the target pattern [80]. He also suggested the use of penalty terms to account for the tone and mask manufacturing requirements (although no

results were presented). Finally, random local variation and analytic gradient based optimization techniques were proposed to solve the optimization problem, and a post-processing clean-up step was employed to simplify the SRAF.

7.2.3 GILT - University of Delaware

In another work recently reported in March 2007, Xu, et al. [50] inspired by our ILT framework, employed the same forward model, sigmoid formulation, and analytic gradient-based technique discussed in this thesis. The only difference is that they propose to estimate complex-valued masks for single exposure lithography. As discussed in Section 5.3 and Appendix B, the above extension is exactly equivalent to estimating two real-valued masks for double exposure lithography (DEL). The authors in [50] also employ the regularization framework to restrict the mask to have four phase values, thereby enabling them to overcome phase-conflicts. The above is also equivalent to estimating two 100% transmission strong PSM as observed in Appendix B and Section 5.4.

7.2.4 Invarium Inc.

Yenikaya and Sezginer [96] (March 2007) use Fourier series expansion of the mask transmission function, and optimize an L_2 norm cost function with respect to the mask Fourier coefficients. The cost function measures the aerial image fidelity (similar to (3.13)) over a uniform-grid (pixels), and the contour fidelity is measured at target-points placed along the edges of the pattern. The mask transmission values

are penalized by OPC or PSM bound constraints (as seen in (2.17)), thereby reducing the problem to the continuous domain. The estimated continuous tone mask is then converted to manufacture-friendly polygons using a post-processing step.

7.2.5 Synopsys Inc.

In another recent work (March 2007), Li, et al.[43] analytically calculate the change in the gradient of the aerial image due to the addition of an assist bar (for an isolated feature). The optimal position of the assist feature is determined by scanning all the positions in the vicinity of the main feature, and choosing the one which maximizes the aerial image gradient. The ILT framework utilizes the analytic gradient of an aerial image-based cost function to directly determine the optimal size and position of the assist feature (as observed in Fig. 3.11).

In conclusion, we observe that there has been an increasing activity in the inverse lithography mask design approach in recent years[1]. The underlying similarity between all these works is the desire to avoid the integer optimization, and focus on continuous function formulation with gradient-based optimization techniques. The systematic exploration of the solution space coupled with distributed computing power have enabled ILT-based techniques to be extendible to full-chip level [44] making them attractive candidates for 45 and 32nm nodes.

[1] The work discussed in the thesis has also been published in several conferences and leading journals [73, 72, 71, 70].

7.3 Future Work and Extensions

In this section, we present our preliminary results for alignment error tolerant DEL-ILT solutions. We also discuss extensions of the ILT mask design problem, and some new avenues in the semiconductor industry, where an inverse imaging approach may be potentially beneficial.

7.3.1 Robustness of DEL to Alignment Error

In practice, the flowchart in Fig. 5.1 indicating the steps involved in double exposure is missing a critical step. Before performing the second exposure, the second mask needs to be positioned at exactly the same place as the first one. The above step called mask-to-mask alignment is an integral part of the DEL procedure. Unfortunately, the mask alignment error (also called *overlay error*) is inherent to every exposure tool. The traditional H and V decomposition (see Fig. 5.4) is very sensitive to overlay, and the mask alignment error directly translates into wafer CD error. This imposes very stringent overlay accuracy requirements. However, recall that DEL-ILT is an ill-posed problem, and there can be infinitely many ways of decomposing the pattern. In this section, we explore the possibility of choosing an alternate pattern decomposition schema, which leverages the effect of mask misalignment. Thus, our goal is to split the pattern in such a way that the effect of mask alignment error on the final wafer CD is minimized, thereby increasing the robustness of the lithography system.

The forward model discussed in (5.5) only accounts for the aerial image calcu-

lations. Since alignment error is measured on the wafer CD, our first step is to account for the resist effects in the DEL-ILT modeling. The augmented forward model is given as,

$$\underline{\mathbf{z}} = \text{sig}(|\mathbf{H}\underline{\mathbf{a}}|^2 + |\mathbf{H}\underline{\mathbf{b}}|^2). \tag{7.1}$$

The resist is modeled using a soft threshold sigmoidal function (defined in Chapter 2), where the latter acts on the sum of the aerial image intensities of the two exposures. Note that $\underline{\mathbf{z}}$ here corresponds to the output wafer pattern.

In our analysis, we assume that the second mask is positioned exactly in the same horizontal mask-plane (X-Y) as the first one, and hence there is no displacement along the Z-direction (or the optical axis)[2]. The mask misalignment is only due to linear translational motion of the second mask in the X-Y plane. The probability distribution of the tool alignment error can be obtained from the production job data. The above can be used to setup the DEL-ILT for the 3σ misalignment case. Alternatively, we can also solve for the worst case alignment error.

The double exposure model in (7.1) can be augmented to account for the displacement of the second mask:

$$\underline{\mathbf{z}} = \text{sig}(|\mathbf{H}\underline{\mathbf{a}}|^2 + |\mathbf{H}\mathbf{S}_\mathbf{x}\mathbf{S}_\mathbf{y}\underline{\mathbf{b}}|^2), \tag{7.2}$$

where $\mathbf{S}_\mathbf{x}, \mathbf{S}_\mathbf{y} \in \Re^{MN \times MN}$ are matrices (operators) which shift the mask $\underline{\mathbf{b}}$ along the horizontal and vertical directions by 1 pixel. Thus, the mask is assumed to be misaligned in both directions. Since our principal objective is to demonstrate the

[2]Displacement along the optical axis causes de-focus errors, which can be accounted for while optimizing for the process window.

existence of alignment error tolerant mask solutions, we currently restrict our study to the simple case of full-pixel motion. The forward model (7.2) can also be extended to the more general cases of sub-pixel motion, but this forms a direction of future research.

The optimization problem can now be formulated as estimating the masks $\widehat{\underline{\mathbf{a}}}$ and $\widehat{\underline{\mathbf{b}}}$ such that the contour fidelity in the presence of alignment error is minimized. Therefore,

$$[\widehat{\underline{\mathbf{a}}}, \widehat{\underline{\mathbf{b}}}] = \arg\min_{\underline{\mathbf{a}},\underline{\mathbf{b}}} \|\underline{\mathbf{z}}^* - \text{sig}\left(|\mathbf{H}\underline{\mathbf{a}}|^2 + |\mathbf{H}\mathbf{S}_\mathbf{x}\mathbf{S}_\mathbf{y}\underline{\mathbf{b}}|^2\right)\|_2^2. \tag{7.3}$$

The gradient of the above function with respect to $\underline{\mathbf{a}}$ and $\underline{\mathbf{b}}$ is given as,

$$\nabla_{\underline{\mathbf{a}}} F(\underline{\mathbf{a}}, \underline{\mathbf{b}}) = -\mathbf{H}^\mathbf{T}\left[(\underline{\mathbf{z}}^* - \underline{\mathbf{z}}) \odot (\underline{\mathbf{1}} - \underline{\mathbf{z}}) \odot \underline{\mathbf{z}} \odot (\mathbf{H}\underline{\mathbf{a}})\right] \in \Re^{MN\times 1} \tag{7.4}$$

$$\nabla_{\underline{\mathbf{b}}} F(\underline{\mathbf{a}}, \underline{\mathbf{b}}) = -\mathbf{S}_\mathbf{y}^\mathbf{T}\mathbf{S}_\mathbf{x}^\mathbf{T}\mathbf{H}^\mathbf{T}\left[(\underline{\mathbf{z}}^* - \underline{\mathbf{z}}) \odot (\underline{\mathbf{1}} - \underline{\mathbf{z}}) \odot \underline{\mathbf{z}} \odot (\mathbf{S}_\mathbf{x}\mathbf{S}_\mathbf{y}\mathbf{H}\underline{\mathbf{b}})\right] \in \Re^{MN\times 1} \tag{7.5}$$

Note that $\mathbf{S}_\mathbf{x}^\mathbf{T}$ and $\mathbf{S}_\mathbf{y}^\mathbf{T}$ are 2-D matrix operators representing shift along the horizontal (left) and vertical (up) direction by 1 pixel. The gradient calculation can be carried out without the raster-scan operation, and the optimization problem can be solved using one of the two methods discussed in Section 5.3.

We now present some results demonstrating the working of our approach. Fig. 7.1 illustrates the resist contours obtained for the pattern in Fig. 5.14 when the ILT problem is solved without any alignment error considerations. In other words, the problem was solved using the forward model (7.1) to bring the contours on target. As expected, the pattern error is excellent (equal to zero) when mask $\underline{\mathbf{b}}$ is perfectly aligned. However, the contour fidelity is very poor if mask $\underline{\mathbf{b}}$ is shifted right or down

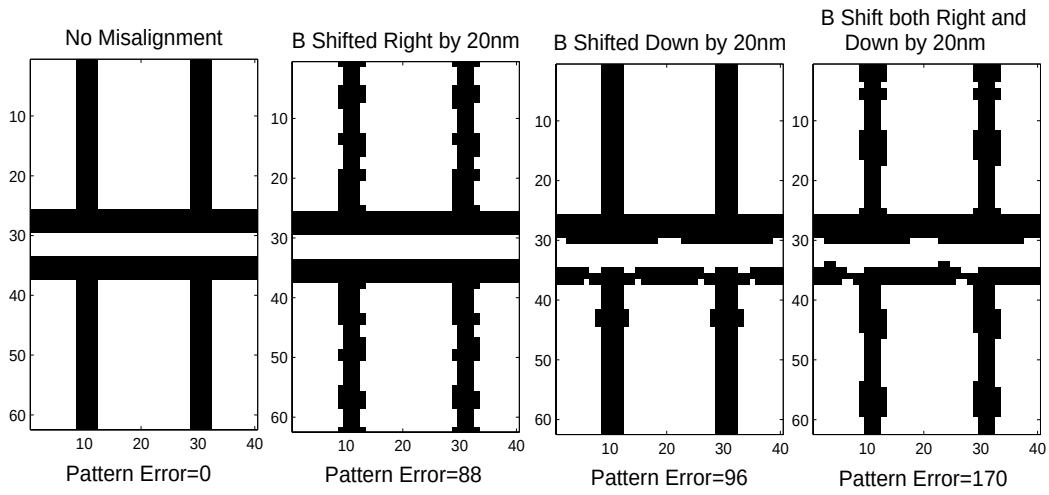

Figure 7.1: Left to right: The resist contours for no misalignment, $\underline{\mathbf{b}}$ shifted right by 20nm, $\underline{\mathbf{b}}$ shifted down by 20nm, and $\underline{\mathbf{b}}$ shifted both right and down by 20nm, where the ILT problem is solved for the contour fidelity with no misalignment.

or both by 1 pixel as observed in the graphs in Fig. 7.1. The pattern error shoots up from zero to 88, 96, and 170 for the above cases. Note that the mask is sampled at 20nm resolution, therefore one pixel shift corresponds to a displacement of 20nm.

Fig. 7.2 illustrates the aerial image decomposition and the resist contours at $t_r = 0.5$, where the ILT problem is solved considering that the second mask $\underline{\mathbf{b}}$ is shifted both right and down by 20nm. The aerial images indicate that only a fraction of the total energy is deposited in the second exposure to minimize the effect of misalignment. The decomposition is also very different from the one observed in Fig. 5.14. The contours in Fig. 7.2 indicate that the contour fidelity is very good for all four placements of mask $\underline{\mathbf{b}}$. The pattern error is very low (less than or equal to 4) for all the above cases, and has drastically improved compared to Fig. 7.1. Thus, DEL-ILT has succeeded in estimating masks which have made the double exposure

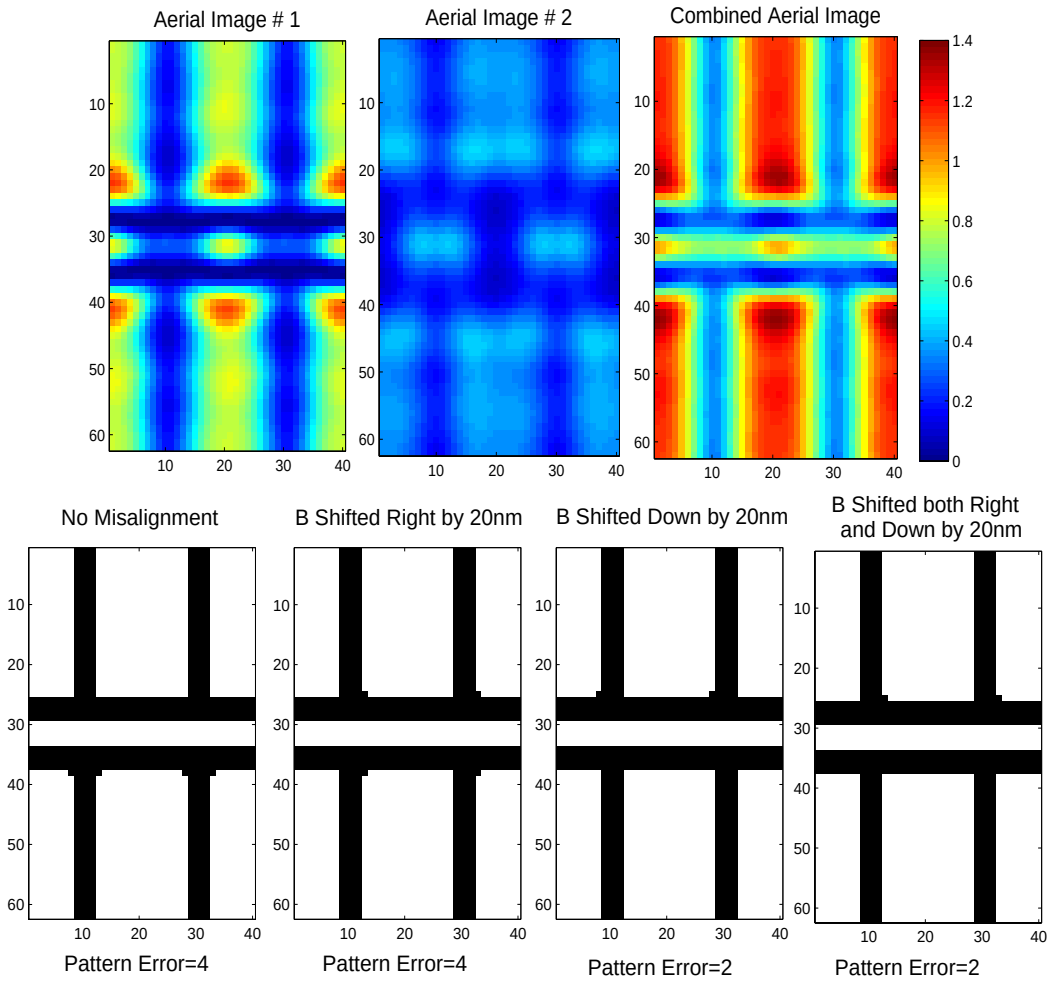

Figure 7.2: The top row indicates the aerial images from the first exposure (left), second exposure (center), and the combined aerial image (right). The bottom row indicates the resist contours with no misalignment, and the second mask is shifted right or down or both by 20nm. The DEL-ILT problem is solved only for the last case (**b** misaligned in both directions).

lithography system robust to alignment error. It is important to note that the order of exposure needs to be maintained in this case.

In conclusion, preliminary results (using the simple alignment error model in (7.2)) indicate the existence of *alignment error tolerant* pattern decomposition schema[3]. This considerably eases the tight alignment CD budgets faced by the tool manufacturers, which might help reduce the tool cost as well as increase the yield. DEL-ILT can thus be possibly extended to obtain *aerial image aware*, *mask misalignment aware*, patterning solutions.

7.3.2 Thick-Mask Effects

The framework proposed in this thesis assumes a thin-mask or Kirchhoff approximation. The mask is considered to be a 2-D object (infinitely thin) where the electric field exiting the mask is given by the ideal transmission function of the mask pattern (see Fig. 7.3). Although the above approximation provides reasonably accurate results with mask features larger than the wavelength of light used (as was the case for 65nm and higher nodes), it turns out to be very inaccurate as the mask feature sizes shrink below 193nm. The mask topography effects become non-negligible, and both amplitude and phase of the exit electric field are non-ideal as observed in Fig. 7.3. Furthermore, strong PSM (AltPSM and CPL) consists of etching the mask creating trenches (with depths in the order of the wavelength), to cause the 180 degree phase-shift. Therefore, the simplified Kirchhoff model is insufficient, and rigorous 3D

[3]Similarly other user-defined requirements like process window size, MEEF, etc might lead to a different pattern decomposition.

electromagnetic field simulations (accounting for the mask topography, mask side-wall angle, material property, geometry, and incident field conditions) have become necessary for accurate aerial image and CD prediction.

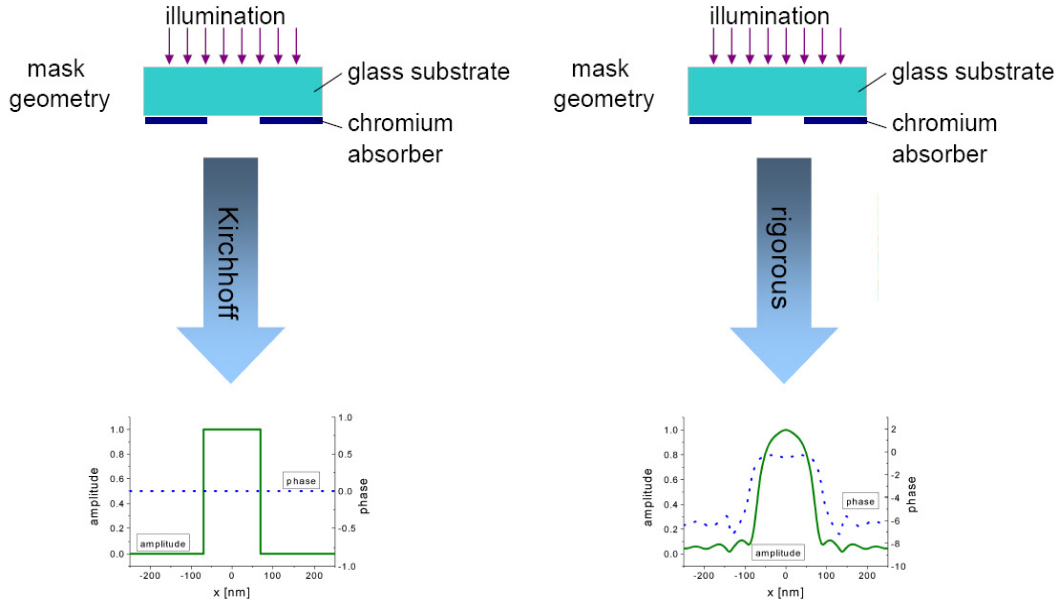

Figure 7.3: Kirchhoff (thin) versus rigorous (thick) mask modeling (from [25]).

There has been no reported work in the literature which performs inverse lithography while also accounting for the thick mask effects. The reason is that EMF solvers are computationally so taxing that even the forward modeling problem, is practically a very challenging one. Recently [89, 88] proposed using boundary layer models and lookup tables to simplify the thick-mask modeling. Any practical application of ILT would require an accurate and computationally efficient approximation to the rigorous electric field computations, perhaps using an empirical or boundary-layer model.

From Fig. 7.3, we observe that the exit electric field (for thick-masks) is a

perturbation of the ideal (thin) mask transmission function. Therefore, the ILT mask solutions calculated using thin mask approximation can serve as good initialization for thick-mask ILT. In this way, we hope that the framework proposed in this thesis will point to a solution path for mask-topography aware ILT.

7.3.3 Mask Design for 3-D Resist and Etch Profiles

The ILT framework discussed in this thesis only matches the 2-D aerial image and the 2-D resist contours (at the best focus point). However, the resist layer has finite thickness (100-200nm) giving rise to 3-D resist profiles. As discussed in Section 2.1, the resist exposure process consists of steps like bulk (or latent) image formation, photo-active compound (PAC) generation, post-exposure bake, and resist development [13], each of which has a potential impact on the resist profile and side-wall angle. Fig. 7.4 illustrates some commonly arising errors in 3-D resist profiles. Continued shrinking of dimensions have led to poor aspect ratio, thereby increasing the risk of a pattern collapse. Finally, a good resist profile is also important for a good etch profile, and acceptable device performance.

Therefore, an important direction of future work consists of estimating masks which guarantee both 2-D and 3-D contour fidelity. One way to achieve this is to optimize the 2-D contour fidelity for several horizontal slices along the thickness of the photo-resist. Physical or parametric models can be used to capture the degradations caused by the resist exposure process, and accounted in the forward lithography modeling. The above approach was also very recently explored by Yenikaya, et al. [96].

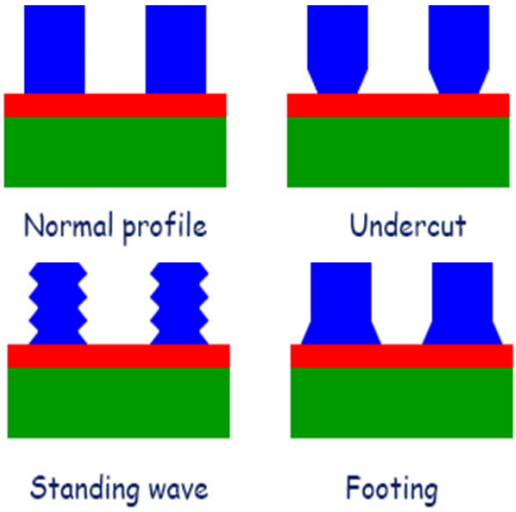

Figure 7.4: Example of 3-D resist profiles (from [39]).

7.3.4 Applications to E-beam Lithography

We now discuss two possible extensions of our work to e-beam lithography.

Proximity Effect Correction

The fabrication of devices at nanometer scale is performed using electron-beam lithography, which consists of directly firing a focused beam of electrons onto a photo-resist coated substrate (see Fig. 7.5). Unlike microlithography, a mask is not used. Instead, the pattern is directly carved on the substrate using the electron beam. Unfortunately, once the electrons enter the resist, they tend to scatter before depositing their energy, giving rise to proximity effects. That is, the electron scattering causes the energy to spread throughout the substrate which results in a blurring effect. The point spread function of the above system is commonly modeled as the sum of two Gaussian kernels accounting for the forward and backscattering of the electrons. The

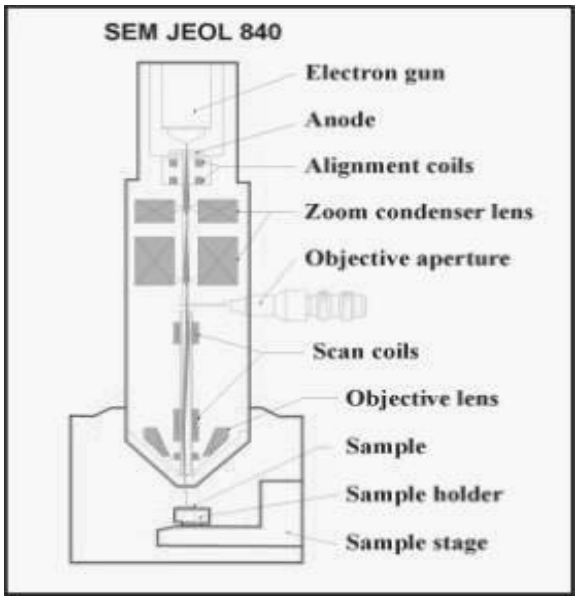

Figure 7.5: E-Beam Lithography (from [18])

above effect is counteracted using electron-beam proximity correction, a predistortion technique commonly utilized by the nanolithography community, [18], [66]. E-beam writers are commonly employed to manufacture the masks used for optical lithography.

The e-beam lithography process can be modeled similarly to microlithography, as shown in Fig. 2.3. The pattern is carved one pixel at a time by the e-beam writer, and hence pixel-based pattern representation is well suited for this purpose. E-beam proximity effect correction techniques fall into two sub-categories of dose modification and shape modification. Unlike before, here the dose (or energy) of the electrons can also be modified, and hence the estimated pattern is not restricted to be binary anymore. Each pixel of $\underline{\mathbf{m}}$ now represents the dose given to the corresponding location. The allowable dose values may be continuous within a finite range, or fixed to some discrete levels, depending on the properties (switching time, power, resolution,

etc) of the e-beam writer.

Peckerar, et al. [65, 68, 67] formulated the dose optimization problem using an L_2 norm cost function, and employed gradient-descent as well as regularization technique to solve the above problem. However, their forward model only accounted for the energy deposited by the e-beam, and does not account for the photo-resist response (contour fidelity). For mask writing applications, this may lead to mask CD variations, which is unacceptable for high mask-error-enhancement-factor (MEEF) systems. The ILT framework proposed in this thesis may be extended to account for the mask CD fidelity, mask side-wall angle, chrome etch depth, etc. Controlling the mask topography is critical for accurate simulation of the thick-mask effects (discussed in the Section 7.3.2). Therefore, it is quite possible that the next generation e-beam writing tools will perform a mask topology aware e-beam proximity effect correction.

Integrating E-beam and Optical Lithography Distortions

It is worth noting that the reticles used for microlithography are themselves fabricated using e-beam lithography (see Fig. 7.6). In a typical process flow, the mask engineer first synthesizes the OPC pattern to pre-compensate for the process distortions arising in optical microlithography process. The synthesized pattern is then forwarded to the mask fabrication laboratory, but it cannot be directly fed to the e-beam writer. Zheng Cui in [19] carried out a combined simulation and demonstrated that the mask distortions due to e-beam proximity effects play an important role in worsening the optical proximity effects. Therefore, the mask manufacturer has to run

one more round of optimization procedure to precompensate for the e-beam process distortions, and determine the correct dose and path for the e-beam writer.

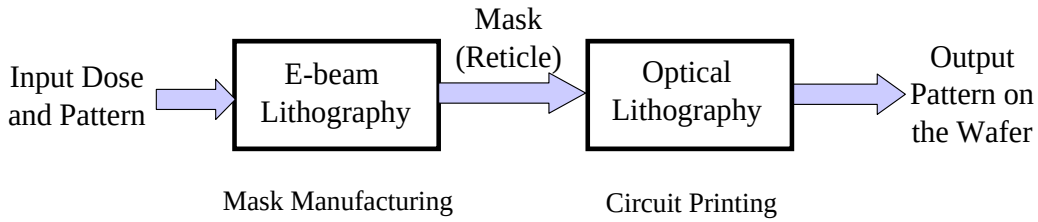

Figure 7.6: Mask manufacturing to circuit printing process

Perhaps an alternate approach is to view the above system as a cascade of e-beam and optical lithography, and to build a lumped process distortion model. We can then directly solve for the optimum e-beam dose/pattern required to obtain the desired pattern $\underline{\mathbf{z}}^*$ on the silicon wafer. The actual reticle is just treated as an intermediate by-product. The penalty on reticle costs, storage, complexity, time, etc are actually incurred at the mask fabrication level, and hence the above approach gives us a more direct control on the above factors.

7.3.5 RET for Optical Maskless Lithography and DUV Laser Writing

Optical maskless lithography is an attractive alternative to mask-based lithography due to the increasing reticle costs. Fig. 7.7 indicates the functional flow of traditional mask-based versus maskless lithography systems. In the latter case, the embedded mask writer converts the data to an embedded reticle, which is an image created by spatial light modulators (SLM). Thus the photomask is replaced with an addressable array of light-modulating elements which can be controlled in real-time.

The two commonly employed modulating elements for SLM are piston or tilting micro-mirrors.

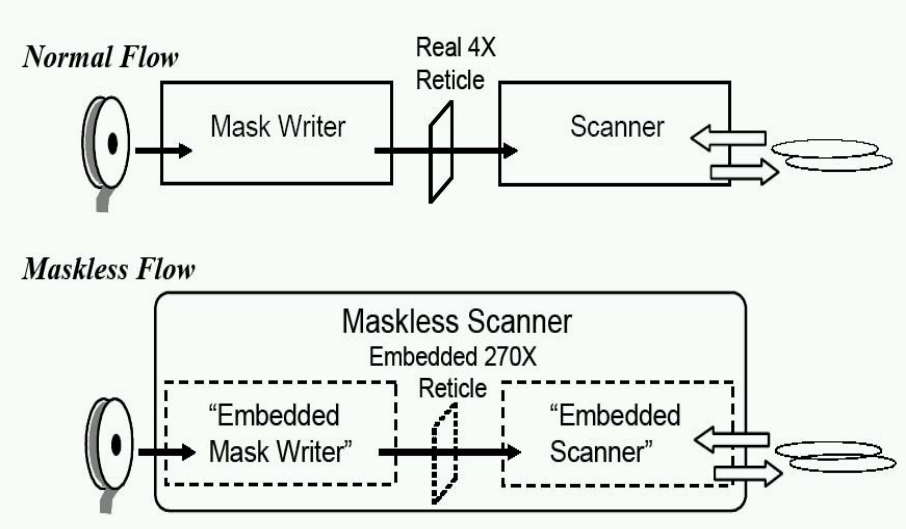

Figure 7.7: Traditional mask-based versus maskless flow for optical lithography (from [78]).

Maskless lithography suffers from the same distortions arising in traditional (reticle-based) optical lithography. The pixel-size (at wafer plane) for maskless projection system (from Micronic Laser Systems) is 30nm [78]. SLM can be used to emulate binary, attenuated, alternating, and chromeless phase shift masks. Unlike traditional systems, gray-scale masks and multiple phase levels can also be simulated using SLM. As observed in [85], an RET based approach is required to determine the optimal arrangement of each mirror such that the resist contours are on target. The ILT approach discussed in this thesis uses a pixel-based mask representation, and is readily extendible to maskless lithography application. Note that pixel size dictates the density of the micro-mirrors, whereas multiple tones increase the data storage and processing time. It will also be interesting to study the trade-off between the pixel-

size versus multiple tone and phase values while designing the maskless lithography system.

Another possible extension of the ILT framework is for Deep Ultra-Violet (DUV) laser writers. The rise in mask complexity has increased the mask writing time for e-beam based mask writers, making DUV-based mask writers an attractive proposition. The Sigma 7500 DUV mask writer uses a 248nm pulsed laser and a programmable SLM with one million micro-mirrors to write the mask pattern. Osterberg, et al [62] recently proposed an embedded optical proximity correction technique to bring the mask contours on target. Once again, the proposed ILT framework should be extendible to correct the DUV mask writer distortions.

7.3.6 Pattern Splitting and Stitching for DPL

As seen from Fig. 5.2, double pattering requires decomposing the target pattern (with pitch p) into two patterns (with pitch $2p$). Hsu, et al. [36] proposed color line method (CLN) and color space method (CSP) for splitting 1-D features. However, pattern splitting for complex 2-D array structures or random logic lines that are connected to each other is more challenging. The above requires *cutting* (and *stitching*) the complex 2-D structure into two separate patterning layers. Park, et al. [26] observed that pattern stitching depends on the proximity environment in order to avoid breaking and necking of the features, and proposed a model-based approach to achieve the above.

The DEL-ILT framework for $k_1 \geq 0.25$ (discussed in Chapter 5) provided

a novel paradigm of splitting the *gray-level* aerial image between the two exposures. However (as discussed in Section 6.4), we had little success in decomposing the patterns for $k_1 < 0.25$. We suspect this might be due to poor initialization, which causes the gradient-based search technique to get stuck in a poor local minimum. Perhaps, a better heuristic-based initialization may help overcome the above issue. For example, all polygons spaced closer than half the pitch must be separated a priori and assigned to two different layers. An inverse problem can then be solved to determine the best way of splitting the interconnections between the features (for the two patterning layers), and estimating the corresponding mask structures.

7.3.7 Practical Extensions

We now discuss some more extensions of the proposed ILT framework geared towards increasing its practical utility for the lithography industry.

In our approach, we only optimize for the best focus and dose conditions. However, an important extension would be to optimize over the process window (a range of dose and de-focus values), in order to increase the robustness of the lithography process. We are also modeling the resist behavior using a CTR model, but it does not accurately model the real world resist systems. More sophisticated resist models like the variable threshold resist (VTR)[75, 17] or other empirical models must be employed to add practical value.

The current ILT problem is solved in the spatial domain where all MN pixels of the estimated masks are treated as unknown variables. Therefore, we have

to search the vast MN dimensional space in order to arrive at the optimal mask solution. One way to reduce the dimensionality of search space (which forms part of the ongoing research at the University of Texas Austin) is to transform the mask pattern to frequency space using DCT. The high frequency components are dropped (set to zero), and only the low frequency components are considered as free parameters to form the solution space. Finally, there is also a possibility of using Quasi-Newton optimization algorithm, which uses the second order derivative of the cost function, to search the solution space more efficiently.

Appendix A

Derivation of the Gradient for Incoherent Imaging Systems

Here we present the derivation for the gradient of the cost function for an incoherent imaging system. We also observe that the gradient calculation can be carried out directly on the 2-D matrices, with no need for the raster-scan operation.

From (2.15) and (2.9), the L_2 norm cost function in terms of the unconstrained parameter vector $\underline{\boldsymbol{\theta}}$ is gives as,

$$F_1^i(\underline{\boldsymbol{\theta}}) = \sum_{k=1}^{MN} (z_k^* - z_k(\underline{\boldsymbol{\theta}}))^2 \tag{A.1}$$

where,

$$z_k(\underline{\boldsymbol{\theta}}) = z_k^i(\underline{\boldsymbol{\theta}}) = \frac{1}{1 + \exp\left[-a\left(\sum_{j=1}^{MN} \tilde{h}_{kj} \frac{1+\cos(\theta_j)}{2}\right) + at_r\right]}$$

The partial derivative of (A.1) with respect to θ_p is given as,

$$\begin{aligned}
\frac{\partial F_1^i(\underline{\boldsymbol{\theta}})}{\partial \theta_p} &= 2\sum_{k=1}^{MN}\left(z_k^* - z_k(\underline{\boldsymbol{\theta}})\right)\left(-\frac{\partial z_k(\underline{\boldsymbol{\theta}})}{\partial \theta_p}\right) \\
&= 2\sum_{k=1}^{MN}\left(z_k^* - z_k(\underline{\boldsymbol{\theta}})\right)\ z_k^2(\underline{\boldsymbol{\theta}})\ \frac{\partial}{\partial \theta_p}\exp\left(-a\sum_{j=1}^{MN}\tilde{h}_{kj}\frac{1+\cos(\theta_j)}{2} + at_r\right) \\
&= 2\sum_{k=1}^{MN}\left(z_k^* - z_k(\underline{\boldsymbol{\theta}})\right) z_k(\underline{\boldsymbol{\theta}})(1 - z_k(\underline{\boldsymbol{\theta}}))\frac{\partial}{\partial \theta_p}\left(-a\sum_{j=1}^{MN}\tilde{h}_{kj}\frac{1+\cos(\theta_j)}{2} + at_r\right) \\
&= a\sum_{k=1}^{MN}\left(z_k^* - z_k(\underline{\boldsymbol{\theta}})\right) z_k(\underline{\boldsymbol{\theta}})(1 - z_k(\underline{\boldsymbol{\theta}}))\sin(\theta_p)\tilde{h}_{kp} \\
&= a\sin(\theta_p)\sum_{k=1}^{MN}\tilde{h}_{kp}\left(z_k^* - z_k(\underline{\boldsymbol{\theta}})\right) z_k(\underline{\boldsymbol{\theta}})(1 - z_k(\underline{\boldsymbol{\theta}}))
\end{aligned}$$

Therefore, the gradient vector can be calculated as,

$$\nabla F_1^i(\underline{\boldsymbol{\theta}}) = \left[\frac{\partial F_1^i(\underline{\boldsymbol{\theta}})}{\partial \theta_1}, \ldots, \frac{\partial F_1^i(\underline{\boldsymbol{\theta}})}{\partial \theta_{MN}}\right]^T = a(\tilde{\mathbf{H}}^{\mathbf{T}}\ [(\underline{\mathbf{z}}^* - \underline{\mathbf{z}}^{\mathbf{i}}) \odot \underline{\mathbf{z}}^{\mathbf{i}} \odot (\mathbf{1} - \underline{\mathbf{z}}^{\mathbf{i}})]\ ) \odot\ \sin(\underline{\boldsymbol{\theta}}).$$

In the above equation, $\odot$ resembles element-by-element multiplication of the vectors. The matrix multiplication with $\tilde{\mathbf{H}}^{\mathbf{T}}$ corresponds to a convolution operation with $\tilde{h}(-x,-y)$. Both these operations can be implemented directly on the 2-D matrices as demonstrated in Fig. 3.1.

Appendix B

Double Exposure Theorem

The double exposure theorem states that the aerial image obtained using a complex-valued mask function can be also achieved using successive exposure of two real valued masks. Let mask $\underline{\mathbf{m}} = \underline{\mathbf{a}} + i\underline{\mathbf{b}}$, where $\underline{\mathbf{a}}, \underline{\mathbf{b}} \in \Re^{MN \times MN}$ and $i = \sqrt{(}-1)$. Then for a coherent imaging system,

$$\underline{\mathbf{z}} = |\mathbf{H}\underline{\mathbf{m}}|^2 = |\mathbf{H}(\underline{\mathbf{a}} + i\underline{\mathbf{b}})|^2 = |\mathbf{H}\underline{\mathbf{a}} + i\mathbf{H}\underline{\mathbf{b}}|^2$$

Square of the magnitude of a complex number is equal to the sum of the squares of the real and imaginary parts. Hence,

$$\underline{\mathbf{z}} = |\mathbf{H}\underline{\mathbf{a}}|^2 + |\mathbf{H}\underline{\mathbf{b}}|^2$$

A real-valued mask consists of only two phases (0 and 180 degrees). Therefore any image obtained using a continuously varying phase in one exposure, can also be obtained by means of two exposures, each using a mask with only two phases 0 and π. Furthermore, the two real-valued mask can be identified as the real and imaginary

parts of the complex-valued mask.

Pati and Kailath in [64] noted that the number of discrete phase-levels used in PSM design should be at least four, in order to avoid conflicts and ambiguities in phase assignment (four-phase theorem). Xu and Arce in [50] proposed to used generalized ILT by estimating single exposure PSM masks to estimate both phase and magnitude. The phase values were restricted to 45, 135, -135, and -45 degrees, which enabled them to pattern 2-D Manhattan geometry patterns. Note that in our DEL-ILT framework we employ the regularization term to estimate two CPL masks (transmission values -1 and +1). This is equivalent to restricting the transmission values of the corresponding complex-valued mask to be $1+i$, $1-i$, $-1-i$ and $1-i$. Thus, the single exposure mask is indirectly restricted to have only four distinct phase values 45, 135, -135, and -45 degrees, respectively. The aerial images obtained using DEL-ILT have good contrast (even in the phase-conflict regions) as observed in Fig. 5.14 and Fig. 5.16.

Appendix C

Relationship Between ILS and Dose Magnification.

Here we prove that for a given incoming dose magnification p, the ILS of the transmitted dose (D_t) is upper-bounded by p times the ILS of the original incoming dose D_i (at the given edge location).

Using (6.2), the transmitted CEL dose for the magnified incoming dose (pD_i) can be calculated as follows,

$$D_t = f(D_i) = pD_i - \frac{t_c}{c} + \frac{\log[1 + \exp(-acpD_i + at_c)]}{ac} \tag{C.1}$$

Let t_r be the resist threshold which determines the desired CD (edge location). Therefore, the original ILS can be calculated as,

$$ILS_{orig} = \frac{1}{D_i} \left.\frac{\partial D_i}{\partial x}\right|_{D_i=t_r} = \frac{1}{t_r} \left.\frac{\partial D_i}{\partial x}\right|_{D_i=t_r} \tag{C.2}$$

Similarly, the ILS of the CEL transmitted dose D_t at the same edge location is given

as,

$$\begin{aligned} ILS_{new} &= \frac{1}{D_t}\frac{\partial D_t}{\partial x}\Big|_{D_t=t_r} = \frac{1}{t_r}\frac{\partial f(D_i)}{\partial x}\Big|_{f(D_i)=t_r} \\ &= \frac{1}{t_r}\frac{\partial f(D_i)}{\partial D_i}\frac{\partial D_i}{\partial x}\Big|_{f(D_i)=t_r} \\ &= \frac{1}{t_r}\left(\frac{p}{1+\exp(-acpD_i+at_c)}\right)\frac{\partial D_i}{\partial x}\Big|_{f(D_i)=t_r} \end{aligned} \tag{C.3}$$

Note that new ILS is evaluated for the same CD (and edge location) as before, i.e., $D_i = f(D_i) = t_r$. As seen earlier in Section 6.2.2, if $p = 2$ and we use the same CEL parameters ($t_c = 0.15$, $a = 30$), $f(D_i) = t_r$ gives the desired CD for $c = 0.45$. Therefore,

$$\begin{aligned} ILS_{new} &= \frac{1}{t_r}\left(\frac{p}{1+\exp(-acpD_i+at_c)}\right)\frac{\partial D_i}{\partial x}\Big|_{(D_i=t_r,t_c=0.15,c=0.45,a=30)} \\ &= p\left(\frac{1}{1+\exp(-acpD_i+at_c)}\right)\frac{1}{t_r}\frac{\partial D_i}{\partial x}\Big|_{(D_i=t_r,t_c=0.15,c=0.45,a=30)} \\ &= p\left\{\frac{1}{1+\exp(-acpt_r+at_c)}\right\}_{(t_c=0.15,c=0.45,a=30)}\left\{\frac{1}{t_r}\frac{\partial D_i}{\partial x}\Big|_{D_i=t_r}\right\} \\ &= p\left\{\frac{1}{1+\exp(-acpt_r+at_c)}\right\}_{(t_c=0.15,c=0.45,a=30)} \times ILS_{orig} \end{aligned} \tag{C.4}$$

For the above case, we observe that $ILS_{new} = 2 \times 0.9734 \times ILS_{orig} = 1.95 \times ILS_{orig}$. Thus, for the given CEL material, magnifying the dose twice gives 1.95 times improvement in the ILS.

Furthermore, we also observe that $\frac{1}{1+\exp(-acpt_r+at_c)}$ is upper bounded by 1. Therefore, $ILS_{new} \leq p \times ILS_{orig}$; magnifying the incoming dose by a factor p, can lead to a maximum of p times improvement in the ILS.

Appendix D

Intel Fab Results

Here we present the printed wafer results for single and double exposure inverse lithography. The experiments were performed in the Intel fab using a 193nm dry lithography system with NA=0.93 and $\sigma = 0.2$.

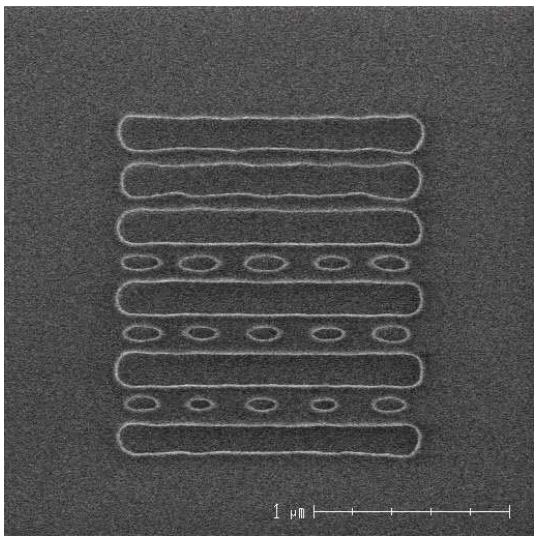

Figure D.1: SEM image of the fabricated wafer pattern obtained using the single exposure ILT mask pattern from Fig. 4.12 (courtesy Intel Corporation).

Fig. D.1 is the SEM image of the wafer pattern fabricated using the single exposure CPL-ILT mask from Fig. 4.12. We observe that the lines, spaces, and the contacts are all distinguishable with no pattern collapse, which is very encouraging. But the printed size of the individual contacts is different. This can be attributed to the model discrepancy. The mask design problem was solved for a (practically infeasible) fully coherent imaging system, but the experiments in the fab were performed using a scanner with $\sigma = 0.2$ (partially coherent). Furthermore, our forward model does not account for the thick mask effects discussed in Section 7.3.2. Any practical application of ILT will need to incorporate the above effects into the forward model.

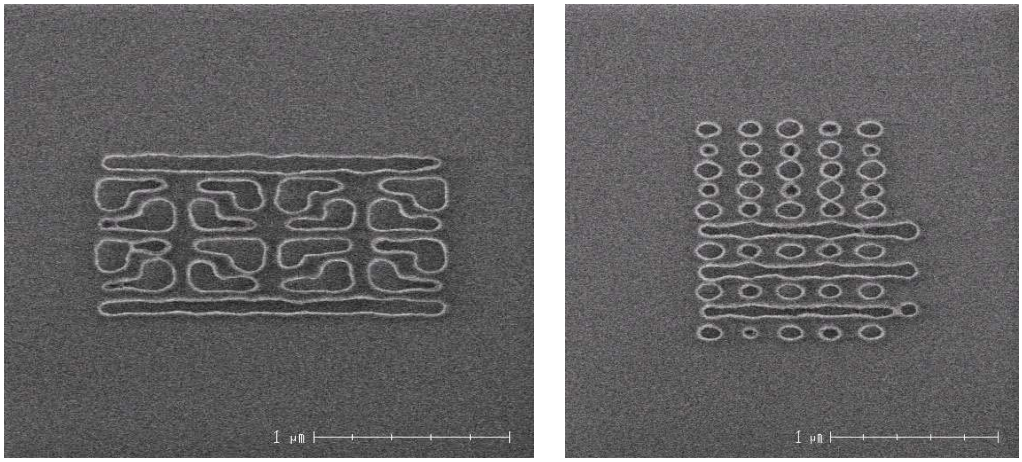

Figure D.2: SEM images of the fabricated patterns using the DEL-ILT mask patterns from Fig. 5.13 (left) and Fig. 5.17 (right). The figure is courtesy Intel Corporation.

Fig. D.2 consists of SEM images of the wafer patterns for two DEL-ILT examples from Chapter 5 (Fig. 5.13 and Fig. 5.17). We observe that the printed wafer patterns resemble their respective target patterns, thereby validating the aerial image decomposition paradigm of DEL-ILT. The contour fidelity can be improved by

accounting for partial coherence and mask topography effects, as discussed earlier.

Bibliography

[1] *International technology roadmap for semiconductors*, `http://public.itrs.net/` (2003).

[2] C. Ahn, H. Kim, and K. Baik, *Novel approximate model for resist process*, Optical Microlithography, Proc. SPIE, vol. 3334, 1998, pp. 752–763.

[3] M. Albota and et al, *Design of organic molecules with large two-photon absorption cross sections*, Science **281** (1998), 1653–1656.

[4] A. K. Bates, M. Rothschild, T. M. Bloomstein, T. H. Fedynyshyn, R. R. Kunz, V. Liberman, and M. Switkes, *Review of technology for 157-nm lithography*, IBM Journal of Research and Development **45** (2001), 605–614.

[5] M. Bazaraa and C. Shetty, *Nonlinear Programming Theory and Algorithms*, John Wiley and Sons, 1979.

[6] UC Berkeley, *Lithography analysis using virtual access,* `http://cuervo.eecs.berkeley.edu/Volcano/`.

[7] P. Berman, A.B. Kahng, D. Vidhani, E.H. Wang, and F.A. Zelikovsky, *Optimal phase conflict removal for layout of dark field alternating phase shifting masks*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems **19** (2000), 175–187.

[8] D. Bernard and J. Li, *Clear field dual alternating phase shift mask lithography*, Optical Microlithography, Proc. SPIE, vol. 4691, 2002, pp. 999–1008.

[9] A. Biswas, J. Li, J. Hiserote, and L. Melvin III, *Extension of 193nm dry lithography to 45-nm half-pitch node: Double exposure and double processing technique*, BACUS Symposium on Photomask Technology, Proc. SPIE, vol. 6349, 2006.

[10] J. Bjorkholm, *EUV lithography — the successor to optical lithography?*, Intel Technology Journal (1998), no. Q3, 8.

[11] M. Born and E. Wolfe, *Principles of Optics*, Cambridge University Press, 1999.

[12] T. Chiou, A. Chen, S. Hsu, M. Eurlings, and E. Hendrick, *Development of automatic OPC treatment and layout decomposition for double dipole lithography for low-k1 imaging*, Advanced Microlithography Technologies, Proc. SPIE, vol. 5645, 2005, pp. 21–31.

[13] P. Choudhury, *Handbook of Microlithography, Micromachining and Microfabrication*, SPIE Press, 1997.

[14] N. Cobb and D. Dudau, *Dense OPC and verification for 45nm*, Optical Microlithography, Proc. SPIE, vol. 6154, 2006.

[15] N. Cobb and A. Zakhor, *Fast, low-complexty mask design*, Optical Microlithography, Proc. SPIE, vol. 2440, 1995, pp. 313–327.

[16] ———, *Fast sparse aerial image calculation for OPC*, BACUS Symposium on Photomask Technology, Proc. SPIE, vol. 2621, 1995, pp. 534–545.

[17] ———, *Mathematical and CAD framework for proximity correction*, Optical Microlithography, Proc. SPIE, vol. 2726, 1996, pp. 208–222.

[18] B. Cook, *Pyramid - a hierarchical rule-based proximity effect correction scheme for electron-beam lithography: Generalization and optimization for homogeneouse substrates*, Ph.D. thesis, Cornell University, 1996.

[19] Z. Cui, *Computer simulation from electron beam lithography to optical lithography*, International Conference on Modeling and Simulation of Microsystems, 2000, pp. 87–90.

[20] M. Drobizhev, A. Karotki, A. Rebane, and C. Spangler, *Dendrimer molecules with record large two-photon absorption cross section*, Optics Letters **26** (2001), 1081–1083.

[21] M. Drobizhev, Y. Stepanenko, Y. Dzenis, A. Karotki, A. Rebane, P. Taylor, and H. Anderson, *Understanding strong two-photon absorption in π-conjugated porphyrin dimers via double-resonance enhancement in a three-level model*, Journal of the American Chemical Society **126** (2004), 15352–15353.

[22] W. Duch and N. Jankowski, *Survey of neural transfer functions*, Computing Surveys **2** (1999), 163–213.

[23] A. Erdmann, R. Farkas, T. Fuhner, B. Tollkuhn, and G. Kokai, *Towards automatic mask and source optimization for optical lithography*, Optical Microlithography, Proc. SPIE, vol. 5377, 2004, pp. 646–657.

[24] A. Erdmann and W. Henke, *Simulation of optical lithography*, Optical Microlithography, Proc. SPIE, vol. 3729, 1999, pp. 480–494.

[25] Andreas Erdmann, *Rigorous electromagnetic field modeling: Challenges and solution for lithography simulation.*

[26] J. Park et al, *Application challenges with double patterning technology beyond 45nm*, BACUS Symposium on Photomask Technology, Proc. SPIE, vol. 6349, 2006.

[27] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, *Fast and robust multi-frame super-resolution*, IEEE Transcation on Image Processing **13** (2004), 1327–1344.

[28] M. Fritze, B. Tyrrell, D. Astolfi, R. Lambert, D. Yost, A. Forte, S. Cann, and B. Wheeler, *Subwavelength optical lithography with phase-shift-mask*, Lincoln Laboratory Journal **14** (2003), 237–250.

[29] R. Gonzalez and R. Woods, *Digital Image Processing*, Prentice Hall, 2002.

[30] Y. Granik, *Illuminator optimization methods in microlithography*, Optical Microlithography, Proc. SPIE, vol. 5524, 2004, pp. 217–229.

[31] ———, *Solving inverse problems of optical microlithography*, Optical Microlithography, Proc. SPIE, vol. 5754, 2005, pp. 506–526.

[32] Y. Granik, K. Sakajiri, and S. Shang, *On objectives and algorithms of inverse methods in microlithography*, BACUS Symposium on Photomask Technology, Proc. SPIE, vol. 6349, 20076.

[33] L.F. Halle, *A water soluble contrast enhancement layer*, Journal of Vaccuum Science and Technology B **3** (1985), 323–326.

[34] M. Hasegawa, A. Suzuki, K. Saitoh, and M. Yoshii, *New approach for realizing $k_1 = 0.3$ optical lithography*, Symposium on Photomask and X-Ray Mask Technology, Proc. SPIE, vol. 3748, 1999, pp. 278–287.

[35] M. Himel, R. Hutchins, J. Colvin, M. Poutous, A. Kathman, and A. Fedor, *Design and fabrication of customized illumination patterns for low k1 lithography: A diffractive approach*, Optical Microlithography, Proc. SPIE, vol. 4346, 2001, pp. 1436–1442.

[36] S. Hsu, J. Park, D. Broeke, and J. Chen, *Double exposure technique for 45nm and beyond*, BACUS Symposium on Photomask Technology, Proc. SPIE, vol. 5992, 2005.

[37] W. Huang, C. Lin, C. Kuo, C. Huang, J. Lin, J. Chen, R. Liu, Y. Ku, and B. Lin, *Two threshold resist models for optical proximity correction*, Optical Microlithography, Proc. SPIE, vol. 5377, 2001, pp. 1536–1543.

[38] C. Hung, Q. Liu, K. Sakajiri, S. Shang, and Y. Granik, *Model based insertion of assist features using pixel inversion method: Implementation in 65nm node*, Photomask and Next-Generation Lithography Mask technology - XIII, Proc. SPIE, vol. 6283, 2006.

[39] H.W.Lin and L. Voelkel, *Challenges of selecting bottom anti-reflective coatings.*

[40] P. Flanner III, *Two-dimensional optical imaging for photolithography simulation*, Master's thesis, University of California, Berkeley, 1986.

[41] C.T. Kelly, *Iterative Methods for Optimization*, SIAM, 1999.

[42] M. Levenson, N.S. Vishwanathan, and R.A. Simpson, *Improving resolution in photolithography with a phase-shifting-mask*, IEEE Transactions on Electronic Devices **29** (1982), 1828–1836.

[43] J. Li, Q. Yan, and L. Melvin III, *Rapid search for the optimum placement of assist feature to improve the aerial image gradient in iso-line structure*, Advanced Lithography, Proc. SPIE, vol. 6520, 2007.

[44] B. Lin, M. Shieh, J. Sun, J. Ho, Y. Wang, X. Wu, W. Leitermann, O. Lin, J. Lin, Y. Liu, and L. Pang, *Inverse lithography technology at chip scale*, Optical Microlithography XIX, Proc. SPIE, vol. 6154, 2006.

[45] Y. Liu, D. Abrams, L. Pang, and Andrew Moore, *Inverse lithography technology principles in practice: Unintuitive patterns*, BACUS Symposium on Photomask Technology, Proc. SPIE, vol. 5992, 2005, pp. 231–238.

[46] Y. Liu and A. Zakhor, *Optimal binary image design for optical lithography*, Optical Microlithography, Proc. SPIE, vol. 1264, 1990, pp. 401–412.

[47] ———, *Binary and phase shifting mask design for optical lithography*, IEEE Transactions on Semiconductor Manufacturing **5** (1992), 138–151.

[48] ———, *Computer aided phase shift mask design with reduced complexity*, IEEE Transcation on Semiconductor Manufacturing **9** (1996), 170–184.

[49] W. Loong and H. Pan, *A direct approach to the modeling of polydihexisilane as a contrast enhancement material*, Journal of Vaccuum Science and Technology B **8** (1990).

[50] X. Ma and G. R. Arce, *Generalized inverse lithography methods for phase-shifting mask design*, Advanced Lithography, Proc. SPIE, 2007.

[51] C. Mack, *Contrast enhancement techniques for submicron optical lithography*, Journal of Vaccuum Science and Technology A **5** (1987).

[52] _____, *Development of positive photoresist*, Journal of Electrochemical Society **134** (1987), 148–152.

[53] _____, *The new, new limits to optical lithography*, Optical Microlithography, Proc. SPIE, vol. 5374, 2004, pp. 1–8.

[54] _____, *The problems with PSM*, Yield Management Solutions Magazine **7** (2005).

[55] A. Moore, T. Lin, Y. Liu, G. Russell, L. Pang, and D. Abrams, *Inverse lithography technology at low k1: Placement and accuracy of assist features*, BACUS Symposium on Photomask Technology, Proc. SPIE, vol. 6349, 2006.

[56] K. Nashold and B. Saleh, *Image construction through diffraction-limited high contrast imaging system: An iterative approach*, Journal of Optical Society of America A - Optics Image Science and Vision **2** (1985), 635–643.

[57] Electronics Design Strategy News, *IBM to roll out 45nm immersion process by end of year,* `http://www.edn.com/article/CA6419180.html/`, Feb 27,2007.

[58] C. Noelscher, M. Heller, B. Habets, M. Markert, U. Scheler, and P. Moll, *Double line shrink lithography at* $k_1 = 0.16$, Microelectronic Engineering **83** (2006), 730–733.

[59] Y. Oh, J. C. Lee, and S. Lim, *Resolution enhancement through optical proximity correction and stepper parameter optimization for 0.12-μ mask pattern*, Optical Microlithography, Proc. SPIE, vol. 3679, 1999, pp. 607–613.

[60] W. Oldham, *The use of contrast enhancement layers to improve the effective contrast of positive photoresist*, IEEE Transactions on Electron Devices **34** (1987), 247–251.

[61] E. Ong, B. Singh, R. Ferguson, and A. Neureuther, *Comparison of proximity effects in contrast enhancement layer and bilayer resist processes*, Journal of Vaccuum Science and Technology B **5** (1987), 443–448.

[62] A. Osterberg, L. Ivansen, H. Ahlfeldt, H. Fosshaug, and T. Newman, *Embedded optical proximity correction for the sigma DUV mask writer*, Photomask and Next Generation Lithography Mask Technology, Proc. SPIE, vol. 6607, pp. , year =.

[63] L. Pang, Y. Liu, and D. Abrams, *Inverse lithography technology (ILT), what is the impact on the photomask industry?*, Sympoisum on Photomask and Next Generation Mask Technology VII, Proc. SPIE, vol. 6283, 2006.

[64] V.C. Pati and T. Kailath, *Phase-shifting masks for microlithography : Automated design and mask requirements*, Journal of Optical Society of America A - Optics Image Science and Vision **9** (1994), 2438–2452.

[65] Y. Pati, A. Teolis, D. Park, R. Bass, K-W Rhee, B. Bradie, and M.C. Peckerar, *An error measure for dose correction in e-beam nanolithography*, Journal of Vaccuum Science and Technology B **8** (1990), 1882–1988.

[66] Y.C. Pati, A. Teolis, D. Park, R. Bass, K. Rhee, B. Bradie, and M.C. Peckerar, *An error measure for dose correction in e-beam nanolithography*, Journal of Vaccuum Science and Technology - B **8** (1990), 1882–1888.

[67] M.C. Peckerar, S. Chang, and C.R.K. Marrian, *Proximity correction algorithms and a co-processor based on regularized optimization. Part I description of the algorithm*, Journal of Vaccuum Science and Technology B **13** (1995), 2518–2525.

[68] M.C. Peckerar and C.R.K. Marrian, *Error measure comparison of currently employed dose-modulation schemes for e-beam proximity effect control*, Electron-Beam, X-Ray, EUV and Ion-Beam Submicrometer Lithographies for Manufacturing, Proc. SPIE, vol. 2437, 1995, pp. 222–238.

[69] J. Peterson, *Analytical description of anti-scattering and scattering bar assist features*, Optical Microlithography, Proc. SPIE, vol. 4000, 2000, pp. 77–89.

[70] A. Poonawala and P. Milanfar, *Double exposure mask synthesis using inverse lithography*, Submitted to the Journal of Microlithography, Microfabrication and Microsystems (Preprint available at http://www.soe.ucsc.edu/ amyn/publications.htm).

[71] ———, *A pixel-based regularization approach to inverse lithography*, Accepted for publication in Microelectronic Engineering.

[72] ———, *OPC and PSM design using inverse lithography: A non-linear optimization approach*, Optical Microlithography, Proc. SPIE, vol. 6154, 2006, pp. 114–127.

[73] ———, *Fast and low-complexity mask design in optical microlithography - an inverse imaging problem*, IEEE Transactions on Image Processing **16** (2007), 774–788.

[74] Q. Qian and F. Leon, *Fast algorithms for 3d high NA lithography simulation*, Optical Microlithography, Proc. SPIE, vol. 2440, 1995, pp. 372–380.

[75] J. Randall, *Variable threshold resist models for lithography simulation*, Optical Microlithography, Proc. SPIE, vol. 3679, 1999, pp. 176–182.

[76] J. Randall, K. Ronse, T. Marschner, M. Goethals, and M. Ercken, *Variable threshold resist models for lithography simulation*, Optical Microlithography, Proc. SPIE, vol. 3679, 1999, pp. 176–182.

[77] S. Robertson, C. Mack, and M. Maslow, *Toward a universal resist dissolution model for lithography simulation*, Lithography for Semiconductor Manufacturing II, Proc. SPIE, vol. 4404, 2001, pp. 111–122.

[78] T. Sandstrom and H. Martinsson, *RET for optical maskless lithography*, Optical Microlithography, Proc. SPIE, vol. 5377, 2004, pp. 1750–1763.

[79] S. Sayegh and B. Saleh, *Image design: Generation of a prescribed image at the output of a bandlimited system*, IEEE Transcation on Pattern Analysis and Machine Intelligence **5** (1983), 441–445.

[80] S. Sayegh, B. Saleh, and K. Nashold, *Image design: Generation of a prescribed image through a diffraction limited system with high-contrast recording*, IEEE Transaction on Acoustics, Speech and Signal Processing **33** (1985), 460–465.

[81] F. Schellenberg, *Resolution enhancement with OPC/PSM*, Future Fab Intl **9** (2000).

[82] ———, *A little light magic*, IEEE Spectrum **40** (2003), 34–39.

[83] ———, *Resolution enhancement technology: The past, the present, and extensions for the future*, Optical Microlithography (Bruce W. Smith, ed.), Proc. SPIE, vol. 5377, 2004, pp. 1–20.

[84] S. Sherif, B. Saleh, and R. Leone, *Binary image synthesis using mixed linear integer programming*, IEEE Transactions on Image Processing **4** (1995), 1252–1257.

[85] Y. Shroff, Y. Chen, and W. Oldham, *Image optimization for maskless lithography*, Emerging Lithography Technologies, Proc. SPIE, vol. 5374, p. 637.

[86] R. J. Socha, D. Van Den Broeke, S. Hsu, J. F. Chen, T. Laidig, N. Corcoran, U. Hollerbach, K. Wampler, and X. Shi, *Contact hole reticle optimization by using interference mapping lithography (IML)*, Photomask and Next-Generation Lithography Mask Technology XI, Proc. SPIE, vol. 5446, 2004, pp. 516–534.

[87] M. Terai, T. Kumada, T. Ishibashi, and T. Hanawa, *Newly developed resolution enhancement lithography assisted by chemical shrink process and materials for next-generation devices*, Japanese Journal of Applied Physics **45** (2006), 5434.

[88] J. Tirapu-Azpiroz, P. Burchard, and E. Yablonovitch, *Boundary layer model to account for thick mask effects in photolithography*, Optical Microlithography, Proc. SPIE, vol. 5040, 2003, pp. 1611–1619.

[89] J. Tirapu-Azpiroz and E. Yablonovitch, *Fast evaluation of photomask near-fields in sub-wavelength 193nm lithography*, Optical Microlithography, Proc. SPIE, vol. 5377, 2004, pp. 1528–1535.

[90] C. Vogel, *Computational Methods for Inverse Problems*, SIAM Press, 2002.

[91] Y. Wang, Y. Pati, H. Watanabe, and T. Kailath, *Automated design of halftoned double-exposure phase-shifting masks*, Optical Microlithography, Proc. SPIE, vol. 2440, 1995, pp. 290–301.

[92] R.G. Wilson, *Fourier Series and Optical Transform Techniques in Contemporary Optics*, John Wiley and Sons, 1995.

[93] A. Wong, *Microlithography: Trends, challenges, solutions, and their impact on design*, IEEE Micro **23** (2003), 12–21.

[94] Alfred Kwok-Kit Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, 2001.

[95] E. Yablonovitch and R. Vrijen, *Optical projection lithography at half the rayleigh resolution limit by two-photon exposure*, Optical Engineering **38** (1999), 334–338.

[96] B. Yenikaya and A. Sezginer, *Model-based assist feature generation*, Advanced Lithography, Proc. SPIE, 2007.