Modeling of Optical Alignment and Metrology in VLSI Manufacturing

Chi-Min Yuan

Electrical and Computer Engineering

12/17/89

Paul Wymbsatt

1/10/90
MODELING OF
OPTICAL ALIGNMENT AND METROLOGY
IN VLSI MANUFACTURING

A DISSERTATION
SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY
in
ELECTRICAL ENGINEERING

by

CHI-MIN YUAN

Pittsburgh, Pennsylvania
October, 1989
To Mannling and Yiheng
Abstract

The objective of this thesis is to develop rigorous models to simulate optical images of semiconductor structures detected by various optical alignment and metrology schemes. Two specific areas we chose to study are optical alignment for mask and wafer, and optical metrology for line-width measurement. Theoretical and experimental verification efforts have been made to demonstrate the validity of the models developed. A simulator based on these models have been implemented to study various aspects of alignment and metrology problems.
Acknowledgements

I would like to acknowledge my gratitudes to my advisor Andrzej Stojwas who led me into this exciting field four years ago. His guidance and confidence on me have made my graduate student years enjoyable.

Also, I would like to thank Dr. Rick Dill who served in my committee and set up an industrial connection for me. Thanks go to Dr. Jerry Shaw for his many valuable discussions and generosity for supplying optical image data, and Dr. Alexander Starikov for supplying resist profile data.

Finally, I would like to thank my parents, who have been offering me great studying environment during the past twenty years, and my wife Mannling, who sacrificed her career so I could concentrate on my research in my final student year.
## Contents

1 Introduction
   1.1 Problem Formulation ............................................. 2
   1.2 Modeling Approach ............................................... 3
   1.3 Outline .......................................................... 5

2 Characterization of Coated Resist Profiles ................. 6
   2.1 Previous Work .................................................... 7
   2.2 Characterization of Coated Resist Profiles ................. 7
   2.3 An Example ....................................................... 9

3 Modeling of Light Scattering .................................... 13
   3.1 Fresnel Model ..................................................... 14
   3.2 Our Wave-Guide Model versus Traditional Wave-Guide Model ... 15
   3.3 Previous Work on Wave-Guide Model ........................... 17
   3.4 TM Polarization .................................................. 21
   3.5 Non-Vertical Incidence ........................................... 25
   3.6 Polarization of Incident Waves ................................. 26
   3.7 Numerical Stability ............................................ 27
      3.7.1 Junction Spike ............................................. 27
      3.7.2 Highly Dissipative Layers ................................ 29
   3.8 Comparison Between Scattering Models ......................... 31
      3.8.1 Rayleigh Hypothesis Model ................................. 31
      3.8.2 Differential Model ......................................... 33
      3.8.3 Integral Model ............................................. 35
      3.8.4 Comparisons Between Scattering Models ................. 38
CONTENTS

4 Modeling of Imaging
   4.1 Introduction to Abbe’s Theory ........................................... 41
   4.2 Light Projection and Sampling ........................................... 43
   4.3 Reduction System ........................................................... 45
   4.4 Scanning ................................................................. 46
   4.5 Defocus ................................................................. 48
   4.6 Laser Beam Incidence .................................................... 50

5 Modeling of Signal Detection and Processing ............................. 53
   5.1 Noise ................................................................. 53
   5.2 Detectors ............................................................... 55
   5.3 Signal Processing Algorithms ........................................... 57

6 Verifications of Models ............................................................. 62
   6.1 Theoretical Verifications .................................................. 62
       6.1.1 Conservation of Energy ........................................... 62
       6.1.2 Reciprocity Theorem ............................................. 64
       6.1.3 Comparison with Abbe’s Theory .................................. 66
       6.1.4 Comparison with Fresnel Model .................................. 68
       6.1.5 Comparison with Differential Model ............................. 69
   6.2 Experimental Verifications ............................................... 72
       6.2.1 Nikon - Optimum L/S Ratio ....................................... 72
       6.2.2 Ultratech - Alignment Images on Al Level ...................... 74
       6.2.3 Phase Contrast Images ........................................... 76

7 Applications .............................................................................. 81
   7.1 Alignment Schemes ......................................................... 81
       7.1.1 Bright Field ......................................................... 82
       7.1.2 Dark Field Scanning .............................................. 85
       7.1.3 Moire Interference Scanning ..................................... 90
   7.2 Metrology Schemes .......................................................... 92
       7.2.1 Confocal Scanning .................................................. 93
       7.2.2 Differential Interference Contrast ............................... 95
       7.2.3 Differential Phase Contrast ..................................... 100

8 Summary .................................................................................. 103
<table>
<thead>
<tr>
<th>CONTENTS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A Two Dimensional Laser Slit</td>
<td>105</td>
</tr>
<tr>
<td>B Energy Carried by an Inhomogeneous Plane Wave</td>
<td>108</td>
</tr>
<tr>
<td>C Simple Model for DIC Scheme</td>
<td>110</td>
</tr>
<tr>
<td>D Simple Model for DPC Scheme</td>
<td>113</td>
</tr>
</tbody>
</table>
## List of Tables

3.1 Summary of comparisons between scattering models. .................................................. 39

6.1 Comparison between the wave-guide model and the differential model
for $h = 0.173 \mu m$. ............................................................................................... 71

6.2 Comparison between the wave-guide model and the differential model
for $h = 0.240 \mu m$. ............................................................................................... 72

6.3 Comparison between the measured and the simulated line-widths from
the DPC scheme. ................................................................................................. 76

6.4 Comparison between the measured and the simulated line-widths from
the differential DIC scheme. ........................................................................... 79
List of Figures

1.1 The problem is formulated using grating approach. Isolated features on reticle and wafer are approximated by gratings with a large period to avoid interference. ........................................... 3
1.2 The simulation flow. ............................................ 4

2.1 Modeled coated resist profile $f_{PR}(x)$ on topography $f(x)$. ........... 8
2.2 Measured and modeled coated resist profiles on grooves with widths equal to (a)1.0 (b)2.0 (c)4.0 (d)6.0 (e)8.0 (f)10.0 (g)15.0 $\mu$m. .......... 11
2.3 The curve-fitted parameters $\alpha_1$, $\alpha_2$, $x_o$, $\beta_1$, $\beta_2$ are plotted versus groove width. ...................................................... 12

3.1 The reflected $E$ field, according to the Fresnel model, is the summation of the Fourier series of two regions multiplied by the reflectivity coefficients. ...................................................... 14
3.2 Comparison between the traditional wave-guide model and the wave-guide model used in this chapter. ...................................................... 16
3.3 Topography stratified into layers in the wave-guide model. ................. 18
3.4 (a) Traditional way to characterize a junction. (b) Our approximation for a junction. ...................................................... 28
3.5 Simulation results of full field images of an oxide window using different $L$’s are shown to demonstrate the convergence. ....................... 29
3.6 A grating of silicon grooves with a period shown is used to illustrate the Rayleigh hypothesis model, the differential model, and the integral model. For simplicity, the illumination is a TE-polarized, vertically incident plane wave. ...................................................... 32

4.1 Basic optical configuration of the imaging model. .......................... 41
4.2 Abbe’s theory of image formation ............................... 42
LIST OF FIGURES

4.3 Different optical configurations generate incident light differently. 44
4.4 An Mx reduction system and a 1x system 45
4.5 (a) Scanning model (b) Defocus model 47
4.6 (a) Defocus mode from the wave aberration theory 49
4.7 Two dimensional laser slit scanning and three dimensional laser spot scanning 50
4.8 Comparisons between 2D slit and 3D spot intensities under various defocus conditions. Pictures on the left - MSS=1μm; on the right - MSS=0.5μm 52

5.1 The flow of signal detection and processing 53
5.2 The incidence of TE and TM waves on the detector surface 57
5.3 (a) Bright field image (b) Dark field image of an oxide line on the silicon substrate 58
5.4 (a) Detected noisy dark field image; subsequently processed by (b) amplitude selective algorithm (c) auto-convolution algorithm (d) bicell cross-convolution algorithm 60

6.1 The difference between the incident energy and the (reflected + transmitted) energies is plotted against the incident angle, represented by the NA. The illumination is under TM polarization. In general, better result is obtained when a larger number of waves for the scattering calculation is utilized 64
6.2 Images simulated by using different orders 65
6.3 Illustration of the reciprocity theorem 66
6.4 The deviation from the reciprocity theorem is plotted against the NA 67
6.5 Simulated mask or reticle images on the photoresist surface using the Abbe's theory and the wave-guide model 68
6.6 Comparison between the images generated by the Fresnel model and the wave-guide model. Figures on the left are from coherent illumination and figures on the right are from Kohler illumination 70
6.7 A sinusoidal grating illuminated by a normally incident plane wave of unit amplitude 71
6.8 Left: The Nikon Laser Step Alignment scheme scans the alignment grating over the stationary elliptic laser spot to generate the alignment signal. Right: A typical alignment signal 73
LIST OF FIGURES

6.9 The detected signal intensity is plotted against different space widths, while the period is kept to be 8μm. ........................................ 74
6.10 The model used to simulate the Ultratech key-target convolution alignment scheme which is operated on the aluminum level. .............. 75
6.11 The detected images (top row) are compared with the simulated images using the Fresnel model (middle row) and the wave-guide model (bottom row). ........................................ 77
6.12 Comparison between the measured and the simulated images from the DPC scheme. ........................................ 78
6.13 Comparison between the measured and the simulated images from the differential DIC scheme. ........................................ 80

7.1 Typical images obtained using the bright field alignment scheme. ... 82
7.2 A highly asymmetric wafer mark is used to study the alignment. .... 83
7.3 Simulated alignment images under e-line illumination. ............... 84
7.4 Simulated alignment images under d-line illumination. ............... 85
7.5 The sum of the images obtaining by using e-line and d-line. ........ 86
7.6 The model used to simulate the alignment images of the GCA steppers. 87
7.7 The simulated alignment images from the GCA DFAS. ............... 88
7.8 The simulated alignment images from the dark field scanning scheme with only the large-angle scattering light collected. ............... 89
7.9 A simple model to calculate the alignment offset due to oblique scanning. 90
7.10 The simulated image of the oblique dark field scanning scheme. .... 90
7.11 The wafer feature used to study the Moire interference alignment scheme. 91
7.12 The alignment images from the Moire interference scanning scheme with the photoresist profile shifted to the left by 0.0, 0.2, and 0.4μm. 92
7.13 (a) The operating principle of the confocal scanning scheme. (b) The shadowing effect. ...................................................... 93
7.14 The confocal scanning images of silicon line and groove of 2μm-thick and 4μm-wide. Upper: focus on top surface; Lower: focus on bottom surface. ...................................................... 94
7.15 The operating principle of the differential interference contrast scheme. 95
7.16 The simulated DIC images of silicon lines and grooves of different heights and depths are shown. The illumination wavelength is 0.6328μm, the NA’s are 0.9, the spot size and the sheared distance δ are 0.4μm. 96
7.17 The reflectivity effect shown in the DIC images can be alleviated by adopting the differential and normalized DIC schemes. .................. 97

7.18 Simulated images of resist-coated alignment structures observed by the DIC, the differential DIC, and the normalized DIC schemes. 1st row: dip=0.00µm; 2nd row: dip=0.04µm; 3rd row: dip=0.08µm; 4th row: dip=0.12µm. ........................................... 99

7.19 Simulated DPC images of a 1µm wide, 1000Å thick silicon line and groove on silicon are shown. These two structures closely resemble phase objects. .............................................. 101

7.20 Simulated DPC images of a 1µm wide, 1000Å thick oxide line and groove on silicon are shown. These two structures closely resemble amplitude objects. ........................................... 101

7.21 Simulated images of resist-coated alignment structures observed by the DPC scheme are shown. .............................................. 102

C.1 A simplified model for the scanning DIC scheme. ......................... 110

C.2 The dark field, DIC, and bright field images of an isolated silicon line on silicon are simulated with the β set to be 0, π, and π/2, respectively.112

D.1 A simplified model for the scanning DPC scheme. ......................... 113
Chapter 1

Introduction

During integrated circuit manufacturing, many high precision measurements and inspections are performed to enable monitoring over the processes so that process control can be facilitated. Typically, these in line measurements and inspections are carried out for sheet resistance, film defect, feature line-width, etc. As the feature size in modern IC processes becomes smaller, it is crucial that the correct feature size to be manufactured can be guaranteed by proper lithography and metrology techniques.

There are several alternatives for performing pattern transfer and line-width metrology. In pattern transfer, optical lithography currently still surpasses other competitors such as x-ray or e-beam lithography in a real production environment mainly due to its mature technology and high throughput. This trend must be still maintained as more and more deep-UV steppers and photoresists become available on the market. In line-width metrology, the common methods used in a fabrication line are the scanning electron microscope (SEM), electrical, and optical measurements. The SEM measurement in general has the best accuracy despite its low speed and destructive nature. The electrical measurement can offer relatively short measuring time but with worse accuracy, and the processing level must have a conductive layer on top to facilitate the measurement. The optical measurement which is nondestructive can offer the highest measurement throughput. However, due to the wave nature of light, the reflectivity and interference effects may degrade the measurement accuracy so that special care must be taken in calibrating the optical tool with other known standards.

In photolithography and optical metrology, there are two major obstacles that often hinder process engineers and equipment manufacturers from adopting smaller design rules. One obstacle is the optical alignment between the reticle and the wafer
which is usually performed by the stepper before each photolithography exposure step. The minimum alignment accuracy required is around a third of the minimum feature size manufactured, and this accuracy is not always achievable due to the complicated wafer topography encountered in IC manufacturing. Different stepper manufacturers adopt their own proprietary schemes to perform alignment and these schemes impose different constraints on the processes in order to cope with the alignment step. The other obstacle is the uncertainty of the measurement accuracy from various optical metrology tools. This uncertainty is mainly due to the wave nature of light, and it is especially severe when the feature size to be measured is comparable to the measurement wavelength used. For these sizes the scalar wave theory no longer holds and the intuition metrology engineers have built up from measurements of larger features gradually deviates from the truth. Thus, there is a need to model these optical schemes rigorously.

1.1 Problem Formulation

In this dissertation, our aim is to develop rigorous models in order to study various aspects of the optical alignment and metrology problems. To achieve this goal, the problem formulation has to be general enough to include most of the alignment and metrology schemes used at present.

Almost all the optical detection schemes adopt the same approach. An incident beam is formed by the projection part of the optical configuration. This incident beam, which may contain reticle information in alignment, can have different degree of coherence depending on the numerical apertures (NA’s) of the configuration. It is projected onto the surface of the object (i.e., wafer structure in alignment and metrology) to be detected and light scattering occurs. The reflected or transmitted light from the object is collected by the collecting portion of the configuration and imaged onto the surface of an optical detector. The portion of the light collected also depends on the NA’s of the optical configuration. The detector transforms the optical signal into electrical signal and proper signal processing is subsequently performed to acquire the information of the object. The whole detection process can be static if the optical configuration and the object remain stationary, or dynamic if the incident beams moves relative to the object with time and the optical signal is detected dynamically.

Our formulation is based on the following sequence:
1. Calculate the incident light pattern on the surface of the object using an imaging model.

2. Calculate the reflected light pattern on the surface of the object using a scattering model.

3. Calculate the collected light pattern on the surface of the detector using the same imaging model.

4. Calculate the final alignment or metrology result using a signal detection model and a suitable signal processing model.

1.2 Modeling Approach

![Diagram of reticle and wafer with isolated and grating features](image)

Figure 1.1: The problem is formulated using grating approach. Isolated features on reticle and wafer are approximated by gratings with a large period to avoid interference.

The formulation is set up in two dimensions, i.e., the optical configuration, the reticle and the wafer features are all assumed to extend indefinitely in the $y$ direction as shown in Fig. 1.1. The optics theories used in the dissertation range from the simple ray-tracing model to the scalar wave theory to the vector Maxwell’s equations. The ray-tracing model is only used in places where qualitative explanations are given. For the light imaging, the scalar theory is adopted. For the light scattering, a vector scattering model which calculates the scattering matrix under different polarizations is utilized. For both the imaging and scattering models, we adopt the grating approach where the reticle and the wafer features are periodical. If isolated features are
modeled, the period of the grating is chosen to be large enough so that interference between neighboring features can be neglected.

![Diagram showing the simulation flow.](image)

Figure 1.2: The simulation flow.

The simulation flow, which is in a slightly different order than the true optical signal flow, is given in Fig. 1.2. The scattering matrix elements $S_{ij}$'s of the wafer feature are calculated first with the topography and the wavelength to be the input parameters. The calculated scattering matrix is then verified by the conservation of energy law and the reciprocity theorem. If the accuracy is not satisfactory, the scattering calculation is redone by increasing the number of waves used to approximate the fields. Once the correct scattering matrix is obtained, the imaging and detection models are subsequently utilized to model various alignment and metrology schemes with the NA's of the optical configuration to be the input parameters, and the simulated optical image can be obtained. By using a proper signal processing algorithm,
the final alignment or metrology result is thus generated.

1.3 Outline

This dissertation is organized as follows:

- Chapter 2 - An analytical equation is developed in order to delineate the asymmetric profile of the coated photoresist layer on wafer topography. This anomalous coating phenomenon is commonly seen in alignment problems.

- Chapter 3 - The wave-guide model, which is a rigorous scattering model, has been enhanced so that more realistic simulations can be performed. Also, other scattering models available in literature are discussed and compared with the wave-guide model.

- Chapters 4 and 5 - The imaging and detection models which are used to simulate the light projection, sampling, and detection due to different configurations of optical tools are given.

- Chapter 6 - Our models introduced in the previous chapters are verified theoretically and experimentally.

- Chapter 7 - Simulation images from different optical alignment and metrology schemes are shown to demonstrate the modeling capabilities of our models.

- Chapter 8 - Our work is summarized and possible extensions are given.
Chapter 2

Characterization of Coated Resist Profiles

In this chapter, we introduce an analytical function developed to characterize the coated resist profiles on non-planar surfaces. Before aligned with the reticle, the alignment mark on the wafer is coated with a layer of photoresist. This coated layer, due to its viscous nature, may not be planar but conform to the topography of the alignment mark. The thickness of this layer varies across the alignment mark, making the reflectivity from the wafer and thus the alignment signal less tractable. Moreover, the coated profile over a symmetric mark may not be symmetric due to the fluid nature of the photoresist when spun over the mark. Very often, this “coating anomaly” causes the alignment signal to slant toward one side and results in alignment offset. Therefore, there is a need to characterize such anomalous coating profiles in order to gain more insight into the alignment problem.

We have developed an analytical function to characterize the coated profiles on alignment marks with vertical sidewalls. Five parameters are required to characterize a particular profile. The advantages of using this analytical expression to delineate the profiles are two-fold. Firstly, after characterizing several coated profiles by making measurements at critical sites of a wafer and by performing curve-fitting, we can easily find out profiles everywhere else on a wafer using interpolation. Secondly, from the simulation point of view, the user can adjust the profile promptly by changing the parameters and observe the profile on the screen. Then, the effect of coating anomaly can be simulated by using these five parameters for fast input. Sets of measured profiles have been used to curve-fit this analytical function, and we have demonstrated that good agreement can be obtained.
2.1 Previous Work

Modeling efforts of the spin-coated resist thickness on planar surfaces have been made by several researchers [59, 18]. It has been shown that to develop a physically-based model, resist solvent loss during spinning, internal force between resist molecules, and non-Newtonian behavior of resist have to be considered. With all these complicated factors, a large amount of CPU time has to be devoted to numerical modeling.

Two empirical methods have been also proposed to characterize the coated resist profile on nonplanar surface. Lavergne et. al.[33] use one parameter to characterize the coated resist profile on nonplanar surface, and a mathematical model to simulate the film shrinkage. White [60] models the coated profile by using the wafer topography while attenuating the high frequency component to obtain the rounded profile. Thus, if the topography profile \( T(y) \) can be expressed by a Fourier series:

\[
T(y) = \sum_{n=0}^{j} \left( a_n \cos \frac{ny}{N} + b_n \sin \frac{ny}{N} \right),
\]

(2.1)

the coated profile \( T'(y) \) can then be expressed by this profile with the high frequency components attenuated:

\[
T'(y) = \sum_{n=0}^{j} \left( w_n a_n \cos \frac{ny}{N} + w_n b_n \sin \frac{ny}{N} \right).
\]

(2.2)

The attenuation coefficients \( w_n \)'s can be defined arbitrarily, and are determined by two parameters in White’s paper. Although both these techniques have been proven to work, they suffer from the incapability of modeling asymmetric profile over a symmetric topography.

2.2 Characterization of Coated Resist Profiles

An analytical function has been developed in this work to characterize the coated photoresist profiles on topography based on White’s high-frequency attenuation concept. The topography can either be an isolated line/groove or a grating, and is assumed to be symmetric with vertical sidewalls. Five parameters are employed in the function so that the coating anomaly can be represented.

Consider an isolated groove with depth \( h \) and width \( w \) having a profile \( f(x) \), as shown in Fig. 2.1. Since the profile is assumed to be symmetric, it can be represented solely by its even components of the Fourier integral:
Figure 2.1: Modeled coated resist profile \( f_{PR}(x) \) on topography \( f(x) \).

\[
f(x) = \int_{0}^{\infty} a(\omega) \cos(\omega x) d\omega,
\]

(2.3)

where

\[
a(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(x) \cos(\omega x) dx.
\]

(2.4)

The profile of the coated photoresist \( f_{PR}(x) \) is then characterized by introducing a high spatial frequency filter \( W(\omega) \):

\[
f_{PR}(x) = \int_{0}^{\infty} W(\omega) a(\omega) \cos(\omega x) d\omega + T_{flat},
\]

(2.5)

where \( W(\omega) \) is arbitrarily defined as

\[
W(\omega) = \exp(-\alpha \omega),
\]

(2.6)

and \( T_{flat} \) is the coated resist thickness on flat surface. After algebraic manipulation, Eq. 2.5 becomes

\[
f_{PR}(x) = \frac{h}{\pi} \left[ \tan^{-1} \left( \frac{\frac{4}{\alpha} + \frac{\pi}{2}}{\alpha} \right) + \tan^{-1} \left( \frac{\frac{4}{\alpha} - \frac{\pi}{2}}{\alpha} \right) \right] + T_{flat}.
\]

(2.7)

Note that the profile generated by this function is always symmetric, and is completely determined by the parameter \( \alpha \).

To account for the asymmetric coating, we split the profile into two halves at \( x = x_o \) so that:

\[
f_{PR}(x) = \begin{cases} 
\frac{h}{\pi} \beta_1 \left[ \tan^{-1} \left( \frac{\frac{4}{\alpha_1} + \frac{\pi - 2\delta}{2}}{\alpha_1} \right) + \tan^{-1} \left( \frac{\frac{4}{\alpha_1} - \frac{\pi - 2\delta}{2}}{\alpha_1} \right) \right] + T_{flat} + \delta & \text{if } x \leq x_o \\
\frac{h}{\pi} \beta_2 \left[ \tan^{-1} \left( \frac{\frac{4}{\alpha_2} + \frac{\pi - 2\delta}{2}}{\alpha_2} \right) + \tan^{-1} \left( \frac{\frac{4}{\alpha_2} - \frac{\pi - 2\delta}{2}}{\alpha_2} \right) \right] + T_{flat} & \text{if } x \geq x_o
\end{cases}
\]

(2.8)
Parameters $\beta_1$ and $\beta_2$ are added to the function to adjust for the “heights” of the profile on both sides. To make the profile continuous and smooth at the junction $x_0$, a coefficient $\delta$ is added which can be easily determined by knowing $\alpha_1, \alpha_2, \beta_1, \beta_2$. Thus, a coated profile can be characterized by Eq. 2.8 with five parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, x_0$.

A grating profile with period $P$ can be treated as superposition of many individual lines/grooves separated by $P$. Assuming that the coated photoresist profile still retain the periodic nature of the underlying grating, we can represent the coated profile $f_{PR}(x)$ by Eq. 2.8 which satisfies:

$$f_{PR}(x) = f_{PR}(x \pm nP),$$

where $n$ is an arbitrary integer.

### 2.3 An Example

Preliminary measurements on a set of coated photoresist profiles have been performed[55] and the results are shown in Fig. 2.2. A layer of resist of approximately 0.95$\mu$m-thick is spin-coated on 0.37$\mu$m-deep grooves. The nominal widths of the grooves are 1.0, 2.0, 4.0, 6.0, 8.0, 10.0 and 15.0$\mu$m, respectively. A stylus-based profiler is used to trace the latitude of the resist profile so that the coated profile can be recorded. Leveling and centering have been carried out; however, the accuracy of centering is not well known yet. Note that the errors of the measurements may be up to 0.1-0.2$\mu$m.

The measurements are curve-fitted by using Eq. 2.8, and the results are shown in the same figure. For this particular combination of the resist material and spin-coating process, the coating anomaly does not appear to be significant until the groove width goes up to 8$\mu$m. In Fig 2.2-(g), resist pile-up can be observed on the right side (i.e., away from the wafer center) of the groove. The largest error between measured and modeled profiles in this example is approximately 0.02$\mu$m, which is less than a tenth of the ordinary alignment wavelength ($\frac{\lambda}{n} \sim \frac{0.5}{1.68} \sim 0.3\mu$m) inside the resist.

The curve-fitted parameters $\alpha_1$, $\alpha_2$, $x_0$, $\beta_1$, $\beta_2$ are plotted versus groove width, as shown in Fig. 2.3. Intuitively, as the groove width increases, $\alpha_i$’s should decrease because the coated profile tends to conform to the topography, and $\beta_i$’s should increase because the resist “dip” tends to become larger. However, Fig. 2.3 shows that the trend can only be observed with the width above 6$\mu$m, except for $\alpha_2$ which increases with increasing groove width. One obvious explanation for this discrepancy is that there is no simple intuitive description for the behavior of the parameters in this
mathematical model. However, more data is necessary before coming to a definitive conclusion.
Figure 2.2: Measured and modeled coated resist profiles on grooves with widths equal to (a) 1.0 (b) 2.0 (c) 4.0 (d) 6.0 (e) 8.0 (f) 10.0 (g) 15.0 μm.
Figure 2.3: The curve-fitted parameters $\alpha_1$, $\alpha_2$, $x_o$, $\beta_1$, $\beta_2$ are plotted versus groove width.
Chapter 3

Modeling of Light Scattering

The essence of alignment and metrology modeling lies in the accurate solutions of the light scattering problem. The features in most of the advanced IC processes are not much larger than the illumination wavelength used so that traditional scalar theory cannot be applied to this problem and the Maxwell's equations have to be solved rigorously.

The aim of this chapter is to introduce the “wave-guide” model [10, 28, 45, 46, 70] which is able to handle this problem. The wave-guide model, which is a vector model, is adopted to solve the light scattering problem. In the previous work, the model is restricted to calculating the scattering matrix of TE (i.e., $E$ field oscillates in parallel to the wafer structure) polarized light illuminating vertically. In this chapter, the wave-guide model is extended to include TM (i.e., $H$ field oscillates in parallel to the wafer structure) polarized light so that light incidence with arbitrary polarization including random polarization can be modeled. Also, an efficient way to calculate the scattering matrix with non-vertical illumination is introduced. This numerical technique significantly reduces the computation time and makes modeling of illumination with large NA possible. Some numerical instability issues frequently encountered in the scattering problem are also discussed and solutions are given. However, problems still exist when dealing with highly dissipative materials such as metals. Finally, a comparison between the wave-guide model and other scattering models is given to show the pros and cons of these models in various aspects. Theoretical and experimental verifications of the wave-guide model are presented in Chapter 6.
3.1 Fresnel Model

The "Fresnel model" is presented in this section before going into the wave-guide model. The Fresnel model was, to our knowledge, first introduced by Nygysonen[45] in order to solve the scattering problem from coherent illumination. Kirk[31] implemented this model to account for vertical incidence of TE beam. In this section, we extend this model to incorporate illumination at arbitrary angles for both TE and TM polarizations. The model is less rigorous than the wave-guide model. However, because of its accuracy in some applications, its ease of implementation, and its efficiency in numerical computation, it deserves some attention.

Consider a periodic feature, with only one period $P$ shown in Fig. 3.1. The feature is constructed by two regions, each with a stack of homogeneous layers. An incident plane wave of order $n_0$ (i.e., incident angle $= \sin^{-1} \frac{n_0 \lambda_0}{P}$) impinges on the feature. The objective here is to find the reflected field due to this incident $E$ field.

![Diagram](image)

Figure 3.1: The reflected $E$ field, according to the Fresnel model, is the summation of the Fourier series of two regions multiplied by the reflectivity coefficients.

If the feature is treated as a combination of two individual samples, each composed of a stack of infinitely large and homogeneous layers from a region, then the reflectivity coefficients (i.e., the ratio of the reflected to the incident $E$ fields) $r_1$ and $r_2$ from each sample for both TE and TM polarizations can be obtained using the Fresnel equations[23]. Also, the Fourier series for these two regions can be easily obtained:

$$f_1(x) = \sum_{n=-N}^{N} a_{1,n} e^{i2\pi nx},$$

and
$$f_2(x) = \sum_{n=-N}^{N} a_{2,n} e^{i2\pi n b x}.$$  

Then, the Fresnel model approximates the reflected $E$ field by the summation of the reflected fields from these two samples, i.e., the reflected $E$ field is:

$$E_{ref}(x) = \sum_{n=-N}^{N} \left( r_1 a_{1,n-n_o} e^{i2\pi n b x} + r_2 a_{2,n-n_o} e^{i2\pi n b x} \right),$$

where $N$ is determined by the NA of the lenses involved.

The Fresnel model does not rigorously treat the scattering problem at the junctions of regions. Therefore, it cannot serve as a rigorous model to observe the edge ringing commonly existing in optical images. However, it does generate correct contrast for bright-field type images because the reflectivity of each region is solved using Fresnel equation, which is originally derived from the Maxwell's equations. Also, this simple model often generates results with the correct trend, as will be seen in Chapter 6. Therefore, it can prove to be powerful for some applications.

As a rule of thumb, the Fresnel model gives accurate results for thin film structures, correct contrast for bright-field type images, and right trend for images illuminated by small NA’s. The model is not able to generate accurate ringing details of thick film structures, and images of features of small widths (in submicron range).

### 3.2 Our Wave-Guide Model versus Traditional Wave-Guide Model

The terminology “wave-guide” model may be misleading in the sense that no wave is intentionally “guided” inside the wafer features. The terminology came from the fact that the wave-guide model has some mathematical similarities to the traditional dielectric wave-guide model[67] which deals mainly with guided waves. A comparison between the traditional wave-guide model and our wave-guide model is shown in Fig. 3.2 to clarify the similarities and the differences between these two models.

A simple sandwiched structure which extends indefinitely in $\pm y$ directions is used. For simplicity, only the TE case is discussed in this section. The traditional dielectric wave-guide model concentrates on finding the modes of guided waves that propagate along the $+z$ direction, and the incident beam is mainly along the $+z$ direction as well. Our wave-guide model calculates the reflected $E$ field above the structure surface, and the incident beam comes from various directions according to the objective NA. The
<table>
<thead>
<tr>
<th>geometry</th>
<th>traditional wave-guide model</th>
<th>our wave-guide model</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x₁</td>
<td>ε₁</td>
<td>ε₁</td>
</tr>
<tr>
<td>x₂</td>
<td>ε₂</td>
<td>ε₁</td>
</tr>
<tr>
<td>ε₁</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| objective | guided mode(s) | reflected E field |

| solution form | \( E(x,z) = X(x) \) \( e^{i\alpha z} e^{-i\omega t} \) |

Maxwell eq. (TE):
\[
\frac{d^2 E}{dx^2} + \frac{d^2 E}{dz^2} - \mu_0 \varepsilon(x) \frac{d^2 E}{dt^2} = 0
\]
\[
\rightarrow \frac{d^2 X}{dx^2} + [k_0^2 \varepsilon(x) - \alpha^2] X = 0
\]

| how to find modes | 1. set \( X_1 = C_1 e^{-k_1 x} \)  
|                  | \( X_2 = C_2 e^{-i k_2 x} \)  
|                  | \( X_3 = C_3 e^{-k_3 x} \)  
|                  | 2. match B.C. at \( x_1 \) and \( x_2 \)  
|                  | 3. finite # of modes obtained  |
|                  | 1. set \( X(x) = \sum_n B_n e^{-i k_n x} \)  
|                  | 2. match coeff. of same order  
|                  | 3. infinite # of modes obtained  |

Figure 3.2: Comparison between the traditional wave-guide model and the wave-guide model used in this chapter.
Maxwell's equation governing the behavior of wave propagation is the same in both cases, and it evolves into an eigenvalue problem:

\[ \frac{\partial^2 X(x)}{\partial x^2} + \left[ k_0^2 \varepsilon(x) - \alpha^2 \right] X(x) = 0. \]

Here, \( \alpha \)'s are the modes, which may either be guided or evanescent. The goal now is to find the solution for \( X(x) \). Because the traditional wave-guide model is mainly concerned with the guided modes, solutions of special forms can usually be acquired and finite number of modes can be obtained by matching the boundary conditions. On the other hand, our wave-guide model calculates the scattering problem in general, so both guided and evanescent modes have to be included and the solution must have a general form. Because the number of the guided modes and evanescent modes is infinite, some higher order evanescent modes must be discarded in our wave-guide model during the numerical calculation.

### 3.3 Previous Work on Wave-Guide Model

The wave-guide model is a numerical model that calculates the amplitudes of the reflected/transmitted waves from the illuminated wafer structure, as shown in Fig. 3.3. Burckhardt[10] first developed the wave-guide model to calculate the diffraction patterns from sinusoidal dielectric gratings and applied it to hologram study. The incident angle can be arbitrary, and the polarization can either be TE or TM. Kaspar[28] modified the model to include absorbing materials and nonsinusoidal grating structures. Nyyssonen[45] adopted this model to calculate the reflected fields from semiconductor structures. In her model, the illumination was limited to be vertical, and the polarization was limited to be TE. Kirk[30, 46] extended the model to account for topography with an arbitrary profile using stratified layer technique. In summary, the wave-guide model was capable of calculating the reflected/transmitted fields from topography with an arbitrary profile by the TE mode, vertically incident illumination.

A TE plane wave is incident onto the surface of the structure as shown in Fig. 3.3. Normal incidence is assumed here. The objective is to calculate the amplitudes \( E_i^0 \)'s of the reflected waves.

The Maxwell's equation governing the electric field \( E^j(x, z) \) (\( = y E^j(x, z) \)) inside any layer \( j \) is:
\[ \nabla^2 E^j - \mu_o \varepsilon_o \varepsilon^j(x) \frac{\partial^2 E^j}{\partial t^2} + \nabla \left( E^j \cdot \frac{\nabla \varepsilon^j(x)}{\varepsilon^j(x)} \right) = 0, \]

where \( \varepsilon^j(x) \) is the dielectric constant, which is a function of \( x \), for the \( j \)th layer. For the TE mode, the above equation becomes:

\[ \nabla^2 E^j - \mu_o \varepsilon_o \varepsilon^j(x) \frac{\partial^2 E^j}{\partial t^2} = 0. \]  

(3.1)

From now on, the universal constants \( \mu_o \) and \( \varepsilon_o \) will be omitted for simplicity. The relation between the fields is therefore given by \( H_x = -\partial E_y / \partial z \).

In layer 0, which is air with \( \varepsilon(x) \) being equal to one, the \( E \) and \( H \) fields can be expressed by:

\[ E^0_y = e^{i k_0 x} + \sum_{l=-L}^{+L} E^0_l e^{i 2 \pi i b x} e^{-i k_0 \sqrt{1 - (b \lambda_o)^2} z}, \]  

(3.2)
\[ H_x^0 = -i k_0 e^{i k_0 z} + \sum_{l=-L}^{+L} i k_0 \sqrt{1 - (l b \lambda_o)^2} E_l^0 e^{i 2 \pi l b x} e^{-i k_0 \sqrt{1 - (l b \lambda_o)^2} z}, \]

where \( b \) is the inverse of the structure period, and \( \lambda_o \) is the wavelength in air.

In the substrate layer \( s \), the waves are assumed to travel downward with no reflection from the wafer bottom. The dielectric constant \( \varepsilon(x) \) is equal to \( \varepsilon_s \), and the \( E \) and \( H \) fields can be expressed as:

\[ E_y^s = \sum_{l=-L}^{+L} E_l^s e^{i 2 \pi l b x} e^{i k_0 \sqrt{\varepsilon_s - (l b \lambda_o)^2} z}, \]

\[ H_z^s = -\sum_{l=-L}^{+L} i k_0 \sqrt{\varepsilon_s - (l b \lambda_o)^2} E_l^s e^{i 2 \pi l b x} e^{i k_0 \sqrt{\varepsilon_s - (l b \lambda_o)^2} z}, \]

where \( \varepsilon_s \) is the dielectric constant of the substrate.

In an intermediate layer \( j \), the slab is composed of different materials and is characterized by its dielectric constant \( \varepsilon^j(x) \). The wave equation Eq. 3.1 can be decomposed into two ordinary differential equations using the separation of variables technique (i.e., \( E^j(x, z) = X^j(x) Z^j(z) \)):

\[ \frac{\partial^2 X^j}{\partial x^2} + \left[ k_0^2 \sum_q \varepsilon_q^j e^{i 2 \pi q b x} + (\alpha^j)^2 \right] \cdot X^j = 0, \]

\[ \frac{\partial^2 Z^j}{\partial z^2} - (\alpha^j)^2 Z^j = 0, \]

where

\[ \varepsilon^j(x) = \sum_q \varepsilon_q^j e^{i 2 \pi q b x}. \]

The solution of \( X^j(x) \) has a general form

\[ X^j(x) = \sum_l B_{l,m}^j e^{i 2 \pi l b x}, \]

according to the Floquet theorem[61]. Then, the \( E \) and \( H \) fields can be expressed by:

\[ E_y^j = \sum_{m=-L}^{+L} \left[ \left( A_m^j e^{\alpha_m^j z} + A_m^j e^{-\alpha_m^j z} \right) \sum_{l=-L}^{+L} B_{l,m}^j e^{i 2 \pi l b x} \right], \]

\[ H_x^j = -\sum_{m=-L}^{+L} \left[ \left( A_m^j e^{\alpha_m^j z} - A_m^j e^{-\alpha_m^j z} \right) \sum_{l=-L}^{+L} B_{l,m}^j e^{i 2 \pi l b x} \right]. \]
Physically, the $E$ and $H$ fields can be thought of as linear combinations of waves with discrete wave-guide modes.

The amplitudes $E_i^0$'s of the reflected waves can be found by employing the boundary conditions that $E_x$ and $H_z$ are continuous across any interface of two neighboring slabs. By matching the fields at $z = 0$, one obtains

$$E_i^0 = \sum_{m=-L}^{+L} \left( A_m^1 + A_m^n \right) B_{i,m}^1 - \delta_{i,0}, \quad (3.6)$$

$$\sqrt{1 - (l \beta \lambda_o)^2} E_i^0 = -\frac{1}{ik_0} \sum_{m=-L}^{+L} \left( A_m^1 - A_m^n \right) \alpha_m^1 B_{i,m}^1 + \delta_{i,0},$$

where $\delta_{i,j}$ is the Kronecker delta which is equal to one if $i = j$, and is equal to zero otherwise. These two equations can also be written as

$$\sum_{m=-L}^{+L} \left( \sqrt{1 - (l \beta \lambda_o)^2 + \frac{\alpha_m^1}{ik_0}} \right) B_{i,m}^1 A_m^1 = \sum_{m=-L}^{+L} \left( \sqrt{1 - (l \beta \lambda_o)^2 - \frac{\alpha_m^1}{ik_0}} \right) B_{i,m}^1 A_m^n = 2\delta_{i,0},$$

i.e., in matrix notation,

$$\begin{bmatrix} C_{11}^0 & C_{12}^0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_1^1 \\ A_1^n \end{bmatrix} = \begin{bmatrix} R \\ 0 \end{bmatrix}, \quad (3.7)$$

where

$$C_{11}^0(l, m) = \left( \sqrt{1 - (l \beta \lambda_o)^2 + \frac{\alpha_m^1}{ik_0}} \right) B_{i,m}^1,$$

$$C_{12}^0(l, m) = \left( \sqrt{1 - (l \beta \lambda_o)^2 - \frac{\alpha_m^1}{ik_0}} \right) B_{i,m}^1,$$

$$R_i = 2\delta_{i,0}.$$

Similarly, by matching the fields at $z = z_j$, one obtains an equation that relates the fields between layers $j$ and $j + 1$

$$\begin{bmatrix} C_{11}^j & C_{12}^j \\ C_{21}^j & C_{22}^j \end{bmatrix} \begin{bmatrix} A_i^j \\ A_i^{j+1} \end{bmatrix} = \begin{bmatrix} E_{i1}^{j+1} \\ E_{i2}^{j+1} \end{bmatrix}, \quad (3.8)$$

where

$$C_{11}^j(l, m) = e^{\alpha_m^j z_j} B_{i,m}^j \quad C_{12}^j(l, m) = e^{-\alpha_m^j z_j} B_{i,m}^j,$$

$$C_{21}^j(l, m) = \alpha_m^j e^{\alpha_m^j z_j} B_{i,m}^j \quad C_{22}^j(l, m) = -\alpha_m^j e^{-\alpha_m^j z_j} B_{i,m}^j,$$

$$E_{11}^{j+1}(l, m) = e^{\alpha_m^{j+1} z_j} B_{i,m}^{j+1} \quad E_{12}^{j+1}(l, m) = e^{-\alpha_m^{j+1} z_j} B_{i,m}^{j+1},$$

$$E_{21}^{j+1}(l, m) = \alpha_m^{j+1} e^{\alpha_m^{j+1} z_j} B_{i,m}^{j+1} \quad E_{22}^{j+1}(l, m) = -\alpha_m^{j+1} e^{-\alpha_m^{j+1} z_j} B_{i,m}^{j+1}. \quad (3.9)$$
By matching the boundary conditions at $z = T$, one obtains
\[
\begin{bmatrix}
C^n_{11} & C^n_{12} \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
A^n \\
A'^n
\end{bmatrix} =
\begin{bmatrix}
0 \\
0
\end{bmatrix},
\] (3.10)

where
\[
C^n_{11}(l, m) = \left(\sqrt{\varepsilon_s - (lb \lambda_0)^2} - \frac{\alpha_m}{ik_0}\right)e^{\alpha_m T} B^m_{i,m},
\]
\[
C^n_{12}(l, m) = \left(\sqrt{\varepsilon_s - (lb \lambda_0)^2} + \frac{\alpha_m}{ik_0}\right)e^{-\alpha_m T} B^m_{i,m}.
\] (3.11)

Now, by combining Eqs. 3.8 and 3.10, we get
\[
\begin{bmatrix}
0 \\
0
\end{bmatrix} =
\begin{bmatrix}
C^n_{11} & C^n_{12} \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
E^n_{11} & E^n_{12} \\
E^n_{21} & E^n_{22}
\end{bmatrix}^{-1}
\begin{bmatrix}
C^{n-1}_{11} & C^{n-1}_{12} \\
C^{n-1}_{21} & C^{n-1}_{22}
\end{bmatrix}
\begin{bmatrix}
E^{n-1}_{11} & E^{n-1}_{12} \\
E^{n-1}_{21} & E^{n-1}_{22}
\end{bmatrix}^{-1}
\begin{bmatrix}
C^0_{11} & C^0_{12} \\
C^0_{21} & C^0_{22}
\end{bmatrix}
\begin{bmatrix}
A^1 \\
A'^1
\end{bmatrix}
\]
\[
= \begin{bmatrix}
C^0_{21} & C^0_{22} \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
A^1 \\
A'^1
\end{bmatrix}.
\] (3.12)

Furthermore, Eq. 3.12 can be combined with Eq. 3.7 to yield a linear system
\[
\begin{bmatrix}
C^0_{11} & C^0_{12} \\
C^0_{21} & C^0_{22}
\end{bmatrix}
\begin{bmatrix}
A^1 \\
A'^1
\end{bmatrix} =
\begin{bmatrix}
R \\
0
\end{bmatrix}.
\] (3.13)

Here, matrix $C$ contains only the topography information (e.g., topography dimensions, refractive indices of materials, etc.) and vector $R$ contains only the illumination information. Vector $[A^1, A'^1]$ contains the scattering results, and can be solved for by using any standard linear system solver.

Finally, the amplitudes of the reflected waves can be calculated from Eq. 3.6.

### 3.4 TM Polarization

In this section, the wave-guide model is extended to calculate the reflected waves from an arbitrarily shaped structure due to TM illumination. The incidence is assumed to be vertical.

In the TM case, it is simpler to solve for the magnetic field $H$, and the $E$ field can be obtained from $E_x = (1/e(x))(\partial H_y/\partial z)$. The Maxwell's equation governing the TM scattering problem inside layer $j$ is
\[ \nabla^2 \mathbf{H}^j - \mu_o \varepsilon_o \epsilon^j(x) \frac{\partial^2 \mathbf{H}^j}{\partial t^2} + \frac{\nabla \epsilon^j(x)}{\epsilon^j(x)} \times \left( \nabla \times \mathbf{H}^j \right) = 0. \] (3.14)

Mathematically, it is more difficult to deal with TM illumination because an extra term (i.e., the third term) appears in the above equation as compared with the equivalent Maxwell’s equation (Eq. 3.1) for TE illumination. This extra term increases the complexity of the equation formulation, and it is also the source of some numerical stability problems as will be mentioned in this chapter.

By substituting the Fourier series for \( \epsilon^j \) and \( 1/\epsilon^j \) into the above equation, it becomes

\[
\frac{\partial^2 H^j}{\partial x^2} + \frac{\partial^2 H^j}{\partial z^2} + k_o^2 \sum_q \epsilon_q^j e^{i2\pi qbx} H^j - \sum_q (i2\pi qb) e_q^j e^{i2\pi qbx} \sum_q \epsilon_q^j e^{i2\pi qbx} \frac{\partial H^j}{\partial x} = 0,
\]

where

\[ \epsilon^j(x) = \sum_q \epsilon_q^j e^{i2\pi qbx}, \quad \epsilon^j(x) = \frac{1}{\epsilon^j(x)} = \sum_q \epsilon_q^j e^{i2\pi qbx}, \]

have been used. By using the separation of variable \( H^j = X^j(x) Z^j(z) \), one obtains

\[
\frac{\partial^2 X^j}{\partial x^2} - \sum_q C_q^j e^{i2\pi qbx} \cdot \frac{\partial X^j}{\partial x} + \left[ k_o^2 \sum_q \epsilon_q^j e^{i2\pi qbx} + \left( \alpha^j \right)^2 \right] \cdot X^j = 0, \]

\[
\frac{\partial^2 Z^j}{\partial z^2} - \left( \alpha^j \right)^2 Z^j = 0,
\]

where

\[
\sum_q C_q^j e^{i2\pi qbx} = \sum_q (i2\pi qb) \epsilon_q^j e^{i2\pi qbx} \cdot \sum_q \epsilon_q^j e^{i2\pi qbx}.
\]

Assuming that \( X^j(x) \) can be expressed by a Fourier series as in the TE case, i.e., \( X^j(x) = \sum_i B_i^j e^{i2\pi lbx} \), Eq. 3.15 becomes

\[
\sum_i (2\pi lb)^2 B_i^j e^{i2\pi lbx} + \sum_q C_q^j e^{i2\pi qbx} \cdot \sum_i (i2\pi lb) B_i^j e^{i2\pi lbx} - \left[ k_o^2 \sum_q \epsilon_q^j e^{i2\pi qbx} + \left( \alpha^j \right)^2 \right] \cdot \sum_i B_i^j e^{i2\pi lbx} = 0.
\]

Equating the coefficients of the same order \( j \), one finds that \( (\alpha^j_m)^2 \) and \( B_{i,m}^j \) are the eigenvalues and eigenvectors of a \((2L + 1) \times (2L + 1)\) matrix \( D \), where

\[
D_{pq} = \begin{cases} 
(2\pi pb)^2 - k_o^2 \epsilon_q^j + C_0^j (i2\pi pb) & p = q \\
-k_o^2 \epsilon_{p-q}^j + C_{p-q}^j (i2\pi qb) & p \neq q.
\end{cases}
\]
Therefore, \( \alpha_m^j \) and \( B_{i,m}^j \) for layer \( j \) can be obtained.

In layer 0, the \( H \) and \( E \) fields can be expressed by:

\[
H_0^j = e^{i k_0 z} + \sum_{l=-L}^{+L} H_{l}^0 e^{i 2 \pi l b x} e^{-i k_0 \sqrt{1 - (l b \lambda_0)^2} z}, \tag{3.16}
\]

\[
E_0^j = i k_0 e^{i k_0 z} - \sum_{l=-L}^{+L} i k_0 \sqrt{1 - (l b \lambda_0)^2} H_{l}^0 e^{i 2 \pi l b x} e^{-i k_0 \sqrt{1 - (l b \lambda_0)^2} z}.
\]

The objective now is to find the amplitudes \( H^0_l \)'s of the reflective waves.

In an intermediate layer \( j \),

\[
H_j^l = \sum_{m=-L}^{+L} \left[ (A^m_j e^{\alpha_m^j z} + A_{-m}^j e^{-\alpha_m^j z}) \sum_{l=-L}^{+L} B_{i,m}^j e^{i 2 \pi l b x} \right],
\]

\[
E_j^l = \sum_{m=-L}^{+L} \left[ (A^m_j e^{\alpha_m^j z} - A_{-m}^j e^{-\alpha_m^j z}) \alpha_m^j \sum_{l=-L}^{+L} B_{i,m}^j e^{i 2 \pi l b x} \right],
\]

where \( \sum_l B_{i,m}^j e^{i 2 \pi l b x} = \sum_l B_{i,m}^j e^{i 2 \pi l b x} \cdot \sum_q e^{i 2 \pi q b x} \) has been used.

In the substrate layer \( s \),

\[
H_s^l = \sum_{l=-L}^{+L} H_{l}^s e^{i 2 \pi l b x} e^{i k_0 \sqrt{\epsilon_s - (l b \lambda_0)^2} z}, \tag{3.17}
\]

\[
E_s^l = \frac{1}{\epsilon_s} \sum_{l=-L}^{+L} i k_0 \sqrt{\epsilon_s - (l b \lambda_0)^2} H_{l}^s e^{i 2 \pi l b x} e^{i k_0 \sqrt{\epsilon_s - (l b \lambda_0)^2} z}.
\]

By matching the fields at \( z = 0 \), one obtains

\[
H_l^0 = \sum_{m=-L}^{+L} (A_m^1 + A_{-m}^1) B_{i,m}^1 - \delta_{l,0}, \tag{3.18}
\]

\[
\sqrt{1 - (l b \lambda_0)^2} H_l^0 = -\frac{1}{i k_0} \sum_{m=-L}^{+L} (A_m^1 - A_{m}^1) \alpha_m^1 B_{i,m}^1 + \delta_{l,0},
\]

or equivalently:

\[
\sum_{m=-L}^{+L} \left( \sqrt{1 - (l b \lambda_0)^2} B_{i,m}^1 + \frac{\alpha_m^1}{i k_0} B_{i,m}^1 \right) A_m^1 +
\]

\[
\sum_{m=-L}^{+L} \left( \sqrt{1 - (l b \lambda_0)^2} B_{i,m}^1 - \frac{\alpha_m^1}{i k_0} B_{i,m}^1 \right) A_m^1 = 2 \delta_{l,0},
\]
and in matrix notation:

\[
\begin{bmatrix}
C_{11}^0 & C_{12}^0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
A^1 \\
A'^1
\end{bmatrix}
= 
\begin{bmatrix}
R \\
0
\end{bmatrix},
\]

where

\[
C_{11}(l, m) = \sqrt{1 - (lb\lambda_o)^2} B_{l,m}^1 + \frac{\alpha_m}{ik_0} B_{l,m}'^1,
\]

\[
C_{12}(l, m) = \sqrt{1 - (lb\lambda_o)^2} B_{l,m}^1 - \frac{\alpha_m}{ik_0} B_{l,m}'^1,
\]

\[R_l = 2\delta_{l,0}.\]

By matching the fields at \(z = z_j\), one obtains

\[
\begin{bmatrix}
C_{11}^j & C_{12}^j \\
C_{21}^j & C_{22}^j
\end{bmatrix}
\begin{bmatrix}
A^j \\
A'^j
\end{bmatrix}
= 
\begin{bmatrix}
E_{11}^{j+1} \\
E_{21}^{j+1}
\end{bmatrix},
\]

\[
\begin{bmatrix}
C_{11}^j & C_{12}^j \\
C_{21}^j & C_{22}^j
\end{bmatrix}
\begin{bmatrix}
A^j \\
A'^j
\end{bmatrix}
= 
\begin{bmatrix}
E_{12}^{j+1} \\
E_{22}^{j+1}
\end{bmatrix},
\]

where

\[
C_{11}^j(l, m) = e^{\alpha_m z_j} B_{l,m}^j,
\]

\[
C_{12}^j(l, m) = e^{-\alpha_m z_j} B_{l,m}^j,
\]

\[
C_{21}^j(l, m) = \alpha_m e^{\alpha_m z_j} B_{l,m}^j,
\]

\[
C_{22}^j(l, m) = -\alpha_m e^{-\alpha_m z_j} B_{l,m}^j,
\]

\[
E_{11}^{j+1}(l, m) = e^{\alpha_m z_j} B_{l,m}'^{j+1},
\]

\[
E_{12}^{j+1}(l, m) = e^{-\alpha_m z_j} B_{l,m}'^{j+1},
\]

\[
E_{21}^{j+1}(l, m) = \alpha_m e^{\alpha_m z_j} B_{l,m}'^{j+1},
\]

\[
E_{22}^{j+1}(l, m) = -\alpha_m e^{-\alpha_m z_j} B_{l,m}'^{j+1}.
\]

(3.19)

By matching the boundary conditions at \(z = T\), one obtains

\[
\begin{bmatrix}
C_{11}^n & C_{12}^n \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
A^n \\
A'^n
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0
\end{bmatrix},
\]

where

\[
C_{11}^n(l, m) = \left(\sqrt{\epsilon_s - (lb\lambda_o)^2} B_{l,m}^n - \frac{\alpha_m}{ik_0} \epsilon_s B_{l,m}'^n\right) e^{\alpha_m T},
\]

\[
C_{12}^n(l, m) = \left(\sqrt{\epsilon_s - (lb\lambda_o)^2} B_{l,m}^n + \frac{\alpha_m}{ik_0} \epsilon_s B_{l,m}'^n\right) e^{-\alpha_m T}.
\]

(3.20)

Then, by following the same procedure as in the TE case, one can obtain a linear system as in Eq. 3.13 except that \(C_{ij}^0\)'s are different. The amplitudes of the reflective waves can be finally derived from Eq. 3.18.
3.5 Non-Vertical Incidence

The scattering problem discussed so far is restricted to normal incidence. In reality, the structure is illuminated by an objective lens with finite NA. Therefore, the scattering due to non-vertical incidence has to be taken into account.

In the previous work [10], a linear system of equations

\[
\begin{bmatrix}
  C_{11}^0 & C_{12}^0 \\
  C_{21}^0 & C_{22}^0 \\
\end{bmatrix}
\begin{bmatrix}
  A_1^1 \\
  A_1^1 \\
\end{bmatrix}
= \begin{bmatrix} R \\ 0 \end{bmatrix}
\]

is derived to describe illumination at a particular incident angle. For different incident angles, vector \( R \) remains unchanged while the matrix \( C \) is altered. Therefore, in order to solve the scattering problem due to illumination at different angles, different \( C \) matrices have to be set up, which is computationally expensive.

We have made the following observation in the equation formulation that can drastically reduce the computational effort. If the incident plane wave is of order \( l_o \), i.e. the incident angle equals \( \sin^{-1}(l_o b \lambda_o) \), the incident term in Eq. 3.2 and 3.16 will become \( e^{ik_o \sqrt{1-(l_o b \lambda_o)^2}} e^{i2\pi l_o b x} \) instead of \( e^{ik_0 x} \). This results in replacing vector \( R \) in Eq. 3.13 by

\[
R_l = 2\sqrt{1 - (l_o b \lambda_o)^2} \delta_{l,l_o}
\]

while matrix \( C \) remains unchanged. Also, if the incident angle does not belong to any of the diffraction angles, the incident plane wave can still be represented by a linear combination of these diffraction waves. Mathematically, if the incident wave can be expressed by

\[
E(x) = \sum_l E_n e^{i2\pi l b x},
\]

the element of vector \( R \) becomes

\[
R_l = 2\sqrt{1 - (l_o b \lambda_o)^2} E_l.
\]

The advantage of arranging the linear system (i.e., Eq. 3.13) in this way is that only the excitation vector \( R \) is altered for different incident angles, while matrix \( C \) is kept unchanged. Numerically, matrix \( C \) needs to be set up only once to solve the scattering problem at an arbitrary incident angle. Thus, there is a considerable saving in CPU time.
3.6 Polarization of Incident Waves

Having the capability of calculating the scattering fields for waves with either TE or TM polarization, we can, in theory, calculate for waves of arbitrary polarizations. This is possible because an arbitrarily polarized plane wave may be expressed by the linear combination of a TE and a TM waves\[5\]. In the following, it is shown how the polarization of the illumination beam for alignment or metrology is characterized.

Laser (e.g., He-Ne laser, $\lambda_o = 0.6328 \mu m$) is a popular source of the illumination beam. The output of many lasers is linearly polarized with the ratio of the light polarized in one direction exceeding that polarized in the orthogonal direction by three orders of magnitude\[49\]. This is due mainly to the Brewster surface used within the laser in order to reduce energy loss. Some lasers are built differently to produce light beams that are composed of two or more linearly polarized waves of different wavelengths. In any case, the laser illumination in general can be treated as linearly polarized and characterized by a linear combination of TE and TM waves.

Another popular source is the natural light source (e.g., arc lamp) which consists of a very large number of randomly oriented atomic emitters. The emitters generate waves of the same frequency which form a single polarized wave that lasts for about $10^{-8}$ second. The variation of the polarization is thus much faster than the sampling rate (say, in milliseconds) of the alignment or metrology signal so that the incident wave can be treated as randomly polarized. Random polarization can be modeled as the results of superposition of many waves with arbitrary polarizations, and a random number generator is used to facilitate this computation.

The polarization state of the incident waves discussed above is restricted to the one immediately generated by the source. After the light is sampled and projected onto the wafer surface by optical devices (i.e., lens, mirrors, etc.), the polarization state may be altered slightly resulting in partially polarized light. The change of the degree of polarization depends strongly on the specific optical devices and configuration employed. Here, we assume that the polarization of the incident waves will not be altered by these optical devices since, to our knowledge, no manufacturers intentionally polarize/depolarize the alignment beams through these devices.
CHAPTER 3. MODELING OF LIGHT SCATTERING

3.7 Numerical Stability

Several numerical stability problems have been discovered during calculations. They can be divided into two categories - one due to “junction spike” and the other one due to highly dissipative layers. A solution is given to overcome the junction spike problem. However, for highly dissipative layers, several methods have been tried only with limited success.

3.7.1 Junction Spike

As shown previously, the solutions for the $E$ and $H$ fields in the $j$th layer have the following form:

$$E^j_y, H^j_y \sim \sum_{m=-L}^{+L} \left[ \left( A^j_m e^{i \alpha m z} + A^j_m e^{-i \alpha m z} \right) \sum_{l=-L}^{+L} B^j_{lm} e^{i 2 \pi l k x} \right].$$

As the summation limit $L$ increases, more evanescent waves and more Fourier terms for the dielectric constant $\varepsilon(x)$ are included in the calculation so that simulation should become more accurate and eventually converge to a true answer. However, this convergence is not observed in the TE case, but not in the TM case.

The reason for this instability problem uniquely for TM case is due to the extra term (i.e., the third term) in the TM Maxwell’s equation (Eq. 3.14):

$$\nabla^2 H^j - \mu_0 \varepsilon_0 \varepsilon^j(x) \frac{\partial^2 H^j}{\partial t^2} + \frac{\nabla \varepsilon^j(x)}{\varepsilon^j(x)} \times \left( \nabla \times H^j \right) = 0.$$ 

Consider a layer composed of two materials with the dielectric constants $\varepsilon_1$ and $\varepsilon_2$. In the previous work, the dielectric constant $\varepsilon(x)$ used to describe the junction between two materials is a step function, as shown in Fig. 3.4(a). Therefore, $\nabla \varepsilon(x)$ in the Maxwell’s equation becomes a delta function, which is divided by a finite function $\varepsilon(x)$ to become a “delta-like” function. This delta-like function, tentatively referred as junction spike, needs to be expressed by its Fourier series which will never reach a convergent value no matter how many terms are included. Therefore, this junction spike causes a nonconvergence problem for TM case.

Our solution is to replace this step function by a smoothly varying function to avoid the junction spike. The assumption made here is that accurate result should still be obtained as long as the smoothly varying distance is small compared with the wavelength. A good candidate for this smoothly varying function is the arc tangent function as shown in Fig. 3.4(b). Then, $\varepsilon(x)$ and $\nabla \varepsilon(x)$ can be expressed by
Figure 3.4: (a) Traditional way to characterize a junction. (b) Our approximation for a junction.

\[
\varepsilon(x) = \frac{\varepsilon_2 - \varepsilon_1}{\pi} \tan^{-1} \gamma x + \frac{\varepsilon_2 + \varepsilon_1}{2},
\]

\[
\nabla\varepsilon(x) = \hat{x} \frac{\varepsilon_2 - \varepsilon_1}{\pi} \frac{\gamma}{1 + \gamma^2 x^2},
\]

where \( \gamma \) controls the "abruptness" of these two functions. The function \( \nabla\varepsilon(x)/\varepsilon(x) \) is now a finite function which can be approximated by its Fourier series with a finite number of terms. Numerically, a standard fast Fourier transform package is used to calculate Fourier coefficients. The criteria used to guarantee the abrupt change of this function is that the transition region between 10% and 90% of the step high is a tenth of the wavelength \( \lambda_n \) in the relevant materials. A unique value of \( \gamma \) can be determined using this criteria:

\[
\gamma = \frac{20}{\lambda_n} \tan (0.4\pi).
\]

Fig. 3.5 shows the simulated images of a 1\( \mu \)m-thick, 4\( \mu \)m-wide oxide window on silicon substrate under TM coherent illumination. Two arc tangent functions with the one tenth wavelength criteria and a less stringent quarter wavelength criteria are adopted here. The collecting NA is 0.6. Different values of summation limit \( L \) from 20 to 50 are adopted to show the convergence. The value of \( L \) commonly used in our simulation is between 40 to 50. Note that although little difference is observed for bright field images by using those two junction approximations, significant difference can be obtained when simulating dark field type of images.
Figure 3.5: Simulation results of full field images of an oxide window using different $L$'s are shown to demonstrate the convergence.

3.7.2 Highly Dissipative Layers

The other numerical problem occurring both in the TE and TM cases is the under- and/or over-flow errors due to highly dissipative layers. During the matrix formulation, terms like $e^{\pm \alpha_m z_j}$'s in Eqs.3.9,3.11,3.19, and 3.20 need to be evaluated. If the relevant material is highly dissipative, i.e., its dielectric constant has a large imaginary part, some of the modes $\alpha_m$'s will have large real parts and $e^{\pm \alpha_m z_j}$'s become extremely large or small resulting in numerical problems. Also, for layers with materials of moderate values of dielectric constants but with large depth (i.e., large $z_j$), the same problem exists.

This numerical difficulty cannot be removed simply by assigning zeroes to small
numbers near under-flow. Neither can it be removed if highly dissipative topography are replaced by dielectric topography coated with a thin layer of that highly dissipative material, because numerically a few highly dissipative modes with large real part still exist even though many other such modes are removed.

A solution, called “relative z-coordinate”, is given here that partially solves the problem. Consider a dissipative layer located in the lower portion of the topography so that its \( z_j \) value is large. The combination of the layer’s dielectric constant and \( z_j \) is such that the drop of the field amplitude inside this layer is within numerical tolerance, while numerical problem may still exist because the large value of \( z_j \) creates under/over-flow errors when evaluating \( e^{\pm \alpha_m(z_j-z_{j-1})} \). However, if the equation formulation is modified so that in each layer \( j \), the \( z \)-coordinate is referred to the upper surface \( z_{j-1} \) of that layer, then instead \( e^{\pm \alpha_m(z_j-z_{j-1})} \)’s are evaluated and this problem can be overcome. The \( E \) and \( H \) fields will now have the form:

\[
E^j_y, H^j_y \sim \sum_{m=-L}^{+L} \left[ \left( A^j_m e^{\alpha_m(z_j-z_{j-1})} + A^j_m e^{-\alpha_m(z_j-z_{j-1})} \right) \sum_{l=-L}^{+L} B^j_{l,m} e^{i2\pi l bx} \right].
\]

This relative \( z \)-coordinate formulation yields minor changes of the matrix formulation by replacing Eq. 3.9 with

\[
C^{i1}_{11}(l, m) = e^{\alpha_m(z_j-z_{j-1})} B^i_{l,m} \quad C^{i2}_{12}(l, m) = e^{-\alpha_m(z_j-z_{j-1})} B^i_{l,m}
\]
\[
C^{i1}_{21}(l, m) = \alpha_m e^{\alpha_m(z_j-z_{j-1})} B^i_{l,m} \quad C^{i2}_{22}(l, m) = -\alpha_m e^{-\alpha_m(z_j-z_{j-1})} B^i_{l,m}
\]
\[
E^{i+1}_{11}(l, m) = B^{i+1}_{l,m} \quad E^{i+1}_{12}(l, m) = -\alpha_m B^{i+1}_{l,m} + \frac{\alpha_m}{i\kappa_0} B^{i+1}_{l,m}
\]
\[
E^{i+1}_{21}(l, m) = \alpha_m B^{i+1}_{l,m} \quad E^{i+1}_{22}(l, m) = -\alpha_m^2 B^{i+1}_{l,m},
\]

by replacing Eq. 3.11 with

\[
C^{n1}_{11}(l, m) = \left( \sqrt{\varepsilon_2 - (l\beta_0)^2} - \frac{\alpha_m}{i\kappa_0} \right) e^{\alpha_m(T-z_{n-1})} B^n_{l,m},
\]
\[
C^{n2}_{12}(l, m) = \left( \sqrt{\varepsilon_2 - (l\beta_0)^2} + \frac{\alpha_m}{i\kappa_0} \right) e^{-\alpha_m(T-z_{n-1})} B^n_{l,m},
\]

by replacing Eq. 3.19 with

\[
C^{i1}_{11}(l, m) = e^{\alpha_m(z_j-z_{j-1})} B^i_{l,m} \quad C^{i2}_{12}(l, m) = e^{-\alpha_m(z_j-z_{j-1})} B^i_{l,m}
\]
\[
C^{i1}_{21}(l, m) = \alpha_m e^{\alpha_m(z_j-z_{j-1})} B^i_{l,m} \quad C^{i2}_{22}(l, m) = -\alpha_m e^{-\alpha_m(z_j-z_{j-1})} B^i_{l,m}
\]
\[
E^{i+1}_{11}(l, m) = B^{i+1}_{l,m} \quad E^{i+1}_{12}(l, m) = B^{i+1}_{l,m}
\]
\[
E^{i+1}_{21}(l, m) = \alpha_m B^{i+1}_{l,m} \quad E^{i+1}_{22}(l, m) = -\alpha_m B^{i+1}_{l,m},
\]

and by replacing Eq. 3.20 with
\[ C_{11}^n(l, m) = \left( \sqrt{\varepsilon_s - (k_0 l \lambda_0)^2} B_{i,m}^n - \frac{\alpha_m^n}{i k_0} \varepsilon_s B_{i,m}^n \right) e^{\alpha_m^n(T - z_{n-1})}, \]
\[ C_{12}^n(l, m) = \left( \sqrt{\varepsilon_s - (k_0 l \lambda_0)^2} B_{i,m}^n + \frac{\alpha_m^n}{i k_0} \varepsilon_s B_{i,m}^n \right) e^{-\alpha_m^n(T - z_{n-1})}. \]

As a rule of thumb, with the relative z-coordinate correction, the wave-guide model can handle metallic layer (e.g., aluminum) of approximately 1 \( \mu \text{m} \) thick with visible light illumination, no matter where this layer is located in.

### 3.8 Comparison Between Scattering Models

Three other scattering models - Rayleigh hypothesis model, differential model, and integral model are introduced in this section. A comparison between these models as well as Fresnel and wave-guide models is given at the end.

#### 3.8.1 Rayleigh Hypothesis Model

A simple topography of silicon groove which extends indefinitely in \( \pm y \) directions as shown in Fig. 3.6 is used to illustrate the Rayleigh hypothesis model. The model also adopts the grating approach, but only one period is shown in the figure. The topography is divided into two regions - air and silicon regions. A TE plane wave \( E_{\text{in}}^m \) is incident vertically onto the groove, and the wave equation can be written as

\[
\frac{\partial^2 E(x, z)}{\partial x^2} + \frac{\partial^2 E(x, z)}{\partial z^2} + k_0^2 \varepsilon_i E(x, z) = 0, \tag{3.21}
\]

where \( \varepsilon_i \) can be \( \varepsilon_{\text{air}} \) or \( \varepsilon_{\text{Si}} \) depending on the region of interest. The solution we look for is of the form

\[ E(x, z) \sim e^{i k_x x} e^{i k_z z}. \]

This solution has to satisfy two conditions: it is periodic in \( \pm x \)-direction such that \( E(x \pm P, z) = E(x, z) \), and it satisfies the wave equation Eq. 3.21. These two conditions uniquely determine \( k_x, k_z \) so that the solution now becomes

\[ E(x, z) \sim e^{\pm i k_0 \sqrt{\varepsilon_i - (l b x)^2}} e^{i 2 \pi l x}, \]

where \( l \) is an arbitrary integer. The reflected field \( E_r(x) \) in the air region and the transmitted field \( E_t(x) \) in the silicon region can thus be written as
\[ E^r(x, z) = \sum_i \left[ E_i^r e^{+ik_0 \sqrt{\varepsilon_{\text{air}} - (lx)^2} z} + E_i^r e^{-ik_0 \sqrt{\varepsilon_{\text{air}} - (lx)^2} z} \right] e^{i2\pi lx}, \]

\[ E^t(x, z) = \sum_i \left[ E_i^t e^{+ik_0 \sqrt{\varepsilon_{\text{si}} - (lx)^2} z} + E_i^t e^{-ik_0 \sqrt{\varepsilon_{\text{si}} - (lx)^2} z} \right] e^{i2\pi lx}. \]

The Rayleigh hypothesis states that the above expressions, called Rayleigh expansions, are valid not only in regions above \( z = z_1 \) and below \( z = z_2 \), but also in between \( z_1 \) and \( z_2 \). Then, the coefficients \( E_i \)'s can be obtained by matching the boundary conditions of the \( E \) and \( H \) fields at the interface of those two regions.

The Rayleigh hypothesis was first introduced by Lord Rayleigh in 1907. Its validity was not doubted by researchers until 1950's\[35\]. Since then, many efforts have been devoted into studying its rigorousness and range of validity. It is now believed that the Rayleigh hypothesis gives rigorous expressions in regions above and below the groove (i.e., \( z > z_1 \) and \( z < z_2 \)). Inside the groove, the Rayleigh hypothesis may not give a convergent answer\[40\] and therefore cannot represent the correct \( E \) field. However, when the groove depth is small, the Rayleigh hypothesis model was verified\[52\] to give accurate results.

![Figure 3.6](image-url)  

Figure 3.6: A grating of silicon grooves with a period shown is used to illustrate the Rayleigh hypothesis model, the differential model, and the integral model. For simplicity, the illumination is a TE-polarized, vertically incident plane wave.

The Rayleigh hypothesis model has been applied to modeling the images of semiconductor structures [20] and dark/bright field images of alignment marks\[4\]. In semiconductor problems, the range of the groove depth for the hypothesis to hold was not reported.
3.8.2 Differential Model

The differential model[41, 51] is a numerical model that decomposes the differential Maxwell’s equation (Eq. 3.21) into a linear system of ordinary differential equations, and further adopts a multistep method to solve the linear system. The same topography and illumination condition in Fig. 3.6 are used to illustrate the differential model.

The electric fields $E^0$ above $z = z_1$, $E^1$ between $z = z_1$ and $z_2$, and $E^s$ below $z = z_2$ can be expressed rigorously as

$$E^0 (x, z) = e^{i k_0 z} + \sum_l E_l^0 e^{i 2 \pi l b x} e^{-i k_0 \sqrt{1 - (l b \lambda_o)^2} z},$$

$$E^1 (x, z) = \sum_l u_l (z) e^{i 2 \pi l b x}, \quad (3.22)$$

$$E^s (x, z) = \sum_l E_l^s e^{i 2 \pi l b x} e^{-i k_0 \sqrt{1 - (l b \lambda_o)^2} z}.$$

The dielectric constant $\varepsilon (x, z)$ in between $z_1, z_2$ can be expressed as

$$\varepsilon (x, z) = \sum_l \varepsilon_l (z) e^{i 2 \pi l b x}. \quad (3.23)$$

The objective is to obtain the amplitudes $E_l^r$'s of the reflected field.

The heart of the differential model lies in the calculation of $E^1 (x, z)$ field in between $z_1$ and $z_2$. To achieve it, we substitute Eqs. 3.22 and 3.23 into the wave equation Eq. 3.21 to obtain a linear system of ordinary differential equations

$$\frac{d^2 u (z)}{dz^2} + A u (z) = 0,$$

where $u (z)$ is vector of functions

$$u^t (z) = [\cdots, u_{-i} (z), \cdots, u_{-1} (z), u_0 (z), u_1 (z), \cdots, u_i (z), \cdots].$$

Matrix $A$ is completely determined by the dielectric constants $\varepsilon_i$'s of the topography. Also, by matching $E^0$ and $E^1$ at $z = z_1$, and by matching $E^1$ and $E^s$ at $z = z_2$, two constraints can be derived:

$$\frac{du (z_1)}{dz} + B u (z_1) = g, \quad (3.24)$$

$$\frac{du (z_2)}{dz} + C u (z_2) = 0. \quad (3.25)$$
Matrices $B$ and $C$ are determined by the period of the topography, and vector $g$ is determined by the illumination angle which is zero here. The game plan is to solve for $u(z)$ of Eq. 3.22 with these two constraints. Then, $E^r_i$'s can be obtained by matching the fields at $z = z_1$.

Defining the vectors $e^{in}, e^r, e^t$ to be the incident, reflected and transmitted vectors:

$$e^{in}(z) = [\cdots, 0, 0, 1, 0, 0, \cdots],$$

$$e^r(z) = [\cdots, E^r_{-1}(z), \cdots, E^r_0(z), E^r_1(z), \cdots, E^r_i(z), \cdots],$$

$$e^t(z) = [\cdots, E^t_{-1}(z), \cdots, E^t_0(z), E^t_1(z), \cdots, E^t_i(z), \cdots],$$

we can write three linear relations

$$e^{in} = M_{in}e^t,$$  \hspace{1cm} (3.26)

$$e^r = M_re^t,$$  \hspace{1cm} (3.27)

$$e^r = M_rM^{-1}_{in}e^{in}. \hspace{1cm} (3.28)$$

The assumption made here is that all materials of interest are isotropic and their dielectric constants are field independent. Now, by assigning an arbitrary vector to $e^t(z_2)$ (or equivalently, $u(z_2)$), we obtain $du(z_2)/dz$ from Eq. 3.25. By using a suitable multistep method (e.g., Adams-Moulton method) in Eq. 3.22, $u(z_1)$, or equivalently $e^{in}$ and $e^r$, can be calculated. Recall that this particular set of $E^r_i$'s is the solution for the arbitrarily defined vector $e^t(z_2)$. If the $p$th element of the arbitrary vector $e^t(z_2)$ is $\delta_{p0}$, then the calculated $e^{in}$ and $e^r$ are the $p$th columns of the matrices $M_{in}$ and $M_r$, as can be seen from Eqs. 3.26 and 3.27. By repeating such calculation $(N - 1)$ times where $N$ is the order of the matrices, $M_{in}$ and $M_r$ can be completely specified and the reflected field $E^r$ can be derived from Eq. 3.28.

Chandezon et al.\[12\] developed another formalism to cope with gratings with highly dissipative materials as the substrate. Nonplanar interfaces are mapped into new coordinate system to become parallel planes. The Maxwell’s equations in this new coordinate system are solved using Floquet theorem, and the derivation is similar to that of the wave-guide model.

The differential model was first proposed in the late 60’s. It has been developed and modified by several researchers since then, and has been considered a rigorous model. Lately, Depine and Simon\[15\] objected to its validity when applying to the
CHAPTER 3. MODELING OF LIGHT SCATTERING

TM case due to the incapability of common multistep methods to approximate discontinuous functions. This objection was refuted by Neviere and Vincent[44] with supportive simulation results using Cray II computer. The conclusion seems to be that the model is rigorous theoretically, but some numerical problems may need to be overcome when implementing it.

Yeung[69] utilized the differential model to perform the calculation of two dimensional photoresist bleaching, as opposed to one dimensional bleaching in SAMPLE[48].

3.8.3 Integral Model

The integral model adopts the Green’s function approach to solve the wave equation (Eq. 3.21). The wave equation in this way is reduced to an integral equation which has one unknown function to be calculated numerically. Conceptually, it is more difficult to apprehend than any other model that has been introduced. In the following derivation, intermediate functions (i.e., $\psi, \phi$) which are not necessarily equivalent to the true field will be used. To avoid confusion, the true field is always meant whenever the symbol “$E$” is utilized.

An important theorem of the integral model is given here without proof[51]. Consider two semi-infinite media, both having the same dielectric constant $\epsilon_j$, are separated by the profile $z = f(x)$. If a function $\psi(x, z)$ is continuous everywhere except on $z = f(x)$, and satisfies the following two conditions

$$\frac{\partial^2 \psi(x, z)}{\partial x^2} + \frac{\partial^2 \psi(x, z)}{\partial z^2} + k_0^2 \epsilon_j \psi(x, z) = 0,$$

$\psi(x, z)$ be bounded when $y \rightarrow \pm \infty$,

then the solution $\psi$ can be expressed by

$$\psi(x, z) = \int_{-P/2}^{P/2} G(x - x', z - f(x')) \eta(x') \, dl + \int_{-P/2}^{P/2} \frac{dG}{dn}(x - x', z - f(x')) \tau(x') \, dl,$$

(3.29)

where the line integral $dl$ is along the profile $z = f(x)$. The Green’s function $G(x, z)$ is

$$G(x, z) = \frac{1}{i2P} \sum_i \frac{1}{k_0 \sqrt{\epsilon_j - (ib\lambda_0)^2}} \frac{1}{|z|} e^{ik_0 \sqrt{\epsilon_j - (ib\lambda_0)^2}|z|} e^{i2\pi l bx}$$
and the jump of $\psi$ and the jump of its normal derivative $d\psi/dn$ can be denoted by

$$\tau(x) = \psi(x, f(x) + \delta) - \psi(x, f(x) - \delta),$$

$$\eta(x) = \frac{d\psi}{dn}(x, f(x) + \delta) - \frac{d\psi}{dn}(x, f(x) - \delta).$$

Furthermore, the limits of $\psi$ and $d\psi/dn$ approaching the interface can be found to be

$$\psi(x, f(x) \pm \delta) = \pm \frac{\tau}{2} + \int_{-P/2}^{P/2} G(x, x') \eta(x') dx' +$$

$$\int_{-P/2}^{P/2} \frac{dG}{dn}(x, x') \tau(x') dx'$$

(3.30)

$$\frac{d\psi}{dn}(x, f(x) \pm \delta) = \pm \frac{\eta}{2} + \int_{-P/2}^{P/2} \frac{dG}{dn}(x, x') \tau(x') dx' \quad \text{if } \tau = 0.$$  

(3.31)

Each set of $\tau(x)$ $\eta(x)$ uniquely determines the solution $\psi(x, z)$, and the correct set can be found by matching the boundary conditions at $z = f(x)$. Very often, only one of these two unknown functions exists in the final integral equation to be solved.

The same topography and illumination conditions in Fig. 3.6 are used to illustrate the integral model. The air and silicon regions are now denoted by region 1 and 2. The profile of the interface between these two regions is represented by $z = f(x)$. In order to facilitate the manipulation, two intermediate function $\psi$ and $\phi$ are introduced. The equation formulation can be divided into four steps:

- **step 1:** If a function $\psi(x, z)$ satisfies the following conditions

  $$\psi(x, z) = E^r(x, z) \quad \text{in region 1},$$

  $$\frac{\partial^2 \psi(x, z)}{\partial x^2} + \frac{\partial^2 \psi(x, z)}{\partial z^2} + k_o^2 \epsilon_1 \psi(x, z) = 0, \quad \text{in region 2},$$

  $$\psi(x, z) \text{ be continuous across } z = f(x),$$

  $$\psi(x, z) \text{ be bounded when } y \to \pm \infty,$$

then $\psi$ and its normal derivative right above the interface $z = f(x)$ can be obtained from Eq. 3.30,3.31

$$\psi(x, f(x) + \delta) = \int_{-P/2}^{P/2} G(x, x') \eta(x') dx',$$
\[
\frac{d\psi}{dn}(x, f(x) + \delta) = \frac{\eta}{2} + \int_{-P/2}^{P/2} \frac{dG}{dn}(x, x') \tau(x') \, dx'.
\]

**Step 2:** Since the functions \(\psi(x, f(x) + \delta)\) and \(\frac{d\psi}{dx}(x, f(x) + \delta)\) derived above represent the true reflected field, the true field and its normal derivative right below the interface can be found by using the boundary conditions that the true field and its normal derivative are continuous across the interface:

\[
E_t^r(x, f(x) - \delta) = e^{ik_0 f(x)} + \int_{-P/2}^{P/2} G(x, x') \eta(x') \, dx',
\]  \hspace{1cm} (3.32)

\[
\frac{dE_t^r}{dn}(x, f(x) - \delta) = \frac{d}{dn} e^{ik_0 f(x)} + \frac{\eta}{2} + \int_{-P/2}^{P/2} \frac{dG}{dn}(x, x') \tau(x') \, dx'.
\]

**Step 3:** A function \(\phi(x, z)\) is introduced that satisfies the following conditions:

\[
\phi(x, z) = 0 \text{ in region 1},
\]

\[
\phi(x, z) = E_t^r(x, z) \text{ in region 2},
\]

\[
\phi(x, z) \text{ be bounded when } y \to \pm\infty,
\]

then the true field \(E_t^r(x, z)\) in region 2 can be obtained from Eq. 3.29 by using the results of step 2 as \(\tau(x)\) and \(\eta(x)\):

\[
\tau(x) = -E_t^r(x, f(x) - \delta) = -e^{ik_0 f(x)} - \int_{-P/2}^{P/2} G(x, x') \eta(x') \, dx',
\]

\[
\eta(x) = -\frac{dE_t^r}{dn}(x, f(x) - \delta) = -\frac{d}{dn} e^{ik_0 f(x)} - \frac{\eta}{2} - \int_{-P/2}^{P/2} \frac{dG}{dn}(x, x') \tau(x') \, dx'.
\]

Thus, an expression for the true field \(E_t^r(x, z)\) in region 2 with an unknown function \(\eta(x)\) from step 1 is derived.

**Step 4:** Substituting \(z\) in \(E_t^r(x, z)\) obtained above with \(f(x)\), and equating it with Eq. 3.32, we obtain an integral equation with one unknown function \(\eta(x)\) (from step 1) in it. By solving this integral equation numerically, \(\eta(x)\) can be calculated and the reflected field \(E_t^r\) can be obtained.
CHAPTER 3. MODELING OF LIGHT SCATTERING

The integral model was first proposed in the 60's, and the development effort has been carried on since [51, 41]. It is believed to be a rigorous model, and its numerical stability is believed to be superior than the differential model which has been developed almost in parallel. However, its difficulties of apprehension and implementation have made the model less attractive than the differential model. To our knowledge, nobody has employed this model to perform simulation in the optical lithography or metrology field.

3.8.4 Comparisons Between Scattering Models

Before comparing the five models that have been introduced, it is instructive to observe the different expressions used for the \( E \) field in the wave-guide, differential and Rayleigh hypothesis models. These expressions will be repeated in this section. In the wave-guide model, the \( E \) field inside an inhomogeneous slab is expressed by

\[
E_{WG}(x, z) = \sum_{l} \left[ \sum_{m=-L}^{+L} \left( A_m e^{\gamma_m z} + A^*_m e^{-\gamma_m z} \right) B^2_{l,m} \right] e^{i2\pi lx}.
\]

The field variation in \( \pm z \) directions is determined by waves in the wave-guide modes that are derived from the Maxwell's equations. In the differential model, the \( E \) field inside an inhomogeneous slab is

\[
E_{DM}(x, z) = \sum_{l} u_l(z) e^{i2\pi lx}.
\]

The field variation in \( \pm z \) directions is calculated by numerically solving for the unknown functions \( u_l(z) \)'s. In the Rayleigh hypothesis model, the field in a homogeneous region of any shape is approximated by

\[
E_{RH}(x, z) = \sum_{l} \left[ E^{+}_l e^{+ik_0 \sqrt{\epsilon-(l_b x)^2} z} + E^{-}_l e^{-ik_0 \sqrt{\epsilon-(l_b x)^2} z} \right] e^{i2\pi lx}.
\]

As far as the numerical accuracy is concerned, the Fresnel model gives satisfactory results when dealing with shallow topography under small NA illumination. Otherwise, it shows the correct trend in many cases. The Rayleigh hypothesis model gives accurate results for shallow topography, and may be unreliable for the deep one. The wave-guide, differential and integral models seem to generate accurate results.

The Fresnel model is very stable because no evanescent wave is included in the calculation. The Rayleigh hypothesis may give unstable results because of possible divergence of the Rayleigh expansion. The wave-guide and differential models seem
to be stable in many cases. The integral model is more stable than the differential model[41] and can be applied to diffraction study for all wavelengths of practical interest. The floating point problem due to highly dissipative layers is expected to exist in all these models.

The Fresnel model is a very efficient model because of its simplicity. The waveguide model is efficient for topography with vertical sidewalls, and its CPU time is linearly proportional to the number of stratified layers. The CPU times of the Rayleigh and integral models depend linearly on the number of deposited layers of the topography, and the CPU time of the differential model depends linearly on the thickness of the topography.

The above comparisons are simplified and summarized in Table. 3.1.

<table>
<thead>
<tr>
<th></th>
<th>Fresnel</th>
<th>Wave-guide</th>
<th>Rayleigh</th>
<th>Differ.</th>
<th>Integral</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>non-rigorous</td>
<td>rigorous</td>
<td>semi-rigorous</td>
<td>rigorous</td>
<td>rigorous</td>
</tr>
<tr>
<td>Accuracy</td>
<td>fair</td>
<td>good</td>
<td>fair-good</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>Stability</td>
<td>excellent</td>
<td>good</td>
<td>fair-good</td>
<td>fair-good</td>
<td>excellent</td>
</tr>
<tr>
<td>CPU</td>
<td>fast</td>
<td>slow</td>
<td>slow</td>
<td>slow</td>
<td>slow</td>
</tr>
<tr>
<td>Used in Lithography?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Lib [45]</td>
<td>[45, 46, 70]</td>
<td>[20, 4]</td>
<td>[69]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Summary of comparisons between scattering models.
Chapter 4

Modeling of Imaging

In the previous chapter, it is shown that the scattering matrix of a semiconductor structure can be calculated by using the wave-guide model. In this chapter, we develop imaging models so that in conjunction with the calculated scattering matrix optical images formed on the surface of the optical detector can be simulated.

Imaging is referred to the projection and sampling of optical waves through optical devices (e.g., lenses, field stops). Various optical imaging configurations, depending on their specific applications, are used by manufacturers to perform alignment and metrology. The variables of these imaging configurations may simply be the NA's of lenses, or complicated schemes (e.g., DIC scanning) of different types of microscopes. The aim of this chapter is to develop a general model to calculate the amplitudes of the projected light onto the sample surface, and of the collected light reflected from the sample surface. Then, various optical configurations can be handled by utilizing this model in conjunction with the previously calculated scattering matrix.

The basic optical configuration of our model, which was adopted by O'Toole [50] and Kirk[32], is shown in Fig. 4.1. It is general enough to include most of the commercial alignment and metrology schemes in that almost all these schemes project incident light onto the sample surface through the objective lens, and the reflected or transmitted light which contains the information of the sample structure is gathered by the collecting lens and analyzed by a detector. The objective and collecting lenses are usually the same lens when operating in the reflection mode. The theory adopted in our model is the Abbe's theory[24] in Fourier optics, and a brief introduction to the theory is given here. Modeling of light projection and imaging (including laser incidence) is followed to show the feasibility of dealing with various alignment and metrology schemes. Imaging of scanning systems can be modeled by taking into
account the phase variation in the scanning direction. Defocus effect, traditionally treated by using the paraxial approximation, is modeled here by considering the phase variation in the vertical direction. It is shown that this simple but rigorous defocus model utilizing the phase variation approach is equivalent to the one utilizing the wave aberration approach[25]. Since our imaging model is two dimensional while a laser spot is three dimensional, discrepancies exist when thick features are imaged by small scanning spots. Modeling of various optical configurations used in commercial alignment and metrology tools will be shown in Chapter 7.

Figure 4.1: Basic optical configuration of the imaging model.

4.1 Introduction to Abbe’s Theory

The Abbe’s theory, proposed in 1873, is a simple but powerful imaging theory which enables calculations of the images formulated by lenses. The 2D Abbe’s theory can be demonstrated by referring to Fig. 4.2. A grating located on the object plane \( \Sigma_o \) is illuminated by a vertically incident plane wave of unit amplitude. This plane wave is diffracted by the grating resulting in many diffracted plane waves traveling in different directions. Some of them are intercepted by the transform lens \( L_t \) and the Fraunhofer pattern of the grating is formed on the transform plane \( \Sigma_t \) (also known as Fourier plane, conjugate plane, or back focal plane). The Fraunhofer pattern on the transform plane serves as point emitters of Huygens’ wavefront, and the resultant waves reach the imaging plane \( \Sigma_i \) and form an interference pattern. Mathematically, if the grating on the object plane is represented by \( f_o(x) \) where
\[ f_o(x) = \begin{cases} 1, & \text{if transparent} \\ 0, & \text{if opaque} \end{cases} \]

and its Fourier series is

\[ f_o(x) = \sum_{n=-\infty}^{+\infty} a_n e^{i2\pi nx}, \]

then the Abbe's theory states that each Fourier component represents a diffracted plane wave traveling in a specific direction and the amplitude of the interference pattern \( f_i(x) \) on the imaging plane becomes

\[ f_i(x) = \sum_{n=-L}^{+L} a_n e^{i2\pi nx}. \]

The number of the intercepted waves \( L \) is determined by the NA of the transform lens, and is equal to one in the simple example in Fig. 4.2.

![Figure 4.2: Abbe’s theory of image formation](image)

The image of the grating is always more blurry than the grating itself because some of the high spatial frequency components of its Fourier series are filtered out by the transform lens. Therefore, a transform lens can sometimes be thought of as a high frequency filter of the optical image. Another way to remove the high frequency components is to place a field stop on the transform plane to block out high frequency spots. On the contrary, if the field stop is placed in the center of the plane so that low frequency components are removed, the transform lens serves as a low frequency filter. High and low orders may contain different information of the wafer object, and
the choice of the sizes of lens and the aperture stop is important in optical design. For simplicity, the term NA used in the dissertation is referred to the "combined NA" due to the lens and aperture.

The Abbe’s theory can be derived from the Fraunhofer formula [6], which is a result of the scalar theory. Therefore, the Abbe’s theory is valid provided that the incident wavelength is much smaller than the grating feature size, which is the reticle feature size in alignment problem. The wavelength commonly used in alignment is approximately equal to or smaller than 0.65μm and the reticle feature is usually larger than 10μm in a 5x or 10x system so that the Abbe’s theory is manipulated in its valid range. However, if the reticle feature size is comparable to the wavelength as is the case in the 1x submicron photolithography exposure, the Abbe’s theory may not be valid and a more rigorous vector model (e.g., the wave-guide model) may need to be adopted. The range of validity of the Abbe’s theory will be discussed in Chapter 6.

4.2 Light Projection and Sampling

The projected or collected light shown in Fig. 4.1 is determined by the specific optical configuration of the measurement tool. In our modeling effort, various optical configurations are categorized according to the application and the illumination source used as illustrated in Fig. 4.3. They cover most of the alignment and metrology schemes as will be shown in Chapter 7. In the figures, the reflected or transmitted light from the sample is omitted for simplicity. Monochromatic light incidence is temporarily assumed.

The first type of metrology is the traditional Kohler illumination[7] that uses arc lamp as the illumination source. The objective lens is illuminated in such a way that it can be treated as an incoherent source with uniformly distributed intensity. This greatly simplifies the modeling work, for each point on the incoherent source can be modeled as an independent source and the final image is just the scalar summation of image intensities due to different points. The NA of the objective lens determines the degree of coherence of the illumination and coherent illumination can be achieved by making the NA small.

The second type of metrology is laser slit/spot scanning. Metrology that adopts scanning in general achieves better resolution than traditional Kohler full field style [63], with the sacrifice of complicating the optical design. A laser slit/spot with Gaussian intensity distribution enters the entrance pupil of the objective lens. This
<table>
<thead>
<tr>
<th></th>
<th><strong>Metrology</strong></th>
<th><strong>Alignment</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arc lamp</strong></td>
<td><img src="image" alt="incoherent source" /></td>
<td><img src="image" alt="incoherent source" /></td>
</tr>
<tr>
<td></td>
<td><strong>objective lens</strong></td>
<td><img src="image" alt="reticle objective lens" /></td>
</tr>
<tr>
<td><strong>Laser</strong></td>
<td><img src="image" alt="objective lens" /></td>
<td><img src="image" alt="objective lens L₁" /></td>
</tr>
<tr>
<td></td>
<td><img src="image" alt="reticle objective lens L₂" /></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.3: Different optical configurations generate incident light differently.

slit/spot is then imaged onto the sample surface and scans over it. The Gaussian beam is Fourier-transformed into a linear combination of diffracted waves to cope with the calculation of the Abbe's theory.

In order to measure the relative position between reticle and wafer for alignment, it is common to project the reticle image onto the wafer. The model used for the first type of alignment, which is proposed and implemented by Kirk[32], adopts this concept. Here, arc lamp is utilized as the light source. Kohler illumination is assumed for the objective lens of the reticle so that the reticle is modeled as being illuminated by an incoherent source.

The second type of alignment uses laser source, which is commonly employed in modern steppers. The laser slit/spot can either be very large, in which case it can be approximated by coherent illumination in the first type of alignment, or be small to scan over the reticle and wafer to find their edges. In the latter case, little error is usually introduced when searching for the reticle edges. The remaining task is to locate the edges of the wafer mark and it becomes equivalent to the second type of
CHAPTER 4. MODELING OF IMAGING

metrology configuration. There are other alignment schemes which project the wafer image onto the reticle (e.g., ASM) with laser source, and will be treated in Chapter 7.

Broad band sources are sometimes used to reduce interference effect. They are commonly treated as a combination of incoherent sources under different wavelengths. The image due to broad band sources can be calculated by taking the summation of the image intensities from these incoherent sources.

4.3 Reduction System

Very often in alignment problem, the alignment pattern on the reticle is projected onto the wafer through an $Mx$ ($M = 5$ or 10, usually) reduction lens used for lithography exposure. Therefore, the reticle feature is $Mx$ larger than the expected reticle image size on the wafer. It is shown in this section that, provided the Abbe’s theory holds, the image of the $Mx$ reticle pattern on the wafer by using an $Mx$ lens is the same as the image of the $1x$ reticle pattern by using a $1x$ lens. Then, $1x$ system will be used from now on to model all imaging systems.

![Image of Mx and 1x systems](image.png)

Figure 4.4: An $Mx$ reduction system and a $1x$ system

An $Mx$ system and a $1x$ system are shown in Fig. 4.4, with the pattern on the object plane $\Sigma_o$ in the $Mx$ system $M$ times larger than that in the $1x$ system. In the $Mx$ system, if a plane wave on the object side travels at an angle $\theta_M$ and is collected by the lens and travels on the image side at an angle $\theta_1$, then the sine condition[8] gives
\[ M = \frac{\sin \theta_1}{\sin \theta_M}, \]  

(4.1)

If the Fourier series \( f(x) \) of the object feature is expressed by

\[ f(x) = \sum_{n=-\infty}^{+\infty} a_n \exp \left( i2\pi n \frac{x}{MP} \right), \]

where \( MP \) is the period, then the amplitude of the image on the reticle side that is intercepted by the \( Mx \) lens is

\[ f_\circ(x) = \sum_{n=-L}^{+L} a_n \exp \left( i2\pi n \frac{x}{MP} \right). \]

The summation limit \( L \) is equal to \( \frac{MP}{\lambda_0} \sin \theta_M \), which corresponds to the highest order of the diffracted waves intercepted by the lens. Then, the amplitude \( f_i(x) \) of the image on the image plane \( \Sigma_i \) becomes

\[ f_i(x) = \sum_{n'=-L}^{+L} a_{n'} \exp \left( i2\pi n' \frac{x}{P} \right), \]  

(4.2)

where the sine value of the travelling angle of each order is reduced by a factor of \( M \) from Eq. 4.1. However if the 1x reticle pattern is used on the 1x system, the image obtained on the image plane will be the same as in Eq. 4.2. Thus, any reduction system can be modeled by an equivalent 1x system.

The above result is valid provided that the Abbe's theory holds for both \( Mx \) and 1x systems. However, small reticle features close to the wavelength are sometimes projected by 1x systems, and more rigorous model may need to be used to model the projected images.

### 4.4 Scanning

In optical metrology or alignment, the wafer structure is often scanned by a focused beam and part of the reflected light is collected and analyzed. The scanning can be performed either by moving the wafer mark across the stationary focused beam or by moving a flying beam across the stationary wafer. In the former case, the optics is easier to design while undesirable mechanical movement is involved. In the latter case, the beam uses different part of the lens so that aberration errors may occur. The scanning model we adopt implies the equivalence of these two cases by assuming that no errors are introduced from lens aberrations and mechanical movements.
The model uses simple phase advance and delay techniques, as shown in Fig. 4.5(a). To our knowledge, the first implementation of this model was done by Kirk[31]. In Chapter 3, it is shown that once the coordinates and refractive indices of the topography (i.e., wafer of dotted line) is specified, the scattering matrix elements $S_{l,m}$'s can be calculated using the wave-guide model. During scanning, however, the wafer will be moved to different $x$ locations (i.e., wafer in solid line) with time, so that the $S_{l,m}$'s calculated under the original wafer location no longer apply to this new wafer location. Consider the wafer is displaced in the $x$-direction by $\Delta x$. Then, an incident plane wave of order $l$ and incident angle $\theta$ reaches the new location later than it reaches the original location and the optical path difference $d_i$ is

$$d_i = (\Delta x) \cdot \sin \theta = (\Delta x) \cdot lb\lambda_0.$$ 

(4.3)

With the same argument, the reflected plane wave of order $m$ and at the reflected angle $\phi$ leaves the new wafer location earlier than it leaves the original location. The optical path difference $d_r$ is

$$d_r = - (\Delta x) \cdot \sin \phi = - (\Delta x) \cdot mb\lambda_0.$$ 

Therefore, the scattering matrix element $S'_{l,m}$ for the new wafer location is

$$S'_{l,m} = S_{l,m} \cdot e^{i2\pi \frac{d_i}{\lambda_0}} \cdot e^{i2\pi \frac{d_r}{\lambda_0}} = S_{l,m} \cdot e^{i2\pi \frac{(l-m)\Delta x}{\lambda_0}}.$$ 

Figure 4.5: (a) Scanning model (b) Defocus model
4.5 Defocus

Traditionally, defocus is modeled[21] by making the paraxial approximation which assumes that waves travel nearly in parallel to the lens axis. Therefore, it may not give accurate results when lenses with large NA’s are involved. Our defocus model, with no such approximation being made, is similar to the scanning model in that it takes into account the phase variations of the incident and reflected waves in the \(z\)-direction.

The defocus model is illustrated in Fig. 4.5(b). The scattering matrix elements \(S_{l,m}'s\) are first calculated for the wafer (i.e., dotted wafer) in perfect focus. If the wafer is now displaced away from the lens by \(\Delta z\), the incident wave of order \(l\) and incident angle \(\theta\) reaches this new location later and the optical path difference \(d_i\) is

\[
d_i = (\Delta z) \cdot \cos \theta = (\Delta z) \cdot \sqrt{1 - (lb\lambda_o)^2}.
\]  

(4.4)

The reflected wave of order \(m\) and reflected angle \(\phi\) also leaves the wafer later and \(d_r\) is

\[
d_r = (\Delta z) \cdot \cos \phi = (\Delta z) \cdot \sqrt{1 - (mb\lambda_o)^2}.
\]

Thus, the scattering matrix element \(S_{l,m}'\) for the defocused wafer is

\[
S_{l,m}' = S_{l,m} \cdot e^{i\frac{2\pi}{\lambda_o}d_i} \cdot e^{i\frac{2\pi}{\lambda_o}d_d} = S_{l,m} \cdot e^{i\frac{2\pi}{\lambda_o}\sqrt{1-(lb\lambda_o)^2}\Delta z} \cdot e^{i\frac{2\pi}{\lambda_o}\sqrt{1-(mb\lambda_o)^2}\Delta z}.
\]  

(4.5)

If the wafer was moved toward the lens, the above expression still holds provided that the sign of \(\Delta z\) is reversed.

The validity of this defocus model is verified in two ways. Firstly, the scattering matrix for the defocused wafer in Fig. 4.5(b) can be obtained either from Eq. 4.5, or from the direct computation for the wafer with a layer of air added onto it. Simulation results show that the scattering matrices derived from these two approaches are the same. Secondly, the result obtained in Eq. 4.4 is found to be equivalent to the result obtained from the wave aberration theory[25]. The wave aberration theory states that the defocus effect can be modeled as wavefront aberration of the incident wave.

As shown in Fig. 4.6, \(O_1\) is the focal point of the perfectly focused wafer surface, and \(O_2\) is the focal point of the defocused wafer surface. Arcs \(AR_1\) and \(AR_2\) are part of the reference spheres of the focal points \(O_1\) and \(O_2\), respectively. When the wafer is defocused, it is treated as the incident waves originated from the reference sphere of arc \(AR_1\), focused onto \(O_1\), and forming an image on the defocused wafer surface.
Or, it is equivalent to treating the incident waves as if originated from the reference sphere of arc $AR_2$ with wavefront aberrated by $R_1R_2$ and are focused onto $O_2$ that form an image on the defocused wafer surface. Therefore, when taking the second approach we need to calculate the wavefront aberration $R_1R_2$.

If the defocus $\Delta z$ is much smaller than $R_1O_1$ and $R_2O_2$, which is usually the case, then

$$R_2O_1 = R_2O_2 - (\Delta z) \cos \theta.$$  

Also, knowing that the difference between two radii is $\Delta z$ gives:

$$R_1O_1 = R_2O_2 - \Delta z.$$  

From the above two equations,

$$R_1R_2 = (\Delta z) \cdot \cos \theta - (\Delta z),$$

where the constant phase term $-(\Delta z)$ can be dropped without affecting the result. Comparing the above equation with Eq. 4.4 yields the conclusion that our defocus model is equivalent to the one derived from the wave aberration theory. Also note that when $\theta$ is small, the above equation reduces to

$$R_1R_2 = -(\Delta z) \cdot \frac{\sin^2 \theta}{2},$$

which is the result[50, 30] of the paraxial approximation.

![Figure 4.6: (a) Defocus mode from the wave aberration theory.](image-url)
4.6 Laser Beam Incidence

As shown in Fig. 4.3, laser may be the illumination source for the scanning beam. The scanning pattern of the laser beam on the wafer can either be a slit whose intensity distribution is two dimensional, or be a spot with the three dimensional intensity distribution, as shown in Fig. 4.7. In both cases, we restrict our discussion to the Gaussian intensity distribution which is the most widely encountered type of laser beam used in optics.

![Diagram of 2D slit scanning and 3D spot scanning](image)

Figure 4.7: Two dimensional laser slit scanning and three dimensional laser spot scanning.

The $E$ field of the 2D laser slit derived in Appendix 1 can be expressed by the fundamental Gaussian-beam solution:

$$E_{2D}(x, z) = E_o \left(1 + \frac{x^2}{z_o^2}\right)^{-\frac{1}{2}} \cdot \exp \left(-i \left[k_o z - \frac{\eta(z)}{2}\right] - x^2 \left[\frac{1}{\omega^2(z)} + \frac{i k_o}{2R(z)}\right]\right). \quad (4.6)$$

At the plane of $z = 0$, the laser slit is considered focused and is

$$E_{2D}(x, z = 0) = E_o \cdot \exp \left(-\frac{x^2}{\omega_o^2}\right). \quad (4.7)$$
where $\omega_o$ is the minimum slit size (MSS). There are two ways to calculate the scattering matrix for the defocused wafer under laser slit incidence. The first one is to adopt the focused field from Eq. 4.7 together with the phase variation method from Eq. 4.5. This is the approach introduced in the previous chapter. The second one is to use the defocused field $E_{2D}(x, \Delta z)$ from Eq. 4.6 directly together with the original scattering matrix. Simulation results show that these two approaches yield the same result, which manifests the validity of Eq. 4.6.

The fundamental Gaussian-beam solution for the 3D laser spot is[67]:

$$E_{3D}(x, y, z) = E_o \left( 1 + \frac{z^2}{z_o^2} \right)^{-\frac{\mu}{2}} \cdot \exp \left( -i \left[ k_o z - \eta(z) \right] - r^2 \left[ \frac{1}{\omega^2(z)} + \frac{ik_o}{2R(z)} \right] \right). \quad (4.8)$$

where $r^2 = x^2 + y^2$. To model the 3D spot scanning by using our 2D imaging model, we approximate the incident field to be a 2D slit with the field $E_{2D}(x, z)$ equal to the "central" field of the 3D spot:

$$E_{2D}(x, z) = E_{3D}(x, y = 0, z).$$

A discrepancy occurs when making this approximation. When the slit/spot is out of focus (e.g., when scanning over the valley part of the structure while the beam is focused on the top surface, as in Fig. 4.7), the rate of intensity drop with defocus is higher for 3D spot than for 2D slit. Intuitively, this can be realized by imagining the energy of the 3D defocused spot spreading in the $x$ and $y$ directions, while the 2D defocused slit only spreads in the $x$ direction. As shown in Fig. 4.7, intensities $I_{2D}(x, z)$ of a 2D slit and $I_{3D}(x, y = 0, z)$ of a 3D spot under different defocus $\Delta z$ are plotted. Pictures on the left are for a slit and spot with MSS = 1 $\mu$m, and on the right are for MSS = 0.5 $\mu$m. The wavelength used is 0.6328 $\mu$m. Slit or spot with smaller size suffers intensity drop more severely; but if used properly, it can also serve as a scheme to discriminate depth variation of the topography as is used in confocal microscopy. From the figure, it can be seen that for MSS = 1 $\mu$m or larger, the discrepancy is negligible if the topography depth does not exceed, say 3 $\mu$m. The size of the scanning spot and the thickness of the wafer structure used in alignment usually fall in this range. However, if the spot size is below 0.5 $\mu$m as may be used in some scanning microscopes or if the topography is relatively thick, the results from our model may have incorrect image contrasts once the topography exceeds 1 $\mu$m although the correct trend prediction is still expected.
Figure 4.8: Comparisons between 2D slit and 3D spot intensities under various defocus conditions. Pictures on the left - MSS=1μm; on the right - MSS=0.5μm.
Chapter 5

Modeling of Signal Detection and Processing

In the previous chapter, the optical image formed by the collecting portion of the configuration is modeled. Subsequently, an optical detector is used to transform the optical signal into electrical signal. A suitable signal processing algorithm is utilized to make a decision on the alignment/metrology from the electrical signal. During this signal detection and processing steps, as shown in Fig. 5.1, noises which may severely degrade the accuracy of the final alignment/metrology results occur. The aim of this chapter is to develop models for different types of noise and for optical detectors, and to implement several signal processing algorithms to demonstrate their capabilities of improving alignment/metrology resolution.

![Diagram of signal detection and processing]

Figure 5.1: The flow of signal detection and processing.

5.1 Noise

In this section, four types of noise (i.e., noise from granular surface, photon noise, shot noise, and thermal noise) are considered to be the dominant ones in degrading the alignment/metrology accuracy. Others like 60Hz pick-up, light source fluctuations,
and noises in different frequency spectra are assumed to be removable by engineering and are not considered here.

The surfaces of materials like polysilicon and aluminum are often granular and their optical images are distorted by this structural irregularity. The grain size of these “noisy surfaces” strongly depends on the material type and the processing step. For example, the grain size of polysilicon can range from 0.03 to 3\,\mu\text{m}[58] mainly depending on the thermal cycle. Although the physical structures of these grains have been studied, not much effort has been devoted into investigating their optical properties. Kirk[46] modeled the optical images of granular surfaces by randomly perturbing the dielectric constant function $\epsilon^i(x)$ of the “smooth” layer $i$ so that the dielectric constant function $\epsilon''(x)$ of the granular layer becomes

$$
\epsilon''(x) = \epsilon^i(x) + \sum_{j=1}^{J} C_j^i \delta \left( x - X_j^i \right).
$$

The parameters $C_j^i$ and $X_j^i$ are the amplitude and position of the $j$th disturbance, respectively.

Here, we adopt a different approach to modeling materials with small grain sizes by disturbing the calculated scattering matrix elements $S_{i,j}$’s of the smooth topography represented by $\epsilon^i(x)$’s. If, for the smooth topography, the calculated scattering matrix element $S_{i,j}$ is

$$
S_{i,j} = r_{i,j} \cdot \exp \left( i \phi_{i,j} \right),
$$

then the scattering matrix element $S'_{i,j}$ of the noisy topography is modeled by independently perturbing the amplitude $r_{i,j}$ and phase $\phi_{i,j}$ of $S_{i,j}$ so that

$$
S'_{i,j} = r_{i,j} (1 + RN_1) \cdot \exp \left[ i \phi_{i,j} (1 + RN_2) \right].
$$

The random numbers $RN_1$ and $RN_2$ are assumed to have uniform distributions between $\pm SNR_1$ and $\pm SNR_2$, respectively. The signal-to-noise ratios $SNR_1$, $SNR_2$ are the input parameters used to adjust the noise level, or rather the degree of granularity, of the topography. The advantage of using this model is that simulations of optical images due to different grain sizes only require the computation of the waveguide scattering problem once. However, for samples with large grain size, the optical images tend to be deformed in a less random manner. Then, one needs to resort to Kirk’s model by modifying the topography and solving the scattering problem to observe this effect.
Due to the quantum nature of photons, the number of photon arrivals at the optical detector’s surface fluctuates. The noise due to this nonuniform arrivals is called photon noise. Here, the photon noise is modeled by disturbing the amplitude of the scattering matrix element only. The assumption made here is that only the number of photon arrivals (i.e., the amplitude of the scattering wave) fluctuates while the time of arrival (i.e., the phase of the scattering wave) is unchanged. Thus,

$$S'_{i,j} = r_{i,j} (1 + RN) \cdot \exp (i\phi_{i,j}).$$

$RN$ is assumed to follow uniform distribution although Poisson distribution is more realistic. Its range of value is also confined by a user-given $S/N$ ratio.

Shot noise is generated by the formation and flow of mobile charge carriers in the optical detector due to light incidence. Its equivalent noise generator[66] shows that the noise power generated is proportional to the incident energy received by the detector. To model the shot noise, we superimpose an extra term in the detected signal $I_o$ so that

$$I' = I_o + I_o \cdot RN.$$  \hspace{1cm} (5.1)

The power of thermal noise, however, does not depend on the incident energy but on the temperature which causes thermal motions of the charged carriers[66]. It is modeled here by

$$I' = I_o + I_t \cdot RN$$  \hspace{1cm} (5.2)

where $I_t$ is an input parameter that does not depend on $I_o$.

### 5.2 Detectors

The functionality of an optical detector is to transform the received optical signal into an electrical signal which is to be analyzed. The commonly used detectors include photomultiplier, photodiode, or even video camera. All these detectors do not utilize the same mechanisms to generate electrical signals; and even for the same type of detector (e.g., photomultiplier), many factors may affect its specific signal generating mechanism (e.g., photoelectric effect) so that slightly different results may be obtained. It is pointed out[16] that the signal measured by optical detectors may not necessarily reflect the arrival rate of photons, but rather depend on the incident angle, the polarization, the detection efficiency of the detector, etc.
To model various detectors, we make a simplifying assumption that all detectors react to the electric field $\mathbf{E}$ and the measured signal intensity is proportional to $|\mathbf{E}|^2$ at the surface of the detectors. In the scanning schemes, detectors with large receiving area are commonly used to collect as much energy as possible. In Fig. 5.2, if the electric field $\mathbf{E}(x)$ of the TE case at the surface is

$$\mathbf{E}(x) = \hat{y} \sum_l E_l \cdot \exp(\imath 2\pi lbx),$$

the generated signal $I$ by the detector is proportional to

$$I \propto \int_{-P/2}^{P/2} \left| \sum_l E_l \cdot \exp(\imath 2\pi lbx) \right|^2 \, dx = \sum_l |E_l|^2. \quad (5.3)$$

Note that since we adopt the grating approach, to obtain an intensity of finite value the integration can only be performed within a period. In the TM case, if the magnetic field at the surface is

$$\mathbf{H}(x) = \hat{y} \sum_l H_l \cdot \exp(\imath 2\pi lbx),$$

the corresponding electric field $\mathbf{E}(x)$ is

$$\mathbf{E}(x) = \sum_l (\hat{x}H_l \cos \theta_l + \hat{z}H_l \sin \theta_l) \cdot \exp(\imath 2\pi lbx),$$

and the signal $I$ generated is

$$I \propto \int_{-P/2}^{P/2} \left| \sum_l (\hat{x}H_l \cos \theta_l + \hat{z}H_l \sin \theta_l) \cdot \exp(\imath 2\pi lbx) \right|^2 \, dx$$

$$= \sum_l \left( |H_l|^2 \cos^2 \theta_l + |H_l|^2 \sin^2 \theta_l \right) = \sum_l |H_l|^2. \quad (5.4)$$

In the full field schemes, if the electric field $\mathbf{E}(x)$ of TE mode at the detector surface is

$$\mathbf{E}(x) = \hat{y} \sum_l E_l \cdot \exp(\imath 2\pi lbx),$$

and the detected signal $I(x)$ is

$$I(x) \propto \left| \sum_l E_l \cdot \exp(\imath 2\pi lbx) \right|^2.$$

Similarly, if the magnetic field $\mathbf{H}(x)$ of the TM case at the surface is
Figure 5.2: The incidence of TE and TM waves on the detector surface.

\[ H(x) = \hat{y} \sum H_l \cdot \exp(i2\pi lx), \]

and the detected signal \( I(x) \) becomes

\[ I(x) \propto \left| \sum H_l \cdot \cos \theta_l \cdot \exp(i2\pi lx) \right|^2 + \left| \sum H_l \cdot \sin \theta_l \cdot \exp(i2\pi lx) \right|^2. \]

It may be instructive to point out that from Eqns.5.3 and 5.4 the total energy flow is simply the sum of individual energy flows independent of the presence of others. This orthogonality property is equivalent to the one seen in the traditional waveguide problem[68]. However, the orthogonality property in our case is a result of the periodicity of the grating, while the one in the traditional waveguide problem is derived from the Lorentz reciprocity theorem.

### 5.3 Signal Processing Algorithms

After the electrical signal is generated, a suitable signal processing algorithm is required to make the final decision on the reticle-wafer misalignment in alignment problem or on the feature line-width in metrology problem. Different optical schemes generate different kinds of electrical signal patterns so that various algorithms need to be devised to cope with the need of tools. The desirable algorithm must be not only accurate, but also efficient so that the signal processing does not seriously decrease the throughput. The aim of this section is to introduce several standard algorithms
and to give an example to demonstrate the abilities of three algorithms in discerning the structure's edges from noisy signal.

The detected electrical signals of typical bright field and scanning dark field images of an oxide line on the silicon substrate are shown in Fig. 5.3(a)(b), respectively. The oxide line is 1000Å thick and it extends from -3μm to 1μm. The signals are collected with a 0.025μm interval. These two signals can be regarded as metrology signals in which case the distance (i.e., 4μm) between two edges must be determined, or they can be alignment signals where the reticle extends from -5μm to 5μm in which case the misalignment (i.e., -1μm) between the oxide line and the reticle is to be found. The detection circuit is assumed to be noisy so that shot and thermal noises are superimposed onto the original signals according to Eq. 5.1 and 5.2. Higher noise level is used for the dark field image.

![Figure 5.3: (a) Bright field image (b) Dark field image of an oxide line on the silicon substrate](image)

For the bright field image as in Fig. 5.3(a), thresholding[45] is a simple and commonly used algorithm in which case a threshold (say, 50%) of the maximum and minimum intensities' difference is adopted that intercepts the signal at the "sample's edges". Theoretical threshold values are derived[45] and signals from different samples (e.g., feature on silicon substrate, transparent line on mask, etc.) require different thresholds[62]. However, the contrast of bright field signals vary drastically with sample thickness, and coherent illumination on thick samples generates fringes so that sometimes the thresholding algorithms yield incorrect results. Fourier analysis such as autocorrelation and auto-convolution algorithms are also used. If the detected
image is \( I(x_i) \), then the autocorrelated signal \( I_{\text{corr}}(x_i) \) is
\[
I_{\text{corr}}(x_i) = \sum_j I(x_j) \cdot I(x_i + x_j),
\]
and the auto-convoluted signal \( I_{\text{conv}}(x_i) \) is
\[
I_{\text{conv}}(x_i) = \sum_j I(x_j) \cdot I(x_i - x_j).
\]
Nyyssonen[47] proposed a similar algorithm called mean square integration that detects the edges from signals of coherent illumination
\[
I_{\text{msi}}(x_i) = \sum_j [I(x_j) \cdot I(x_i + x_j)]^2.
\]
A more complicated Fourier analysis approach is adopted by Guillaume et. al.[22] to determine the line-widths from bright field type signals. The zeroes in the frequency spectrum of the signal are used in pairs to determine the line-width assuming that the dark lines in the signal correspond to the edges. Measurements of thick layers are achieved by decomposing the image into incoherent and partially coherent components.

For the dark field image as in Fig. 5.3(b), noise is often a major obstacle in accurately detecting the edges. To discern the two edges, a peak-fitting algorithm can be used that adopts several points near the peaks and performs quadratic fitting. Another simple approach is to differentiate the signal and find the zeroes. However, the accuracy of these simple approaches is often degraded by the noise, so that a noise suppression technique is often needed to be incorporated into the algorithm. Amplitude selective algorithm[14] allows nonlinear amplification of the signal so that noise with small amplitude can be suppressed. Fourier analyses such as adaptive synthesis[14] and auto-convolution inherently possessing averaging effect to remove noise offer better resolution with the sacrifice of complicating the calculation.

The dark field signal in Fig. 5.3(b) is processed here by three algorithms to demonstrate their edge discerning abilities. The signal is reproduced in Fig. 5.4(a). The first algorithm is amplitude selective in which the signal \( I(x) \) is amplified to become \( I^5(x) \), as shown in Fig. 5.4(b). Noise away from the peaks are significantly suppressed but the edge locations are still ambiguous; extra effort still has to be made. The second algorithm is auto-convolution in which the signal is convoluted with itself , and the result is shown in Fig. 5.4(c). The distance between two peaks corresponds to the line-width in metrology, or the distance between the major peak and the reticle
center (i.e., \(x = 0\)) corresponds to twice the misalignment. From this algorithm the line-width is determined to be 3.875\(\mu\)m and the misalignment to be -1.025\(\mu\)m, with no ambiguity in locating the peaks. The third algorithm is cross-convolution in which the signal is cross-convoluted with a “bi-cell”\[1\] type function \(B(x)\) where

\[
B(x) = \begin{cases} 
-1 & -1\mu m \leq x < 0\mu m \\
+1 & 0\mu m < x \leq +1\mu m \\
0 & \text{otherwise}
\end{cases}
\]  

\( (5.5) \)

Figure 5.4: (a) Detected noisy dark field image; subsequently processed by (b) amplitude selective algorithm (c) auto-convolution algorithm (d) bi-cell cross-convolution algorithm.

The zeroes of the cross-convoluted signal correspond to the sample edges as shown in Fig. 5.4 (d). There is also no ambiguity in determining the location of the edges. The line-width measured with this algorithm is 3.98\(\mu\)m and the misalignment is -1.00\(\mu\)m.
From the above results, it can be seen that the second and third algorithms offer better resolution in this case, with the sacrifice of more computation time. Also note that a "uniform" noise is imposed onto the dark field signal so that it can still be resolved accurately. If the sample itself is deformed (say, due to non-uniform etching or coating with a non-symmetric photoresist layer), the signal generated often yields offset and it is much more difficult to devise an algorithm to intelligently detect this error.
Chapter 6

Verifications of Models

The validity of the scattering, imaging and detection models developed in the previous chapters is verified in this chapter. The verification efforts are divided into the theoretical and experimental parts.

In the theoretical part, the wave-guide model is first verified by examining the calculated scattering matrix with the conservation of energy law and the reciprocity theorem. Then, the simulated images are compared with the ones from the Abbe's theory in the short wavelength limit and the ones from the Fresnel model in the thin film limit, and further to study their validity ranges. Finally, a direct comparison between the diffraction efficiencies calculated from the wave-guide model and those calculated from the differential model found in the literature[52] is given.

In the experimental part, some published experimental results[54, 11] for Nikon and Ultratech steppers are used to verify the models. Also, phase contrast images of oxide lines and grooves on silicon are simulated and compared with the experimental data, and good agreements are obtained.

6.1 Theoretical Verifications

6.1.1 Conservation of Energy

Very often, the wafer structures under study are composed of dielectric materials (e.g., oxide under visible light; photoresist under alignment wavelength) that do not dissipate energy, except for the silicon substrate whose dielectric constant usually has a negative imaginary part under visible and UV light. In this case, if we are able to calculate the energy carried by the transmitted plane waves just below the substrate
CHAPTER 6. VERIFICATIONS OF MODELS

surface (i.e., for \( z = T \) in Fig. 3.3) before their amplitudes start to decay, we can verify the validity of the calculated scattering matrix by the conservation of energy law. For an incident plane wave of unit amplitude and order \( l_o \) illuminated onto the wafer surface, its energy is equal to the energies carried by the reflected waves of amplitude \( E_l^0 \)'s and the energies carried by the transmitted waves of amplitude \( E_l^t \)'s right below the substrate surface.

Mathematically, it can be expressed by

\[
\cos \theta_{l_0} = \sum_{l=-K}^{+K} E_l^0 + \sum_{l=-L}^{+L} E_l^t
\]

where \( E_l^0 \) and \( E_l^t \) represent the energies carried by the \( l \)th-order reflected wave in air and the \( l \)th-order transmitted wave in the substrate. The number \( K \) is the maximum diffraction order of the traveling wave, and its value is equal to the largest integer smaller than \( P/\lambda_o \). As derived in Appendix B, the reflected waves of orders greater than \( K \) are evanescent waves that do not carry energy, as can be seen from Eqs. B.3 and B.4. All the transmitted waves, however, carry energy because the substrate dielectric constant is a complex number. Therefore, the transmitted waves of all orders have to be taken into account and the summation limit \( L \) is the number of order used for the calculation. Substituting the results of Eqs. B.3 and B.4 into the above equation, we obtain, in the TE case,

\[
\cos \theta_{l_0} = \sum_{l=-K}^{+K} |E_l^0|^2 \cdot \cos \theta_l + \sum_{l=-L}^{+L} |E_l^t|^2 \cdot Re \left[ \sqrt{n_s^2 - (l \lambda_o)^2} \right]. \tag{6.1}
\]

In the TM case, it becomes

\[
\cos \theta_{l_0} = \sum_{l=-K}^{+K} |H_l^0|^2 \cdot \cos \theta_l + \sum_{l=-L}^{+L} |H_l^t|^2 \cdot Re \left[ \sqrt{n_s^2 - (l \lambda_o)^2} \right]. \tag{6.2}
\]

Numerically, the scattering problem due to TM polarization is more difficult to solve than TE polarization. An example of TM polarization is given below. The scattering matrices for a 2\( \mu \)m thick, 4\( \mu \)m wide oxide line on the silicon substrate are repetitively calculated using the order \( L = 23, 30, 40, \) and 50. The period and the wavelength used are 15.0\( \mu \)m and 0.6328\( \mu \)m so that the largest traveling order in air is 23. The results are shown in Fig. 6.1 where the error between the incident energy and the reflected/transmitted energy sum (i.e., \( \cos \theta_{l_0} - (\sum_{l=-K}^{+K} E_l^0 + \sum_{l=-L}^{+L} E_l^t) \)) in percentile is plotted against the incident angle \( \theta_{l_0} \) (represented by the NA). Two general trends can be seen from these results. Firstly, for the calculation under the
Figure 6.1: The difference between the incident energy and the (reflected + transmitted) energies is plotted against the incident angle, represented by the NA. The illumination is under TM polarization. In general, better result is obtained when a larger number of waves for the scattering calculation is utilized.

same number of order, the error tends to become larger for larger incident angles. This is probably because more energy is diffracted into the higher order inhomogeneous waves when the incident angle is large, and a larger number of order should be adopted to gain better accuracy. Secondly, the accuracy in general improves when a larger number of order is utilized although order 40 behaves as an exception. The error under the calculation of order 23 can be as large as 10\% while that of order 50 is under 1\%. Thus the accuracy obtained from our scattering calculations is a trade-off with the computation time and it is desirable to have such an energy conservation checker to aid us in doing accurate and efficient simulations. The corresponding full field images are shown in Fig. 6.2 where a small number of waves, namely 15, was also employed. Note that the errors due to TE polarization for the same scattering problem discussed are well below 0.01\% for $L=50$!

6.1.2 Reciprocity Theorem

Another way to verify the calculated scattering matrix is to utilize the reciprocity theorem. It is more general than the conservation of energy checker in that the reciprocity theorem holds even if the structure is lossy, i.e., the dielectric constant
of the structure may have a negative imaginary part. In this case, Eqns.6.1 and 6.2 derived for the conservation of energy checker can still offer useful information such as the portion of the energy reflected from a lossy structure. The reciprocity theorem, however, does not hold when energy-generating materials (i.e., dielectric constants having positive imaginary part) exist.

The reciprocity theorem in the grating theory is illustrated in Fig. 6.3. If a wafer structure is illuminated by a plane wave with an incident angle $\theta_1$ and amplitude $E_1$, part of the diffracted energy will be carried by the diffracted plane wave with angle $\theta_2$ and amplitude $E_2$. The diffraction efficiency $\beta_{1 \rightarrow 2}$ is defined as

$$\beta_{1 \rightarrow 2} = \frac{|E_2|^2 \cdot \cos \theta_2}{|E_1|^2 \cdot \cos \theta_1}.$$ 

Now, if the same structure is illuminated by a plane wave of incident angle $\theta_2$ and amplitude $E'_2$ and part of the diffracted energy is carried by the diffracted plane wave of angle $\theta_1$ and amplitude $E'_1$, the reciprocity theorem states that the diffraction efficiencies in both cases is the same$[42]$. Mathematically, it is expressed by

$$\frac{|E_2|^2 \cdot \cos \theta_2}{|E_1|^2 \cdot \cos \theta_1} = \frac{|E'_1|^2 \cdot \cos \theta_1}{|E'_2|^2 \cdot \cos \theta_2}. \quad (6.3)$$
An example is given below to illustrate the convergence of the numerical calculation examined by the reciprocity theorem. The same TM illumination and structure used in the previous section are employed here, except that the oxide is replaced by a lossy photoresist material whose refractive index is $1.60 - j0.02$. Roughly, 75% of the energy is lost if a plane wave travels into the resist line and is reflected back from the substrate surface. The scattering matrices are calculated with the order 23, 30, 40, and 50. Within a given illumination NA, all pairs of orders are examined by the reciprocity theorem and their deviations from Eq. 6.3 are averaged and plotted in percentile against NA in Fig. 6.4. During the averaging process, the deviation is weighted by the square of the wave amplitude. As can be seen, the same trends as in the energy conservation case exist in the reciprocity theorem. The accuracy improves with increasing number of order used and higher order illumination has poorer accuracy than lower one. Again, the errors calculated for the TE polarization are much smaller.

In our simulation flow, both the conservation of energy and the reciprocity theorem checkers are used to examine the accuracy after every scattering calculation. From our experience, the results of these two checkers seem to indicate the same trend all the time; i.e., the accuracy improves when a proper number of order is chosen.

### 6.1.3 Comparison with Abbe's Theory

During the photolithography exposure, a clear pattern on the mask or reticle is illuminated by a condenser lens and its image is projected onto the photoresist surface by an objective lens. Traditionally, this imaging step is modeled by the Abbe's theory as introduced in Chapter 4. However, the Abbe's theory is valid provided that the
Figure 6.4: The deviation from the reciprocity theorem is plotted against the NA.

mask or reticle feature is large as compared with the illumination wavelength, and its validity is challenged as the feature size becomes comparable to the wavelength. As mentioned in Chapter 3, the rigorous wave-guide model can also be used to calculate the transmitted field of the scattering structure. It is the aim of this section to verify the wave-guide model by comparing it with the Abbe’s theory when the feature size is much larger than the wavelength, and to examine the validity range of the Abbe’s theory as the feature size decreases.

As shown in Fig. 6.5, the simulated images (i.e., $|E(x)|^2$) of the mask or reticle features are generated using the Abbe’s theory and the wave-guide model. The mask or reticle is assumed to be coated with an 800Å Cr layer with refractive index 2.5 – $j3.4$. The illumination wavelength used is g-line (i.e., 0.4358μm), and the NA of the 1X objective lens used is 0.54 which is the largest projection NA[29] for steppers currently on the market. The illumination type is assumed to be Kohler and the coherence factor is 2/3. The feature sizes range from 5.0μm to 0.5μm, which are approximately 10 to 1 times the wavelength. When the feature size is 2μm or above, the images calculated from both the Abbe’s theory and the wave-guide model agree well. When the feature size goes down to 0.5μm, the image from the wave-guide model under the TE polarization is still approximately equal to the Abbe image while the TM image is significantly weaker. For the objective lenses with smaller NA’s, this “TM anomaly” is expected to be less severe. Therefore, the Abbe’s theory can be safely used down to the mask or reticle feature sizes of about twice the wavelength.
Figure 6.5: Simulated mask or reticle images on the photoresist surface using the Abbe's theory and the wave-guide model.

6.1.4 Comparison with Fresnel Model

It has been proven experimentally that the Fresnel model can accurately predict the coherent images of thin film structures under conventional full field microscopes[45]. As the film thickness increases, the light scattering and interference occurring on the structures become more pronounced and the Fresnel model may not be able to serve as a rigorous model. Also, if the feature width is comparable to the wavelength used, the Fresnel model may also yield incorrect results. The objective of this section is to verify the wave-guide model by comparing the thin film images with the Fresnel model's images, and to examine the validity range of the Fresnel model as the film thickness increases.
As shown in Fig. 6.6, the simulated images of isolated oxide lines on silicon with different heights are generated using both the Fresnel and the wave-guide models. The oxide lines are 2μm wide with vertical sidewalls. Figures on the left show the coherent images with the collection NA = 0.6, and figures on the right show the images from Kohler illumination with the illumination NA = 0.4 and the collection NA = 0.6 (i.e., coherence factor = 0.67). The illumination wavelength used is g-line and random polarization is adopted. For the line height less than 1000Å, the Fresnel model agrees well with the wave-guide model in both the coherent and Kohler images. As the thickness goes up to 2500Å, the offset of edge location from the Fresnel model clearly appears while the overall shape of the images are the same. For the 5000Å case, the coherent images are very different but the Kohler images still possess the same contrast.

From our experience, the Fresnel model yields the same trend when compared with the wave-guide model for many optical alignment and metrology schemes we have studied, and the simplicity of its implementation makes it very attractive.

6.1.5 Comparison with Differential Model

Popov and Mashev[52] employed the differential model to calculate the diffraction efficiencies of a sinusoidal grating exposed in the air. In this section, the same diffraction problems are calculated using the wave-guide model in order to make direct comparisons. The topography used in this problem is shown in Fig. 6.7 where a sinusoidal grating of dielectric constant \( n_s = 2.3 \) is illuminated by a normally incident plane wave of unit amplitude. The illumination wavelength is 0.6328μm. The grating period \( P \) is 0.8μm and the modulation is 2\( h \). From this particular combination of the period, wavelength, and refractive index, only three reflective waves and five transmitted waves exist that carry energy. The efficiency \( \beta_j \) is defined as

\[
\beta_j = |E_j|^2 \cdot n \cdot \cos \theta_j
\]

where \( n \) is the refractive index and \( \theta_j \) is the diffracted angle.

In the first problem, the modulation amplitude \( h \) is 0.173μm and the illumination is TE-polarized. The topography used by the wave-guide model is stratified into 10 and 20 layers. The simulation results are compared with the one extracted from Fig. 5 of reference[52] as shown in Table. 6.1, and good agreement is obtained. Note that the reflected efficiencies are small in this problem and are not plotted in the reference.
Figure 6.6: Comparison between the images generated by the Fresnel model and the wave-guide model. Figures on the left are from coherent illumination and figures on the right are from Kohler illumination.
Figure 6.7: A sinusoidal grating illuminated by a normally incident plane wave of unit amplitude.

<table>
<thead>
<tr>
<th>efficiencies (TE)</th>
<th>( \beta_0 )</th>
<th>( \beta_{\pm 1} )</th>
<th>( \beta_{\pm 2} )</th>
<th>( \beta_0 )</th>
<th>( \beta_{\pm 1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>wave-guide 10 layers</td>
<td>0.101</td>
<td>0.222</td>
<td>0.192</td>
<td>small</td>
<td>small</td>
</tr>
<tr>
<td>wave-guide 20 layers</td>
<td>0.102</td>
<td>0.220</td>
<td>0.194</td>
<td>small</td>
<td>small</td>
</tr>
<tr>
<td>differential model</td>
<td>( \sim 0.10 )</td>
<td>( \sim 0.20 )</td>
<td>( \sim 0.18 )</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Table 6.1: Comparison between the wave-guide model and the differential model for \( h = 0.173 \mu m \).

In the second problem, the modulation amplitude \( h \) becomes 0.24\( \mu m \) and the polarizations used are TE and TM modes. All other parameters are unchanged. Identical simulation steps are repeated in order to compare with the results extracted from Fig. 6 of the same reference, and the comparisons are shown in Table. 6.2. Although the results from the wave-guide model show their convergence with the increasing number of stratified layers, they are totally different from the ones found in the reference. We suspect the reference offered incorrect values of some parameters.
Table 6.2: Comparison between the wave-guide model and the differential model for $h = 0.240 \mu m$.

### 6.2 Experimental Verifications

#### 6.2.1 Nikon - Optimum L/S Ratio

The Nikon Laser Step Alignment (LSA) scheme[54] employs a grating structure to be the wafer alignment mark as shown in Fig. 6.8. This grating is moved across the stationary elliptic laser spot. When the featureless part of the wafer is scanned, most of the reflected energy is carried by the 0th order diffracted wave due to the high coherence of the laser spot. This 0th order wave is blocked by a spatial filter so that little signal is detected. When the wafer is moved to the position where the alignment grating is partly or fully exposed under the elliptic spot, the illumination is diffracted into waves of different orders. The spatial filter picks up the $\pm 1\text{st}$, $\pm 2\text{nd}$, and $\pm 3\text{rd}$ order waves which are then sensed by the detector to generate the alignment signal. It is essentially a dark field type of alignment scheme which adopts a grating structure for the mark to enhance the alignment signal. The wavelength used is 0.6328$\mu m$.

Since the Nikon alignment scheme uses a 3D structure, our 2D models are not able to generate the signal profile as shown in Fig. 6.8. However, when the alignment grating is fully under the exposure of the laser spot, the diffraction problem can be approximated by treating the grating and the spot as extending indefinitely to the $\pm y$ directions. Then, our 2D models can be applied to study the peak intensity of the alignment signal.
Figure 6.8: Left: The Nikon Laser Step Alignment scheme scans the alignment grating over the stationary elliptic laser spot to generate the alignment signal. Right: A typical alignment signal.

The specific problem under study in this section is the optimum line-space ratio of an oxide grating on silicon in order to generate the maximum peak intensity, and the simulation result will be compared with the experimental data published in Table B-1 of reference [54]. In the experiment, gratings with 8 \( \mu m \) period and three different line-space ratios are scanned and the maximum signal intensities are recorded. The ratios are 3/5 (line=3 \( \mu m \), space=5 \( \mu m \)), 4/4 (line=4 \( \mu m \), space=4 \( \mu m \)), and 5/3 (line=5 \( \mu m \), space=3 \( \mu m \)), and the one found to yield the largest signal intensity is 4/4. In the simulation, we approximate the incident elliptic laser spot to be a TE-polarized plane wave, which is reasonable due to its high degree of coherence. The coating profile of the photoresist is unknown and is assumed to be a 1 \( \mu m \) thick planar photoresist. The refractive indices used for the photoresist and the oxide are 1.60 and 1.456, respectively. The scattering matrix is calculated and the energies carried by the \( \pm 1 \)st, \( \pm 2 \)nd, and \( \pm 3 \)rd order waves are added to yield the signal intensity.

This simulation is repeated for different values of the L/S ratio, and the results are shown in Fig. 6.9 where the x-axis represents the space width instead of the ratio. The optimum width found from the simulation is 3.75 \( \mu m \). The peak intensity obtained using 4 \( \mu m \) space is greater than those using 3 \( \mu m \) or 5 \( \mu m \), and this result is in agreement with the experimental result. However, the diffraction in general depends strongly on the exact profile of the topography so that both the experimental and the simulation results mentioned above are only valid for the specific topography discussed. For alignment conditions under different levels and different resist coating processes, the simulation work has to be repeated to obtain the optimum line-space ratio for that specific case.
Figure 6.9: The detected signal intensity is plotted against different space widths, while the period is kept to be 8μm.

6.2.2 Ultratech - Alignment Images on A1 Level

Castel and Shamma[11] performed a series of alignment experiments on the Ultratech 900 stepper in order to specify and optimize the design parameters of the alignment marks. It was discovered that when the alignment was operated on the aluminum level, the detected images degraded with the increasing mark height. In this section, the detected images are compared with the simulated ones using both the Fresnel model and the wave-guide model.

The Ultratech 900 stepper employs a dark field type of alignment scheme as shown in Fig. 6.10. A 3μm clear reticle pattern (i.e., the key) is projected onto the wafer and the alignment mark (i.e., the target) is scanned by, or convoluted with this projected image. In real designs, the central order of the key image is not collected by a reflection lens which projects the key image onto the wafer surface, and this is modeled by placing an obstacle in the middle of the projection lens. On the other hand, only the central order of the reflected image from the wafer is collected, and this is modeled by placing an obstacle with a central hole in front of the imaging lens. The illumination NA on the key feature is 0.20, the projection NA is 0.315 with the central order blocked, and the imaging NA is small enough so that only the 0-th order wave can pass through. The illumination source used is broad-band which ranges roughly from 500 to 650nm.
The alignment level of interest is the aluminum level as shown in the figure. An oxide line of height $h_1$ and width $w = 2\mu\text{m}$ is first manufactured on the wafer surface as the alignment target. A layer of aluminum ($n = 0.4 - j7.8$) is then deposited onto the target, and is subsequently coated with a layer of photoresist ($n = 1.6$). Three different targets with the heights $h_1$’s being equal to 0.75, 1.2, and $1.5\mu\text{m}$ are used to perform the alignment. The topography parameters $h_1$, $h_2$, $h_3$, $h_4$ and $w$ are well-specified. The photoresist layer can have up to $0.1\mu\text{m}$ profile variation and that the aluminum layer conforms less to the target as $h_1$ increases. Also the aluminum layer has a discontinuity on its profile near the mark edges when the thickness is larger than $1.2\mu\text{m}$. In the experiments it was found that as $h_1$ increases, the detected alignment image changes from single-peaked, narrow width, and well-specified signal to multi-peaked, broad width, and ambiguous signal as shown on the top row of Fig. 6.11.

Two assumptions are made in order to facilitate the simulation work. Firstly, although the parameters $h_1$’s are well-specified, the exact curvature of the aluminum and photoresist coatings are not known. For simplicity, we assume that the photoresist surface is planar. Since the aluminum layer is highly reflective, most likely the signal variation is due the change of the aluminum profile. Secondly, the illumination is assumed to be monochromatic under $575\text{nm}$ to simplify the calculation. Since the detection scheme is essentially the dark field type, the image profile obtained is
CHAPTER 6. VERIFICATIONS OF MODELS

expected to be relatively insensitive to the wavelength so that the simulated image from the monochromatic illumination should have the same trend as that from the broad-band illumination.

The simulation results generated by using the Fresnel model are shown in the middle row of Fig. 6.11 in which the curvature of the aluminum layer is adjusted so that similar images to the detected ones are obtained. The same trend is still obtained even if arbitrary curvatures are assigned. The same simulations, however, cannot be repeated using the wave-guide model because of the floating point problem mentioned in Chapter 3. Here, we decrease the imaginary part of the aluminum refractive indices in these samples to the values where the wave-guide model can barely be used without having numerical problems. The indices used are $0.4 - j3.8$, $0.4 - j3.0$, and $0.4 - j1.5$ for the samples with $h_1$ being equal to 0.75, 1.2, and 1.5$\mu$m, respectively. Although different values of the refractive indices are used, these indices still correspond to highly reflective materials except for the 1.5$\mu$m case. The simulated results from the wave-guide model using these new indices are shown in the bottom row of Fig. 6.11 where the same trend can still be observed.

6.2.3 Phase Contrast Images

In this section, the simulated images of two phase contrast metrology schemes are employed to compare with the experimentally measured profiles[53]. These two schemes are the differential phase contrast (DPC) and the “differential” differential interference contrast (differential DIC) schemes and their optical configurations and operating principles will be introduced in Chapter 7. The samples used for the experiments are the Line-Width Standard[43] which has oxide lines and grooves of different widths on the silicon substrate. The thickness of the oxide lines and grooves is approximately 1500Å. The microscope uses TE-polarized He-Ne laser beam as the scanning source. The NA employed is 0.9.

<table>
<thead>
<tr>
<th>Nominal Width</th>
<th>1.07$\mu$m line</th>
<th>1.41$\mu$m line</th>
<th>0.82$\mu$m groove</th>
<th>0.98$\mu$m groove</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>1.00$\mu$m</td>
<td>1.37$\mu$m</td>
<td>0.94$\mu$m</td>
<td>1.07$\mu$m</td>
</tr>
<tr>
<td>Simulated</td>
<td>1.03$\mu$m</td>
<td>1.33$\mu$m</td>
<td>0.92$\mu$m</td>
<td>1.06$\mu$m</td>
</tr>
</tbody>
</table>

Table 6.3: Comparison between the measured and the simulated line-widths from the DPC scheme.
Figure 6.11: The detected images (top row) are compared with the simulated images using the Fresnel model (middle row) and the wave-guide model (bottom row).
Figure 6.12: Comparison between the measured and the simulated images from the DPC scheme.

The DPC scheme is first utilized to measure the line-widths of four samples - 1.07\,\mu\text{m} line, 1.41\,\mu\text{m} line, 0.82\,\mu\text{m} groove and 0.98\,\mu\text{m} groove. The measured and the simulated images are shown in Fig. 6.12 where the scanning spot size used for the simulations is 0.5\,\mu\text{m}.

Also, the measured and simulated line-widths of these four samples are compared in Table. 6.3. The edge locations are assumed to coincide with the peaks when analyzing the measured and simulated images. For these particular samples, the measured line-widths are smaller than the true widths when line structures are detected, and larger when groove structures are detected. The simulations are able to predict the same trend.

The differential DIC scheme[17] is also used and the measured and simulated
images and line-widths are compared in Fig. 6.13 and Table. 6.4, respectively. In the simulations, the spots’ shear and the spot size used are 0.5µm. Again, simple peak-finding algorithm is utilized to locate the edges from these images. It is more difficult to make comparisons with the differential DIC scheme in that this scheme inherently generates asymmetric profiles and thus it is difficult to determine from the experiments the best focused setting. Highly asymmetric profiles are measured for the line structures while these are not observed in the simulated ones. However, good agreement can still be obtained for the groove images.

<table>
<thead>
<tr>
<th>Nominal Width</th>
<th>1.07µm line</th>
<th>1.41µm line</th>
<th>0.82µm groove</th>
<th>0.98µm groove</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>1.14µm</td>
<td>1.52µm</td>
<td>0.87µm</td>
<td>0.96µm</td>
</tr>
<tr>
<td>Simulated</td>
<td>1.13µm</td>
<td>1.44µm</td>
<td>0.84µm</td>
<td>0.96µm</td>
</tr>
</tbody>
</table>

Table 6.4: Comparison between the measured and the simulated line-widths from the differential DIC scheme.
Figure 6.13: Comparison between the measured and the simulated images from the differential DIC scheme.
Chapter 7

Applications

The models developed in our work can be also applied to areas other than those discussed in the previous chapter. In this chapter, they are categorized into three groups - the fundamental factors, the alignment schemes, and the metrology schemes.

The fundamental factors that can be studied utilizing the models are the topography (e.g., see Fig. 6.11), the polarization (e.g., see Fig. 6.6), the NA (e.g., see Fig. 6.5), and the defocus (e.g., see Fig. 4.8). Although not very extensively, their effects on the detected images were explored in the previous chapters and will not be discussed here.

Beside the Nikon and the Ultratech schemes introduced before, the alignment schemes that will be modeled here are the bright field, various dark field scanning, and Moire interference scanning schemes. The metrology schemes that can be modeled other than the conventional full field scheme are the confocal scanning, the differential interference contrast (DIC, also known as Nomarski scheme) and its derivatives, and the differential phase contrast schemes.

7.1 Alignment Schemes

The alignment problem is different from the metrology problem in that in alignment only the relative distance between the mask and the wafer is relevant, regardless of the absolute size of the mask and wafer features. In a way, the alignment problem is easier to deal with than the metrology problem. However, the alignment has to be performed for all process levels so that the wafer alignment mark can be made of various materials that have various thicknesses. It is further complicated by the spin-coated photoresist layer whose asymmetry may cause alignment offset or even
failure. In this section, we apply our models to study three types of commonly used alignment schemes - bright field, dark field scanning, and Moire interference schemes in order to demonstrate the tradeoffs of different alignment schemes.

### 7.1.1 Bright Field

The bright field alignment scheme under study adopts a pair of isolated mask and wafer features. The wafer feature is a long strip of line or groove of approximately 2-5\(\mu\)m wide, and the corresponding mask feature is a transparent strip larger than the wafer feature. The mask feature is first illuminated by using an arc lamp and its image is projected onto the wafer feature, and the reflected image is collected and analyzed in order to determine the misalignment between the mask and wafer. The imaging model for this type of alignment has been shown in Fig. 4.3, and the typical images detected are shown in Fig. 7.1 where the images of the wafer feature (circled) are inside the broader images of the mask. By employing a suitable algorithm (e.g., two dark fringes corresponding to the two edges of the wafer feature) to find the relative position of the mask and wafer from the image, the alignment error can be determined and corrected. Due to the thickness variation of the photoresist and the underlying layers, the reflectivity from the wafer changes which results in different profiles. It is this profile variation that makes the detection of wafer position difficult, while the detection of the mask position is usually much easier. Thus, we will concentrate on studying the images of the wafer features only.

![Figure 7.1: Typical images obtained using the bright field alignment scheme.](image)

The wafer feature used in our study is a 4\(\mu\)m wide and 0.4\(\mu\)m thick silicon groove coated with a layer of photoresist of 1\(\mu\)m nominal thickness, as shown in Fig. 7.2. The hypothetical photoresist layer is nonplanar, and the “dip” (i.e., the distance between the bottom of the photoresist groove and the photoresist surface) used in our simulations varies from zero (i.e., planar layer) to 0.20\(\mu\)m. The dip shown in the figure has been exaggerated to be approximately 0.4\(\mu\)m to illustrate the profile. It
Figure 7.2: A highly asymmetric wafer mark is used to study the alignment.

embodies two photoresist coating anomalies commonly encountered - abrupt profile change on the right side and smooth profile variation on the left side. The effect of these anomalies on the detected alignment images will be examined.

In the simulations, the wafer feature is illuminated by the mask image through the objective lens of NA=0.4 and the reflected image is collected by the same lens. The alignment wavelength used is e-line (0.546\(\mu\)m). The simulated images with different photoresist dips are shown in Fig. 7.2. When the photoresist layer is relatively planar (i.e., the dip is less than, say 0.04\(\mu\)m), the image is symmetric and the wafer location can easily be discerned. As the dip increases, the intensity on the right side of the image remains unchanged because the resist thickness on the right side remains the same, while the intensity on the left and above the wafer groove varies dramatically. The symmetric property of the image completely disappears and most of the signal detection algorithms (e.g., threshold, minima detection) will give significant offsets in the alignment result.

Censor SRA 100 (10x g-line) and 200 (5x h-line) steppers adopt this through-the-lens bright field alignment scheme\[37, 38, 39\]. The recommended width from the manufacturer for the wafer feature is 2.5-4.0\(\mu\)m, and the recommended mask width is such that its image on the wafer is around 40\(\mu\)m. To avoid ambiguous alignment result as shown in Fig. 7.3, two alignment wavelengths - e-line and d-line (0.578\(\mu\)m) are used. For each wavelength, an alignment image is generated and evaluated separately. The stepper then automatically selects the image with the best properties for alignment and focusing. A simple threshold algorithm is utilized to calculate the misalignment. Hitachi LD-5010i (5x i-line) steppers also adopt the bright field scheme\[26\] to be one of the alignment schemes available on the steppers. The same wavelengths - e-line and d-line are used for alignment except that here the images generated by these two wavelengths are summed to be the alignment image. Simulations for the same
samples used in Fig. 7.3 are performed but with a different wavelength (d-line), and the results are shown in Fig. 7.4. Totally different images are obtained because the reflectivity from the wafer feature strongly depends on the wavelength. Neither of these two wavelengths can yield accurate alignment images when the dip is larger than 0.04μm. The sum of the images (Fig. 7.3 and Fig. 7.4) obtained by using e-line and d-line are shown in Fig. 7.5 where the reflectivity effect is alleviated and the edges of the wafer feature can now be easily discerned by locating the darkest fringes with relatively little errors when the dip is 0.12μm or less. As the dip goes above 0.16μm, the left edge of the wafer feature is still difficult to locate and other schemes may need to be adopted.

In summary, the bright field scheme offers high alignment signal intensity and relatively simple optical configuration. However the image profile is strongly depen-
dent on the films’ thicknesses, and a popular remedy is to adopt multi-wavelength or broad-band illumination to “average out” this reflectivity effect.

7.1.2 Dark Field Scanning

Various dark field scanning schemes are widely adopted by stepper manufacturers. The high contrast of the signals and the simplicity of the signal processing algorithms over the bright field scheme make them attractive even though the dark field signal intensity is much lower.

The Dark Field Alignment System (DFAS)[1] in the GCA steppers belongs to this category. The operating principle is shown in Fig. 7.6. For each edge of the wafer feature, there is a corresponding transparent feature on the reticle. The reticle
Figure 7.5: The sum of the images obtained by using e-line and d-line.

feature is imaged onto the wafer surface through a mirror with a central hole so that the effective aperture is half of the objective aperture. The reflected signal from the wafer is collected by the same mirror and only the light outside of the mirror hole is collected and directed into a photomultiplier tube. Therefore, only the scattered light is sensed by the photomultiplier tube while the specular light goes through the mirror hole and is not collected. As the wafer is scanned across the stationary reticle image, maximum intensity is obtained when alignment is achieved.

In our simulations, the reticle feature is chosen to have two transparent lines of $1\mu m$ width separated $4\mu m$ apart. The alignment wavelength used is d-line, while the one used by GCA steppers is g-line. The objective NA on the GCA steppers is 0.35, so that the projection NA used is 0.18 and the collecting NA is from 0.18 to 0.35. The same samples in Fig. 7.2 are utilized, and the simulation results are shown in Fig. 7.7.
Figure 7.6: The model used to simulate the alignment images of the GCA steppers.

The alignment images are immune from the reflectivity effect as expected, and they are of the same shape which greatly simplifies the signal processing task. However, the coating anomaly still causes alignment offset, and the largest error occurs at the dip equal to 0.16\(\mu\)m is -0.18\(\mu\)m. In reality a bi-cell type of detection scheme (Eq. 5.5) is used by the GCA steppers so that zero-crossing detection has to be performed. Also, features are replicated on the reticle and wafer so that signal intensity is enhanced.

The dark field scanning scheme collects the scattering light only. However, the scattering light at various angles contains different information on the wafer topography. In alignment, the scattering is from both the photoresist layer and the alignment mark. The coated photoresist layer is usually much smoother and more planar than the alignment mark so that intuitively the light scattered by the mark edges has a larger portion that goes into larger angle. In the GCA case, the collecting NA is from 0.18 to 0.35 so that the light scattered by the photoresist surface may also have a significant portion collected which causes alignment offset. Here, another dark field scheme is utilized in which a focused He-Ne laser spot scans across the same samples and only the scattered light at larger angle is collected. The laser spot is focused by the objective NA=0.40 such that the spot size on the wafer is 0.5\(\mu\)m, and the collecting NA is from 0.60 to 0.80. The simulated images are shown in Fig. 7.8. Small errors are introduced using this scheme and the largest error occurring at the right edge of the dip=0.20\(\mu\)m sample is 0.1\(\mu\)m. However, the signal intensity obtained using this scheme is very weak so that any glitch in the detection circuit may significantly alter the alignment result. Thus, there is a trade-off between the signal intensity and accuracy. Canon FPA-1500 and 1550 steppers[57] adopt this type of beam scanning scheme except that the wafer mark is oriented 45° to the scanning direction in order to calculate the misalignment as well as the rotational error. Also, the alignment source used is the He-Cd laser (0.442\(\mu\)m).

Another scheme is the oblique dark field scanning in which the beam scans over
the structure at an angle to the optical axis of the telecentric lens. This can be achieved by utilizing a portion of the objective lens that is off the optical axis, or simply by employing an auxiliary lens. The advantage of using an auxiliary lens is that different optics can be designed only for the alignment wavelength so that little chromatic aberration will be introduced. However, offset due to the obliquity may occur depending on the scanning angle and the photoresist thickness. As illustrated in Fig. 7.9, the offset $\delta$ can be estimated by using a simple equation

$$\delta = d \cdot \tan \left[ \sin^{-1} \left( \frac{n_1}{n_2} \sin \theta_1 \right) \right] \quad (7.1)$$

where $d$ is the photoresist's nominal thickness, $\theta_1$ is the incident angle of the scanning beam, and $n_1$ and $n_2$ are the refractive indices of air and photoresist, respectively.

A simulated alignment image due to oblique scanning is shown in Fig. 7.10. The
Figure 7.8: The simulated alignment images from the dark field scanning scheme with only the large-angle scattering light collected.

0.5\(\mu\)m spot size He-Ne laser beam scans over the dip=0.0\(\mu\)m sample at a 40° incident angle. The scattering light within NA=0.40 is collected. Although the topography of the alignment mark is symmetric, the two peaks are of different heights because the scattering conditions at the two edges are different due to the obliquity. The simulated offsets at the left and right edges are -0.38\(\mu\)m and -0.50\(\mu\)m respectively. The offsets calculated using the approximated equation derived above are -0.44\(\mu\)m and -0.52\(\mu\)m.

In summary, the dark field alignment schemes offer images with high contrast and simple profile to analyze. The scattered light at different angles may contain different topography information. Its signal intensity can be very small so that noise is often a problem, and laser source is frequently employed to overcome this difficulty.
CHAPTER 7. APPLICATIONS

Figure 7.9: A simple model to calculate the alignment offset due to oblique scanning.

Figure 7.10: The simulated image of the oblique dark field scanning scheme.

7.1.3 Moire Interference Scanning

Another alignment technique to be discussed is the Moire interference scanning scheme. A Moire pattern is an interference pattern which results from the overlap of two periodic patterns. The relative position of two periodic patterns usually can be determined very accurately by analyzing their Moire pattern with a proper detection scheme. In alignment, if two periodic gratings are located on the reticle and the wafer, the image of either one can be imposed onto the other optically and a Moire pattern is thus generated. By analyzing this pattern, the alignment can be achieved.

An example is given here to demonstrate the Moire interference scanning scheme. A grating of alignment feature on the wafer is composed of silicon lines and grooves as shown in Fig. 7.11. Its period is 8μm and the groove thickness is 0.12μm. A nonplanar but symmetric photoresist layer of 1μm nominal width is coated onto the wafer feature. During the alignment, the wafer feature is illuminated by a He-Ne
Figure 7.11: The wafer feature used to study the Moire interference alignment scheme.

laser of the spot size covering many periods so that it can be approximated as a plane wave incidence. The reflected wafer image is filtered so that only the ±1 orders are collected. This collected image is projected onto the reticle grating so that the transmitted image is a Moire pattern. The transmitted energy is maximized when the wafer and the reticle is aligned. Note that both the periods of the collected wafer image and the reticle feature are half of the wafer grating. The simulated images shown in Fig. 7.12 are the transmitted energy of the Moire pattern when the wafer is scanned over the illumination beam. On the left, a perfectly symmetric image is obtained when the wafer grating is coated with a symmetric layer of photoresist. In the middle, the coated layer is intentionally shifted to the left relative to the wafer by 0.2μm and the detected image generates an alignment offset of about 0.06μm. On the right, the layer is shifted to the left by 0.4μm and the alignment offset is 0.12μm.

This basic principle introduced above is utilized by the ASM steppers[19, 65]. In the real design a dynamic detection scheme is used in which the filtered wafer image is split into two orthogonally polarized beams to each other before overlapping with the reticle. This is accomplished with the aid of a polarizer and a birefringent element. The two beams are sheared by a distance and the intensities of the two transmitted Moire patterns are collected in turn using a modulator. By taking the difference of these two intensities, the alignment is achieved when the difference is zero. An improvement[9] in the spatial filtering was made so that the filtering is performed after the wafer image is overlapped with the reticle feature. The advantage is that there is no need to manufacture the spatial filter on the entrance pupil of the objective lens so that many other lenses commercially available can also be utilized. Makosch[36] also proposed an alignment scheme based on this Moire interference principle. However, the misalignment is calculated by analyzing the signals which are generated by intentionally tilting the reticle at specific angles, and the alignment
is achieved either by moving the wafer or the reticle mechanically or by tilting the reticle to compensate for the misalignment. Sugimoto[56] proposed a Moire alignment scheme by splitting the reticle grating into two parts which are 180° out of phase. Then, two Moire patterns can be generated. By taking the difference between them the alignment can be achieved.

![Alignment Images](image)

Figure 7.12: The alignment images from the Moire interference scanning scheme with the photoresist profile shifted to the left by 0.0, 0.2, and 0.4μm.

In summary, the Moire interference scanning schemes can offer simple alignment image with high dynamic range for analysis if coupled with a suitable signal detection scheme. It is relatively insensitive to wafer feature's defect since the diffraction due to grating has an averaging effect. The same alignment image can be obtained if misaligned by a multiple of the period so that the prealignment accuracy has to be well within a period.

### 7.2 Metrology Schemes

Various optical metrology schemes are utilized in IC manufacturing to perform critical dimension control, overlay measurement, and defect inspection. In general, the accuracy of the optical metrology tool employed needs to be a tenth of the minimum line-width manufactured. As the feature size becomes closer to the resolution limit of the illumination light, the task becomes more difficult and new approaches are being devised to keep pace with the miniature. These approaches include the development of new optical schemes and the use of smaller wavelengths. In this section, three schemes with different optical configurations that can be adopted in IC metrology are modeled. They are the confocal scanning, differential interference contrast, and different phase contrast.
7.2.1 Confocal Scanning

The operation principle of the confocal scanning scheme is illustrated in Fig. 7.13(a). The light source (e.g., a laser source) is focused into a diffraction-limited spot on the structure surface. The reflected light is collected by the imaging lens and focused onto the slit detector surface. The slit detector differs from the conventional full field detector in that instead of receiving the total energy collected by the imaging lens, the slit detector only collects a small portion of the light passing through the slit. The structure is scanned over the stationary spot and the slit detector (or vice versa), and a confocal image is obtained.

Figure 7.13: (a) The operating principle of the confocal scanning scheme. (b) The shadowing effect.

One major advantage of using the confocal scanning scheme is its depth discriminating ability. As shown in Fig. 4.8, if a focused spot is displaced away from its focus plane, its central intensity drops significantly. This phenomenon is more pronounced if the spot size is small. Therefore, the detected image will be bright only for the part of the feature in focus, and be dark for the part out of focus. By using this property, the confocal scanning can be employed to measure the thickness variation on the wafer surface provided that the variation is larger than the depth of focus.

However, the accuracy of the confocal image varies with the focus plane chosen. Very often the scanning image is more precise about the edge location when focusing on the top surface than on the bottom surface[2] of a feature. This can be explained by referring to Fig. 7.13(b) where the "shadowing effect" occurs when the focus plane is set to be the bottom surface. This shadowing effect is particularly severe in confocal scanning due to the high NA of the lens used to focus the scanning spot. Simulations of several confocal scanning images of 2µm-thick, 4µm-wide silicon line and space have been performed and the results are shown in Fig. 7.14. The illumination is a
Figure 7.14: The confocal scanning images of silicon line and groove of 2μm-thick and 4μm-wide. Upper: focus on top surface; Lower: focus on bottom surface.

He-Ne laser spot of 0.4μm spot size. The projection and collecting NA's used are 0.9, and the area of the detecting slit is assumed to be small so that only the central intensity is collected. When the scanning spot is focused on the top surface of the line or groove (upper figures), a sharp intensity drop occurs at the structure’s edges. However, when the spot is focused on the bottom surface (lower figures), the intensity drop is more gradual and the edge location is ambiguous.

KLA[2] developed confocal scanning microscopes particularly suitable for photore sist line-width measurement. The Ar laser source (0.488μm) used may cause strong reflection and interference from different layers which in turn degrade the quality of the detected images. This problem is suppressed by utilizing a shorter wavelength (0.325μm) He-Cd laser due to the higher light absorption by the photoresist.
7.2.2 Differential Interference Contrast

The differential interference contrast (DIC) scheme [34] is a phase sensitive detection scheme that is frequently used for the qualitative inspection of wafer thickness variation in IC manufacturing. In this section, only the scanning DIC scheme is discussed although the full field version of it can also be included by our formulation.

![Diagram of DIC scheme](image)

Figure 7.15: The operating principle of the differential interference contrast scheme.

The operating principle of the DIC scheme is illustrated in Fig. 7.15. The incident beam is first polarized in the 45° direction through a polarizer. It is then separated into two orthogonally polarized beams (i.e., TE and TM) of equal intensity sheared by a distance δ by a Nomarski prism. The sheared distance δ is usually around the resolution limit of the optical device or even smaller. These two sheared beams are projected onto the wafer surface and the wafer is scanned across them in the same direction as the sheared direction. The two reflected or transmitted beams are collected and recombined by another Nomarski prism, which is usually the same prism if operated in the reflection mode. The two beams interfere with each other and are analyzed by a 135° polarizer. The Nomarski prisms used can be adjusted so that a phase shift β between the two beams is introduced. The phase shift β is usually chosen to be π/2 so that the resultant beam is very sensitive to the phase variation of the wafer structure. Typically, a pair of bright and dark fringes appear at the two edges of a phase object. A simple mathematical model for the scanning DIC scheme is given in Appendix C which can help us understand the basic features of this scheme.

The DIC scheme is suitable for detecting phase objects. Samples such as silicon lines or grooves on silicon closely resemble phase objects so that they can be clearly
Figure 7.16: The simulated DIC images of silicon lines and grooves of different heights and depths are shown. The illumination wavelength is $0.6328\mu m$, the NA's are 0.9, the spot size and the sheared distance $\delta$ are $0.4\mu m$. 
Figure 7.17: The reflectivity effect shown in the DIC images can be alleviated by adopting the differential and normalized DIC schemes.

Imaged by the DIC scheme. However, the DIC scheme works well only within a range of feature thickness. When the silicon line height or groove depth exceeds this range, factors such as strong edge scattering and defocus come in that destroy the desired DIC image pattern and these factors are especially pronounced when the feature width is small. The simulated images of silicon lines and grooves of 1μm wide on silicon are shown in Fig. 7.16 with the He-Ne laser as the illumination source. When the height or depth is 2500Å or less, clear DIC images with little offset can be observed. When the height and depth increase, degrading factors start to dominate and the DIC images are gradually destroyed.

The DIC image pattern is distorted when observing a structure with amplitude
variation across its dimension. Two samples of oxide line and space of 1μm wide and
1000Å thick on silicon are chosen to illustrate this “reflectivity effect”. The thickness
of the oxide layer is chosen so that the normal reflectivity from the bare silicon is
about 5 times larger than that from the oxide layer. Therefore, they can be regarded
as amplitude objects. The simulated DIC images are shown on the top of Figs. 7.17.
This reflectivity effect can be qualitatively explained by Eq. C.1

\[ I(x) = r_{\text{TE}}^2(x) \left\{ 1 + \left[ \frac{r_{\text{TM}}(x)}{r_{\text{TE}}(x)} \right]^2 - 2 \cdot \frac{r_{\text{TM}}(x)}{r_{\text{TE}}(x)} \cdot \cos [\alpha(x) + \beta] \right\}. \]

In Appendix C, the ideal DIC image (Eq. C.2) is derived under the assumption that
\( r_{\text{TM}}(x) - r_{\text{TE}}(x) = r_{\circ}, \) or equivalently, \( 1 + \left[ \frac{r_{\text{TM}}(x)}{r_{\text{TE}}(x)} \right]^2 = 2 \cdot \frac{r_{\text{TM}}(x)}{r_{\text{TE}}(x)} = 2. \) Apparently,
this assumption is no longer valid when dealing with amplitude objects and thus the
reflectivity effect destroys the ideal DIC profile.

This reflectivity effect can be alleviated by employing the “differential” DIC
scheme and the “normalized” DIC scheme[17]. If half of the beam is directed into
a 45° polarizer before incident onto the 135° polarizer, an additional image \( I'(x) \)
can be obtained behind this 45° polarizer and it is

\[ I'(x) = r_{\text{TE}}^2(x) \left\{ 1 + \left[ \frac{r_{\text{TM}}(x)}{r_{\text{TE}}(x)} \right]^2 + 2 \cdot \frac{r_{\text{TM}}(x)}{r_{\text{TE}}(x)} \cdot \cos [\alpha(x) + \beta] \right\}. \]

The differential DIC scheme takes the difference of the two images to be the final
image, and the normalized DIC scheme takes the ratio of the difference and the sum
to be the final image:

\[ I_{\text{diff}}(x) = I'(x) - I(x) = 4r_{\text{TE}}(x)r_{\text{TM}}(x) \cdot \cos [\alpha(x) + \beta], \]

\[ I_{\text{norm}}(x) = \frac{I'(x) - I(x)}{I'(x) + I(x)} = \frac{4r_{\text{TE}}(x)r_{\text{TM}}(x)}{r_{\text{TE}}^2(x) + r_{\text{TM}}^2(x)} \cdot \cos [\alpha(x) + \beta]. \]

The reflectivity effect is alleviated because the “dc” term (i.e., \( 1 + \left[ \frac{r_{\text{TM}}(x)}{r_{\text{TE}}(x)} \right]^2 \)) disappears. The simulated differential and normalized images on the same structures are
shown in Figs. 7.17. The reflectivity effect is removed and clear DIC images can
be observed so that these two schemes can also be used to detect amplitude objects.
However, the normalized scheme yields significant offsets in this case and we do not
observe any absolute advantage of this scheme over the differential scheme which is
one step simpler.

\[ ^{1}\text{nomenclature not universal} \]
Figure 7.18: Simulated images of resist-coated alignment structures observed by the DIC, the differential DIC, and the normalized DIC schemes. 1st row: dip=0.00μm; 2nd row: dip=0.04μm; 3rd row: dip=0.08μm; 4th row: dip=0.12μm.
Finally, it is interesting to employ these DIC schemes on the resist coated structures (Fig. 7.2) which can be thought of as combinations of phase and amplitude objects. The simulation results are shown in Figs. 7.18. When looking at the planar sample (i.e., dip = 0.0μm), the DIC scheme suffers from reflectivity effect due to the high reflectivity in the middle of the alignment structure, but the differential and the normalized schemes yield clear and accurate images. As the dip becomes larger, DIC scheme fails quickly while the other two schemes still yield observable images (with significant offsets) until the dip becomes larger than, say 0.10μm. Note that these dip thresholds are much smaller than that of any of the alignment schemes introduced earlier.

### 7.2.3 Differential Phase Contrast

The conventional scanning scheme adopts a focused scanning spot and a full field detector to collect the energies of all diffracted waves from the wafer structure. In this case the amplitude information of the scanned structure is obtained. However, if the full field detector is split into two halves with their splitting line in perpendicular to the scanning direction, the differential signal from these two detectors yields the phase information of the structure. A simple model is derived in Appendix D to illustrate the phase detection nature of this differential phase contrast (DPC) scheme\[64, 3\].

The scheme is suitable for detecting the phase variation across the wafer structure. Simulation results of the DPC images of a 1000Å height silicon line and a 1000Å deep silicon groove on silicon are shown in Fig. 7.19. Both samples are of 1μm wide. Clear and symmetric images can be observed for both types of polarizations. However, the peaks in the TM images do not correspond to the true edges while the TE images yield correct line-width information.

It is interesting to note that the DPC image profiles are less affected by the reflectivity variations across the structures than the DIC image profiles. This can be seen from Eqs. C.1 and D.1

\[
I_{DIC}(x) = r_{TE}^2(x) \left\{ 1 + \left[ \frac{r_{TM}(x)}{r_{TE}(x)} \right]^2 - 2 \cdot \frac{r_{TM}(x)}{r_{TE}(x)} \cdot \cos[\alpha(x + \beta)] \right\}
\]

\[
I_{DPC}(x) = r_1^2(x) - r_{-1}^2(x) + 2r_0(x) [r_1(x) \cos \phi_1(x) - r_{-1}(x) \cos \phi_{-1}(x)]
\]

The DIC image is determined by the summation of the “reflectivity term” \(r_{TE}^2(x) + r_{TM}^2(x)\) and the “phase term” \(2r_{TE}(x)r_{TM}(x)\cos[\alpha(x) + \beta]\). The reflectivity term
Figure 7.19: Simulated DPC images of a 1μm wide, 1000Å thick silicon line and groove on silicon are shown. These two structures closely resemble phase objects.

is nonzero and its value is determined by the reflectivity of the structure. The DPC image is also determined by a reflectivity term \( (r_1^2(x) - r_{-1}^2(x)) \) and a phase term \( (2r_0(x) [r_1(x) \cos \phi_1(x) - r_{-1}(x) \cos \phi_{-1}(x))] \). However, the reflectivity term here is nonzero only when the beam scans across a structure's boundaries. Therefore, it is expected that the DPC scheme can be applied to observe amplitude objects from which the DIC scheme suffers. The same amplitude objects used in Fig. 7.17 are adopted here and their DPC images are simulated and shown in Fig. 7.20. The reflectivity effect is significantly alleviated by using the DPC scheme, as compared to the DIC images in Fig. 7.17.

Figure 7.20: Simulated DPC images of a 1μm wide, 1000Å thick oxide line and groove on silicon are shown. These two structures closely resemble amplitude objects.

Finally, the DPC scheme is used to look at the resist-coated alignment structure as shown in Fig. 7.2 and the simulated results are shown in Fig. 7.21. It can be seen that the bright-dark pair is totally destroyed when the dip goes up to, say 0.1μm.
This is approximately the same dip various DIC schemes fail as shown in Fig. 7.18.

Figure 7.21: Simulated images of resist-coated alignment structures observed by the DPC scheme are shown.

In summary, the DIC scheme utilizes two different scanning beams to generate the final signal while the DPC scheme uses one scanning beam to generate two different signals and their difference is taken to be the final signal. Both schemes are suitable for phase detection and the signal patterns they generate are all of a pair of bright-dark fringes. The DIC is sensitive to the reflectivity variation while the DPC is not. The image profile obtained by using the DIC scheme is usually not symmetric due to the different nature of TE and TM beams, while the DPC always yields symmetric profiles as long as the structure is symmetric.
Chapter 8

Summary

In this thesis, scattering, imaging and detection models have been developed in order to simulate the optical images generated by various optical tools under different processing conditions. Many commercial alignment and metrology schemes have been included in our modeling efforts. The major contributions of this work are two-fold. First, the rigorous wave-guide scattering model is enhanced and verified so that accurate calculation of the scattering matrix for the wafer structure is possible. Secondly, the configurations of various optical schemes are modeled so that different aspects can be understood from simulations.

A simulator based on these models has been implemented. In general, two types of tasks can be performed using this simulator. For lithography or metrology engineers who do not have much freedom in altering the optical configuration of the tool they operate, the simulator can help in understanding the effects of different topographies. For optical designers whose goal is to develop tools that can handle various types of wafer structures encountered in IC manufacturing, the simulator can help in understanding the limitations of different optical schemes and in designing better ones.

There are several possible extensions of this work that can be done. The wave-guide model still suffers from structures with highly dissipative materials. Better numerical techniques need to be devised to remove this limitation. Also, our formulation is two dimensional. As demonstrated in the thesis, there are cases where a three dimensional formulation is required to facilitate more realistic simulations. Finally, since the wave-guide model matches the boundary conditions of neighboring layers independently, it is suitable for parallel processing if many stratified layers need to be adopted to delineate the wafer structure. This is especially important if the
wave-guide model is to be applied to solve the 2D photoresist bleaching problem in which massive computation needs to be performed.
Appendix A

Two Dimensional Laser Slit

In this appendix, the fundamental Gaussian-beam solution for the 2D laser slit, as shown in Fig. 4.7, is derived. The derivation follows closely the one[67] for the 3D laser spot.

We restrict our attention to beams traveling in homogeneous media (in particular, air). Then, the vector wave equations of $E$ and $H$ fields are reduced to a scalar wave equation

$$\frac{\partial^2 \Psi(x, z)}{\partial x^2} + \frac{\partial^2 \Psi(x, z)}{\partial z^2} + k_o^2 \Psi(x, z) = 0,$$

where $\Psi(x, z)$ can be any Cartesian component of $E$ and $H$ fields, and $k_o$ is the wave number in air. The solution we look for is that with the flow of energy predominantly along the $z$ direction, so that it can be approximated by the scalar-wave approximation

$$\Psi(x, z) = \psi(x, z) \cdot e^{ik_o z}.$$  

Substituting the above equation into the wave equation, one obtains

$$\frac{\partial^2 \psi(x, z)}{\partial x^2} - 2ik_o \frac{\partial \psi(x, z)}{\partial z} = 0 \tag{A.1}$$

Note that an assumption has been made in which the field variation in $z$ direction is small so that $\partial^2 \psi / \partial z^2$ is much smaller than $\partial^2 \psi / \partial x^2$ and $2ik_o \partial \psi / \partial z$.

The solution that satisfies Eq. A.1 has to be symmetric with $x$. Two complex functions $P(z)$ and $q(z)$ are introduced in the trial solution

$$\psi(x, z) = \psi_o \cdot \exp \left[ -iP(z) + \frac{k_o}{2q(z)} x^2 \right], \tag{A.2}$$
and Eq. A.1 becomes

\[
\left[ \frac{k_0^2}{q^2(z)} - \frac{k_o^2 q'(z)}{q^2(z)} \right] x^2 + \left[ \frac{ik_o}{q(z)} + 2k_o P' (z) \right] = 0.
\]

Since the above equation holds for all \( x \), one has

\[
q'(z) = 1, \quad P'(z) = -\frac{i}{2q(z)},
\]

so that

\[
q(z) = z + q_o, \quad (A.3)
\]

\[
P(z) = -\frac{i}{2} \ln \left( 1 + \frac{z}{q_o} \right) + C, \quad (A.4)
\]

where \( C \) is a constant phase term which does not affect the final result, and is set to be zero. Note that \( q_o \) has to be imaginary in order to have a bounded solution when \( x \to \pm \infty \), and we define

\[
q_o \equiv \frac{i \pi \omega^2}{\lambda_o}.
\]

Finally, by substituting Eqs. A.3,A.4 into Eq. A.2 and by using these two identities

\[
\exp \left[ -i \frac{k_o}{2q(z)} x^2 \right] = \exp \left[ -\frac{x^2}{\omega_o^2 \left( 1 + \frac{z^2}{z_o^2} \right)} \right] + \exp \left[ -i \frac{k_o x^2}{2z \left( 1 + \frac{z^2}{z_o^2} \right)} \right],
\]

\[
\exp \left[ -\frac{1}{2} \ln \left( 1 + \frac{z}{q_o} \right) \right] = \left( 1 + \frac{z^2}{z_o^2} \right)^{-\frac{1}{4}} \exp \left( -\frac{i}{2} \tan^{-1} \frac{z}{z_o} \right),
\]

one obtains the 2D fundamental Gaussian-beam solution

\[
\Psi_{2D}(x,z) = \psi_o \left( 1 + \frac{z^2}{z_o^2} \right)^{-\frac{1}{4}} \cdot \exp \left( -i \left[ k_o z - \eta(z) \right] - x^2 \left[ \frac{1}{\omega^2(z)} + \frac{ik_o}{2R(z)} \right] \right), \quad (A.5)
\]

where

\[
z_o \equiv \frac{\pi \omega_o^2}{\lambda_o}, \quad \eta(z) \equiv \tan^{-1} \frac{z}{z_o}, \quad \omega(z) \equiv \omega_o \left( 1 + \frac{z^2}{z_o^2} \right)^{\frac{1}{2}},
\]

and

\[
R(z) \equiv z \left( 1 + \frac{z^2}{z_o^2} \right).
\]
At the plane of \( z = 0 \), Eq. A.5 reduces to a simple Gaussian form

\[
\Psi (x, z = 0) = \psi_o \cdot \exp \left( \frac{x^2}{\omega_o^2} \right),
\]

and therefore \( \omega_o \) is called the minimum slit size.

The equivalent equation\cite{67} of A.5 for the 3D laser spot is

\[
\Psi_{3D} (x, y, z) = \psi_o \left( 1 + \frac{z^2}{\omega_o^2} \right)^{-\frac{1}{2}} \cdot \exp \left( -i \left[ k_o z - \eta (z) \right] - r^2 \left[ \frac{1}{\omega^2 (z)} + \frac{ik_o}{2R(z)} \right] \right)
\]

(A.6)

where \( \omega_o \) is now the minimum spot size.
Appendix B

Energy Carried by an Inhomogeneous Plane Wave

In the grating problem, a plane wave of order $l$ traveling in an homogeneous energy-dissipative material with complex dielectric constant $\epsilon_s$ is called an inhomogeneous plane wave where the constant phase plane is not the same as the constant amplitude plane. Because the region of concern is homogeneous, the Helmholtz equation can be used to solve for the fields in both the TE and TM cases. The energy carried by this inhomogeneous plane wave decreases as the wave travels along. The objective is to calculate the energy carried by this wave when it first enters the energy dissipative material (i.e., for $z = T$ in our scattering problem in Fig. 3.3).

In the TE case, the expression (Eq. 3.3) of the inhomogeneous plane wave $E_l(x, z)$ can be written as

$$E_l(x, z) = \hat{y}E_l \exp(i2\pi lbx) \exp\left(ik_0\sqrt{\epsilon_s - (lb\lambda_o)^2}z\right).$$

where $E_l$ is the complex amplitude. In the most general case[13], the above equation can be written as

$$E_l(x, z) = \hat{y}E_l \exp(ik_c n_s n \cdot r),$$

where vector $n$ is a constant vector that satisfies

$$n \cdot n = 1, \quad n \cdot E_l = 0, \quad n \cdot H_l = 0,$$

and vector $r$ is the position vector

$$r = \hat{x}x + \hat{y}y + \hat{z}z,$$

(B.1)
and $n$ is the refractive index. Note that $\mathbf{n} \cdot \mathbf{n} = 1$ here does not imply $|\mathbf{n}|^2 = 1$ because $n$ may be complex. Because the vector $\mathbf{n}$ has to satisfy $\mathbf{n} \cdot \mathbf{n} = 1$ and $\mathbf{n} \cdot \mathbf{E}_t = 0$, it can be uniquely determined so that

$$\mathbf{n} = \frac{\hat{x} \sqrt{n_s^2 - (lb\lambda_o)^2}}{n_s} + \frac{\hat{z}}{n_s}.$$  \hfill (B.2)

Then, from the normalized relation $\mathbf{H}_t = n_s \mathbf{n} \times \mathbf{E}_t$

$$\mathbf{H}_t(x, z) = \left(-\hat{x}\sqrt{n_s^2 - (lb\lambda_o)^2} + \hat{z}lb\lambda_o\right) E_t e^{i2\pi lb\lambda_o x} e^{ik_0 \sqrt{\varepsilon_s - (lb\lambda_o)^2}z}.$$  

Finally, the energy $\varepsilon_t$ carried by this inhomogeneous plane wave is the normal (or $z$) component of the Poynter vector[27]

$$\varepsilon_t = Re[\hat{z} \cdot (E_t \times H_t^*)] = |E_t|^2 \cdot Re \left[\sqrt{n_s^2 - (lb\lambda_o)^2}\right].$$  \hfill (B.3)

In the TM case, the equivalence of Eq. 3.17 for the inhomogeneous plane wave $\mathbf{H}_t(x, z)$ can be written as

$$\mathbf{H}_t(x, z) = \hat{y} H_t \exp(i2\pi lb\lambda_o x) \exp\left(i k_0 \sqrt{\varepsilon_s - (lb\lambda_o)^2}z\right),$$

or

$$\mathbf{H}_t(x, z) = \hat{y} H_t \exp(i k_0 n_s \mathbf{n} \cdot \mathbf{r})$$

where $\mathbf{n}$ has to satisfy Eq. B.1 and has the same expression as in Eq. B.2. Then, the electric field $E_t$ can be uniquely determined to be

$$E_t = \left(-\hat{x}\sqrt{n_s^2 - (lb\lambda_o)^2} + \hat{z}lb\lambda_o\right) H_t e^{i2\pi lb\lambda_o x} e^{ik_0 \sqrt{\varepsilon_s - (lb\lambda_o)^2}z}.$$  

The energy $\varepsilon_t$ carried by this inhomogeneous plane wave is

$$\varepsilon_t = Re[\hat{z} \cdot (E_t \times H_t^*)] = |H_t|^2 \cdot Re \left[\sqrt{n_s^2 - (lb\lambda_o)^2}\right].$$  \hfill (B.4)
Appendix C

Simple Model for DIC Scheme

In this section, a simple qualitative model[34] for the DIC scheme is given. Although to explore many detailed information of the DIC scheme, rigorous scattering and imaging models must be employed, this simple model is able to capture many of its features. Therefore, it can be served as a quick guide. For simplicity, we restrict our discussions here to the scanning DIC scheme in the reflection mode although it can be modified easily to account for its derivatives.

![Diagram of TE and TM beams](image)

Figure C.1: A simplified model for the scanning DIC scheme.

The basic optical configuration of the DIC scheme is shown in Fig. 7.15. Two beams with mutually perpendicular polarizations (i.e., TE and TM) and unit intensity are sheared by a distance $\delta$. These two sheared spots are projected onto the wafer structure and the amplitudes of the two reflected beams from the wafer surface are $r_{TE}(x) \exp(i\alpha_{TE}(x))$ and $r_{TM}(x) \exp(i\alpha_{TM}(x))$ where the magnitudes $r_i$'s and the phases $\alpha_i$'s are functions of the scanning position $x$. These two beams are recombined and analyzed by a 135° polarizer so that the resultant field $E(x)$ behind the polarizer is
\[ E(x) = r_{TE}(x) \cdot e^{i\alpha_{TE}(x)} - r_{TM}(x) \cdot e^{i\alpha_{TM}(x)} \cdot e^{i\beta}. \]

The amplitude difference of the two beams is taken because two mutually orthogonal polarizers are used as shown in Fig. 7.15. If the two polarizers used are oriented in the same direction, the resultant field will become the sum of the amplitudes. A variable phase difference \( \beta \) between the two beams is introduced by the Nomarski prism.

Defining \( \alpha_{TM}(x) - \alpha_{TE}(x) = \alpha(x) \), we can write the above equation as

\[ E(x) = r_{TE}(x) \cdot e^{i\alpha_{TE}(x)} \left[ 1 - \frac{r_{TM}(x)}{r_{TE}(x)} \right] \cdot e^{i(\alpha(x)+\beta)}. \]

Then, the detected signal \( I_{DIC}(x) \) is

\[ I_{DIC}(x) = |E(x)|^2 = r_{TE}^2(x) \left\{ 1 + \left[ \frac{r_{TM}(x)}{r_{TE}(x)} \right]^2 - 2 \cdot \frac{r_{TM}(x)}{r_{TE}(x)} \cdot \cos[\alpha(x + \beta)] \right\}. \quad (C.1) \]

If the structure under study is a phase object, the reflectivities \( r_{TE}(x) \) and \( r_{TM}(x) \) will be equal to a constant \( r_o \) and the scanning image \( I_{DIC}(x) \) becomes

\[ I_{DIC}(x) = 2r_o \cdot \left\{ 1 - \cos[\alpha(x + \beta)] \right\}. \quad (C.2) \]

Note that \( \alpha(x) \) represents the phase difference between the two beams and therefore \( I_{DIC}(x) \) changes its value only when the two beams scan across a "phase" boundary.

The phase term \( \beta \) from the Nomarski prism is adjustable and three types of different images can be observed on the same structure by setting \( \beta \) to 0, \( \pi/2 \), and \( \pi \). If \( \alpha(x) \) is restricted to be small as compared with \( 2\pi \), then Eq. C.2 becomes

\[ I_{DIC}(x) \sim \begin{cases} \alpha^2(x) & \beta = 0 \\ 1 + \alpha(x) & \beta = \pi/2 \\ 4 - \alpha^2(x) & \beta = \pi \end{cases}. \]

When \( \beta = 0 \), the image \( I_{DIC}(x) \) is nonzero only at the phase boundary and thus it is a dark field image. When \( \beta = \pi \), \( I_{DIC}(x) \) is large except at the phase boundary so that it is a bright field image. When \( \beta = \pi/2 \), which is commonly used for the DIC scheme, \( I_{DIC}(x) \) observed has a pair of bright and dark fringes on the two edges of the structure. However, if the structure is not a pure phase object, then the assumption that \( r_{TE}(x) = r_{TM}(x) \) does not hold and the images mentioned above may be drastically altered.

In Fig. C.2, the wave-guide model and the imaging model introduced in chapters 4 and 5 are adopted to simulate the DIC images with these three \( \beta \) settings. The
structure used is a 1\( \mu \)m-wide, 500 Å-thick silicon line on silicon so that it closely resembles a pure phase object. The illumination source is He-Ne laser, the spot size is 0.4\( \mu \)m, the projection and the imaging NA’s are 0.9, and \( \delta = 0.4 \mu \)m. The asymmetry shown in these three figures are due to the difference of the reflectivity coefficients from the TE and TM beams. From the simulation results, it can be shown that the simple model (Eq. C.2) is able to serve as a first-order approximation to the DIC scheme.

Figure C.2: The dark field, DIC, and bright field images of an isolated silicon line on silicon are simulated with the \( \beta \) set to be 0, \( \pi \), and \( \pi/2 \), respectively.
Appendix D

Simple Model for DPC Scheme

In this section, a simple model for the scanning DPC scheme is given to qualitatively illustrate its phase detecting ability. The conventional scanning microscope adopts a focused spot to scan across the structure, and the reflected light is collected by a full field detector. The amplitude information of the structure is thus obtained because the detected signal intensity corresponds to the reflectivity of the structure. The full field detector can be thought of as a combination of two half-field detectors where the splitting line is perpendicular to the scanning direction, and the detected signal is the summation of the signals from these two detectors. However, if instead the difference is taken, the phase information of the structure is obtained as will be illustrated.

![Diagram](image)

Figure D.1: A simplified model for the scanning DPC scheme.

As the focused spot scans across the structure, it is diffracted into plane waves of different orders. The waves diffracted to the right (or left) interfere with one another and the resultant wave is collected by the right (or left) detector. For simplicity, only the zeroth and the first orders of the diffracted waves are considered as shown in Fig. D.1. Then, the signal intensities $I_{\text{right}}(x)$, $I_{\text{left}}(x)$ collected by the right and left detectors are
\[ I_{\text{right}}(x) = \left| r_o(x) + r_1(x) \cdot e^{i\phi_1(x)} \right|^2 = r_o^2(x) + r_1^2(x) + 2r_o(x)r_1(x)\cos\phi_1(x), \]
\[ I_{\text{left}}(x) = \left| r_o(x) + r_{-1}(x) \cdot e^{i\phi_{-1}(x)} \right|^2 = r_o^2(x) + r_{-1}^2(x) + 2r_o(x)r_{-1}(x)\cos\phi_{-1}(x), \]
and the differential signal \( I_{DPC}(x) \) is
\[ I_{DPC}(x) = I_{\text{right}}(x) - I_{\text{left}}(x) \]
\[ = r_1^2(x) - r_{-1}^2(x) + 2r_o(x)[r_1(x)\cos\phi_1(x) - r_{-1}(x)\cos\phi_{-1}(x)]. \quad (D.1) \]
Assuming that the structure is a phase object and that the amplitudes of the diffracted waves do not vary much across the structure, i.e., \( r_o(x) = r_o \) and \( r_1(x) = r_{-1}(x) = r_{\pm 1} \), the above equation becomes
\[ I_{DPC}(x) = 2r_o r_{\pm 1} [\cos\phi_1(x) - \cos\phi_{-1}(x)]. \]
Further, assuming the phase change is small, we can rewrite the above equation as
\[ I_{DPC}(x) \approx r_o r_{\pm 1} [\phi_1^2(x) - \phi_{-1}^2(x)]. \]
Thus the phase information of the structure is obtained. A pair of bright-dark fringes much like the one in the DIC scheme can be observed at the edges when scanning over an isolated line or groove.
Bibliography


[53] The microscopes and experiments were built and performed by Dr. Jerry Shaw. Dr. Shaw is with IBM T. J. Watson Research Center.


[55] The resist profile measurements were performed by Dr. Alexander Starikov. Dr. Starikov is with IBM, East Fishkill facility.


