

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Wavefront metrology for high resolution optical systems

Permalink

<https://escholarship.org/uc/item/1dd5j7ss>

Author

Miyakawa, Ryan

Publication Date

2011

Peer reviewed|Thesis/dissertation

Wavefront metrology for high resolution optical systems

by

Ryan H. Miyakawa

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Applied Science & Technology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Attwood, Chair

Dr. Patrick Naulleau

Professor Andrew Neureuther

Professor Avidoh Zakhor

Spring 2011

Wavefront metrology for high resolution optical systems

Copyright 2011

by

Ryan H Miyakawa

Abstract

Wavefront metrology for high resolution optical systems

by

Ryan H. Miyakawa

Doctor of Philosophy in Applied Science & Technology

University of California, Berkeley

Professor David Attwood, Chair

Next generation extreme ultraviolet (EUV) optical systems are moving to higher resolution optics to accommodate smaller length scales targeted by the semiconductor industry. As the numerical apertures (NA) of the optics become larger, it becomes increasingly difficult to characterize aberrations due to experimental challenges associated with high-resolution spatial filters and geometrical effects caused by large incident angles of the test wavefront.

This dissertation focuses on two methods of wavefront metrology for high resolution optical systems.

The first method, lateral shearing interferometry (LSI), is a self-referencing interferometry where the test wavefront is incident on a low spatial frequency grating, and the resulting interference between the diffracted orders is used to reconstruct the wavefront aberrations. LSI has many advantages over other interferometric tests such as phase-shifting point diffraction interferometry (PS/PDI) due to its experimental simplicity, stability, relaxed coherence requirements, and its ability to scale to high numerical apertures [1].

While LSI has historically been a qualitative test, this dissertation presents a novel quantitative investigation of the LSI interferogram. The analysis reveals the existence of systematic aberrations due to the nonlinear angular response from the diffraction grating that compromises the accuracy of LSI at medium to high NAs. In the medium NA regime ($0.15 < NA < 0.35$), a holographic model is presented that derives the systematic aberrations in closed form, which demonstrates an astigmatism term that scales as the square of the grating defocus. In the high NA regime ($0.35 < NA$), a geometrical model is introduced

that describes the aberrations as a system of transcendental equations that can be solved numerically. The characterization and removal of these systematic errors is a necessary step that unlocks LSI as a viable candidate for high NA EUV optical testing.

The second method is a novel image-based reconstruction that characterizes the aberrations of an optical system with arbitrary numerical aperture and illumination coherence. In this method a known pattern is imaged by the test optic at several planes through focus. A computer model is created that iterates through possible sets of wavefront aberrations until the through-focus series of aerial images matches the aerial images from the experiment. Although the sample space of Zernike coefficients is non-convex¹, a hybrid algorithm consisting of pattern search and simulated annealing methods is used to search for the global minimum.

The computation of aerial images from a partially coherent optical system is expedited with a novel decomposition of the Hopkins equation known as the Reduced Optical Coherent Sum (ROCS). In this method, the Hopkins integral is described by an operator S which maps the space of pupil aberrations directly to the space of aerial images. This operator is shown to be semipositive definite and well-approximated by a truncated sum of its spectral components. The ROCS decomposition has a customizable error bound allowing one to tradeoff aerial image fidelity for significant speed improvements. For aerial image errors of 1–3%, the ROCS algorithm can compute aerial images up to 15 times faster than the Hopkins integration.

The ROCS-based wavefront test is extremely versatile since it is applicable in nearly all optical systems that measure aerial image intensity regardless of numerical aperture or illumination coherence and requires little or no experimental modifications. This test is used to characterize the field-dependent aberrations of the SEMATECH Berkeley Actinic Inspection Tool (AIT), and the results match an independent analysis of the astigmatism aberrations to within $\lambda_{EUV}/20$ rms.

¹A set is *convex* if for every pair of points within the set, every point on the straight line segment that joins them is also within the set [2].

*This work is dedicated to my parents who have always supported me, and to my advisors
Dave Attwood and Patrick Naulleau*

Acknowledgements

There are many people that I would like to acknowledge for their support in making this work possible.

First I would like to thank the members of my thesis committee, each one who has guided me in many ways throughout my graduate career. Andy Neureuther, whose insight and suggestions continue to shape my understanding and appreciation of optics research, also taught me that being stumped on your prelim isn't the end of the world. Avidah Zakhor, besides introducing me to the fascinating work of her former graduate student Nick Cobb that would provide the inspiration for a large part of my research, taught me that a little courage can help ordinary people make extraordinary discoveries.

Dave Attwood, whom I have known for nearly 10 years, could have picked many undergraduates to work in his group, but had he not given me the chance, I surely would not be where I am today. I'll never forget the moment in the lobby of Bechtel engineering building when he took the time to explain the dynamics of electrons in a synchrotron — undoubtedly a story he's told thousands of times — but spoken with the conviction to inspire a young sophomore to take a chance on short wavelength optics.

I would especially like to thank my advisor Patrick Naulleau who has many special qualities that supersede his formidable scientific expertise. Chief among these is an unmatched sense of compassion, which manifests in so many positive ways: an open door over a closed one, patience over frustration, sympathy over hostility, humility over strong-headedness, inclusivity over exclusivity. Patrick is a great scientist, and he is also a great leader.

Finally I would also like to thank my family and my friends, who continue to support me and remind me to appreciate all that is not EUV.

The work presented in this thesis was made possible by support from SEMATECH and the National Science Foundation EUV Engineering Research Center. This work was also supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Contents

1	Introduction	1
1.1	Wavefronts	2
1.2	Optical systems	3
1.3	The semiconductor revolution	4
1.4	Photolithography	7
1.5	Resolution limits	9
1.6	Dissertation overview	11
2	Lateral shearing interferometry, part 1	12
2.1	Introduction	13
2.1.1	Interferometry	13
2.1.2	Phase-shifting point diffraction interferometry	14
2.1.3	Scalability of PS/PDI	14
2.2	Overview of lateral shearing interferometry	16
2.3	Extracting the phase from the interferogram fringes	18
2.3.1	Fourier domain analysis	18
2.3.2	Time domain analysis	21
2.4	The Rimmer method	22
2.4.1	Operations on large matrices	24
2.4.2	Sparse representation of matrices	24
2.4.3	Sparse matrix construction	26
2.4.4	Weighted coefficients	28
2.5	Phase unwrapping	30

2.6	LSI with a diffraction grating	33
2.7	Experimental specifications	36
2.7.1	Independent parameters	36
2.7.2	Dependent parameters	37
2.8	The null interferogram	39
2.9	Summary	41
3	Lateral shearing interferometry, part 2	42
3.1	Null interferogram at medium NA	44
3.1.1	Grating as a hologram	44
3.1.2	Functional form of the 1st order diffracted sphere	46
3.1.3	Functional form of the null interferogram	46
3.2	System alignment and the null interferogram at medium NA	48
3.2.1	Functionalizing $N(x)$ along z	49
3.3	Null interferogram at high NA	52
3.3.1	The 2-ray method	53
3.4	Variable shear	56
3.4.1	Adapting the Rimmer method to variable shear	58
3.5	Discussion	60
3.5.1	Simulation results	60
3.5.2	Summary	61
4	ROCS	62
4.1	Background	63
4.1.1	Aerial image modeling	63
4.1.2	Hopkins decompositions	64
4.2	Motivation	65
4.3	The System Cross-coefficients matrix	68
4.3.1	Properties of the SCC matrix	71
4.4	The Reduced Optical Coherent Sum	74
4.4.1	Spectral decomposition of the SCC matrix	74
4.4.2	Truncation of the spectral sum	75

4.4.3	Diagonalizing the SCC	76
4.5	ROCS in 2D	79
4.6	Effect of mask and source complexity	81
4.7	Comparison to conventional aerial image models	83
4.7.1	Aerial image comparison	83
4.7.2	Analytic form of the error in ROCS	84
4.7.3	Comparison of MAST and ROCS computation times	85
4.8	Summary	88
5	ROCS-based reconstruction of wavefront aberrations	89
5.1	Basic idea and experimental setup	90
5.2	Programming a defocus as an aberration	92
5.2.1	Linear transformation from the binomial basis to Zernike basis	94
5.2.2	Approximation error	95
5.3	Search algorithms	96
5.3.1	Pattern search	98
5.3.2	Simulated annealing	99
5.4	Experiment: The SEMATECH Berkeley AIT	104
5.4.1	Setup	104
5.4.2	Search algorithm	105
5.4.3	Results	106
5.5	Discussion	110
6	Conclusion	111
6.1	Lateral shearing interferometry	111
6.2	ROCS-based wavefront reconstruction	112
6.3	Discussion	113

Chapter 1

Introduction

Contents

1.1 Wavefronts	2
1.2 Optical systems	3
1.3 The semiconductor revolution	4
1.4 Photolithography	7
1.5 Resolution limits	9
1.6 Dissertation overview	11

The work presented in this dissertation centers around the measurement of wavefronts in high resolution optical systems. Although the techniques described in the following chapters have a broad range of applications, the principal focus is to characterize the errors in extreme ultraviolet (EUV) microscopes and lithography tools, as these systems will likely play an important role in the manufacturing of next-generation computer processors, memory, and other semiconductor devices [3].

Before discussing the details of these techniques, we present a brief introduction to the importance of correcting for aberrations in an optical system, and the role of EUV light in the semiconductor industry.

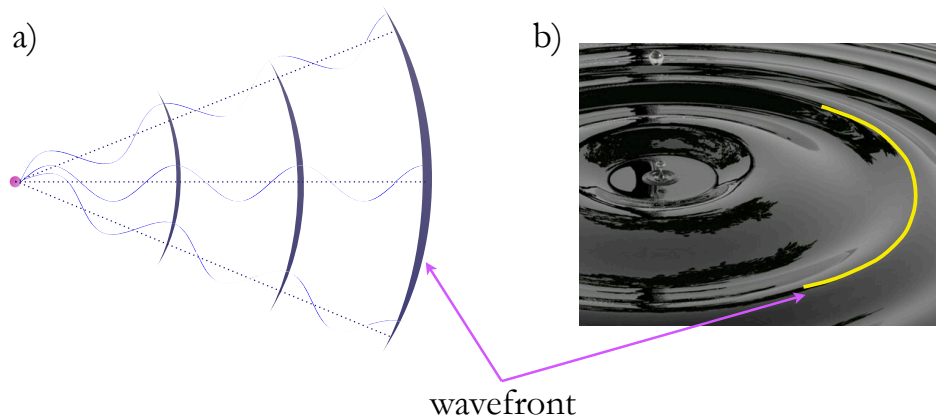


Figure 1.1: a) A wavefront is a surface that specifies points of constant phase. b) A wavefront in water.

1.1 Wavefronts

Light travels in waves that propagate away from its source. Like sound waves or water waves, lightwaves cycle continuously through peaks and troughs, and the number of cycles that have occurred from a given point is known as the *phase* of the wave. The surface that connects points of constant phase is called the *wavefront*, shown in Figure 1.1.

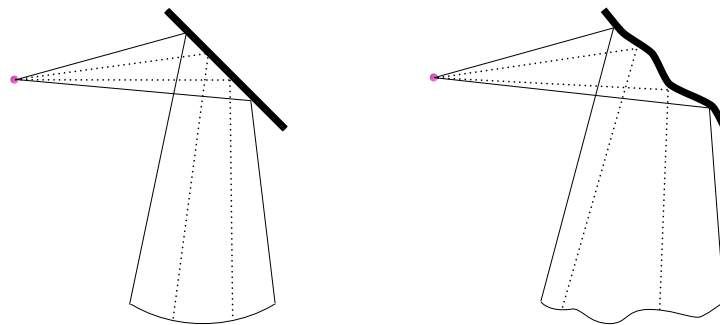


Figure 1.2: Left: A wave reflecting off of a flat mirror preserves its shape. Right: A wave reflecting off of a rough mirror will cause the shape of the wavefront to change.

As light interacts with matter, the shape of its wavefront can change. As shown in Figure 1.2, if a light wave reflects off of a flat mirror, the shape of its wavefront is preserved; however, if the mirror is not perfectly flat, different parts of the wave travel different distances which causes the shape of the wavefront to change.

In this way, measuring a wavefront can give enormous amounts of information about objects that interact with the wave. Biologists, for instance, can measure the light that scatters off of tiny proteins to infer their 3-dimensional structure, a technique that plays an important role in our knowledge of how cells function on a molecular level. Recently a team of scientists at The Scripps Research Institute used this technique and discovered the structure of the protein that delivers HIV to human cells, providing a roadmap for developing drugs that can prevent HIV infection [4].

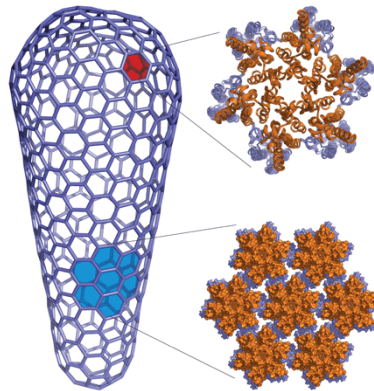


Figure 1.3: The detailed structure of the protein that delivers HIV helps scientists to find a cure [5].

1.2 Optical systems

Measuring wavefronts also has many applications outside of biology. One important application called *optical testing* involves measuring the wavefront that passes through an *optical system*. An optical system is a collection of optical elements such as lenses, mirrors, and diffraction gratings that is designed to manipulate light in a particular way. A camera, for example, is an optical system that is designed to capture the light from a particular plane in space, known as the object plane, onto a separate plane some distance later, known as the image plane. The shape of the wavefronts exiting these systems can give information about whether the devices are performing optimally, and in the case that they are underperforming, information about how to fix them.

One example of an optical system that we all use everyday is the *eye*. If you've been to

the optometrist lately, you may recall a procedure where you stare at a small target as the doctor shines a faint red light into your eye. The doctor is measuring the wavefront of the light that reflects off the surface of your eye which reveals important information about the performance of your eye's optical system. This information can be used to correct your vision by either determining an appropriate prescription for glasses, or more recently by reshaping the surface of your eye with lasers.

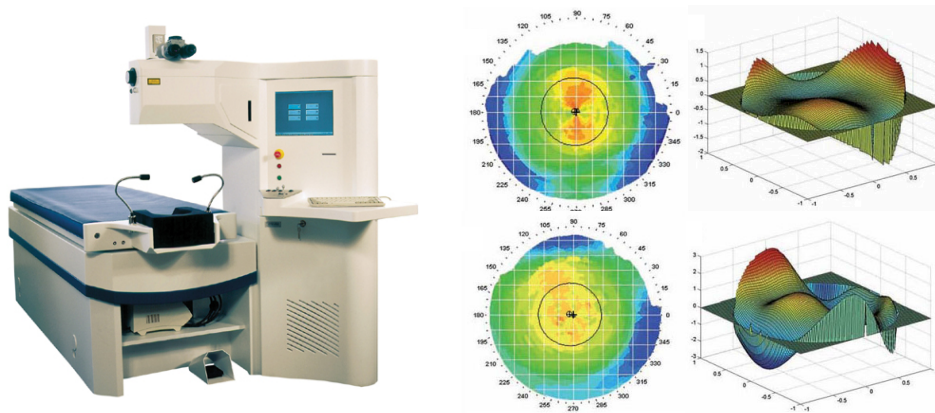


Figure 1.4: Left: The Zyoptix system from Bausch and Lomb has wavefront-sensing capability in addition to its surgical function. Right: Knowledge of the wavefront from an eye gives doctors precise information on how to customize the laser-based procedure on a case by case basis [6],[7],[8].

Your eye isn't the only optical system that needs a prescription. Just as not all eyes see perfectly, not all cameras can take perfect pictures, and not all microscopes can see the smallest features even when it is theoretically possible. This is because small errors in the manufacturing or alignment of the lenses in an optical system known as *aberrations* cause the resulting image to deviate from a perfect image. It is therefore important to measure and correct for these aberrations to maximize the performance of an optical system.

1.3 The semiconductor revolution

The integrated circuit microprocessor, the computational heart of computers and nearly all modern electronic devices, is undoubtedly one of the most important and influential inventions of our time. It is hard to overstate the relevance of the numerous technologies — smart-

phones, credit cards, internet, etc. — that have been made possible by computer processing. The impressive utility of the microprocessor stems in part from its ability perform calculations quickly and reliably, allowing for the automation of tasks that might be tedious or time consuming for a human being.

In 1971 if you opened up a *4004*, Intel’s first single-chip microprocessor, you would find 2,300 tiny structures known as transistors. Transistors are the computational building blocks of the microprocessor, and the power of the processor is proportional to the number of transistors it has. Since the smaller a transistor is, the more of them fit on a single chip, the development of smaller and smaller transistors over the last 40 years has led to an exponential growth in the speed of processors. This is also the reason that computers become outdated so quickly; a computer bought today is approximately 8 times more powerful than one bought just 4 years ago.

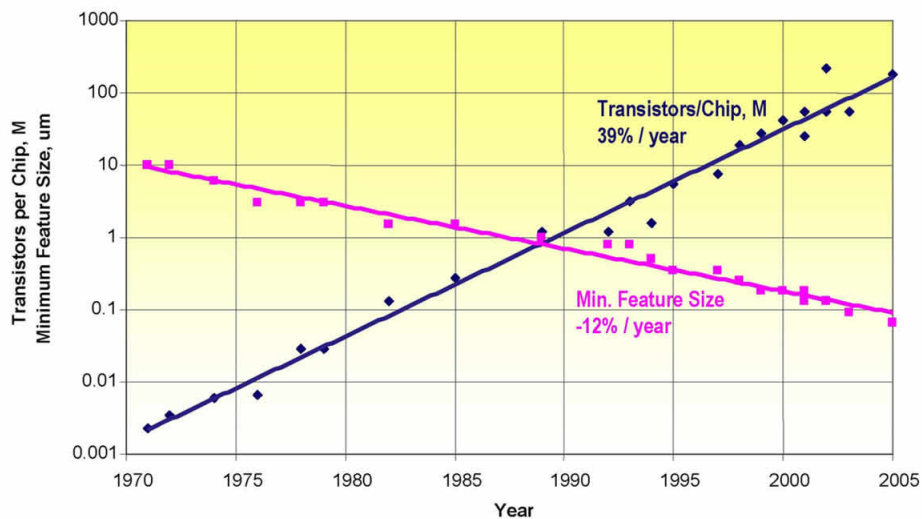


Figure 1.5: Moore’s Law: the size of a transistor halves roughly every 2 years [9].

A transistor in 1971 was about the size of a grain of salt: roughly 100 micrometers, or one tenth of a millimeter square. Since then, transistors have essentially halved in size every 2 years, a trend first predicted by Intel co-founder Gordon E. Moore. This prophetic and arguably self-fulfilling phenomenon known today as *Moore’s Law* held true for the next 40 years. Today, transistors are 100 *nanometers* across, which is nearly a million times smaller by area than their 1971 equivalent (today you could fit *a million* transistors on the top of

a grain of salt). In 2010, Intel released its state of the art Itanium Tukwila microprocessor containing 2 billion transistors.

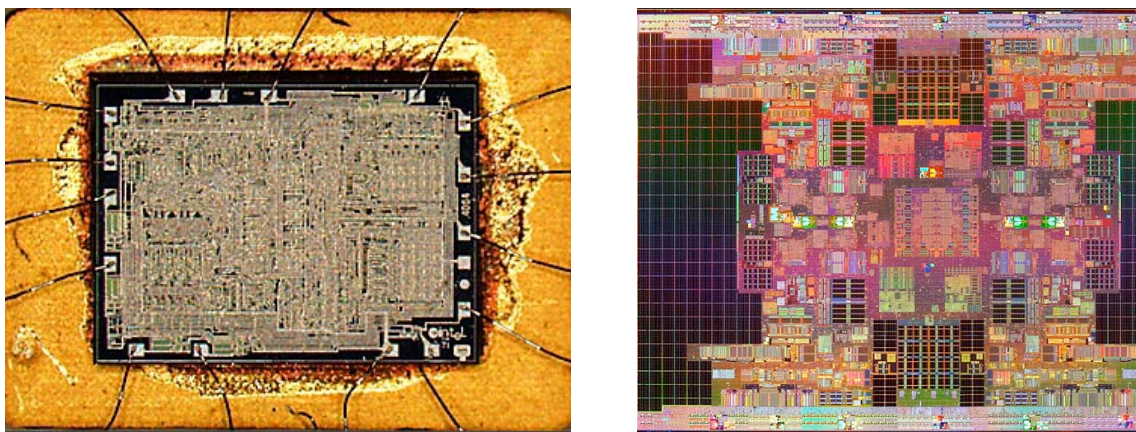


Figure 1.6: Left: The Intel 4004 microprocessor (1971) with 2,300 transistors [10]. Right: The Intel Itanium Tukwila microprocessor (2010) with 2,000,000,000 transistors [11].

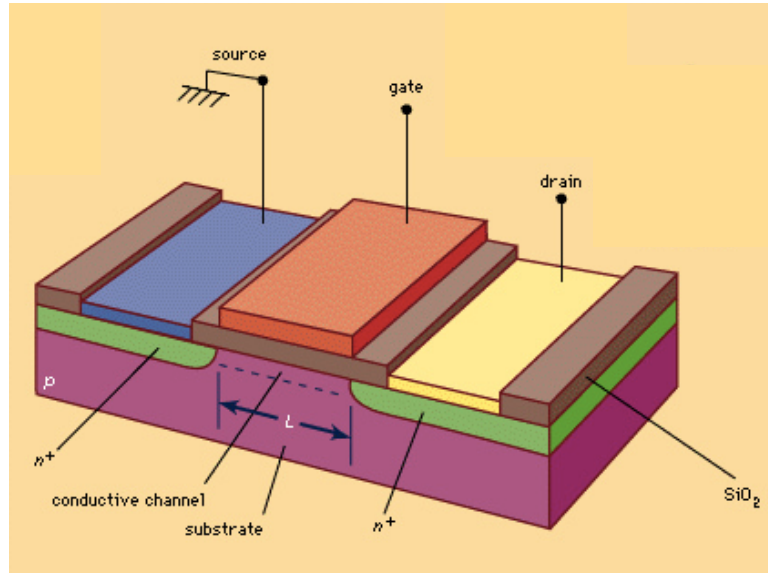


Figure 1.7: Cross-section of a MOSFET transistor [12].

1.4 Photolithography

It is hard to imagine how 2 billion of *anything* can be manufactured in an accurate and timely manner. It doesn't help that transistors are complicated devices requiring a large number of steps to manufacture, and a single defective transistor can ruin the entire chip.

If transistors were to be made one at a time at a rate of 10 transistors per second, making a single modern processor would take over 6 years to complete. So how is Intel able to make processors in a matter of days? The trick is to make all of the transistors at the *same time* using a process known as *photolithography*.



Figure 1.8: Photolithography can be used to add aluminum to a piece of silicon in specific patterns.

Photolithography is a process that uses light to selectively remove parts of an object for the purpose of transferring some geometric pattern to the object. Say for instance we have a

block of silicon pictured below on the left, and we want to add a layer of aluminum on top of it in the pattern shown on the right. Let's see how photolithography can create this structure.

The first step is to apply a uniform layer of aluminum to the top of the silicon (Figure 1.9[a]). Aluminum is generally deposited using evaporation, which means that it will apply evenly over the entire silicon surface. Next, we paint on a thin layer of a chemical called photoresist [b], or just resist for short. Resist has the interesting property that its chemical composition changes when it is exposed to light. The exposed areas in the resist are weakened, whereas the areas that have not been exposed remain unchanged.

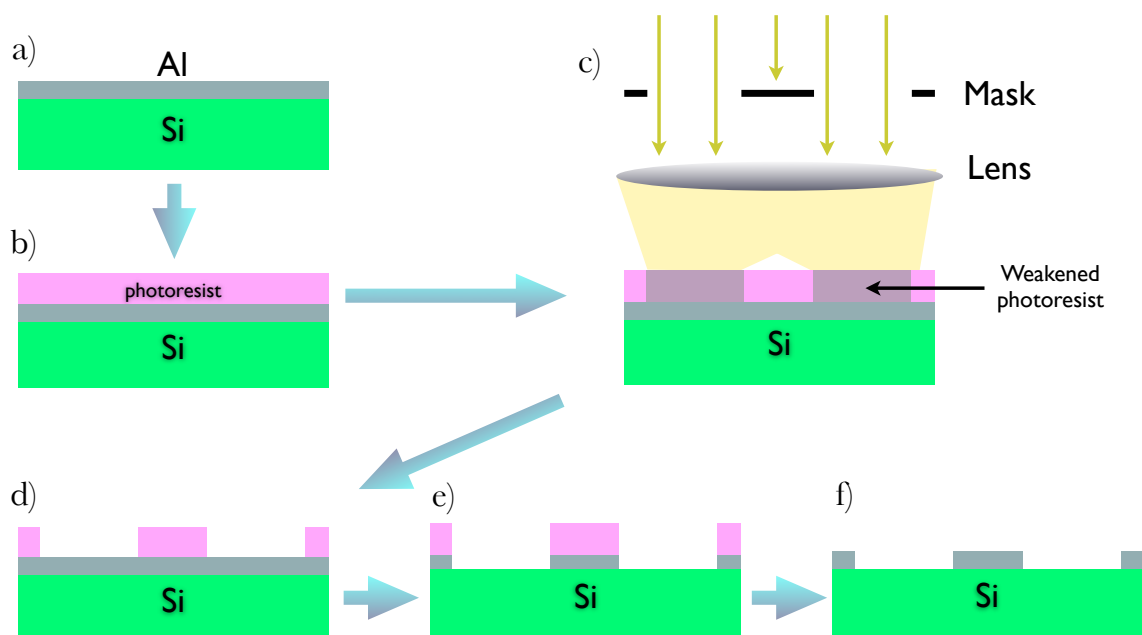


Figure 1.9: (a) Apply aluminum via evaporation, (b) apply photoresist, (c) expose photoresist, (d) wash away exposed photoresist, (e) etch aluminum, (f) remove remaining resist.

Next, a template of the desired pattern, known as the *mask*, is projected onto the resist using a lens just like a camera images a picture onto film [c]. The area of the resist exposed to the light becomes weakened, and washes away when it is treated with developer [d]. With the aluminum now accessible where the notches should be, another acid is applied that dissolves aluminum, leaving the silicon and the unexposed resist intact [e]. Finally the remaining resist is washed away [f].

To add or remove more materials, this photolithographic cycle is repeated over and over

again until the desired device is finished. The fabrication of a modern transistor can take as many as 50 photolithographic cycles to complete the process [13].

In order to make multiple transistors simultaneously, all that is required is a mask that contains multiple copies of the transistor pattern. Modern masks not only have enough patterns to make an entire processor, they have enough to make *multiple* processors at once.

1.5 Resolution limits

If smaller transistors are better, what prevents us from immediately shrinking device dimensions even smaller? There are several reasons why this is challenging, but one of the biggest ones has to do with step [c] in Figure 1.9. The optical system consisting of the mask, lens, and silicon block can only transfer patterns that are larger than its *resolution limit*. This means that once the size of a feature on the mask gets smaller than a certain value, the lens can no longer *see* the feature.

The smallest feature that an optical system can resolve is roughly given by:

$$d_{min} = \frac{\lambda}{2NA} \quad (1.1)$$

where λ is the wavelength of the light illuminating the object, and NA is the *numerical aperture* of the system defined as the sine of the half angle subtended by the lens aperture with respect to the feature. Writing out the numerical aperture into distances:

$$d_{min} = \frac{\lambda L}{D_{lens}} \quad (1.2)$$

where D_{lens} is the diameter of the lens, and L is the distance between the object and the lens. For example, taking the pupil of your eye to have a diameter of 5 mm, and $\lambda = 500$ nm for sunlight, (1.2) says that if you were looking at an open book 50 m away, the text (assuming 12 pt. font) would be smaller than the resolution limit of your eye's optical system. As a result, the words would be illegible even to someone with perfect vision.

Equation (1.1) is extremely important in photolithography because it sets a fundamental limit on how small transistors can be made. Once this limit has been reached, the only way to do better is to either decrease λ or increase NA .

Manufacturers of microprocessors such as Intel and AMD currently use deep ultraviolet (DUV) light ($\lambda = 193$ nm) in their photolithographic processes. In the last few years, in order to keep up with Moore's law, they have been progressively increasing the numerical aperture of their system. However, they are reaching a point where the NA cannot be increased any further [14]. The only remaining option to push down the resolution limit is to decrease the wavelength. For this reason, extreme ultraviolet (EUV) light ($\lambda = 13.5$ nm) is now a leading candidate for next generation of photolithography. Switching from DUV to EUV will provide a substantial leap in the ability to make smaller transistors.

The effect of aberrations on the resolution limit

Equation (1.1) represents the resolution limit of a perfect optical system in the absence of any aberrations. However, when aberrations are present, they can severely degrade the resolution to the point where an aberrated system may have a worse resolution than an aberration-free system operating at a longer wavelength.

In Figure 1.10, the left image represents an image taken by an optical system with sufficient resolution to capture all of the fine detail. The center image is taken through an optical system using a wavelength that is 3 times as long as the original, and cannot support the finest details; notice how the edges of the feathers in the girl's hat no longer come through the system. The final system has the same wavelength as the original system, but has a misaligned lens that is causing aberrations in the system; in this case, the aberrations are so bad that the resolution is no better than the resolution of the long-wavelength system.

It is evident that if the computer industry is going to take advantage of the full capabilities of a photolithographic system using EUV light, it is imperative that there exists a reliable procedure for measuring the aberrations and eliminating them from the system.



Figure 1.10: (a) An image from an optical system with enough resolution to see all of the features; (b) The image from an optical system with a wavelength that is 3 times as long; (c) The image from an optical system with original wavelength in the presence of aberrations.

1.6 Dissertation overview

This dissertation will focus on two different approaches for characterizing the aberrations in an optical system. The first, lateral shearing interferometry (LSI) is a rather large discussion and will be broken up into two chapters. Chapter 2 presents the implementation of LSI in a typical EUV optical system using a diffraction grating to generate the shear. It will also cover the analysis techniques and experimental design considerations. Chapter 3 discusses a novel quantitative investigation of the shearing interferogram in three numerical aperture regimes that demonstrates the existence of systematic aberrations which occur due to nonlinearities in the spread of angles exiting the grating.

The second approach is an iterative technique that is based on a novel decomposition for partially coherent aerial image modeling known as the Reduced Optical Coherent Sum (ROCS). Chapter 4 will be devoted to the derivation and discussion of ROCS. Chapter 5 will demonstrate how ROCS can be used in-situ in an iterative procedure to characterize the aberrations of optical systems with arbitrary numerical aperture and illumination coherence.

Chapter 2

Lateral shearing interferometry, part 1

Contents

2.1	Introduction	13
2.1.1	Interferometry	13
2.1.2	Phase-shifting point diffraction interferometry	14
2.1.3	Scalability of PS/PDI	14
2.2	Overview of lateral shearing interferometry	16
2.3	Extracting the phase from the interferogram fringes	18
2.3.1	Fourier domain analysis	18
2.3.2	Time domain analysis	21
2.4	The Rimmer method	22
2.4.1	Operations on large matrices	24
2.4.2	Sparse representation of matrices	24
2.4.3	Sparse matrix construction	26
2.4.4	Weighted coefficients	28
2.5	Phase unwrapping	30
2.6	LSI with a diffraction grating	33
2.7	Experimental specifications	36

2.7.1	Independent parameters	36
2.7.2	Dependent parameters	37
2.8	The null interferogram	39
2.9	Summary	41

2.1 Introduction

Lateral shearing interferometry (LSI) is an optical test that works by interfering laterally separated or *sheared* copies of a test wavefront. The resulting interference contains information about the wavefront derivative in the direction of the shear. Performing the measurement twice — in the x - and y -directions — gives derivatives in orthogonal directions which can be used to reconstruct the test wavefront.

LSI has a long history [15] as a largely *qualitative* means for characterizing the aberrations in optical systems. This is due partly because of the high computational demand on the shearing interferogram analysis, and partly because of the prevalence of several other interferometric tests that measure wavefronts more directly. However, due to both the emergence of faster computers and high resolution optical systems for which many interferometric techniques become experimentally challenging, LSI is being considered as an important player in the characterization of next-generation optical systems [16].

2.1.1 Interferometry

The shape of a wavefront manifests mathematically as the *phase* of a wave. Since typical detectors measure only the *intensity* of a wave and not its phase, wavefront aberrations cannot be directly measured. Interferometry solves this problem by encoding the phase of the test wave into its intensity by coherently adding it to a second wave, known as the *reference wave*.

If $U_t(x, y)$ and $U_r(x, y)$ are the test and reference waves respectively, where

$$\begin{aligned}
 U_t(x, y) &= A(x, y) \exp \{ikW_t(x, y)\} \\
 U_r(x, y) &= B(x, y) \exp \{ikW_r(x, y)\}
 \end{aligned}
 \tag{2.1}$$

then the intensity of the combined wave will contain a cross-term that contains the test wavefront $W_t(x, y)$:

$$I_T = 2A(x, y)B(x, y)(1 + \cos[k\Delta W(x, y) + \Phi]) \quad (2.2)$$

where $k\Delta W$ is the optical path length difference,

$$k\Delta W(x, y) = k[W_t(x, y) - W_r(x, y)] \quad (2.3)$$

and Φ is the relative phase shift between the two waves. If the reference wavefront W_r is known, it can be subtracted to recover W_t .

2.1.2 Phase-shifting point diffraction interferometry

One type of interferometry that has been successfully used to characterize EUV optics is called phase-shifting point diffraction interferometry (PS/PDI), pioneered by Medecky [17] [18]. In PS/PDI, an initial object-side spatial filter is used to illuminate the test optic with a coherent spherical wave that acquires the aberrations of the optic as it passes through. A diffraction grating effectively splits the test wave into two copies that come to focus in the image plane: one copy is incident on a small image-side spatial filter that creates the spherical reference wave $W_r(x, y)$, and the other copy passes through a large window and interferes with the reference wave at the detector. The interferogram contains an intensity pattern in the form of equation (2.2).

Since the test optic must be coherently illuminated, PS/PDI works best with a light source with a high level of coherent power, such as undulator radiation from a synchrotron sourceⁱ [19]. A schematic of a typical PS/PDI setup is shown in Figure 2.1.

2.1.3 Scalability of PS/PDI

From equation (2.3) it is clear that quality of the spherical reference wavefront $W_r(x, y)$ sets a fundamental lower bound on the error in using PS/PDI to measure $W_t(x, y)$. This quality in turn, is set by the size of the aperture that generates it. It can be shown that the maximum

ⁱThe optic in [17] was tested at beamline 12.0.1 of the Advanced Light Source at Lawrence Berkeley National Laboratory

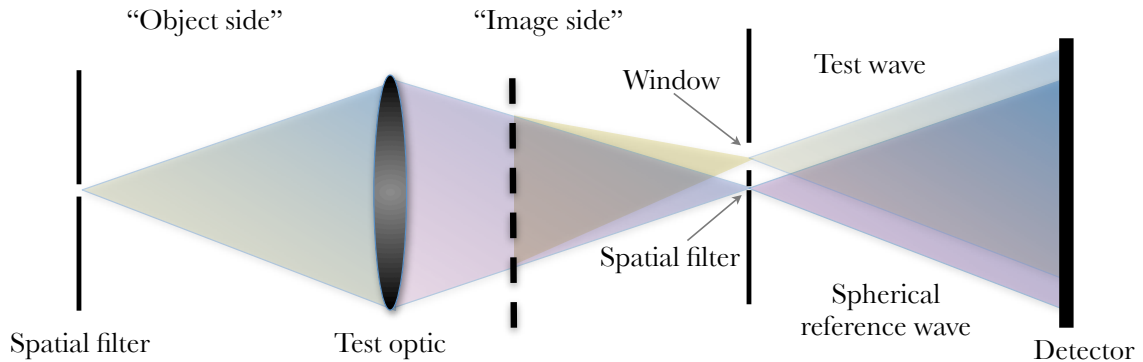


Figure 2.1: Phase-shifting point diffraction interferometry setup.

quality achievable for a spherical wave for PS/PDI can be generated by an aperture whose critical dimension is the coherent diffraction-limited resolution of the optic [20]:

$$d = 0.61 \frac{\lambda}{NA} \quad (2.4)$$

The object-side spatial filter which controls the quality of the input wave also affects the error in the measurement, but in a demagnification systemⁱⁱ the specifications on this aperture are not as strict as the object-side aperture. From (2.4), it becomes evident that for higher resolution optical systems—systems with smaller wavelengths and larger numerical apertures—the dimension of the apertures can be extremely small. Table 2.1 shows the CD of the spatial filter required to perform PS/PDI on various optical systems.

The ease of fabrication and experimental viability of spatial filters rapidly declines as the dimension of the apertures decrease, as small apertures suffer from a host of problems. From a fabrication standpoint, it is difficult to manufacture pinholes once the aspect ratio of the structure is much larger than unity. At a wavelength of 13.5 nm, the gold absorption layer would need to be over 100 nm thick in order to block 98% of the incident light in the opaque regions of the filter [21]. For a 16 nm CD as in the last column of Table 2.1, the structure would have to be 7 times as tall as it is wide.

ⁱⁱEUV lithography production tools use a 5x demagnification

Table 2.1: Critical dimensions of image-side spatial filters for PS/PDI

Optical system	λ	NA	Aperture CD
Visible light	500 nm	0.3	1 μm
DUV Lithography	193 nm	0.5	235 nm
EUV low NA	13.5 nm	0.1	82 nm
EUV high NA	13.5 nm	0.5	16 nm

On the experimental side, small apertures are difficult to locate, since the usable area decreases as the square of the critical dimension. This problem is exacerbated by the fact that small apertures operating in vacuum (as is the case for EUV systems) are prone to collecting contamination buildup, which renders the filter unusable [17]. Achieving a reasonable amount of light that passes through the aperture is also more difficult because of the small transparent area, and because the high numerical apertures commonly associated with high resolution systems result in extremely small depths of focus of the illuminating beam. Reduced flux throughput decreases the intensity of the reference wave which can make the experiment more susceptible to noise.

Due to the fact that the ease and effectiveness of PS/PDI decreases as the critical dimension of the required spatial filters decrease, the implementation of PS/PDI is highly challenging for EUV optical testing at high numerical apertures.

2.2 Overview of lateral shearing interferometry

Lateral shearing interferometry (LSI) is an alternative that scales well to high resolution optical systems because it doesn't require a small image-side spatial filter. LSI works by interfering the test wave with a copy of itself that is shifted laterally by a distance s known as the *shear*. In this case equation (2.3) reads:

$$\Delta W(x, y) = W_t(x + s, y) - W_t(x, y) \quad (2.5)$$

Since neither $W_t(x + s, y)$ nor $W_t(x, y)$ is known, the test wavefront cannot be directly

measured. However, noticing that

$$\Delta_x W(x, y) = s \cdot \frac{W_t(x + s, y) - W_t(x, y)}{s} \approx s \cdot \frac{\partial W_t(x, y)}{\partial x} \quad (2.6)$$

one sees that $\Delta_x W(x, y)$ approximates the partial derivative of $W_t(x, y)$ in the x-direction multiplied by the shear s . Here, the x-subscript on the Δ denotes a shear in the x-direction. Similarly, a shear in the y-direction can be used in a separate measurement to give the partial derivative of $W_t(x, y)$ in the y-direction:

$$\Delta_y W(x, y) = s \cdot \frac{W_t(x, y + s) - W_t(x, y)}{s} \approx s \cdot \frac{\partial W_t(x, y)}{\partial y} \quad (2.7)$$

The test wavefront $W_t(x, y)$ can then be reconstructed from its two partial derivatives using a least squares algorithm known as the Rimmer method, pioneered by Rimmer [22] in 1974, which will be discussed in section 2.4.

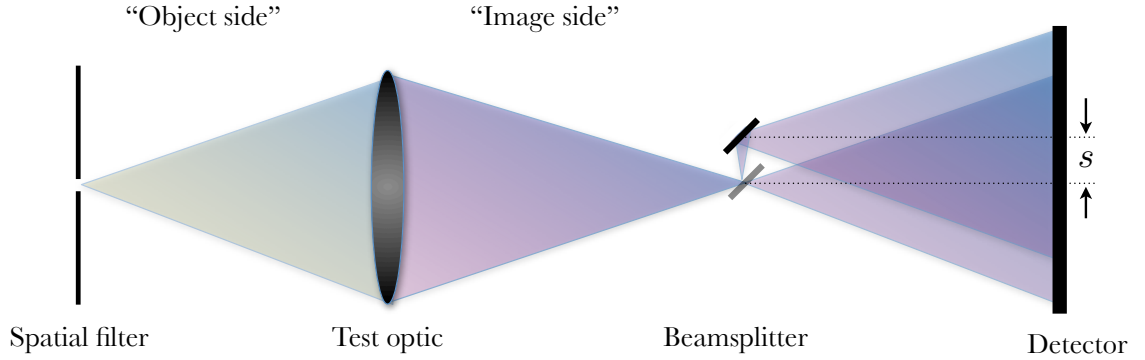


Figure 2.2: A simple realization of LSI using a beamsplitter to create the shear s .

Figure 2.2 shows a simple LSI experimental realization. In this setup, an object-side spatial filter creates a spherical wave that illuminates the test optic. A beamsplitter is then used to divide the wave into two copies that are laterally sheared by a distance s which interfere on the detector.

The light at the detector will consist of two overlapping beams. In the overlap region, the wavefront difference $W(x + s, y) - W(x, y)$ can be extracted from the fringes.

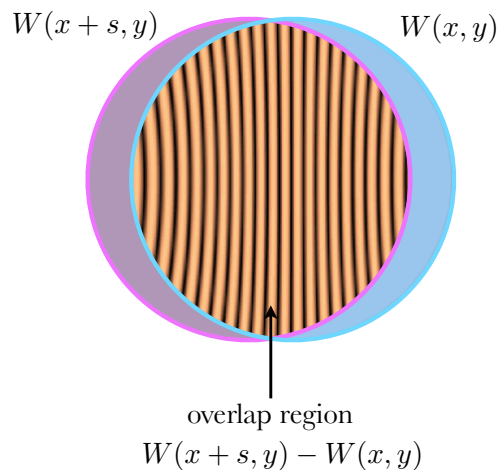


Figure 2.3: Fringes are generated in the overlap region that correspond to the difference equation for $W(x, y)$ in the direction of the shear.

2.3 Extracting the phase from the interferogram fringes

From (2.2), the fringe pattern will contain the difference wavefront as the argument of a cosine function.

$$I_T(x, y) = 2A(x)A(x+s)(1 + \cos[k\Delta W(x, y) + \Phi]) \quad (2.8)$$

There are two standard ways of extracting this argument which will be outlined briefly below. A more thorough explanation of interferogram fringe analysis can be found in [23]. For clarity, the discussion will be presented in 1-D, although the extension to 2-D is straightforward.

2.3.1 Fourier domain analysis

The fringe pattern on the detector is caused by a difference in the optical path lengths between the interfering waves at the detector as shown in Figure 2.4. If the effective source points of the two waves are laterally separated by a distance s , then the fringe peaks are given approximately by

$$\sin \theta = m \frac{\lambda}{s} \quad (2.9)$$

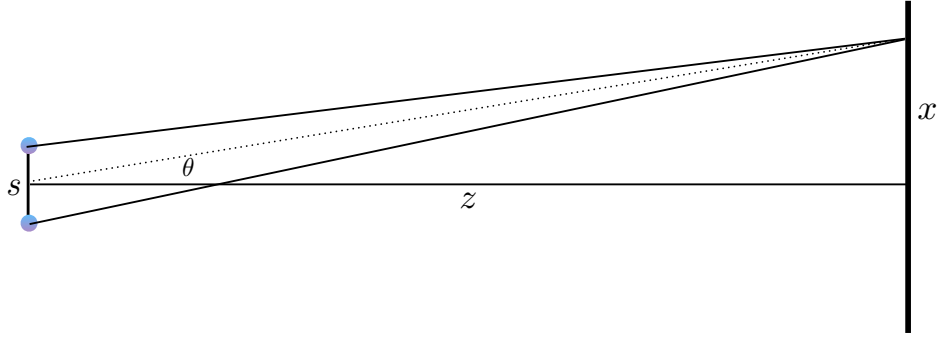


Figure 2.4: The interference between two spherical waves depends on the separation of their effective sources.

or in terms of the detector coordinates,

$$x = \frac{m\lambda z}{s} \quad (2.10)$$

where λ is the wavelength of the light and m is an integerⁱⁱⁱ. These fringes indicate the existence of a linear frequency term $s/\lambda z$ in the argument of the cosine function in (2.8). Rewriting (2.8) in terms of complex exponentials and ignoring the amplitude coefficients, we have:

$$I_T(x) = \left(1 + \frac{1}{2} \exp[ik\Delta W(x) + \Phi] + \frac{1}{2} \exp[-ik\Delta W(x) - \Phi] \right) \quad (2.11)$$

Focusing on a single exponential term, we can separate out the linear frequency term from wavefront difference term:

$$\exp \left[ik\widetilde{\Delta W}(x) + \Phi + ik\frac{sx}{z} \right] = \exp \left[ik\widetilde{\Delta W}(x) + \Phi \right] \cdot \exp \left[ik\frac{sx}{z} \right] \quad (2.12)$$

where $\widetilde{\Delta W}(x, y)$ represents the wavefront difference without its linear frequency term.

Taking the Fourier transform of the interferogram, and invoking the convolution theorem, we have:

$$\begin{aligned} \mathcal{F}\{I_T\} = & \delta(f) + \mathcal{F}\{\exp[ik\widetilde{\Delta W}(x) + \Phi]\} * [\delta(f - f_0)] \\ & + \mathcal{F}\{\exp[-ik\widetilde{\Delta W}(x) - \Phi]\} * [\delta(f + f_0)] \end{aligned} \quad (2.13)$$

ⁱⁱⁱThis solution is equivalent to the interference pattern from the classic double-slit experiment

where $f_0 = s/\lambda z$.

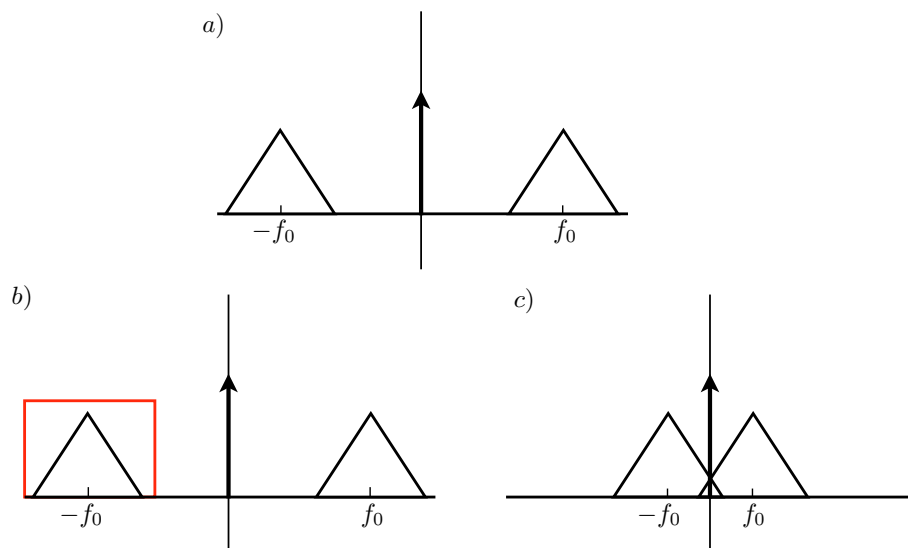


Figure 2.5: a) Spectrum of the interferogram; b) If the extent of the wavefront difference spectrum is smaller than f_0 , it can be isolated using a digital filter; c) This spectrum is not separable in the Fourier domain.

A graphical representation of (2.13) is shown in Figure 2.5. The spectrum of the wavefront difference has been put onto linear frequency carriers whose location in frequency space is determined by the number of fringes on the detector. Provided that the fringe density is sufficient to create a separation between the wavefront difference spectrum and the zeroth order peak, a digital filter can be used to isolate a single sideband. Once the sideband is isolated, the wavefront difference phase is given, up to a bulk phase term, by the argument of the exponential function:

$$k\Delta W(x) = \arg\{\exp[ik\Delta W(x)]\} \quad (2.14)$$

Because of the 2π ambiguity inherent in the \arg function, we can only recover $k\Delta W(x)$ modulo 2π . The method for undoing this modulo—the so-called *unwrapping* problem—will be discussed in section 2.5

2.3.2 Time domain analysis

If f_0 is small compared with the extent of the wavefront difference spectrum, a digital filter cannot be used to isolate the sideband. However, the argument of the cosine can still be recovered using a time domain analysis, provided that the relative phase shift Φ between the two waves in (2.8) can be controlled. In grating-based LSI discussed in section 2.6 for instance, Φ can be manipulated by translating the grating, where the phase shift is given by the ratio of the translation to the grating period.

In the time domain setup, N interferograms are acquired at distinct phase shifts Φ_n , where the Φ_n s are uniformly distributed in the interval $[0, 2\pi - 2\pi/N]$. The wavefront difference can then be recovered by standard phase shifting analysis which will be outlined briefly here, although a thorough description can be found in the literature [24].

At a given phase shift Φ_n , the interferogram intensity will have the form:

$$I_T(x, y) \sim 1 + \cos[k\Delta W(x, y) + \Phi_n] \quad (2.15)$$

where the phase shifts Φ_n are:

$$\Phi_n = \frac{2\pi n}{N}, \quad n = 0, 1, 2, \dots, N - 1 \quad (2.16)$$

Using (2.16), we can rewrite (2.15) as:

$$I_T(x, y) \sim 1 + \cos \left[\frac{\alpha + 2\pi n}{N} \right] \quad (2.17)$$

where $\alpha = Nk\Delta W(x, y)$.

Plotting (2.17) as a function of the phase shift variable n will yield a plot of one period of a cosine function shifted to the left by $Nk\Delta W(x, y)$, as shown in Figure 2.6. This phase shift can be determined by either using an $N + 1$ -point algorithm as in [25], or by taking the phase of the fundamental frequency in the time domain DFT of the signal.

As in the frequency domain analysis, there will be a modulo 2π ambiguity to the recovered wavefront difference.

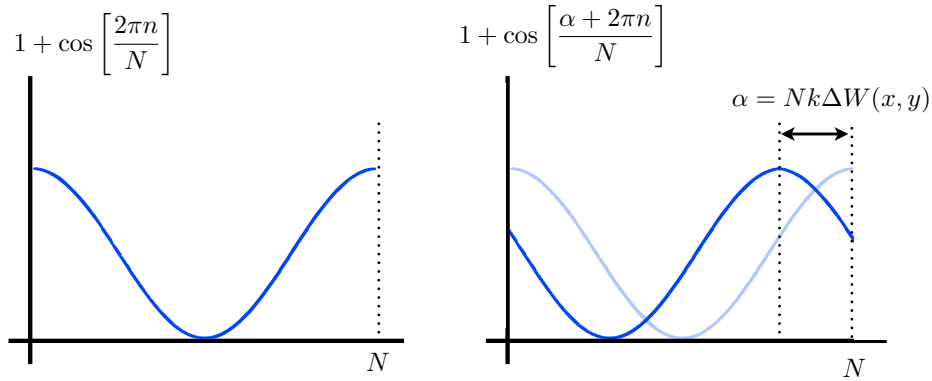


Figure 2.6: The argument of the cosine function is determined by measuring its phase shift.

2.4 The Rimmer method

The fundamental task in LSI is to reconstruct the test wavefront from its respective difference functions. Errors in the experimental data due to noise and other sources prevent a direct integration of the differences, which would lead to conflicting solutions depending on the integration path. However, the system of equations over-specifies the solution since for a grid of $N \times N$, there are N^2 unknown wavefront values, but $2N^2$ difference equations (N^2 x -difference equations and N^2 y -difference equations). A least squares approach therefore can make use of all of the data to compute the result that best fits the experimental data.

Consider the following wavefront W discretized to a 2 dimensional grid shown in Figure 2.7. The matrix elements $W_{i,j}$ correspond to the unknown values of $W(x, y)$ at the i th row and j th column of the grid. The horizontally and vertically sheared wavefronts are shown on the right. For simplicity the shear is taken to be one grid unit.

Defining the $a_{i,j}$ and $b_{i,j}$ to be the measured differences at the grid position (i, j) :

$$\begin{aligned} a_{i,j} &= W_{i,j+1} - W_{i,j} \\ b_{i,j} &= W_{i-1,j} - W_{i,j} \end{aligned} \quad (2.18)$$

An error metric can then be defined as the sum of the individual errors representing discrepancies between the trial wavefront differences and the measured differences:

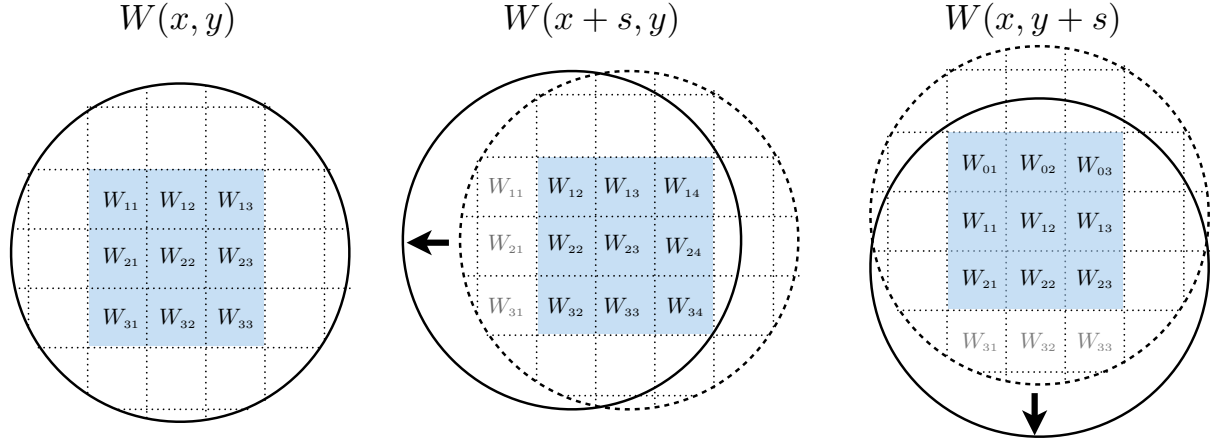


Figure 2.7: Left: original wavefront; Center: horizontally sheared wavefront; Right: vertically sheared wavefront.

$$E = \sum_{i,j}^N E_{i,j} \quad (2.19)$$

where,

$$E_{i,j} = (W_{i,j+1} - W_{i,j} - a_{i,j})^2 + (W_{i,j} - W_{i,j-1} - a_{i,j-1})^2 \\ + (W_{i-1,j} - W_{i,j} - b_{i,j})^2 + (W_{i,j} - W_{i+1,j} - b_{i+1,j})^2 \quad (2.20)$$

Differentiating with respect to each wavefront value $W_{i,j}$ and setting the result to 0 to find the minimum of E gives a system of N^2 equations, each of the form,

$$-4W_{i,j} + W_{i,j+1} + W_{i,j-1} + W_{i+1,j} + W_{i-1,j} = a_{i,j} + a_{i,j-1} + b_{i,j} + b_{i+1,j} \quad (2.21)$$

which can be written in matrix form as:

$$\mathbf{T} \cdot \vec{W} = \mathbf{A} \cdot \begin{bmatrix} \vec{a} \\ \vec{b} \end{bmatrix} \quad (2.22)$$

where \mathbf{T} and \mathbf{A} are $N^2 \times N^2$ and $2N^2 \times N^2$ matrices representing the coefficients of the terms in the LHS and RHS of equation (2.21), and the vectors \vec{W} , \vec{a} , and \vec{b} represent the unknown

wavefront points and the measured differences respectively. Equation (2.22) can be inverted to solve for the \vec{W} :

$$\vec{W} = \mathbf{T}^{-1} \mathbf{A} \cdot \begin{bmatrix} \vec{a} \\ \vec{b} \end{bmatrix} \quad (2.23)$$

2.4.1 Operations on large matrices

When Rimmer first described his least squares method for analyzing LSI interferograms in 1974, computers did not have the resources to solve problems involving large matrices. Solving the matrix equation in (2.23) requires the inversion of \mathbf{T} , which has $N^2 \times N^2 = N^4$ elements for a detector that has $N \times N$ pixels. Even a detector with a modest 200 x 200 pixels would require the inversion of a 40,000 x 40,000 matrix. Just storing a matrix of that size in memory would require over 25 GB of RAM. For reference, the Apple II computer in 1977 had only 4 kB of RAM, barely enough to run the Rimmer reconstruction on a 4 x 4 matrix.

Today (2011), we are entering a turning point in computation where computers are beginning to have the resources and speed to solve problems on the scale that is required for analyzing shearing interferograms and other large-scale matrix calculations. The development of optimized linear algebra toolsets such as LAPACK [26] have also dramatically reduced computation times. Because the feasibility of solving problems is becoming increasingly dependent on the specific algorithms that are used, it is useful to describe some of the techniques employed to facilitate these algorithms.

2.4.2 Sparse representation of matrices

Although high-end computers today can technically support the memory requirements to store reasonably large matrices (such as \mathbf{T} for the 200 x 200 pixel detector in the last section), it is often undesirable to do so. Large matrices take a long time to construct, require a long time to operate on, and often require a contiguous block of memory which may only be a fraction of the total available memory.

Many modern mathematical software tools such as MATLAB contain many linear algebra

Full representation					Sparse representation	
	1	2	3	4		
1	0	4	0	0	$[1,2] = 4$	
2	1	0	0	0	$[2,1] = 1$	
3	0	0	0	6	$[3,4] = 6$	
4	0	1	0	0	$[4,2] = 1$	

Figure 2.8: Sparse matrix representation vs. full matrix representation. The full representation requires 16 numbers to store the matrix whereas the sparse representation only needs 12 (8 indices and 4 numbers).

routines that support sparse representations of matrices. A *sparse* matrix is a matrix that consists primarily of zeros, so that it is generally cheaper from a memory perspective to represent it as a list of the values and locations of its nonzero entries — known as its *sparse representation* — than in the conventional representation where the zeros are also stored. An example of a sparse representation of a matrix is shown in Figure 2.8. Physically, a sparse matrix represents a system that is weakly coupled, meaning that each element in the system is only affected by a handful of other elements. The least squares problem in the Rimmer method is one such system, since each wavefront point $W_{i,j}$ is a function only of its four nearest neighbors. A graphical representation of T and A is shown in Figure 2.9 illustrating their sparsity.

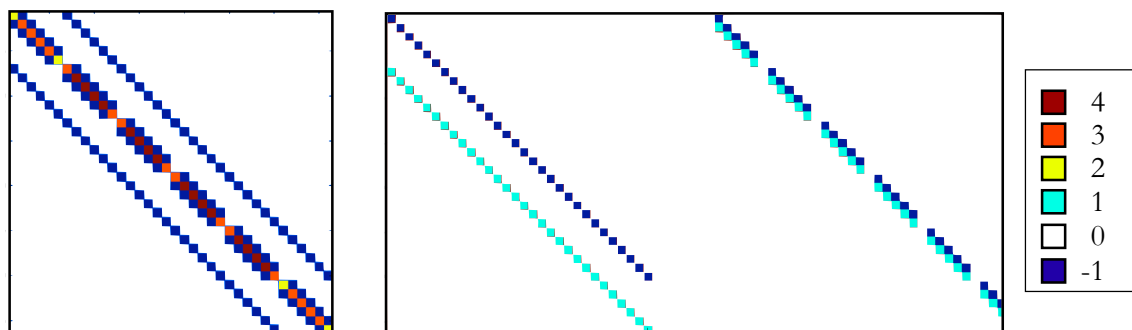


Figure 2.9: A graphical representation of the matrices T (left) and A (right) for an interferogram with 6 x 6 pixels.

The memory savings for highly sparse matrices is substantial; for instance, a sparse representation of T from an interferogram with 100 x 100 pixels is nearly 1200 times smaller than its full representation. Operations on sparse matrices such as eigenvalue decomposition, gaussian elimination, and inversion can also be many times faster than the equivalent operation on a full matrix.

2.4.3 Sparse matrix construction

Often in programs such as MATLAB, where each line of code is interpreted rather than compiled, there is a large computational overhead involved in populating sparse matrices such as T and A when looping through each node and summing up the interactions. This overhead can be occasionally avoided by saving the variable associated with matrices to a file, and recalling it each time the calculation is to be performed. However, since T and A are different for different geometries (e.g. different detector sizes), new versions must be constructed each time a parameter changes in the experiment. This construction, if not optimized, can readily bottleneck the analysis workflow.

The preferred method for constructing sparse matrices is to employ built-in primitive sparse matrices in conjunction with built-in sparse matrix manipulations to piece together the desired matrix by visual inspection. Many programs like MATLAB have built-in sparse matrices — such as the identity matrix — and built-in operations that are some version of the following operations:

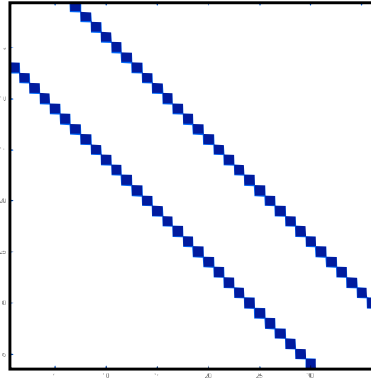
- `shift(N)`: shifts a sparse matrix down N rows
- `∩`: takes the intersection of two sparse matrices
- `∪`: takes the union of two sparse matrices
- `+`: adds two sparse matrices
- `dropRow(N)`: Sets row N to 0.
- `dropRowN(N,M)`: Sets every N th row to 0 starting from the M th row.

- $\text{diag}(\text{vec}, N)$: Takes a $1 \times M$ vector and makes it the $(N + |N - M|)$ -th diagonal of a matrix where all other elements are zero

As an example, let's construct the sparse representation of T for an $N \times N$ pixel interferogram using the built-in matrices and operations. For clarity the illustrations will assume $N = 6$.

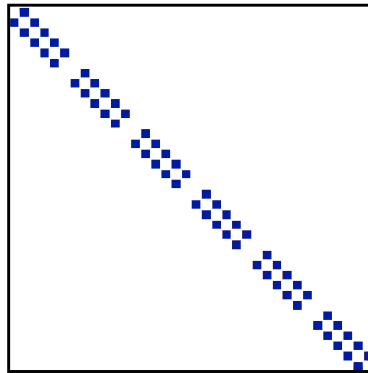
1. Create the outside diagonals:

$$OD = -\text{shift}(\text{identity}, N) \cup \text{shift}(\text{identity}, -N)$$



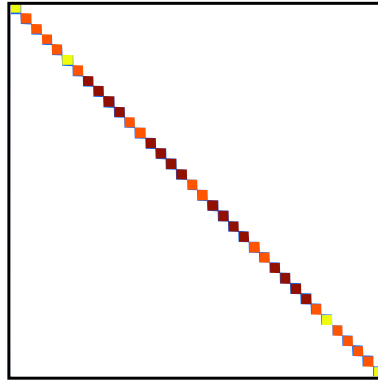
2. Create the inside diagonals:

$$ID = -\text{shift}(\text{dropRowN}(\text{identity}, N), 1) \cup \text{shift}(\text{dropRowN}(\text{identity}, N), -1)$$

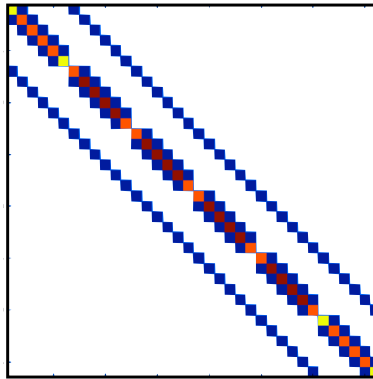


3. Create the diagonal:

$$D = \text{dropRowN}(\text{dropRowN}(\text{identity}, N, 1), N, N - 1) + \\ \text{dropRow}(\text{identity}, 1, \dots, N, N^2 - N + 1, \dots, N) + \\ 2 \times (\text{identity})$$



4. Then combine: $\mathbf{T} = D \cup ID \cup OD$



Although this process was done for the Rimmer matrix on an $N \times N$ pixel detector, it is applicable in countless other situations and even in cases that lack symmetry. All that is required is a recognizable pattern for the nonzero entries of the matrix and a little creativity. Constructing matrices in this manner can save enormous amounts of time over conventional looping; in the case of generating \mathbf{T} for a 512×512 pixel detector for instance, pattern construction can be tens of thousands of times faster.

2.4.4 Weighted coefficients

The last section in the discussion on the Rimmer method focuses on an improvement to Rimmer's original least squares method. There are times where there is a priori information regarding the reliability of the data. For instance, a certain pixel that measures low signal compared with the rest of the image will have a relatively larger uncertainty since it is more

susceptible to Poisson noise. Naturally, if it is known beforehand that certain data points are less reliable than others, one would like to suppress their relative weight in computing the merit function. This can be accomplished by simply inserting coefficients to each error term in equation (2.24)

$$E_{i,j} = \alpha_{i,j}(W_{i,j+1} - W_{i,j} - a_{i,j})^2 + \alpha_{i,j-1}(W_{i,j} - W_{i,j-1} - a_{i,j-1})^2 \\ + \beta_{i,j}(W_{i-1,j} - W_{i,j} - b_{i,j})^2 + \beta_{i+1,j}(W_{i,j} - W_{i+1,j} - b_{i+1,j})^2 \quad (2.24)$$

where $\alpha_{i,j}$ and $\beta_{i,j}$ are coefficients between 0 and 1 that represent the respective weight of the measured differences $a_{i,j}$ and $b_{i,j}$. Differentiating to minimize the error, we have:

$$-(\alpha_{i,j} + \alpha_{i,j-1} + \beta_{i,j} + \beta_{i+1,j})W_{i,j} \\ + \alpha_{i,j}W_{i,j+1} + \alpha_{i,j-1}W_{i,j-1} + \beta_{i,j}W_{i+1,j} + \beta_{i+1,j}W_{i-1,j} \\ = \alpha_{i,j}a_{i,j} + \alpha_{i,j-1}a_{i,j-1} + \beta_{i,j}b_{i,j} + \beta_{i+1,j}b_{i+1,j} \quad (2.25)$$

The matrix equation is then:

$$\mathbf{T}' \cdot \vec{W} = \mathbf{A}' \cdot \begin{bmatrix} \vec{a} \\ \vec{b} \end{bmatrix} \quad (2.26)$$

where \mathbf{T}' and \mathbf{A}' are the least squares matrices that account for the weighted coefficients.

Generally one desires a mapping from a reliability metric of the data to the coefficient that is strong enough to resist large errors from noise, and yet not so aggressive as to throw away too much useful information. For the experiments in this dissertation, the following parametrized weighting function was used:

$$\Omega : [0, 1] \rightarrow [0, 1] \\ \Omega(x; a, b, c) = c + (1 - c) \frac{\tan^{-1}[b(x - a)] - \tan^{-1}(-ab)}{\tan^{-1}[b(1 - a)] - \tan^{-1}(-ab)} \quad (2.27)$$

Here, $a \in (0, 1)$ determines the position of the knee, $b \in (0, \infty)$ is an inflection parameter that controls the abruptness of the knee, and $c \in (0, 1)$ is the initial value. The input metric

x usually corresponded to the fractional intensity of the pixel in question $x = I_{i,j}/I_{max}$, as smaller fractional intensities correspond to lower signal to noise. The parameters (a, b, c) were tuned primarily through trial and error. Figure 2.10 shows a plot of Ω for $(a, b, c) = (0.3, 30, 0.2)$.

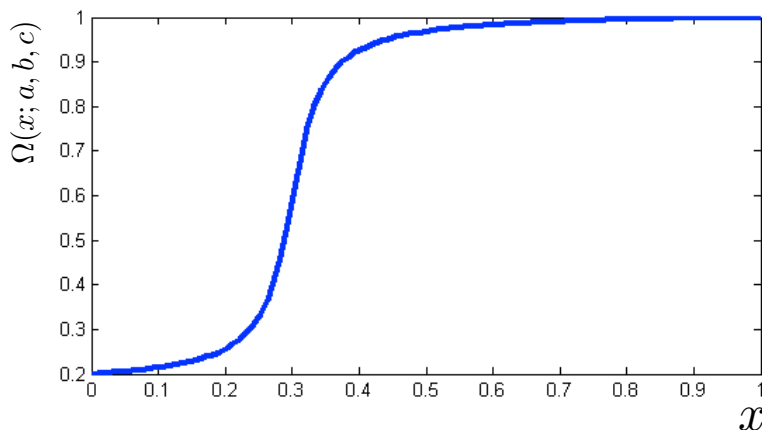


Figure 2.10: A plot of the weight function $\Omega(x; a = 0.3, b = 30, c = 0.2)$.

2.5 Phase unwrapping

In nearly every interferogram analysis, the issue arises that the desired wavefront information is buried inside the argument of a periodic function. Periodic functions such as $\exp\{ix\}$ are not injective^{iv} by definition since they are invariant under a shift by the period T :

$$\exp\{i(x + Tk)\} = \exp\{ix\}, \quad \text{where, } T = 2\pi, k \in \mathbb{Z} \quad (2.28)$$

Therefore, a proper inverse function cannot exist unless we restrict the range to a specific interval such as $[0, 2\pi)$. This has the unwanted effect of mapping the true wavefront $W(x, y)$ to a so-called *wrapped* wavefront $\psi(x, y)$:

^{iv}An function is called *injective* if it never maps distinct elements in the domain to the same element in the codomain.

$$\begin{aligned} \mathbf{w} : \mathbb{R} &\rightarrow [0, 2\pi) \\ \mathbf{w}\{W(x, y)\} &= \text{mod}\{W(x, y), 2\pi\} \equiv \psi(x, y) \end{aligned} \quad (2.29)$$

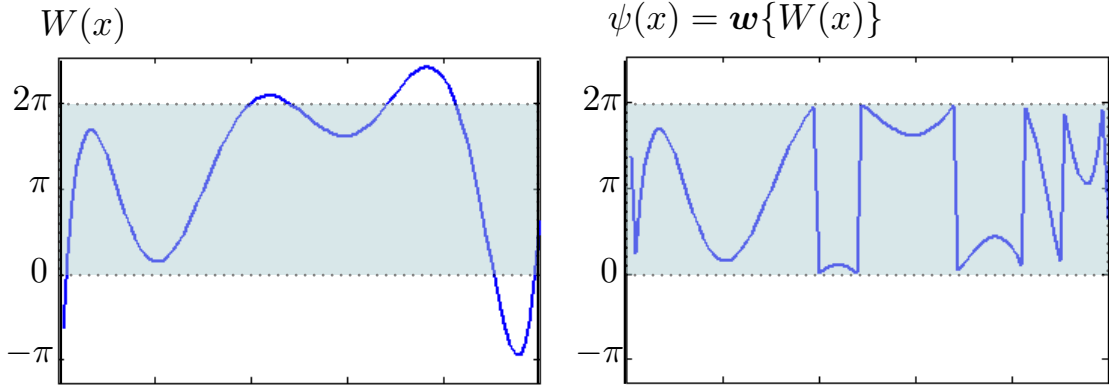


Figure 2.11: The wrapping operator \mathbf{w} maps the range of $W(x)$ into the interval $[0, 2\pi)$.

Figure 2.11 shows a sample wavefront $W(x)$ and its corresponding wrapped function $\psi(x)$. Since ψ is the experimentally measured value, the task becomes to *unwrap* ψ to recover the wavefront W . One might expect that W could be recovered by simply separating continuous regions bordered by 2π step-discontinuities and then adding or subtracting multiples of 2π until they all line up. Such a method would work in the ideal case, however, inconsistencies arise when noise corrupts the signal in a way that makes it unclear whether the phase is actually wrapping or merely changing quickly. Such errors create a scenario where multiples of 2π cannot be added to each point over a 2-dimensional grid to eliminate discontinuities between all adjacent points. Additionally, such an approach does not include valuable information about the reliability of the data points; a low signal at a certain pixel for instance can indicate that the data derived from that pixel is less reliable.

The solution is to create an estimate of W that when wrapped, most closely matches the experimentally measured ψ in the least squares sense. This idea, proposed first by Ghiglia and Romero [27], closely resembles the Rimmer method in the previous section. However, as in the case of the Rimmer method, we can improve the least squares matrix by incorporating coefficients to the error terms that correspond with our confidence in each datapoint. Figure

2.12 illustrates the difference between unwrapping a function by the least squares approach and by a step-discontinuity matching approach. In this example the noise causes inconsistencies that prevent the step-discontinuity matching approach from finding a continuous phase function over the analysis region.

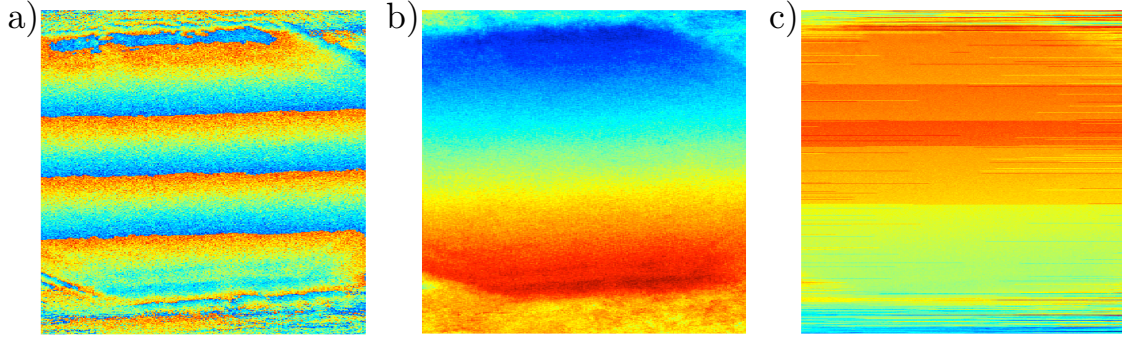


Figure 2.12: a) An experimentally measured wrapped phase function in the presence of noise, b) the function unwrapped by the least squares approach, c) the function unwrapped by a step-discontinuity matching approach.

To formulate the least squares matrix equation, we define the condition for the solution $W_{i,j}$ as follows:

Differences between adjacent points on W should be the same as the differences between adjacent points on the wrapped function ψ if the two points do not straddle a discontinuity of ψ . If the two points do in fact straddle a discontinuity of ψ , then the difference between the adjacent on W should be equal to the difference between adjacent points on ψ modulo 2π .

Defining,

$$\begin{aligned} c_{i,j} &= \mathbf{w}\{\psi_{i,j+1} - \psi_{i,j}\} \\ d_{i,j} &= \mathbf{w}\{\psi_{i-1,j} - \psi_{i,j}\} \end{aligned} \quad (2.30)$$

where \mathbf{w} is the wrapping operator defined in (2.29), and $\psi_{i,j}$ is the wrapped wavefront, we can write the condition out in terms of the following error function:

$$\begin{aligned}
E_{i,j} = & (\psi_{i,j+1} - \psi_{i,j} - c_{i,j})^2 + (\psi_{i,j} - \psi_{i,j-1} - c_{i,j-1})^2 \\
& + (\psi_{i-1,j} - \psi_{i,j} - d_{i,j})^2 + (\psi_{i,j} - \psi_{i+1,j} - d_{i+1,j})^2
\end{aligned} \tag{2.31}$$

Similar to the Rimmer case, reliability coefficients from a weighting function can be added to adjust the impact of each error term according to the reliability of each data point.

$$\begin{aligned}
E_{i,j} = & \gamma_{i,j}(\psi_{i,j+1} - \psi_{i,j} - c_{i,j})^2 + \gamma_{i,j-1}(\psi_{i,j} - \psi_{i,j-1} - c_{i,j-1})^2 \\
& + \zeta_{i,j}(\psi_{i-1,j} - \psi_{i,j} - d_{i,j})^2 + \zeta_{i,j}(\psi_{i,j} - \psi_{i+1,j} - d_{i+1,j})^2
\end{aligned} \tag{2.32}$$

Evidently, the error metric is identical to the Rimmer method error metric in (2.24). This means that the same least squares matrices \mathbf{T}' and \mathbf{A}' that compute a function given its difference functions can be used to solve the phase unwrapping problem. The least squares equation is then:

$$\mathbf{T}' \cdot \vec{W} = \mathbf{A}' \cdot \begin{bmatrix} \vec{c} \\ \vec{d} \end{bmatrix} \tag{2.33}$$

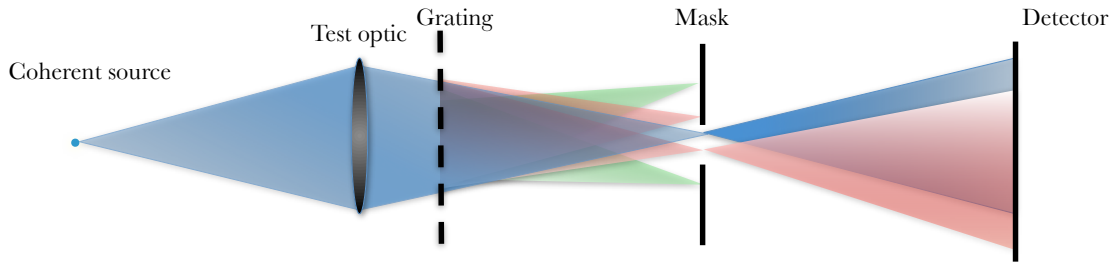
2.6 LSI with a diffraction grating

In the simple LSI schematic shown earlier in Figure 2.2, a beamsplitter was used to create the two sheared copies of the test wavefront. However, beamsplitters are not a viable solution for short wavelength systems due to the lack of materials with suitable mechanical and optical properties. Consequently, LSI setups for EUV optical systems typically use diffractive elements to create the sheared copies.

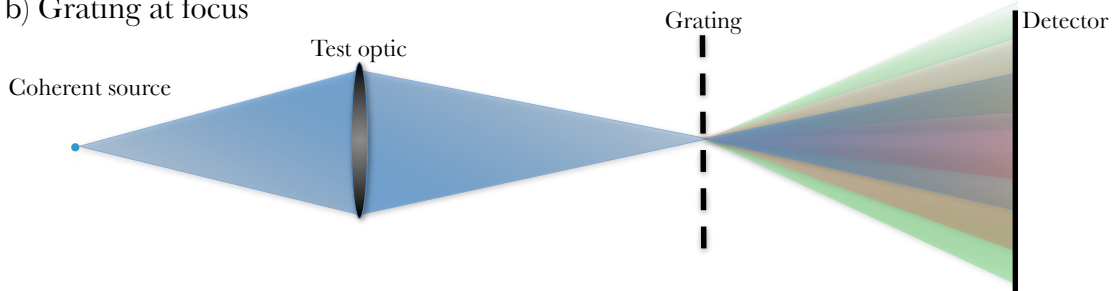
Three typical configurations for LSI on short wavelength systems are shown in Figure 2.13. Each setup has its advantages and disadvantages which will be discussed at the end of the next chapter following the discussion of systematic aberrations. In each setup, the pitch of the grating is chosen to create the desired shear, and the distance to the detector is a function of the detector size and the numerical aperture of the test optic.

There are two very important differences between the implementations of LSI with a beamsplitter and with a grating. The first difference has to do with the nature of the shear

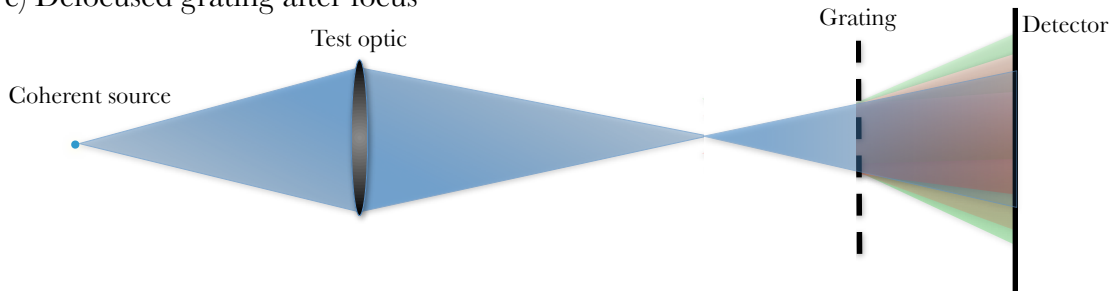
a) Defocused grating before focus with order sorting mask



b) Grating at focus



c) Defocused grating after focus

**Figure 2.13:** Three implementations of LSI using a diffraction grating.

itself. In the case of the beamsplitter setup, both the original wave and the sheared copy propagate in the same direction. In the case of the grating setup, the sheared wave is propagating at an angle defined by diffraction off of the grating. In other words, the beamsplitter creates a *lateral* shear, whereas the grating creates an *angular* shear.

The second difference stems from the nonlinear angular spectrum that the grating imparts on the diffracted wave. Rays that are incident on the grating at an angle θ_i will exit the grating in the first diffracted order at an angle θ_f given by the grating equation:

$$\theta_f = \sin^{-1}\left(\sin \theta_i + \frac{\lambda}{T}\right) \quad (2.34)$$

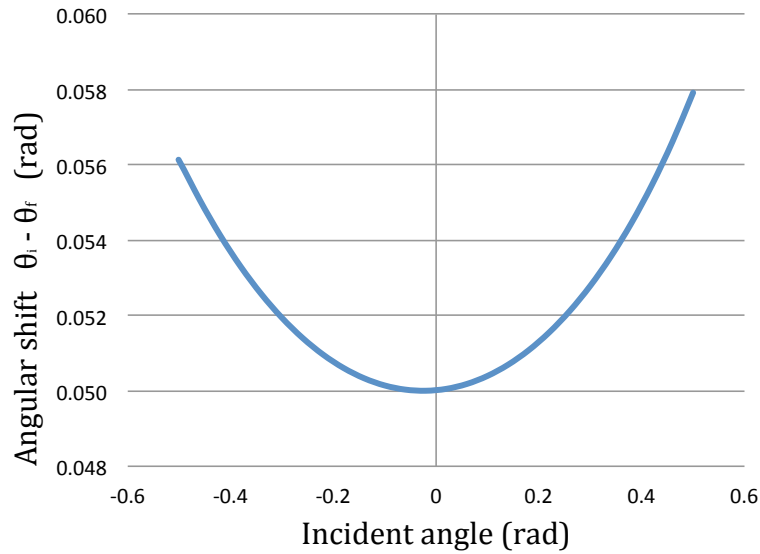


Figure 2.14: Angular shift between the incident ray and the 1st order diffracted ray as a function of incident angle for $\lambda/T = .05$.

A plot of the angular shift $\Delta\theta = \theta_f - \theta_i$ is shown in Figure 2.14. Since a spherical wave contains many different angles, each angle will be diffracted by a different amount. The result is that the shape of the spherical wave in the diffracted order becomes distorted. This is important because LSI is built on the premise that the detector is measuring the difference between two identical, laterally separated waves,

$$\Delta W(x, y) = W_t(x + s, y) - W_t(x, y) \quad (2.5)$$

where this is clearly not the case for grating-based LSI. Nevertheless, the interference can still be characterized if we include an additional path length term $\Lambda(x, y)$ that accounts the distortion to the sphere:

$$\Delta W(x, y) = [W_t(x + s, y) + \Lambda(x, y)] - W_t(x, y) \quad (2.35)$$

If $\Lambda(x, y)$ can be accurately characterized, then it can be accounted for in the interferogram analysis.

2.7 Experimental specifications

2.7.1 Independent parameters

In a typical LSI setup, assuming a specified wavelength λ and numerical aperture NA , and a fixed detector width X , there are 2 independent parameters to be specified:

1. Shear percentage p :

The shear percentage p defined as the ratio of shear s to beam diameter D at the detector determines the sensitivity of the measurement. A small shear value measures the wavefront difference over only a small distance and thus has low sensitivity in detecting wavefront slopes. In the presence of noise this measurement will be more susceptible to errors than a larger shear value. On the other hand, while a large shear value has greater sensitivity to small wavefront derivatives, high frequency wavefront terms may be aliased. Additionally, since data is available only in the overlap region between the interfering waves, the effective numerical aperture of the measurement is decreased by a factor of $(1 - p)$ shown in Figure 2.15

$$NA_{analysis} = NA_{optic}(1 - p) \quad (2.36)$$

2. Number of detector fringes N :

The number of fringes N on the detector determines the fringe density which affects the location of the sideband in the interferogram spectrum. In general, at least 30 fringes are needed to achieve adequate separation between the zeroth order peak and the first order sideband which is required for the Fourier domain analysis to work. Interferograms with less than 30 fringes must be analyzed using the time domain analysis. While increasing the fringe density gives better Fourier domain separation, too high a fringe density may run into aliasing on the detector if the pixel density cannot support the fringe modulation. Additionally, increasing the fringe density requires moving the grating out of focus, which can cause larger systematic errors that will be discussed in the next chapter. For a detector with 1024×1024 pixels, it is generally desired to

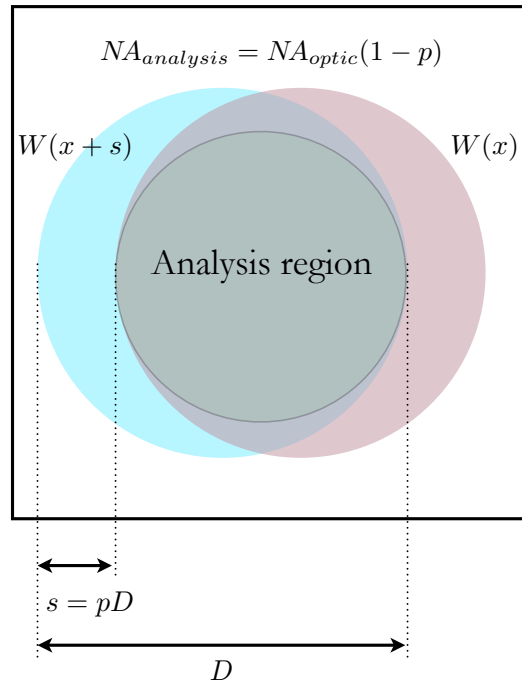


Figure 2.15: The analysis region has a numerical aperture defined by the overlap between the interfering waves.

have no more than 100 fringes on the detector, giving a sampling of 10 pixels per fringe period.

2.7.2 Dependent parameters

Here we outline the dependent experimental parameters with Figure 2.16 as a reference. For clarity, we assume a low numerical aperture regime where small angle approximations are valid, i.e., $\sin \theta = \tan \theta = \theta$. However, the same analysis applies to the general case if all trigonometric functions are retained.

1. T : The grating pitch is set by the percent shear p . Since the first order diffracted angle is given by:

$$\theta = \sin^{-1} \frac{\lambda}{T} \quad (2.37)$$

and the angular range of the nominal wave is:

$$\Delta\theta = 2 \sin^{-1} NA \quad (2.38)$$

The percent shear p can be written as,

$$p = \frac{\sin^{-1} \frac{\lambda}{T}}{2 \sin^{-1} NA} \approx \frac{\lambda}{2T NA} \quad (2.39)$$

which determines the grating pitch:

$$T = \frac{\lambda}{2p NA} \quad (2.40)$$

2. z_1 : The distance between the test optic focus and the grating position is set by the number of fringes N that are desired on the detector. The amount of defocus sets the virtual point source separation d of the diffracted orders:

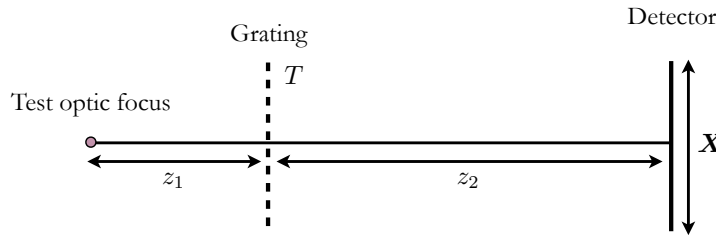


Figure 2.16: Illustration of the dependent parameters in the LSI experimental setup.

$$d = \tan \left[\sin^{-1} \frac{\lambda}{T} \right] z_1 \quad (2.41)$$

The fringe locations at the detector will then be given by:

$$d \sin \theta = n\lambda \quad (2.42)$$

The order of the fringe at the edge of the detector is given by:

$$\tan \left[\sin^{-1} \frac{n_{max} \lambda}{d} \right] = \frac{X/2}{z_2} \quad (2.43)$$

where $n_{max} = N/2$ is the highest order fringe captured on the detector. Noting $NA \approx X/2z_2$, z_1 is given by^v:

$$z_1 = \frac{NT}{2NA} \quad (2.44)$$

3. z_2 : The distance from the test optic focus to the detector is constrained by the beam divergence. The diameter of the beam footprint at the detector is given by:

$$D = 2z_2 \tan [\sin^{-1} NA] \approx 2z_2 NA \quad (2.45)$$

Requiring that the beam diameter is less than the detector width: $D < X$ gives:

$$z_2 < \frac{X}{2NA} \quad (2.46)$$

2.8 The null interferogram

So far the discussion has centered on the recovery of the test wavefront $W_t(x, y)$, since it is the aberrations in this wavefront that compromise the ultimate resolution of an optical system. In a perfect optical system that is free from aberrations, the test wavefront would take the form of a perfect spherical wave:

$$W_t(x, y) = S(x, y) = kr(x, y) \quad (2.47)$$

Since the aberrations in the wavefront are *deviations* from the ideal wavefront, it is convenient to write the test wavefront as a sum of a spherical wave, sometimes referred to as the *base sphere*, and the aberrations $A(x, y)$:

^vThe formulation of z_1 as a function of the number of detector fringes N also gives the number of illuminated grating fringes as N ; we therefore conclude that the number of fringes on the detector is the same as the number of fringes illuminated.

$$W_t(x, y) = S(x, y) + A(x, y) \quad (2.48)$$

Plugging (2.48) into (2.35), we have:

$$W(x, y) = \{A(x + s, y) - A(x, y)\} + \{S(x + s, y) - S(x, y) + \Lambda(x, y)\} \quad (2.49)$$

The first bracketed term is the difference function of the aberrations alone. Since solving for $A(x, y)$ is the ultimate goal, it is this function on which we will perform the interferogram analysis. The second bracketed term is known as the *null interferogram* since it represents the fringe pattern of a perfect spherical wave (i.e. no aberrations):

$$N(x, y) = S(x + s, y) - S(x, y) + \Lambda(x, y) \quad (2.50)$$

Subtracting out the null interferogram from the interferogram allows us to isolate the aberration difference $[A(x + s, y) - A(x, y)]$. Performing the LSI reconstruction without removing $N(x, y)$ will cause systematic aberrations in the wavefront measurement since the interferogram contains the geometrical path length term $\Lambda(x, y)$ which will be mistaken for real wavefront aberrations. A numerical simulation of the total rms systematic aberrations contributed by the failure to remove the null interferogram is shown in Figure 2.17. For low numerical apertures less than 0.15, the systematic error is less than 10 mWaves rms ($\lambda/100$), which is smaller than the error bound on most EUV interferometric tests. As a result, we can safely ignore the effect of the null interferogram in this regime. In the medium to high NA regimes ($NA > 0.15$), failure to remove the null interferogram will create measurable errors in the wavefront reconstruction.

The analysis of the null interferogram $N(x, y)$ in various LSI configurations is the subject of the next chapter. As the numerical aperture of the system increases, $N(x, y)$ becomes increasingly difficult to determine analytically, and we must resort to increasingly complex methods to figure it out. This analysis is crucial to the accuracy and viability of LSI because our ability to correctly remove $N(x, y)$ sets an ultimate limit on the wavefront measurement.

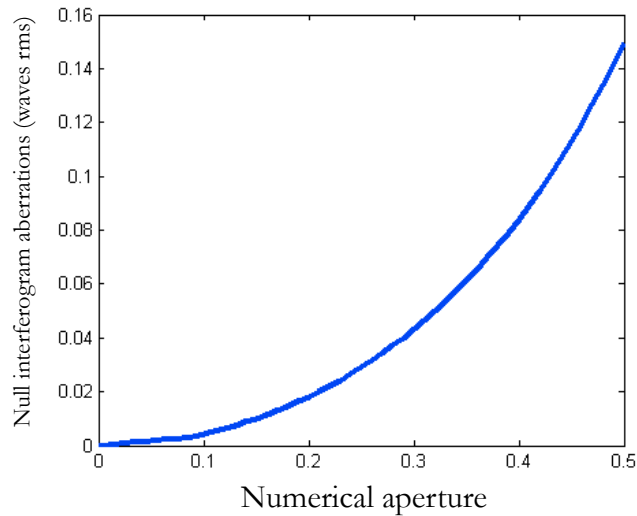


Figure 2.17: Total rms error present in the null interferogram versus numerical aperture. The simulation used $\lambda = 13.5$ nm, 5% shear, 40 fringes at the detector, and a 1 inch format detector.

2.9 Summary

We summarize the discussion in this chapter by outlining the workflow of shearing interferometry:

1. Obtain 2 sets of interferograms with the shear in the x - and y - directions.
2. Extract the phase from the fringes using the Fourier domain analysis or time domain analysis.
3. Unwrap the phase using the least squares phase unwrapping technique
4. Subtract out the null interferogram to obtain the aberration difference function $A(x + s, y) - A(x, y)$.
5. Perform the Rimmer method on the difference function to recover $A(x, y)$.
6. Use knowledge of $A(x, y)$ to align the optical system.
7. Repeat process until $A(x, y)$ is within tolerance.

Chapter 3

Lateral shearing interferometry, part 2

Contents

3.1	Null interferogram at medium NA	44
3.1.1	Grating as a hologram	44
3.1.2	Functional form of the 1st order diffracted sphere	46
3.1.3	Functional form of the null interferogram	46
3.2	System alignment and the null interferogram at medium NA	48
3.2.1	Functionalizing $N(x)$ along z	49
3.3	Null interferogram at high NA	52
3.3.1	The 2-ray method	53
3.4	Variable shear	56
3.4.1	Adapting the Rimmer method to variable shear	58
3.5	Discussion	60
3.5.1	Simulation results	60
3.5.2	Summary	61

It was shown in the previous chapter that in order to make LSI a reliable procedure for aligning EUV systems, it is necessary to characterize and remove the null interferogram $N(x, y)$ from the shearing interferogram. At low numerical apertures ($NA < 0.15$), the

effect of $N(x, y)$ is small and can be neglected without incurring significant errors in the wavefront measurement.

This chapter will present a mathematical investigation of $N(x, y)$ in the medium ($0.15 < NA < 0.35$) and high ($NA > 0.35$) numerical aperture regimes. The medium numerical aperture regime is discussed first, where a novel holographic approach derives an analytical solution to $N(x, y)$ at the detector. This regime is compatible with a defocused grating configuration, and thus the analysis is extended to consider the effects of grating and detector tilt. The second part of the chapter deals with the high numerical aperture regime where the holographic analysis breaks down. In this regime, a geometrical ray analysis is used to predict the form of $N(x, y)$ that cannot be written in closed form, but can be solved numerically. Subsequent numerical simulations of the Huygens-Fresnel propagation integral confirm the solution. A summary of the three regimes is shown in Table 3.1

Table 3.1: Numerical aperture regimes for LSI

Regime	NA range	Type of analysis
Low NA	0 – 0.15	No significant systematic aberrations
Medium NA	0.15 – 0.35	Holographic analysis
High NA	> 0.35	Full geometric ray analysis

To facilitate the clarity of this chapter, the analysis will be presented in 1-dimension; i.e. the single variable null interferogram $N(x)$ will represent the a slice through the 2-dimensional function $N(x) = N(x, y = 0)$. However since the ultimate goal of analyzing the path length error is to correct experimentally relevant wavefronts which are inherently 2-dimensional, the full solution will be presented and discussed as appropriate. The analysis in this chapter will consider a defocused grating configuration where the grating is placed downstream of the test optic focus. This setup is the most general case, and its solution can be readily applied to the other LSI configurations.

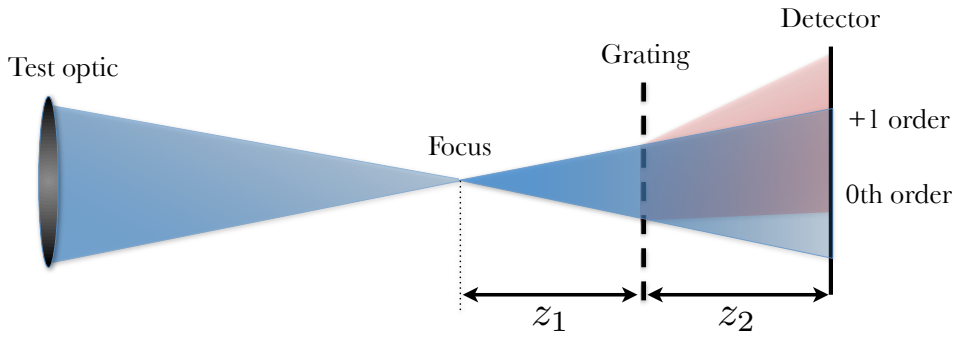


Figure 3.1: LSI configuration: A “leaky prism” grating placed a distance z_1 after the focus of the test optic. A detector located a distance z_2 after the grating capture the two diffracted orders.

3.1 Null interferogram at medium NA

We take the transmission of the diffraction grating to have the form:

$$t_g(x) = 1 + \exp\left(\frac{2\pi i x}{T}\right) \quad (3.1)$$

where T is grating pitch. This grating is a so-called “*leaky prism*,” that contains just two diffracted orders, the 0 and the +1, which have the same amplitude. Such a grating is easier to work with mathematically, and its solution extends easily to a conventional binary amplitude grating.

3.1.1 Grating as a hologram

The general form of the electric field amplitude at the grating is the sum of the base spherical wave with the aberrations of the test optic:

$$U_g(x) = \exp\left(\frac{2\pi i r(x)}{\lambda} + iA(x)\right) \quad (3.2)$$

where $r(x)$ is the distance between the focal spot and the coordinate x , given by:

$$r(x) = \sqrt{x^2 + z^2} \quad (3.3)$$

In general, the aberrations will contribute only a tiny fraction to the overall curvature of the wave. A total rms wavefront error of $\lambda/5$, for instance, would be considered to be

a highly aberrated wavefront; however, the difference between the base sphere and a plane can be thousands of times larger. Therefore, it is safe to conclude that the effect of the grating distortion on the wavefront aberrations contribute negligibly to $N(x)$, as they will be overpowered by the grating distortion of the base sphere. As a result, we set $A(x)$ to 0 in this analysis.

The field at the detector $U_f(x)$ will be the Huygens-Fresnel propagation [20] of the field at the grating modulated by the grating transmission:

$$U_f(x) = \frac{z_2}{i\lambda} \int_{-\infty}^{\infty} dx U_g(x) t_g(x) \cdot \frac{\exp 2\pi i/\lambda}{r(x)^2} \quad (3.4)$$

Equation (3.4) can be computed numerically by a computer, but cannot be solved in closed form. An alternative to this approach is to treat the grating as a hologram recorded by the interference between the spherical zeroth order acting as a “reference” wave, and the unknown first order acting as the “object wave.”

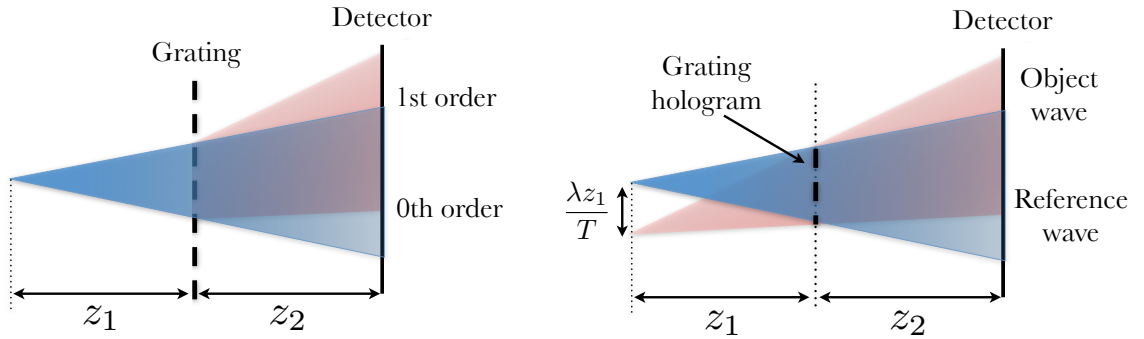


Figure 3.2: Left: A spherical wave diffracts off of a grating. Right: treating the grating as a hologram formed by a spherical wave and an object wave laterally displaced by $\lambda z_1/T$.

A diagram illustrating this approach is shown in Figure 3.2. Here we treat the first order wave to be a spherical wave diverging from a virtual source point laterally displaced a distance $d = \lambda z_1/T$ from the optical axis, summed with a path length error term $P_g(x)$.

$$W_1(x) = \exp ik[r(x + d) + P_g(x)] \quad (3.5)$$

3.1.2 Functional form of the 1st order diffracted sphere

By construction, the phase difference between the two interfering waves in the grating plane must match the phase of the grating transmission function in (3.1).

$$k[r(x+d) + P_g(x) - r(x)] = \frac{2\pi x}{T} \quad (3.6)$$

The path length term $P_g(x)$ can now be solved in closed form:

$$P_g(x) = \frac{\lambda x}{T} - [r(x+d) - r(x)] \quad (3.7)$$

To obtain a better understanding of the functional form of $P_g(x)$, it is useful to write the bracketed terms in a binomial expansion:

$$\begin{aligned} [r(x+d) - r(x)] &= \sqrt{(x+d)^2 + z_1^2} - \sqrt{x^2 + z_1^2} \\ &= z_1 \left[\sqrt{1 + \frac{(x+d)^2}{z_1^2}} - \sqrt{1 + \frac{x^2}{z_1^2}} \right] \\ &= z_1 \left[\left(1 + \frac{(x+d)^2}{2z_1^2} - \frac{(x+d)^4}{8z_1^4} + \dots \right) - \left(1 + \frac{x^2}{2z_1^2} - \frac{x^4}{8z_1^4} + \dots \right) \right] \end{aligned} \quad (3.8)$$

At medium NA, each sum is well approximated by its first three terms. Simplifying, the path length P_g is given as,

$$P_g(x) \approx C + \left(\frac{d}{z_1} - \frac{3d^3}{2z_1^3} \right) x + \left(\frac{3d^2}{4z_1^3} \right) x^2 - \left(\frac{d}{2z_1^3} \right) x^3 \quad (3.9)$$

where C is a bulk phase term that does not manifest experimentally. Equation (3.9) represents the aberrations present in the first order diffracted wave. The cubic term represents a coma in the object wave, which is not surprising; in order to create linear fringes in the grating plane there must be a coma term that is equal and opposite to the coma generated by the interference between the two base spheres. Figure 3.3 illustrates the shape of the distorted wave.

3.1.3 Functional form of the null interferogram

We approximate the effect of the path length term in the detector plane $P_d(x)$ by shifting $P_g(x)$ along the propagation angle $\theta = \sin^{-1} \lambda/T$ and scaling it by a factor z_1/z_2 .

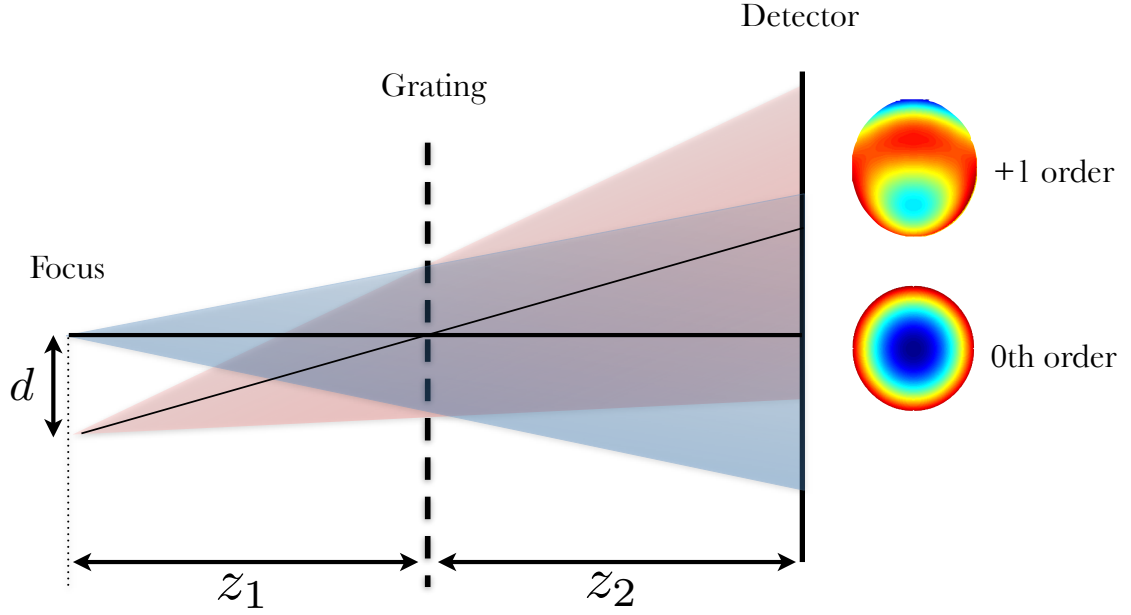


Figure 3.3: The first order diffracted wave is aberrated by a coma term which is equal and opposite to the coma from the interference between the two base spheres in the grating plane.

$$P_d(x) = P_g \left(\frac{z_1}{z_2} x - \frac{\lambda z_2}{T} \right) \quad (3.10)$$

This approximation is valid in the medium NA regime, but breaks down at larger numerical apertures as deviations in the angular spectrum of the phase of the $P_g(x)$ create a distortion of $P_d(x, y)$ as it propagates to the detector plane; this case will be handled in the section on high NA shearing.

At the detector, the fringes will reflect the intensity of the field consisting of the sum of the two spheres and the approximated path length term $P_d(x)$.

$$\begin{aligned} I(x) &= |U(x)|^2 = |\exp[ikr(x)] + \exp[ikr(x+d) + iP_d(x)]|^2 \\ &= 2 + 2 \cos k [r(x+d) - r(x) + P_d(x)] \\ &= 2 + \exp\{ik [r(x+d) - r(x) + P_d(x)]\} \\ &\quad + \exp\{-ik [r(x+d) - r(x) + P_d(x)]\} \end{aligned} \quad (3.11)$$

The argument inside the exponentials are complex conjugates of one another, and up to a

sign, represent the phase of the null interferogram $N(x)$. Using the Fourier domain method described in the previous chapter, we can isolate a single exponential term (either one will do), and take its argument to find $N(x)$. Combining this result with (3.10), we have:

$$\begin{aligned} N(x) &= k [r(x+d) - r(x) + P_d(x)] \\ &= r(x+d) - r(x) + P_g \left(\frac{z_1}{z_2} x - \frac{\lambda z_2}{T} \right) \\ &\approx C + \frac{d}{\lambda} \left(\frac{2z_2 z_1^2 + d^2 z_2 - 3d^2 z_1}{2z_2 z_1^2} \right) \cdot \frac{x}{z_2} + \frac{d^2}{\lambda} \frac{3(z_2 - z_1)}{4z_2 z_1} \cdot \frac{x^2}{z_2^2} \end{aligned} \quad (3.12)$$

Introducing the normalized coordinate $\mathbf{x} = x/NA$ z_2 and taking the far-field limit $z_2 \gg z_1$, we have,

$$N(x) = C + \frac{d}{\lambda} \left(\frac{2z_1^2 + d^2}{2z_1^2} \right) \cdot NA \mathbf{x} + \frac{d^2}{\lambda} \frac{3}{4z_1} \cdot NA^2 \mathbf{x}^2 \quad (3.13)$$

From (3.13), we see that at the detector, the cubic term present in the first order diffracted wave has cancelled the cubic term from the interference between the two base spheres just as it did in the grating plane. Evidently, we can expect perfect linear fringes in the null interferogram despite all of the non-linearities in the angular spread from the grating. In 2D, the equivalent interferogram decomposed on a basis of Zernike polynomials is:

$$N(x, y) = \left(\frac{z_1}{T} \right) NA Z_1 + \left(\frac{z_1 \lambda^3}{2T^3} \right) NA Z_1 - \left(\frac{z_1 \lambda}{2T^2} \right) NA^2 Z_3 - \left(\frac{z_1^2 \lambda^2}{4T^3} \right) NA^2 Z_4 \quad (3.14)$$

From (3.14), we see the presence of an astigmatism term that scales as the square of the grating distance z_1 . Using a numerical aperture of 0.35, 5% shear, and 50 detector fringes, the astigmatism would cause a measurement error of 140 mWaves rms. If this term is not properly corrected for, it will integrate to a coma error in the reconstructed wavefront.

3.2 System alignment and the null interferogram at medium NA

Given the nonlinearity of the angular spectrum of the diffraction grating in (2.34), it is not surprising that a small misalignment in the tilt of the grating θ_g can cause measurable dif-

ferences in the null interferogram $N(x)$ from the null interferogram in (3.13) for an aligned system. $N(x)$ can also be affected by a tilt in the detector plane θ_d , which has the effect of elongating the intensity pattern in the direction of the tilt.

This section will present an analysis of how these errors affect the null interferogram. From this analysis we can predict the error in the wavefront measurement as a function of grating and detector tilt which will play an important role in defining specifications for the shearing interferometer.

3.2.1 Functionalizing $N(x)$ along z

Figure 3.4 shows the relevant geometry, with the grating and detector tilted from the plane normal to the optical axis by angles θ_g and θ_d respectively.

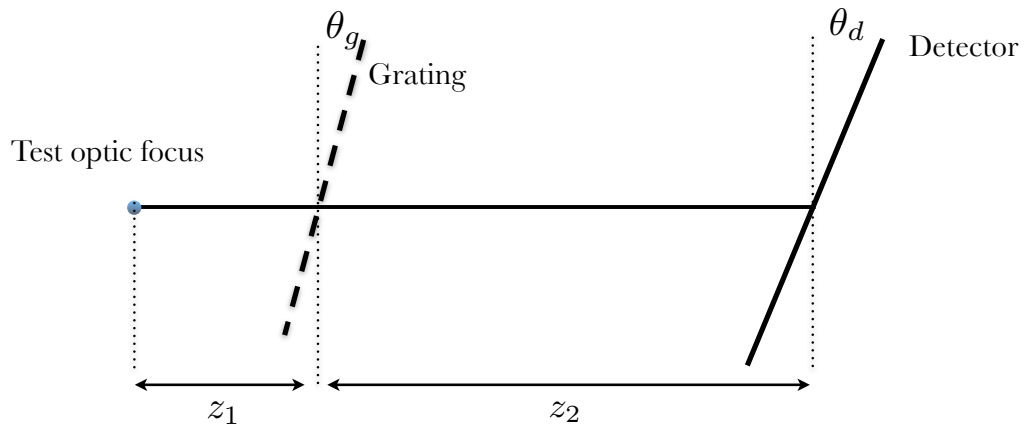


Figure 3.4: Shearing interferometry geometry with grating and detector tilt.

To solve for the field amplitude in the tilted system, we recall that we already have a solution, (3.13) in the case of a 1-dimensional detector, for the null interferogram in the detector plane for a perfectly aligned system ((3.14) represented the 2-D case, however we will again restrict the discussion to 1-D). This solution was obtained by generating a virtual wave that interferes with the test wave to create the transmission of the grating in the grating plane.

A natural extension of this technique is to require that the virtual wave interfere with the

test wave to create the grating transmission in the tilted grating plane. Once the functional form of the virtual wave is defined, the field amplitude $U(x, z)$ can be obtained at any point in space by summing the field contribution from the virtual and test waves. In particular, the field amplitude at the detector can be written as:

$$U(x(\chi, \zeta), z(\chi, \zeta)) \quad (3.15)$$

where χ and ζ are appropriately parametrized coordinates in the detector plane.

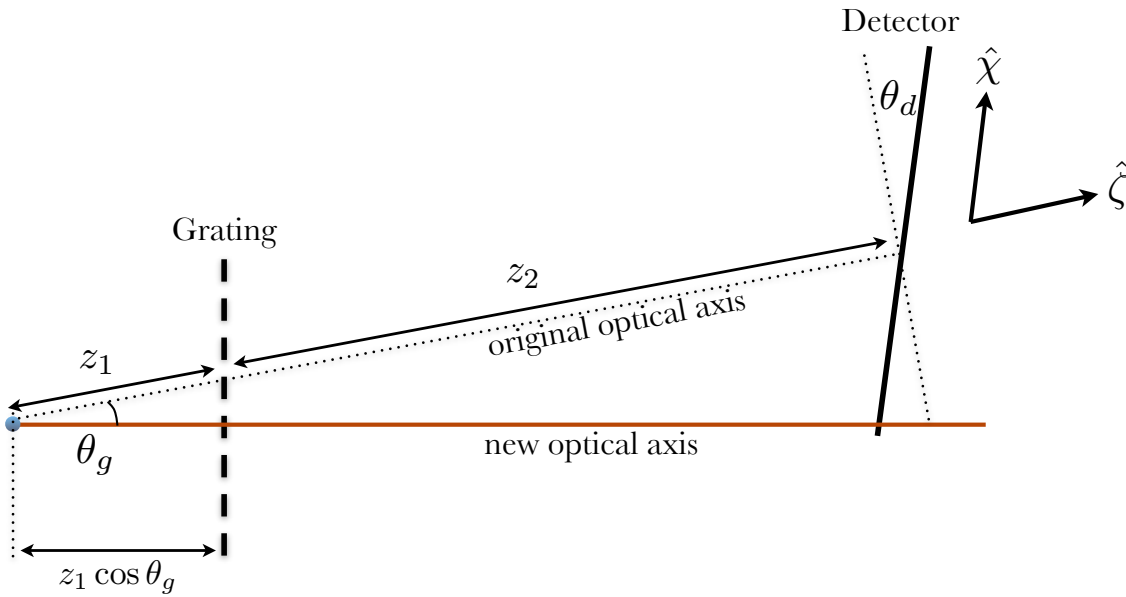


Figure 3.5: Shearing interferometry geometry in a coordinate system rotated by θ_g .

This strategy is easily realized if we rotate the coordinate system by the angle θ_g so that the grating is vertical, shown in Figure 3.5. In effect, this is similar to defining a new “optical axis” that is perpendicular to the grating plane. In practice, the footprint of the beam will restrict the intensity to be centered along the original optical axis, but redefining the coordinate system in this manner is valid since mathematically we do not explicitly rely on the restriction of the spherical wave footprint in the analysis of the virtual wave.

In this new coordinate system there is no grating tilt, so the solution to the field amplitude in (3.13) becomes valid, provided that we shorten the distance between the test wave focus and the grating by a factor of $\cos \theta_g$:

$$z_1 \mapsto z_1 \cos \theta_g \quad (3.16)$$

It then remains to solve for the field at the detector. To do this, we define the new detector coordinates (χ, ζ) that measure the lateral distance along the detector and the distance along the original optical axis respectively. Note that in general, $\hat{\chi}$ and $\hat{\zeta}$ are not orthogonal. Parametrizing (x, z) in terms of the detector coordinates, we have,

$$\begin{aligned} x &= \chi \cos(\theta_g - \theta_d) + \zeta \sin \theta_d \\ z &= \zeta \cos \theta_g - \chi \sin(\theta_g - \theta_d) \end{aligned} \quad (3.17)$$

which can be written in terms of a linear transformation:

$$\begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} \cos(\theta_g - \theta_d) & \sin(\theta_g) \\ -\sin(\theta_g - \theta_d) & \cos(\theta_g) \end{bmatrix} \cdot \begin{bmatrix} \chi \\ \zeta \end{bmatrix} \quad (3.18)$$

Writing the holographic solution in the transformed coordinate system defined by (3.18) and taking $z_2 \gg z_1$, we obtain the field at the detector which takes the form:

$$U_d(\chi, \theta_g, \theta_d) = a_1(\theta_g, \theta_d)\chi + a_2(\theta_g, \theta_d)\chi^2 \quad (3.19)$$

where the coefficients $a_1(\theta_g, \theta_d)$ and $a_2(\theta_g, \theta_d)$ represent linear error (tilt), and quadratic error (astigmatism and defocus) respectively, and are given by:

$$a_1(\theta_g, \theta_d) = \frac{d(2z_1^2 + d^2) \cos \theta_d}{2\lambda z_1^2 \cos \theta_g} + \frac{3d^2 \cos \theta_d \sin \theta_g}{2\lambda z_1 \cos^2 \theta_g} \quad (3.20)$$

$$a_2(\theta_g, \theta_d) = \frac{d(2z_1^2 + d^2) \cos \theta_d \sin(\theta_g - \theta_d)}{2\lambda z_1 \cos^2 \theta_g} + \frac{3d^2[\cos(2\theta_g - \theta_d) - 2 \cos \theta_d] \cos \theta_d}{4\lambda z_1 \cos^3 \theta_g} \quad (3.21)$$

Contour plots of $a_1(\theta_g, \theta_d)$ and $a_2(\theta_g, \theta_d)$ are shown in Figure 3.6 using $\lambda = 13.5$ nm, $NA = 0.35$, 100 fringes, and 5% shear. The linear term is more affected by grating tilt than detector tilt, whereas the quadratic term is affected equally by both alignment errors. In order to ensure a total rms wavefront error of $\lambda_{EUV}/100$, we require that the tilt errors in the grating and detector be smaller than 2.5 mrad or 0.15 degrees each.

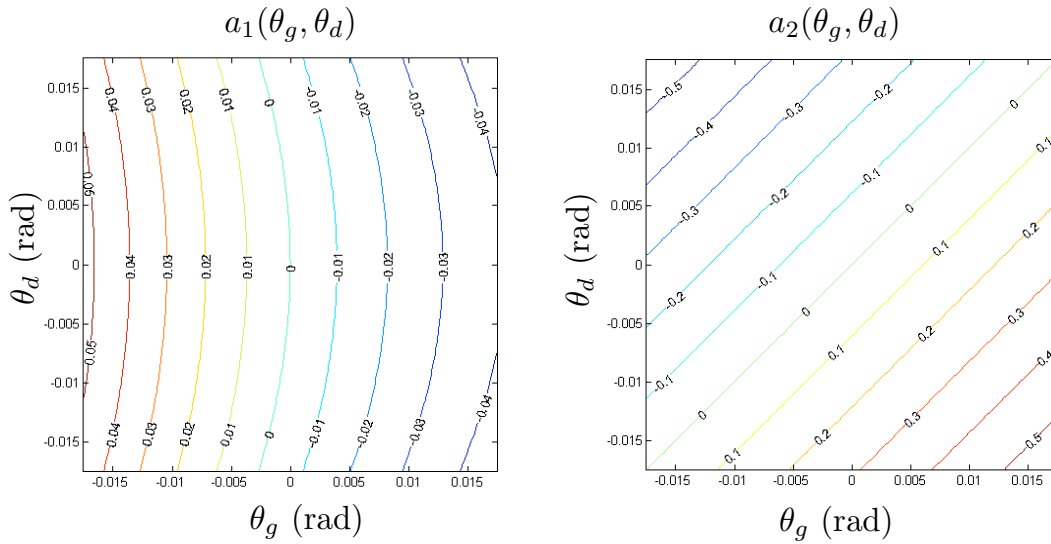


Figure 3.6: Contour plots of the linear and quadratic coefficients of the null interferogram with $\lambda = 13.5$ nm, $NA = 0.35$, 100 fringes, and 5% shear.

3.3 Null interferogram at high NA

One of the key assumptions in the holographic analysis is that the path length term $P_g(x)$ preserves its shape as it propagates to the detector. It is this assumption that allows us to write the effect of this term in the detector plane as:

$$P_d(x) = P_g\left(\frac{z_1}{z_2}x - \frac{\lambda z_2}{T}\right) \quad (3.10)$$

In the high numerical aperture regime ($NA > 0.35$), (3.10) breaks down due to the angular distortion of $P_g(x)$ as it diffracts off of the grating, causing its overall shape to elongate in the x -direction near the edges of the beam profile. As a result, we cannot arrive at a simple closed-form expression for the null interferogram in this regime.

However, we can still derive a mathematical expression in the form of a transcendental equation by using a method that carefully back-traces the geometrical paths of the rays that interfere at the detector. Although this method does not reveal the functional form of the aberrations present in the null interferogram, it still can give a numerical solution to the null interferogram, which is sufficient for analyzing the shearing interferogram.

While the full Huygens-Fresnel propagation can also give a numerical solution, both com-

putational and memory constraints limit its utility for analyzing high numerical aperture systems; a simulation at $NA = 0.5$ for example, requires over 6 days to compute on a 2.4 GHz dual core machine, whereas the equivalent problem can be solved in less than 20 seconds using the back-tracing technique described in this section.

3.3.1 The 2-ray method

To better understand the nature of the interference pattern on the detector, it is useful to consider the plane wave decomposition of the field amplitude at the grating plane. Since this field is simply the product of the grating transmission $t_g(x)$ and the spherical wave $S(x, z)$, we have,

$$U_g = t_g(x) \cdot S(x, z) \quad (3.22)$$

By the convolution theorem, the Fourier transform will be:

$$\begin{aligned} \mathcal{F}\{U_g\} &= \mathcal{F}\{1 + \exp 2\pi i x/T\} * \mathcal{F}\{S(x, z)\} \\ &= \left[\delta(f) + \delta\left(f - \frac{1}{T}\right) \right] * \mathcal{F}\{S(x, z)\} \end{aligned} \quad (3.23)$$

From (3.23), we see that the field amplitude spectrum represents the summation of the Fourier transform of the spherical wave superimposed with a copy of itself shifted in frequency by $1/T$. The spherical wave will contain a broad, continuous distribution of frequencies, corresponding to plane waves propagating at all angles within the numerical aperture of the system. As these waves propagate and disperse, they affect a smaller and smaller relative portion of the beam footprint, and in the far-field limit can be treated as rays.

The value of the phase at any given point along the ray will simply be equal to the acquired phase per unit length times path length L traveled:

$$\Phi(x) = \frac{2\pi L}{\lambda} \quad (3.24)$$

Each point on the detector will therefore be influenced by exactly two rays; one ray is from the un-diffracted zeroth order wave, and the other is from the frequency shifted first order

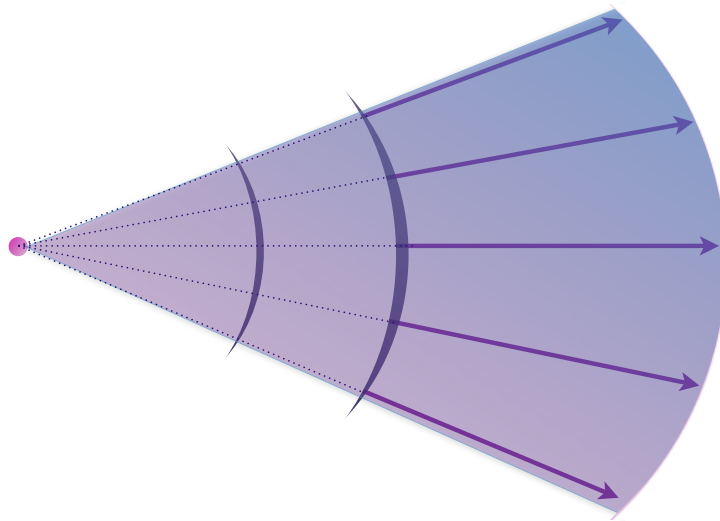


Figure 3.7: As a wave propagates to the far-field, its constituent plane waves disperse and act like rays.

diffracted wave. The interference between these two waves will give the null interferogram $N(x)$.

Figure 3.8 shows a schematic of the geometrical paths taken by a zeroth order ray (red) and a first order diffracted ray (purple) interfering at the detector at a position x . To figure out the value of the field amplitude at the detector coordinate x , we must solve for the phase difference between the interfering rays.

Letting $\Phi_0(x)$ and $\Phi_1(x)$ represent the phases of the zeroth and first order rays at the detector respectively, we have,

$$\begin{aligned} N(x) &= \Phi_1(x) - \Phi_0(x) \\ &= \left[\frac{2\pi}{\lambda} L_1 + \frac{2\pi z_1}{T} \right] - \frac{2\pi}{\lambda} L_0 \end{aligned} \quad (3.25)$$

where the L_0 and L_1 are the respective path lengths of the rays, and the term $2\pi z_1/T$ accounts for the phase acquired by the 1st order ray as it diffracts off of the grating. L_0 and L_1 are determined geometrically to be,

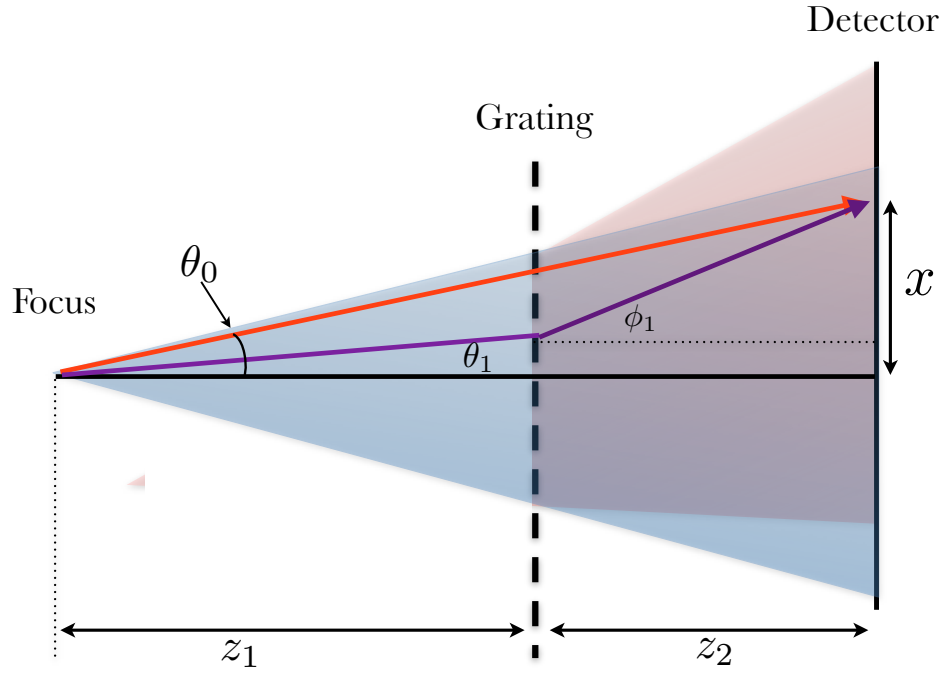


Figure 3.8: Geometrical paths of interfering rays at the detector.

$$\begin{aligned} L_0 &= (z_1 + z_2) \sec \theta_0 \\ L_1 &= z_1 \sec \theta_1 + z_2 \sec(\phi_1) \end{aligned} \quad (3.26)$$

where θ_0 is the angular position of the coordinate x measured from the test wave focus and is given by:

$$\theta_0 = \tan^{-1} \left(\frac{x}{z_1 + z_2} \right) \quad (3.27)$$

In (3.26) ϕ_1 and θ_1 are unknown, but can be determined by noting that:

1. The lateral position of the two rays must coincide at the detector:

$$(z_1 + z_2) \tan \theta_0 = z_1 \tan \theta_1 + z_2 \tan \phi_1 \quad (3.28)$$

2. ϕ_1 and θ_1 must satisfy the grating equation:

$$\sin \theta_1 = \sin \phi_1 + \frac{\lambda}{T} \quad (3.29)$$

The system of equations consisting of (3.28) and (3.29) cannot be simultaneously solved in closed form. However, a solution can be found numerically, given the input parameters: z_1 , z_2 , T , λ and NA . Figure 3.9 shows a plot of the astigmatism in the null interferogram as a function of NA as predicted by the 2-ray method compared with the astigmatism computed using a propagation simulation based on the Huygens-Fresnel diffraction integral.

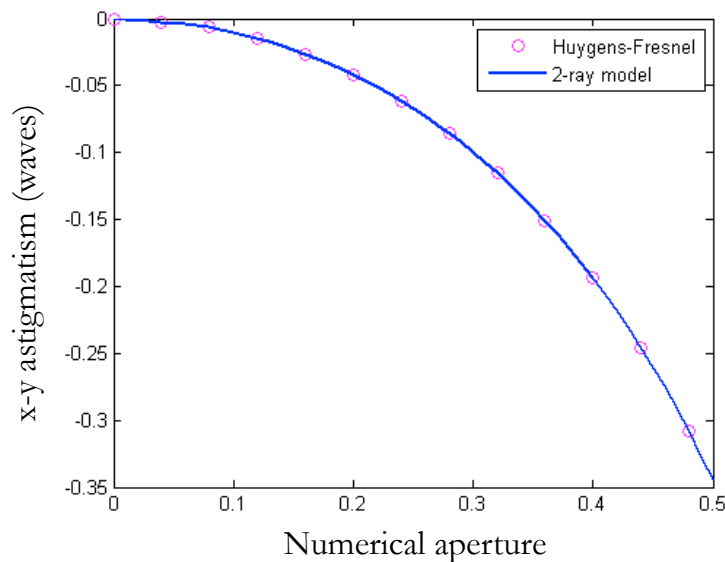


Figure 3.9: A graph comparing the 2-ray method with a numerical simulation based on the Huygens-Fresnel diffraction integral. The shear was taken to be 5%, with 40 fringes on the detector, and $\lambda = 13.5$ nm.

3.4 Variable shear

Up to this point, the discussion of grating-based shearing interferometry has centered primarily around computing the null interferogram $N(x)$. This is because removing $N(x)$ from the field amplitude at the detector in the medium to high numerical aperture regimes is a required step to isolate the aberration function difference:

$$\begin{aligned}
U_a(x) - N(x) &= A(x + s) - A(x) + S(x) - S(x + s) + \Lambda(x) - N(x) \\
&= A(x + s) - A(x) + N(x) - N(x) \\
&= A(x + s) - A(x)
\end{aligned} \tag{3.30}$$

We have also discovered that at high numerical apertures, the aberrations present in the null interferogram can become quite large due to the increasing distortion of the first order diffracted wave due to the grating.

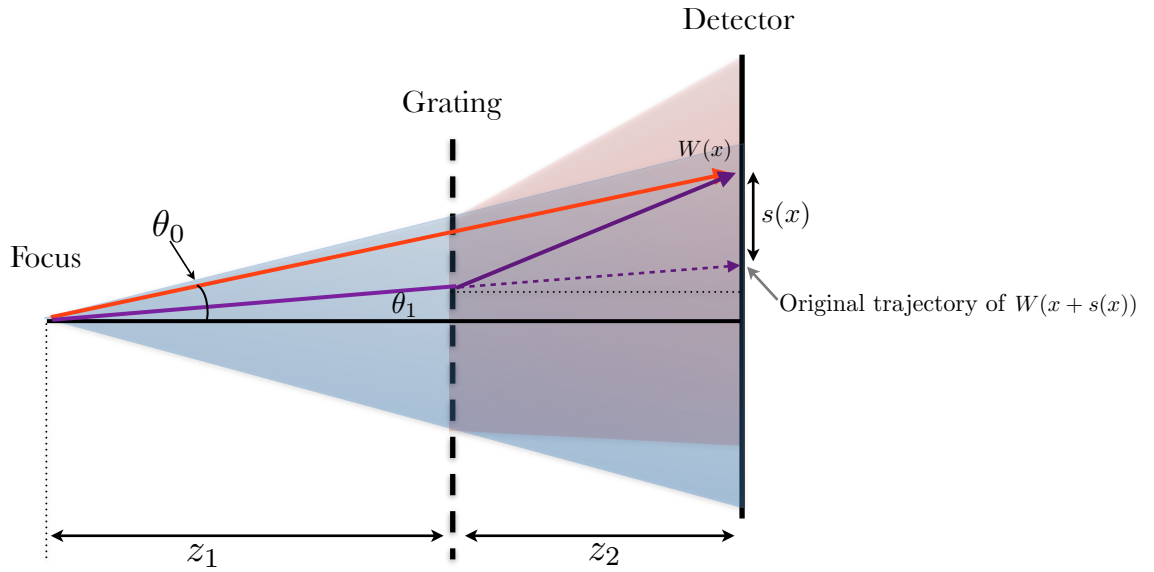


Figure 3.10: The amount of shear depends on the x -coordinate at the detector.

There is one other important effect that is caused by the diffraction grating at high numerical apertures concerning the magnitude of the shear at the detector. Referring to Figure 3.10, we see that the shear s in the grating-based implementation of LSI is given by the tangents of the angles corresponding to the zeroth and first order diffraction off of the grating:

$$s = z_2(\tan \theta_0 - \tan \theta_1) = z_2 \tan \left[\sin^{-1} \left(\theta_1 + \frac{\lambda}{T} \right) \right] \tag{3.31}$$

For low numerical apertures, two approximations to (3.31) can be made: (1), the incident angle θ_0 is small, and (2), the ratio λ/T is small (generally smaller than 0.1) so $\tan \theta \approx \sin \theta \approx \theta$, so that s is approximately given by:

$$s \approx \frac{\lambda z_2}{T} \quad (\text{Low NA})$$

However, in the general case, s is dependent on the input angle θ_0 and thus will vary based on which detector coordinate we are looking at:

$$s \longmapsto s(x) \quad (3.32)$$

As with the solution to the null interferogram at high NA, $s(x)$ cannot be solved in closed form. However, similar to the null interferogram, $s(x)$ can be determined numerically using (3.31) with the solution to θ_0 and θ_1 from (3.28) and (3.29).

3.4.1 Adapting the Rimmer method to variable shear

In Rimmer's original paper, the shear distance is chosen to correspond exactly to the pixel spacing on the detector. In this scenario, the value a_i measured at a detector pixel perfectly corresponds to the difference between the wavefront W_i and the wavefront at the adjacent pixel W_{i+1} . With this configuration, the measured data is an exact representation of the real wavefront differences, and in the absence of noise, the Rimmer method will perfectly reconstruct the unknown wavefront.

If the shear varies from pixel to pixel, it is not feasible to have the shear at every pixel correspond to the pixel separation at that location. In this case, the measured value a'_i will differ from the desired measurement a_i , since it represents the path length difference between $W_i = W(x)$ and the point $W(s(x))$ as shown in Figure 3.11.

Nevertheless, assuming that the function $W(x)$ is slowly varying, i.e. the slope of $W(x)$ does not change appreciably over the shear distance $s(x)$, we can approximate a_i using a linear interpolation,

$$a_i = a'_i \frac{\delta x}{s(x)} \quad (3.33)$$

where δx is the detector pixel spacing. Essentially, this approximation is a Taylor expansion of $W(x)$ around x , where 2nd order and higher derivatives have been neglected, and the first derivative has been approximated by its difference function in $s(x)$:

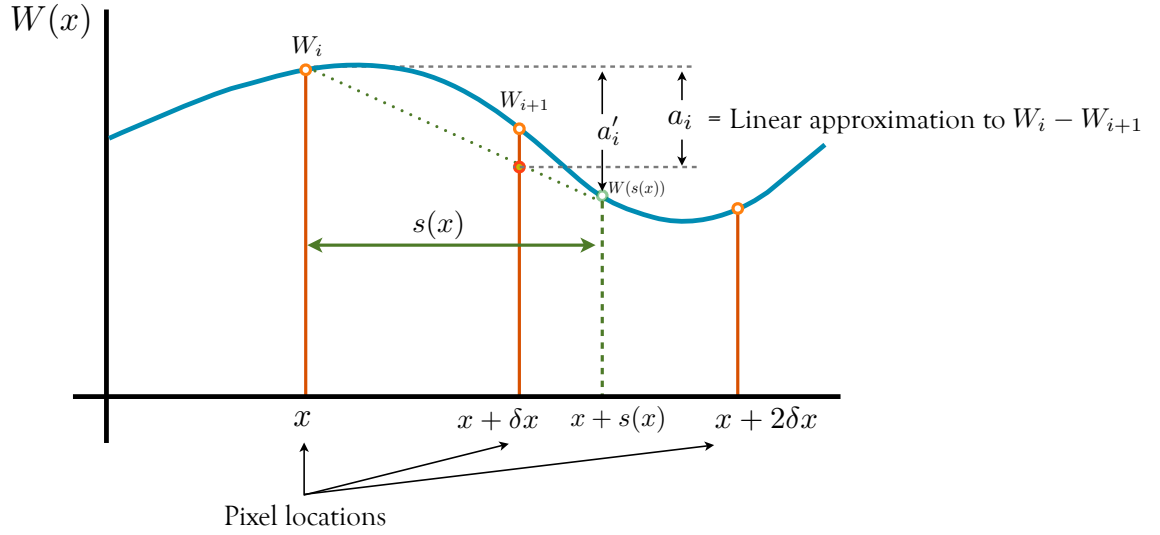


Figure 3.11: The wavefront difference between adjacent pixels is interpolated by using the known shear $s(x)$ and the measured value a_i .

$$\begin{aligned}
 W(x + \delta x) - W(x) &= \delta x \frac{\partial^n W(x)}{\partial x^n} + \frac{\delta x^2}{2} \frac{\partial^2 W(x)}{\partial x^2} + \dots \\
 &\approx \delta x \frac{\partial^n W(x)}{\partial x^n} \\
 &\approx \delta x \left[\frac{W(x + s(x)) - W(x)}{s(x)} \right] \\
 &= a'_i \frac{\delta x}{s(x)}
 \end{aligned} \tag{3.34}$$

Higher order approximations can be implemented in the case that $W(x)$ has non-negligible higher order derivatives over the shear distance. For instance, a second order approximation could be given by writing the second derivative term in (3.34) as a second order difference function:

$$\frac{\partial^2 W(x)}{\partial x^2} \approx \frac{W(x - \delta x + s(x - \delta x)) - W(x - \delta x) + W(x + s(x)) - W(x)}{s(x - \delta x)s(x)} \tag{3.35}$$

In practice, however, wavefront aberrations that result from system misalignment manifest as functions that vary slowly compared with typical shear values, and the linear approximation in (3.33) is sufficient.


3.5 Discussion

3.5.1 Simulation results

The systematic aberrations present in LSI setups for various numerical apertures is shown in Table 3.2. The distances z_1 , z_2 and T were chosen to form 50 fringes on a 1 inch format detector, with a 5% shear. The results in these cases were confirmed by a simulation of the full Huygens-Fresnel propagation integral.

Table 3.2: Systematic aberrations in the LSI interferogram for various NAs (mWaves)

NA	Z_3 (defocus)	Z_4 (x-y stig)	Z_6 (x coma)	Z_8 (spherical)
0.05	-2.0	8.0	0	0
0.25	-21.1	-83.8	0.2	0.1
0.50	-110.5	-430.7	6.5	-1.9
0.75	-491.6	-1,851.5	131.8	24.1



From the holographic analysis, it is evident that all of the systematic aberrations are proportional to powers of the grating defocus z_1 . As the grating is brought closer to the focus of the test optic, the systematic aberrations tend toward 0. However, from the last chapter, decreasing the grating defocus also decreases the number of fringes on the detector which after a certain point restricts our ability to perform a Fourier domain interferogram analysis.

While time domain analysis is still an option, there are many benefits of the Fourier domain approach. Because no phase shifting is required, there is no need for a translation stage on the grating, which can suffer from its own alignment difficulties. The phase shifting setup requires two separate gratings with horizontal and vertical lines, whereas the Fourier domain approach can use a 2-dimensional cross-grating, which can cause shear in both directions simultaneously. Finally, since the Fourier domain approach can take all of the analysis in a single shot, it has looser restrictions on experimental stability.

Nonetheless, accurately removing the null interferograms at high numerical apertures can be a daunting task. Small tilts in the grating and detector can translate into large effects on the systematic aberrations. Therefore the choice of grating defocus must be handled on a case by case basis, taking into account the confidence in both the alignment tolerances and the experimental stability.

3.5.2 Summary

Here we summarize the results from this chapter regarding grating-based LSI. At low numerical apertures ($NA < 0.15$), the systematic aberrations are less than $\lambda/100$ (rms), which means that LSI can be performed with high accuracy without removing the null interferogram. In the medium numerical aperture regime ($0.15 < NA < 0.35$), the null interferogram is given by the holographic solution in (3.14). This solution sheds light on the specific form of the aberrations present in the null interferogram. In particular, the largest term is the astigmatism term, has a z_1^2 dependence, meaning that configurations that bring the grating closer to the focus will create smaller aberrations.

The holographic solution can be extended to describe the effect of grating and detector tilt on the systematic aberrations of the system. For an numerical aperture of 0.35, the grating and detector must be within 2.5 mrad of their nominal tip/tilt positions to ensure a $\lambda/100$ ceiling on tilt-based systematic aberrations.

At high numerical apertures ($NA > 0.35$), the full geometrical back-trace method represented by equations (3.28) (3.29) and (3.25) must be used to accurately determine the systematic aberrations. Although this method cannot provide a closed-form representation of the aberrations, it can still be used to compute the null interferogram numerically, which is sufficient for the interferogram analysis. Variable shear must also be compensated by using an interpolation to find the wavefront differences in terms of the detector pixel spacing.

Once the systematic errors are removed, LSI provides a robust method for characterizing the aberrations in EUV optical systems. Due to its many advantages over conventional methods and its scalability to high numerical apertures, it is a promising candidate for next-generation EUV optical metrology.

Chapter 4

ROCS

Contents

4.1	Background	63
4.1.1	Aerial image modeling	63
4.1.2	Hopkins decompositions	64
4.2	Motivation	65
4.3	The System Cross-coefficients matrix	68
4.3.1	Properties of the SCC matrix	71
4.4	The Reduced Optical Coherent Sum	74
4.4.1	Spectral decomposition of the SCC matrix	74
4.4.2	Truncation of the spectral sum	75
4.4.3	Diagonalizing the SCC	76
4.5	ROCS in 2D	79
4.6	Effect of mask and source complexity	81
4.7	Comparison to conventional aerial image models	83
4.7.1	Aerial image comparison	83
4.7.2	Analytic form of the error in ROCS	84
4.7.3	Comparison of MAST and ROCS computation times	85
4.8	Summary	88

In the next two chapters, the discussion will shift to a different method for wavefront characterization that is based on an in-situ image based iterative procedure. This procedure is enabled by a novel decomposition of the Hopkins equation called the Reduced Optical Coherent Sum (ROCS). In ROCS, the partially coherent imaging system is described by an operator S which is written in low rank approximation, similar to the Cobb's SOCS decomposition for optical proximity correction [28]. This chapter presents the mathematical foundation for ROCS and demonstrates how it is well-suited to rapidly approximate the aerial images from a partially coherent optical system. In the case that the object and the source coherence remain unchanged from image to image, ROCS will have speed improvements of up to 15 times over conventional aerial image computations. The error bound on the approximation will be shown to be well-characterized, allowing the user to trade off between accuracy and computation time.

The iterative procedure for characterizing the aberrations in the optical system will be handled in the next chapter.

4.1 Background

4.1.1 Aerial image modeling

Although the development of ROCS is focused on wavefront measurement, the computation of aerial images from partially coherent optical systems has many other applications related to the semiconductor industry, especially in the EUV lithography community. Developers of photoresist, for example, require aerial image models in order to simulate the dynamics of photoacid generation and subsequent patterning. Manufacturers of EUV masks are interested in studying how defects in masks transfer to the resist. And the makers of the lithographic tools to be used in production must learn how aberrations in the optical systems affect the eventual printed patterns.

Software packages written to perform aerial image modeling typically work by directly computing the Hopkins equation, which governs partially coherent imaging. The problem

with this approach is that Hopkins is notoriously slow to compute since it requires integration over four dimensions. As a result, there has been a significant amount of work done in finding optimizations and decompositions of the Hopkins equation to reduce the computational load.

4.1.2 Hopkins decompositions

One particular branch of decompositions involves separating the source and pupil parts of the integral from the mask part by writing them in a matrix known as the transmission cross-coefficients (TCC) matrix. This formulation is particularly useful to applications such as optical proximity correction (OPC), where a series of aerial images is desired in which the source and the pupil are constant with each image. In this case the TCC matrix will be the same for each image, and can be computed once and stored to drastically reduce the computations for successive images.

Cobb & Zakhor, who pioneered OPC algorithms in the late 90s, demonstrated that the TCC matrix is diagonalizable in most cases, and that a low-rank approximation could be implemented to exchange small errors in aerial image fidelity for significant gains in computation time [28][29]. Cobb's algorithm however can only be applied to relatively small regions of a mask, since the number of elements in the $N \times N$ TCC matrix roughly scales as the 4th power of the grid sampling of the mask ($N = k^2$ where k is the grid sampling of the mask). A TCC operating on a mask that is modeled with 256 x 256 pixels, for example, would require nearly 100 GB of memory to store, not including the memory required to do the eigenvalue decomposition.

A major leap recently occurred as Yamazoe demonstrated [30] that it is possible to obtain the eigenvalues and eigenvectors of the TCC without ever needing to build the TCC, hence avoiding the memory constraint on the size of the mask. He accomplished this by constructing an $M \times N$ matrix P , such that $TCC = P^t P$, which in effect, showed that TCC is semipositive definite. The eigenvalues of TCC are then the square of the singular values of P , the eigenvectors of TCC are the right-singular vectors of P , and M is an upper bound on $rank\{TCC\}$, where M is the number of points in the source sampling grid.

4.2 Motivation

The Hopkins equation can be visualized as a system that produces an aerial image from three inputs: the mask, the pupil aberrations, and the source coherence functions, shown in Figure 4.1a. In the TCC framework, the mask and the source are effectively built into the system, creating a new system that produces an aerial image from the single input of the mask function, shown in Figure 4.1b.

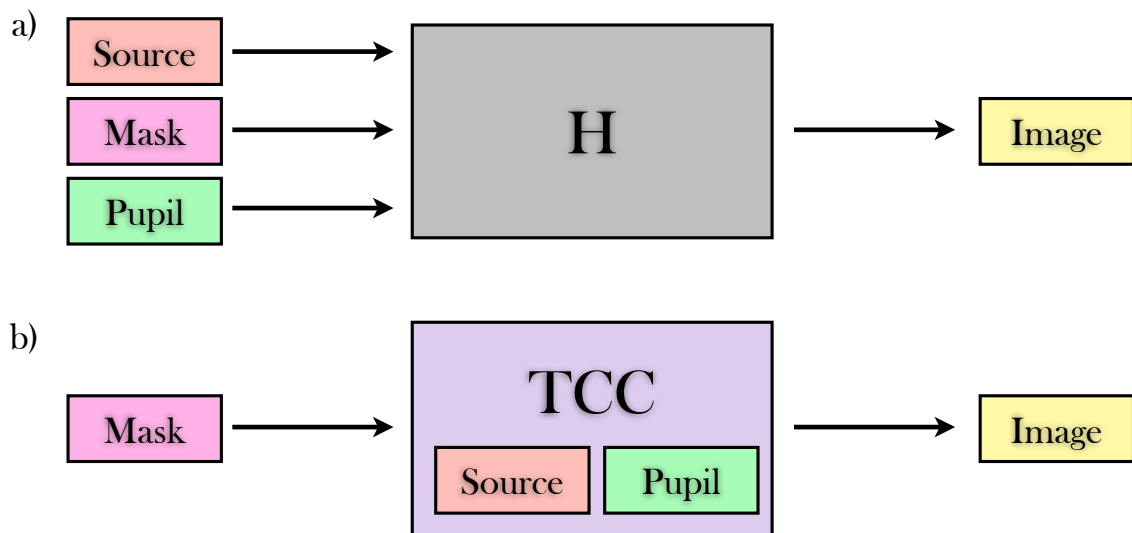


Figure 4.1: a) The Hopkins equation as a system that takes the source, mask, and pupil functions as inputs. b) The TCC approach takes the mask function as the input, and the source and pupil functions are built into the system.

Essentially, the TCC matrix lumps together the effects of source coherence and pupil aberrations into a single effect. Using the TCC approach is efficient when computing a series of images where the source and pupil aberration functions remain constant from image to image. This is quite often the case, given that real imaging tools generally have a fixed source coherenceⁱ which depends on the illumination configuration, and a fixed set of pupil aberrations which depend on the quality and alignment of the optics. The goal of OPC, for example, is given a fixed source coherence and pupil aberration function, to find the mask that produces an image most closely resembling a target aerial image. This problem can be

ⁱSome imaging systems, such as the SEMATECH Berkeley MET have a programmable coherence

solved using an iterative technique similar to the one shown in Figure 4.2.

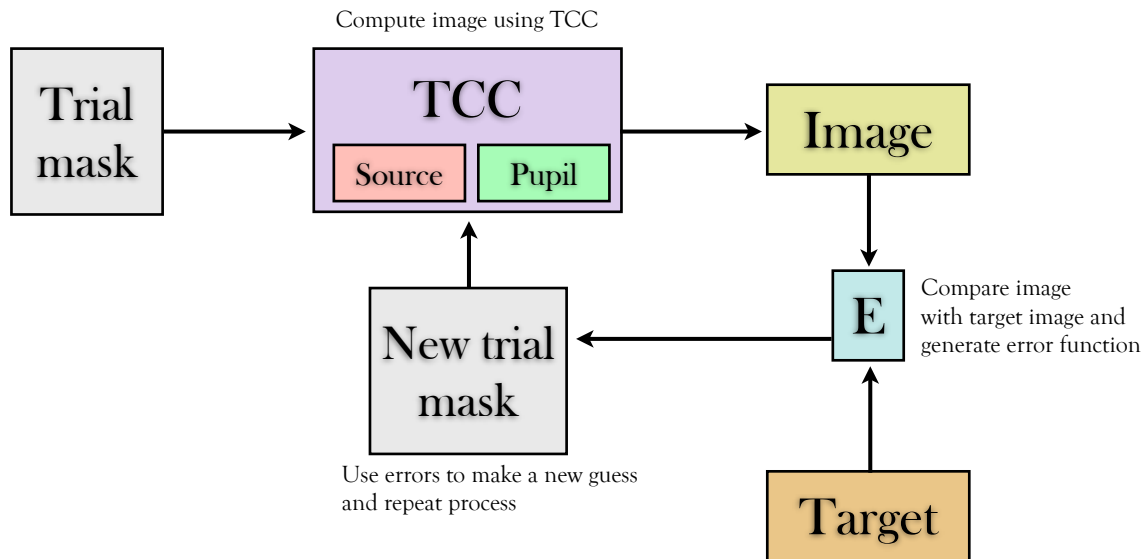


Figure 4.2: An iterative process for performing OPC.

In this technique, a trial mask is fed into the TCC system which computes an aerial image. This image is compared with a target image, and an error function is produced. If this error is below a certain tolerance, then the trial mask is the solution. Otherwise, the error function is used to create a new trial mask. The process is repeated iteratively until a solution is found. The details of exactly how the new trial mask is chosen can be complicated, but a good discussion can be found in [28].

We would like to adapt this idea into a method for determining the unknown aberrations in an optical system. In this adapted scenario, the mask and source coherence functions would be given, and the goal would be to find the unknown aberrations in the system.

Such a method is feasible because specific aberrations have a very characteristic signature in terms of the way they effect aerial image, especially when the image is viewed *through-focus*, or at different image planes. Astigmatism, for example, blurs features horizontally on one side of the focus, and blurs features vertically on the other side of the focus, as shown in Figure 4.3

It seems reasonable then to create a method that can examine the through-focus behavior of the imaging system on a known mask, and infer the aberrations that created that behav-

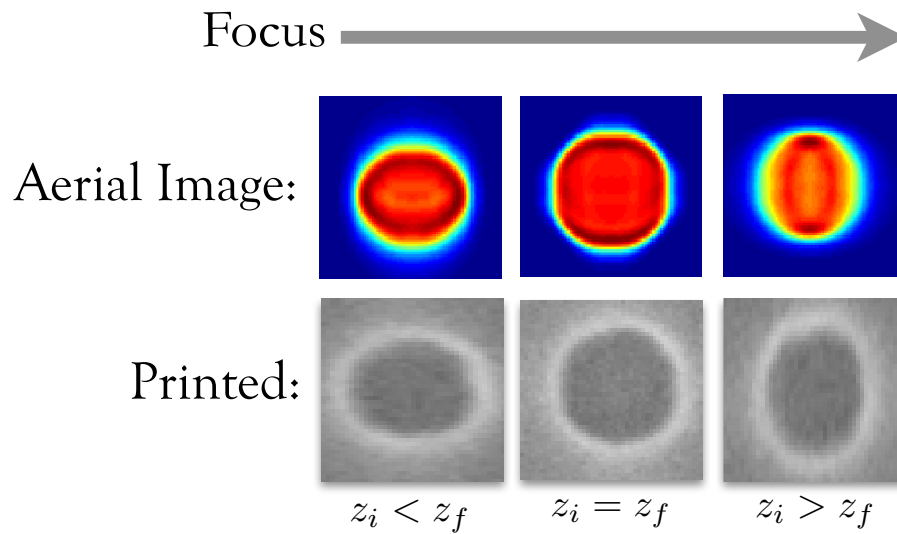


Figure 4.3: The effect of astigmatism on a contact pattern where the image plane z_i is taken as less than, equal to, and greater than the focused plane z_f .

ior. This algorithm will be similar to the OPC method outlined above, except that it would optimize on the pupil rather than the mask. As shown in Figure 4.4, a trial set of aberrations would be fed into some system that computes a set of through-focus aerial images using a fixed mask and source coherence. These images are compared to a set of target images obtained experimentally, and an error function is computed. If the error is below a certain tolerance, then the trial aberrations are claimed to be the aberrations of the system. Otherwise, a new set of aberrations is chosen and the process repeats iteratively.

While TCC approach is highly effective for cases where the source and pupil stay fixed, it is not well-suited for this particular application since the TCC matrix must be recomputed any time the source coherence or the pupil aberrations changes. To make this new algorithm efficient we need a system analogous to the TCC system that can rapidly compute an aerial image from a pupil function given fixed source and mask functions to fill in the question mark in Figure 4.4. The demonstration and derivation of such a system will be presented in the next section.

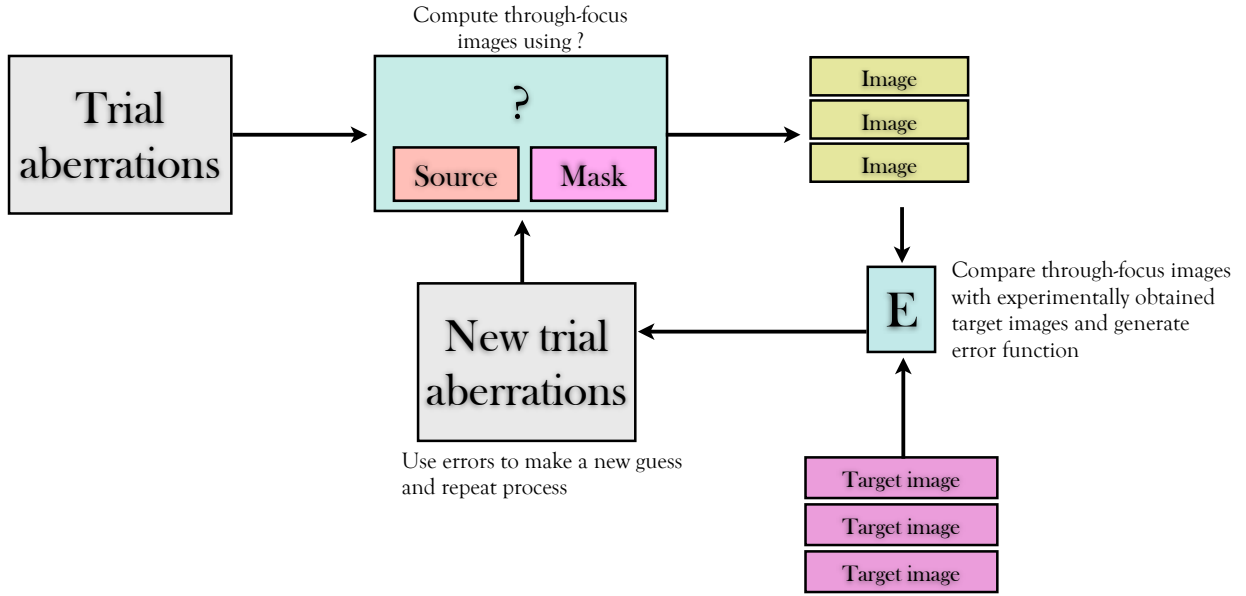


Figure 4.4: An iterative procedure for determining the aberrations in an optical system with known coherence.

4.3 The System Cross-coefficients matrix

In this section, we introduce the system cross-coefficients (SCC) matrix \mathcal{S} , which plays an analogous role to the TCC matrix, except it operates on pupil functions rather than mask functions. The SCC matrix will lay the foundation for the ROCS decomposition.

In one dimension, the Hopkins equation for the aerial image spectrum is [31]:

$$\mathcal{I}(\nu) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dz dp \mathcal{K}(\nu) \mathcal{K}^*(\nu - z) \mathcal{J}_0(p) \mathcal{T}(\nu + p) \mathcal{T}^*(\nu - z + p) \quad (4.1)$$

Where \mathcal{K} is the coherent transmission function, which is the pupil aberrations spectrum; \mathcal{J}_0 is the mutual intensity function spectrum, which represents the source coherence; and \mathcal{T} is the spectrum of the mask itself.

Consider the Kohler illuminated optical system in Figure 4.5. Conceptually, the Hopkins equation is performing a summation over the source points represented in \mathcal{J}_0 , where each source point generates a separate image that is added incoherently (by summing the squared field amplitude) at the detector. The displacement of the source from the optical axis translates to a relative shift between the mask and the pupil spectra before they are multiplied. In effect, for a source that has N incoherent source points, we can think of Hopkins as the

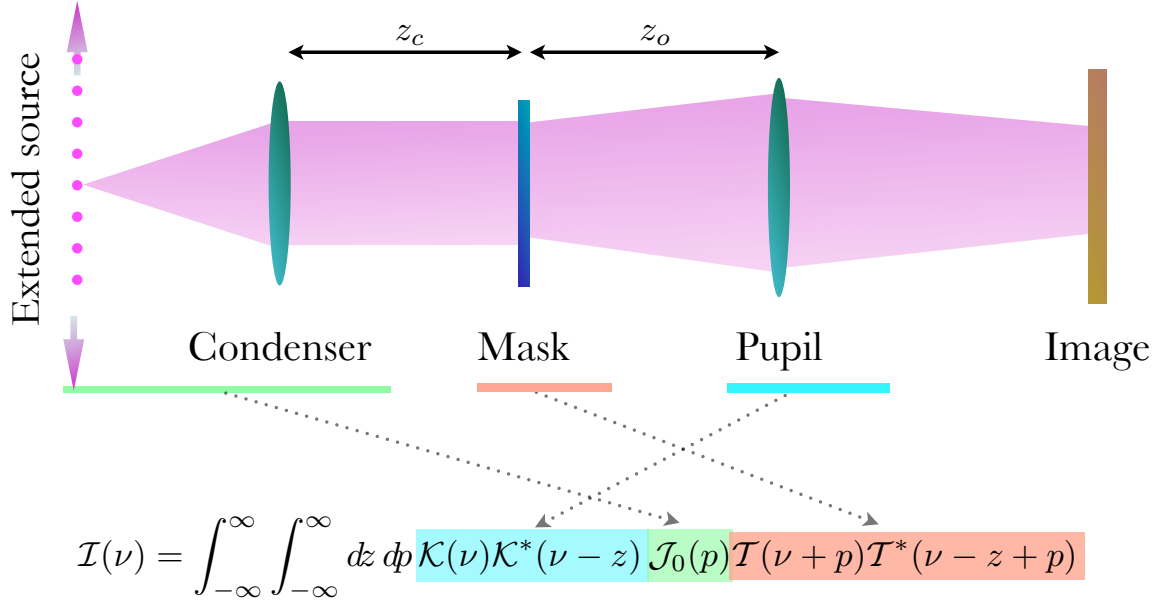


Figure 4.5: A Kohler illuminated optical system. The pupil aberrations and extent specify the coherent transmission function, and the condenser extent specifies the mutual intensity function.

summation of N subsystems shown in Figure 4.6, each of which corresponds to the product of the \mathcal{K} and \mathcal{T} with a relative shift given by \mathcal{J}_0 .

In a Kohler illuminated setup, we can write the coherent transmission function \mathcal{K} in terms of the pupil aberration function $P_o(x)$ as:

$$\mathcal{K}(\nu) = P_o(\lambda z_o \nu) \quad (4.2)$$

And the Fourier transform of the mutual intensity function can be written as a scaled version of the illuminating optic's exit pupil function $P_c(x)$:

$$\mathcal{J}_0(p) = P_c(-\lambda z_i p) \quad (4.3)$$

Since all real lenses have finite extent, we can assume the pupil function goes to zero outside its radius r . The mutual intensity spectrum will then be bandlimited with maximum frequency:

$$p_{max} = \frac{r}{\lambda z_i} \quad (4.4)$$

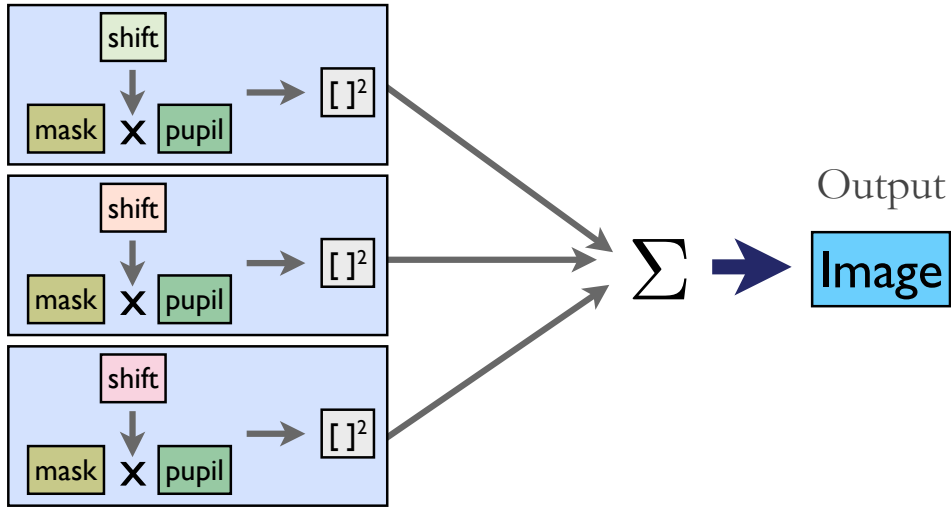


Figure 4.6: Hopkins sums a series of subsystems, each of which is the product of the mask and pupil spectra with a relative shift given by the mutual intensity spectrum.

For the purposes of this discussion, we will assume the illuminating optic exit pupil function is real-valued across the extent of the pupil, and zero otherwise:

$$P_i(x) = \begin{cases} \Re\{P_i(x)\} & : |x| \leq r \\ 0 & : |x| > r \end{cases}$$

The real-valued assumption holds as long as the phase variations in the illuminating optic pupil vary slowly over a region defined by the coherence area A_c , which is generally the case in practice.

We begin by writing (4.1) in the following way,

$$\mathcal{I}(\nu) = \int_{-\infty}^{\infty} dz \mathcal{K}(\nu) \mathcal{K}^*(\nu - z) S(\nu, \nu - z) \quad (4.5)$$

where,

$$S(k, m) = \int_{-p_{max}}^{p_{max}} dp \mathcal{J}_0(p) \mathcal{T}(k + p) \mathcal{T}^*(m + p) \quad (4.6)$$

Here we have separated out the pupil term from the source and mask terms. The function $S(k, m)$ contains information only about the source and the mask, and will be referred to as the *system cross-coefficients*, or SCCs. The SCCs play a role analogous to the TCCs in

OPC; they allow us to build the fixed properties—in this case the source and mask—into the system. We can now think of the SCCs as a system that maps a pupil function to an aerial image directly as shown in Figure 4.7.

In the next two sections we will show that by exploiting the mathematical structure of S , we can approximate it and significantly reduce the number of computations required to calculate the aerial image.

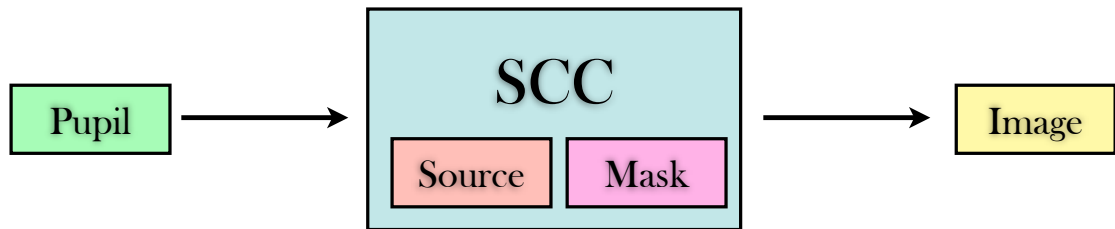


Figure 4.7: A system is needed that takes a pupil function as an input, and produces an aerial image with the fixed mask and source information built into the system.

4.3.1 Properties of the SCC matrix

SCC is Hermitian

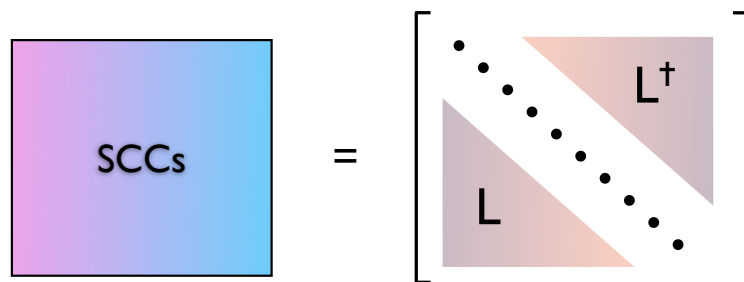


Figure 4.8: A Hermitian matrix S is equal to its conjugate transpose: $S = S^\dagger$.

An operator \mathcal{S} over the field of complex numbers is said to be Hermitian if its matrix representation satisfies $S = S^\dagger$, where \dagger denotes the *Hermitian conjugate*, which for matrices

over the complex field \mathbb{C} is equivalent to the conjugate transpose $S^\dagger = S = (S^*)^T$. Using that \mathcal{J}_0 is real-valued, we have:

$$\begin{aligned} S[k, m]^\dagger = S^*[m, k] &= \int dp [\mathcal{J}_0(p)\mathcal{T}(m+p)\mathcal{T}^*(k+p)]^* \\ &= \int dp \mathcal{J}_0(p)\mathcal{T}(k+p)\mathcal{T}^*(m+p) = S[k, m] \end{aligned} \quad (4.7)$$

so S is hermitian.

In an experiment, the mutual intensity spectrum $\mathcal{J}_0(p)$ will be represented by a discretely sampled function $\mathcal{J}_0[p]$, which corresponds to treating the continuous source as a finite set of incoherent point radiators. In this case S will have a discretized representation:

$$S[k, m] = \sum_p \mathcal{J}_0[p]\mathcal{T}[k+p]\mathcal{T}^*[m+p] \quad (4.8)$$

Since,

$$\begin{aligned} S^*[m, k] &= \sum_p \{\mathcal{J}_0[p]\mathcal{T}[m+p]\mathcal{T}^*[k+p]\}^* \\ &= \sum_p \mathcal{J}_0[p]\mathcal{T}[k+p]\mathcal{T}^*[m+p] = S[k, m] \end{aligned} \quad (4.9)$$

we have that S is also Hermitian in the discrete case.

SCC is low-rank semipositive definite

A Hermitian matrix $S \in \mathbb{C}^{N \times N}$ is said to be semipositive definite if for all vectors $v \in \mathbb{C}^N$,

$$v^\dagger S v \geq 0 \quad (4.10)$$

Among the many properties of semipositive definite matrices is that there exists at least one generating matrix M satisfying:

$$S = M M^\dagger \quad (4.11)$$

Conversely, any matrix S that can be written in the form of (4.11) is semipositive definite. This can be seen easily by writing:

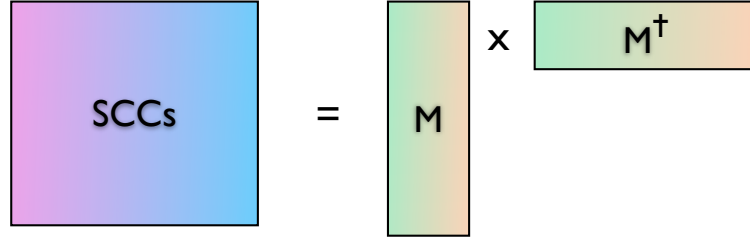


Figure 4.9: A semipositive definite matrix S can be written as the product of a generating matrix M and its Hermitian conjugate: $S = MM^\dagger$.

$$\begin{aligned}
 v^\dagger S v &= v^\dagger M M^\dagger v \\
 &= (M^\dagger v)^\dagger (M^\dagger v) \\
 &= \|M^\dagger v\|^2 \geq 0 \quad \forall v \in \mathbb{C}^N
 \end{aligned} \tag{4.12}$$

Consider the following matrix M and its Hermitian conjugate M^\dagger :

$$M[k, m] = \mathcal{J}_0^{\frac{1}{2}}[m] \mathcal{T}[k + m] \tag{4.13}$$

$$M^\dagger[k, m] = \mathcal{J}_0^{\frac{1}{2}}[k] \mathcal{T}^*[k + m] \tag{4.14}$$

The product of M and M^\dagger is then,

$$\begin{aligned}
 \{MM^\dagger\}[k, m] &= \sum_p M[k, p] M^\dagger[p, m] \\
 &= \sum_p \mathcal{J}_0^{\frac{1}{2}}[p] \mathcal{T}[k + p] \mathcal{J}_0^{\frac{1}{2}}[p] \mathcal{T}[p + m] \\
 &= \sum_p \mathcal{J}_0[p] \mathcal{T}[m + p] \mathcal{T}^*[k + p] \\
 &= S[k, m]
 \end{aligned} \tag{4.15}$$

so S is semipositive definite. From (4.14), the rows of M are copies of the mask Fourier transform shifted by the nonzero values of \mathcal{J}_0 . M has dimensions $[N + P - 1] \times P$, where

N is the number of mask samples, and P is the number of source points. In a typical ROCS decomposition, a reasonable sampling might be $N = 256 \times 256$, and $P = 25 \times 25$, in which case the ratio of the number of rows to columns is roughly 125 to 1. In general M will have many times more rows than columns as shown in Figure 4.9.

The rank of S is then given by the rank of M . The ratio of the rank of S to the number of rows or columns of S is defined as the reduction factor ϵ :

$$\epsilon \equiv \frac{\text{rank}\{S\}}{\# \text{ of rows in } S} = \frac{P}{N + P - 1} \quad (4.16)$$

ϵ represents the the fraction of S that is non-redundant information. The amount of memory required to store S for instance scales as $\text{size}\{S\} \cdot \epsilon$.

4.4 The Reduced Optical Coherent Sum

In this section we introduce ROCS, a novel Hopkins decomposition that enables the rapid computation of aerial images with a fixed mask and source.

4.4.1 Spectral decomposition of the SCC matrix

Spectral theorem guarantees that S is diagonalizable [32] and that we can write S as a sum over the outer products of its eigenvectors weighted by their corresponding eigenvalues:

$$S(k, m) = \sum_{i=1}^P \lambda_i \xi_i(k) \xi_i^*(m) \quad (4.17)$$

where $P = \text{rank}\{S\}$, is the number of discretized source points. Plugging back into (4.5), and swapping the order of the sum and integral gives:

$$\mathcal{I}(\nu) = \sum_{i=1}^P \lambda_i \int_{-\infty}^{\infty} dz [\mathcal{K}(\nu) \xi_i(\nu)] [\mathcal{K}^*(\nu - z) \xi_i^*(\nu - z)]$$

We recognize the integral as the autocorrelation of the product $\mathcal{K}\xi_i$:

$$\mathcal{I}(\nu) = \sum_{i=1}^P \lambda_i [\mathcal{K}\xi_i \star \mathcal{K}\xi_i]$$

Taking the inverse Fourier transform of both sides to get the space-domain aerial image intensity, we have:

$$I(x) = \sum_{i=1}^P \lambda_i |\mathcal{F}^{-1}\{\mathcal{K}\xi_i\}|^2 \quad (4.18)$$

Equation (4.18) is the ROCS decomposition of the aerial image intensity. Once the source and mask properties are defined, the SCCs are computed and diagonalized. The resulting eigenvalues λ_i and corresponding eigenvectors ξ_i are stored and recalled each time the calculation occurs.

A further simplification can be made by invoking the convolution theorem,

$$I(x) = \sum_{i=1}^P \lambda_i [k * \zeta_i]^2 \quad (4.19)$$

where $k = \mathcal{F}^{-1}\{K\}$ and $\zeta_i = \mathcal{F}^{-1}\{\xi_i\}$. However, in practice the formulation in (4.18) is preferred because performing the multiplication and the inverse Fourier transform is generally faster than computing the convolution in (4.19).

Instead of building up the aerial image by summing the contributions from each source point, the ROCS decomposition builds up the aerial image spectrum by summing up the Fourier transforms of the eigenvectors of S , multiplied by the pupil aberration function. Figure 4.10 shows the the eigenvector transforms of the four most significant vectors (i.e., ξ_1 , ξ_2 , ξ_3 and ξ_4) representing a square contact with a disk illumination of $\sigma = 0.6$.

4.4.2 Truncation of the spectral sum

Due to the energy compaction property of spectral decomposition [30], the majority of the weight of S is represented in a relatively small number of terms. An example of the relative weights of the eigenvalues is shown in Figure 4.11. The aerial image intensity is therefore well-approximated by considering only the first $K^* < N$ terms of the sum. In linear algebra, this is known as a low-rank approximation [33].

$$I(x) \approx \bar{I}(x) = \sum_{i=1}^{K^*} \lambda_i |\mathcal{F}^{-1}\{\mathcal{K}\xi_i\}|^2 \quad (4.20)$$

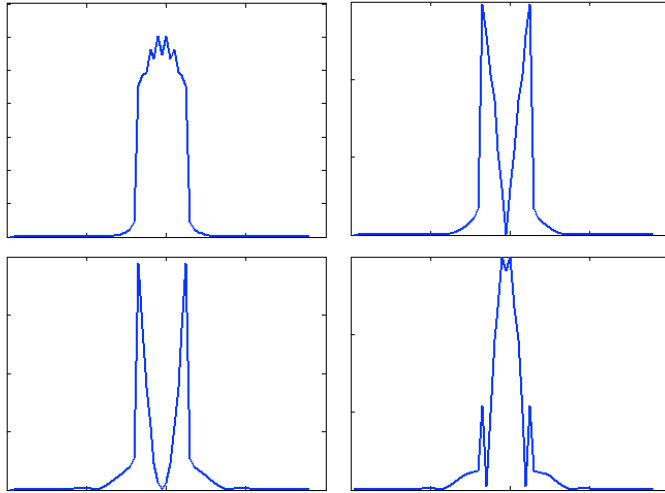


Figure 4.10: An example of the Fourier transforms of the first four eigenvectors of S .

K^* is chosen to give a specified tolerance on the aerial image computation. The normalized error E between $I(x)$ and $\bar{I}(x)$ is bounded by the sum of the remaining eigenvalues:

$$E \leq \sum_{i=K^*+1}^P |\lambda_i| \quad (4.21)$$

In a typical case, to achieve an error less than 1%, $K^*/P \approx .7$ to $.15$, meaning that the aerial image computation is performed 7 to 15 times faster than with conventional methods.

4.4.3 Diagonalizing the SCC

While equation (4.18) provides a means for the rapid computation of aerial images, it requires the eigenvectors and eigenvalues of the SCC matrix. Storing and recalling the eigenvectors from memory is not unreasonable—in the SCC from the example above with $N = 256 \times 256$, $P = 25 \times 25$, and $K^* = 5$, would require

$$5(256 + 25 - 1)^2 \times 16 \frac{\text{bytes}}{\text{complex double}} = 6 \text{ MB of memory} \quad (4.22)$$

By contrast, putting the whole SCC matrix in memory would require

$$(256 + 25 - 1)^4 \times 16 \frac{\text{bytes}}{\text{complex double}} = 95 \text{ GB of memory} \quad (4.23)$$

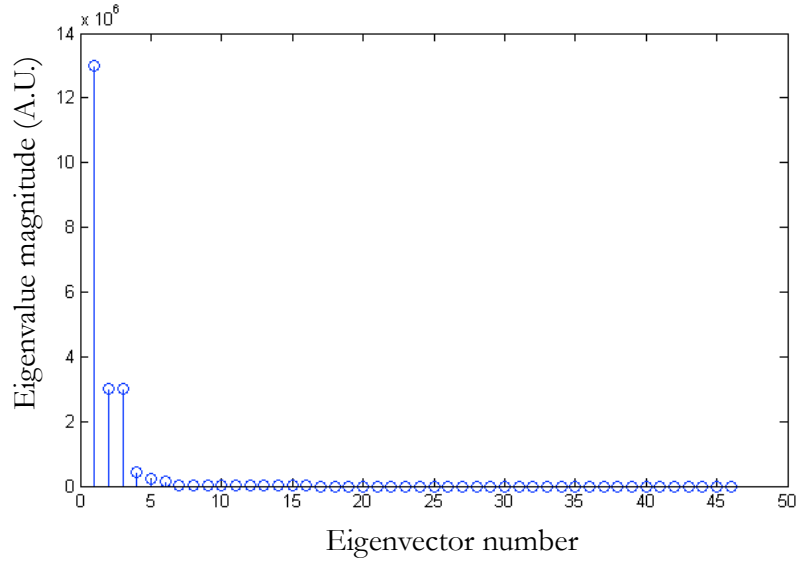


Figure 4.11: An example plot of the relative eigenvalue weights for the first 46 eigenvalues of an SCC decomposition. Most of the weight is concentrated in small number of terms.

Ninety-five gigabytes is a formidable amount of memory even for supercomputers today (2011), and this is not even including the memory and time required to diagonalize the matrix.

Fortunately we can exploit the properties of the S to obtain its eigenvalue decomposition without ever needing to construct the matrix. Since S is semipositive definite, we can write it as the product of its generating matrices M and M^\dagger :

$$S = MM^\dagger \quad (4.24)$$

where M and M^\dagger were defined in (4.14)

$$\begin{aligned} M[k, m] &= \mathcal{J}_0^{\frac{1}{2}}[m] \mathcal{T}[k + m] \\ M^\dagger[k, m] &= \mathcal{J}_0^{\frac{1}{2}}[k] \mathcal{T}^*[k + m] \end{aligned} \quad (4.14)$$

We then compute a singular value decomposition (SVD) of M :

$$\begin{aligned}
M &= U\Sigma V^\dagger \\
M^\dagger &= V\Sigma^*U^\dagger
\end{aligned} \tag{4.25}$$

where U and V are unitary matrices (i.e., $U^\dagger = U^{-1}$ and $V^\dagger = V^{-1}$), and Σ is a diagonal matrix containing the singular values of M . Plugging in for S ,

$$\begin{aligned}
S &= U\Sigma V^{-1}V\Sigma^*U^{-1} \\
&= U\Sigma\Sigma^*U^{-1} \\
&= U\|\Sigma\|^2U^{-1}
\end{aligned} \tag{4.26}$$

and comparing to the standard diagonalization of S ,

$$S = U\|\Sigma\|^2U^{-1} = T\Lambda T^{-1} \tag{4.27}$$

we conclude that the eigenvalues of S are given by the squared absolute value of the singular values of M , and the eigenvectors of S are given by the left singular vectors of M :

$$\begin{aligned}
T &= U \\
\Lambda &= \|\Sigma\|^2
\end{aligned} \tag{4.28}$$

Note that the all of the eigenvalues of S are nonnegativeⁱⁱ, meaning that no term in the sum in (4.18) can make a negative contribution to the aerial image intensity, which is consistent with the fact that an intensity by definition can never have a negative value.

Summary of the ROCS decomposition procedure

Here we summarize the process for computing the eigenvalues λ_n and eigenvectors ξ_n of S :

1. Construct M using $M[k, m] = \mathcal{J}_0^{\frac{1}{2}}[m]\mathcal{T}[k+m]$. The columns of M are simply shifted versions of the mask spectrum \mathcal{T} scaled by the square root of the mutual intensity spectrum \mathcal{J}_0 .

ⁱⁱThis is a property of all semipositive definite matrices

2. Compute the singular value decomposition of M .
3. The eigenvalues of S are given by the square of the absolute value of the singular values of M .
4. The eigenvectors of S are given by the left singular vectors of M .

4.5 ROCS in 2D

In 2 dimensions, the system cross-coefficients take the form of a 4-dimensional matrix:

$$S(k, m, n, l) = \iint dp dq \mathcal{J}_0(p, q) \mathcal{I}(k + p, l + q) \mathcal{I}^*(m + p, n + q) \quad (4.29)$$

The 2-dimensional image intensity spectrum is given by:

$$\mathcal{I}(\nu, \nu) = \iint dz dw \mathcal{K}(\nu, \nu) \mathcal{K}^*(\nu - z, \nu - w) S(\nu, \nu - z, \nu, \nu - w) \quad (4.30)$$

To diagonalize the system cross-coefficients, we need to represent it as an operator that acts on a *vector* in \mathbb{R}^N . This can be accomplished by *unwinding* the pupil, mask, and source matrices into column vectors by using the so-called stacking operator Ψ :

$$\Psi : \mathbb{R}^{N \times N} \mapsto \mathbb{R}^{N^2}$$

$$\Psi \left\{ \left[\begin{array}{c|c|c|c} c_1 & c_2 & \dots & c_N \end{array} \right] \right\} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} \quad (4.31)$$

where c_i is the i th column of a matrix. The SCC matrix acting on the stacked pupil now takes the form of a 2-dimensional matrix:

$$\mathcal{S} = \begin{bmatrix} \mathcal{S}(1, 1, 1, 1) & \mathcal{S}(1, 1, 2, 1) & \dots & \mathcal{S}(1, 1, N, 1) & \mathcal{S}(1, 1, 1, 2) & \dots & \mathcal{S}(1, 1, N, N) \\ \mathcal{S}(2, 1, 1, 1) & \ddots & & & & & \vdots \\ \mathcal{S}(2, 1, 1, 1) & & & & & & \\ \vdots & & & & & & \\ \mathcal{S}(N, 1, 1, 1) & & & & & & \\ \mathcal{S}(N, 1, 1, 1) & & & & & & \\ \mathcal{S}(1, 2, 1, 1) & & & & & & \\ \vdots & & & & & & \\ \mathcal{S}(N, 2, 1, 1) & & & & & & \\ \vdots & & & & & & \\ \mathcal{S}(N, N, 1, 1) & \dots & & & & & \mathcal{S}(N, N, N, N) \end{bmatrix}$$

As before, \mathcal{S} is Hermitian and semipositive definite, and can be written in terms of its generating matrices M and M^\dagger . M will now correspond to shifted copies of the stacked mask spectrum $\tilde{\mathcal{T}} = \Psi \{ \mathcal{T} \}$. As before, the presence of a shift by a frequency coordinate (p, q) is designated by a nonzero value of $\mathcal{J}(p, q)$. Note that shifts can occur in both horizontal and vertical directions even though the stacked mask spectrum $\tilde{\mathcal{T}}$ is 1-dimensional; therefore, the shift operation should be implemented before the matrix is stacked.

The Fourier transform of the eigenvectors will again serve as the building blocks for the aerial image. Figure 4.12 shows the Fourier transforms of the first eight most significant eigenvectors unstacked, representing a square contact with disk illumination of $\sigma = 0.6$. The stacked aerial image is given by,

$$\tilde{I}(x) = \sum_{i=1}^P \lambda_i |\mathcal{F}^{-1} \{ \tilde{\mathcal{K}}_{\xi_i} \}|^2 \quad (4.32)$$

and the final 2-dimensional aerial image is obtained by unstacking $\tilde{I}(x)$:

$$I(x) = \Psi^{-1} \{ \tilde{I}(x) \} \quad (4.33)$$

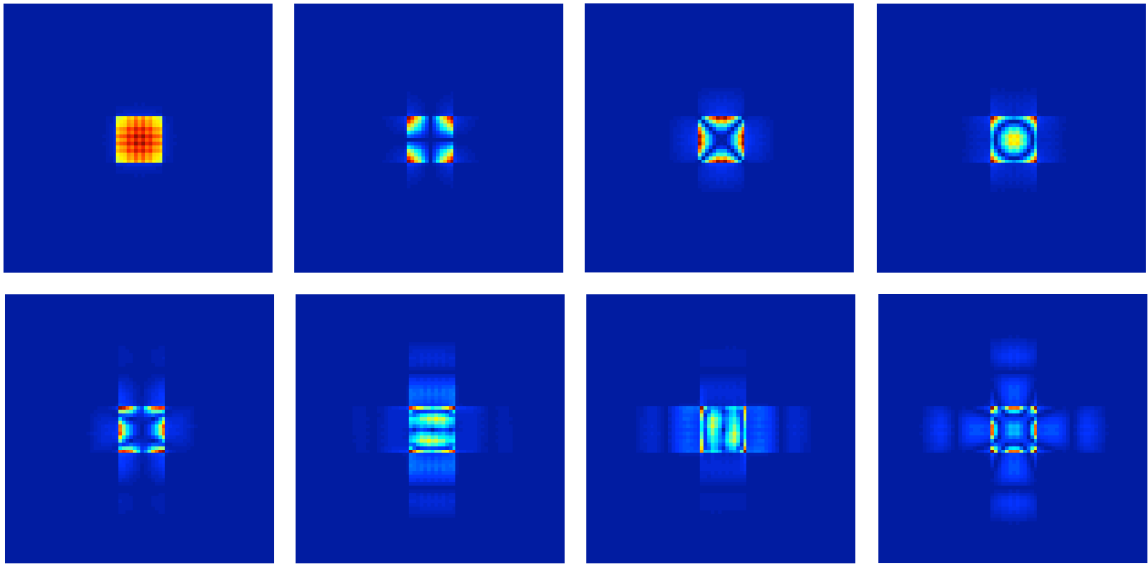


Figure 4.12: An example of the Fourier transforms of the eight most significant eigenvectors S representing a 2-dimensional square contact. Parameters for this series are: $\sigma = 0.6$, $NA = 0.3$, $CD = 150$ nm, and $\lambda = 13.5$ nm.

4.6 Effect of mask and source complexity


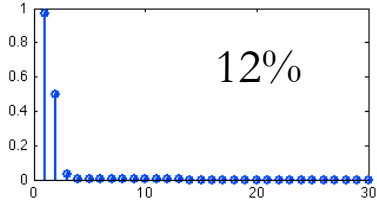
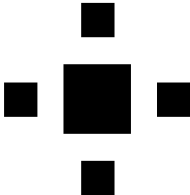
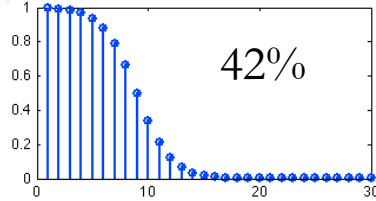
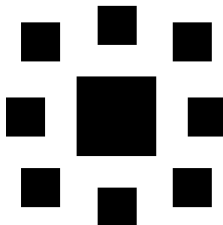
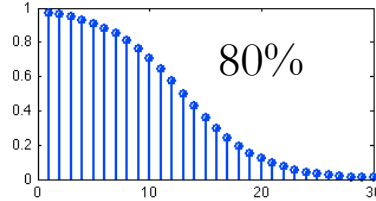
For masks and source configurations that are more complex in nature, the fraction of eigenvectors that must be retained to maintain a fixed error bound increases. This is largely the result of energy compaction property of the singular value decomposition, which groups the largest amount of information into the smallest number of singular vectors. Redundant information, such as a source sampling grid that is much finer than necessary, can be efficiently packaged in a small number of terms.

Often times it is desirable for the source and mask to contain elements that are more complex. Some mask patterns might be relevant to the lithography community, such as an SRAM contact layer. Others may couple into an aberration measurement procedure, where they are designed to specifically target certain aberrations, such as a coma pattern. Sources can be complex as well; in source-mask optimization for instance, a spatial light modulator could be used to define highly complex source profiles in order to achieve a customized coherence function.

In the next chapter, we will describe an algorithm that uses ROCS-generated aerial images

to measure the aberrations in a partially coherent system. In this application, we will be primarily concerned with modeling a standard lithography or imaging system, where the source coherence is likely to be defined by a disk or annular illumination. The complexity of the mask patterns, however, will vary depending on how many Zernike terms are desired. Shown in Table 4.1 are examples of some test patterns used in this algorithm.

Table 4.1: The effect of mask features on ROCS efficiency. Calculation times based on a 2.4Ghz dual core machine. The ROCS error bound is set at 1%.

Feature type	Eigenvalue threshold for 1% error	Typical computation time	Aberration sensitivity
 Simple contact	 12%	0.13s	Astigmatism and focus
 Contact cross	 42%	0.44s	+ Coma and spherical
 Contact octal	 80%	0.82s	+ Higher order

The simple contact pattern has a very small eigenvector sum due to its symmetry and small Fourier domain extent. The contact is a reliable feature when analyzing focus error or astigmatism. The effect of coma on the contact is difficult to measure because the contact can interact with the linear part of the coma which manifests as a translation rather than a distortion of the contact shape. The contact cross requires a larger number of eigenvectors, but gains sensitivity to more aberrations since contacts of different sizes will experience a different relative shift or focus shift in the case of coma and spherical aberration respectively. The contact octal requires the largest number of eigenvectors, and is not much faster than the full Hopkins calculation. Having contacts on the diagonals however gains sensitivity to aberrations with $n > 8$.

The choice of pattern will depend largely on the specific application. For example, in the aberration reconstruction procedure, in order to compensate for the longer computation times of the more complex features, we can increase the error bound to 3 or 5% to gain significant savings in the eigenvector sum. By doing so, we might get a coarse estimate of what the aberrations are, and then push down the error at a later time. This flexibility of the ROCS error allows it to adapt to many different applications.

4.7 Comparison to conventional aerial image models

Because the ROCS decomposition has a customizable error bound, the tradeoff between accuracy and speed can be decided on a case to case basis depending on the nature of the application. Of course, the error bound can always be set to 0, in which case ROCS reduces to the full Hopkins calculation.

4.7.1 Aerial image comparison

To test the aerial images computed by ROCS, they are compared with aerial images generated by MAST, an aerial image modeling software toolbox for MATLAB developed in-house at the Center for X-ray Optics. The optical system is modeled to have a numerical aperture of 0.3, with a coherence function defined by a disk illumination with $\sigma = 0.6$. The mask consists of an isolated contact with a critical dimension of 150 nm.

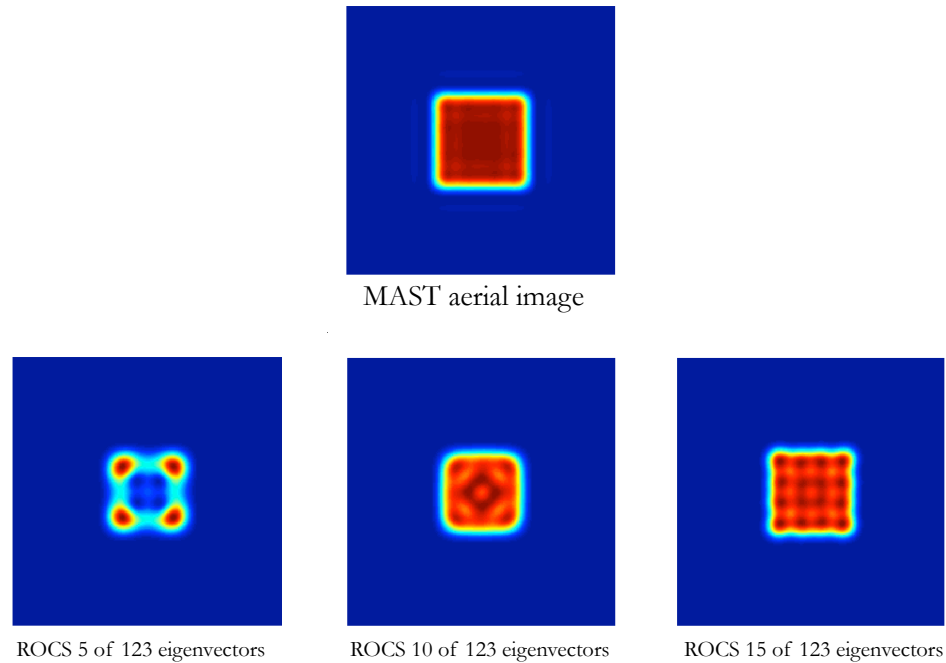


Figure 4.13: Building the aerial image up from its eigenvectors. Even after only 10 of the 123 eigenvectors, the ROCS-approximated image is beginning to resemble the true aerial image.

Figure 4.13 shows how the sum in (4.32) converges to the true aerial image in the absence of aberrations. Due to the large magnitudes of the first few eigenvalues, the eigenvector sum begins looking remarkably similar to the true aerial image even after only the first 10% of the terms have been summed.

Figure 4.14 shows a comparison of aerial images with two separate pupil aberration functions calculated with MAST and with ROCS using a 1% error bound. The first set of images is using a pupil map consisting of astigmatism and coma. The second pupil map has higher order Zernike terms of trefoil and secondary astigmatism.

4.7.2 Analytic form of the error in ROCS

From Figure 4.14, the ROCS approximate aerial image is highly faithful to the true aerial image generated by MAST. One of the unique features of ROCS is that the error in the approximate aerial image is well characterized by the truncated sum of eigenvectors of \mathcal{S} . This allows us to write down the difference function between the approximate and true aerial

images in closed form:

$$I_{MAST}(x, y) - I_{ROCS}(x, y) = \sum_{i=K^*+1}^P \lambda_i |\mathcal{F}^{-1}\{\tilde{\mathcal{K}}\xi_i\}|^2 \quad (4.34)$$

Because this error is described analytically, one can also tailor the ROCS sum to include or discard particular eigenvectors to optimally describe a particular region of interest on the mask. For instance, in a corner-rounding study, one might be interested in the regions surrounding the corner-like features of a pattern, but may not care about the other regions. In this case, the set of eigenvectors of \mathcal{S} can be partitioned into two sets, the relevant eigenvectors and the irrelevant eigenvectors as follows:

$$\{\xi_n\} = E_R \cup E_I \quad (4.35)$$

The ROCS aerial image is then computed as:

$$I(x, y) = \sum_{\xi_i \in E_R} \lambda_i |\mathcal{F}^{-1}\{\tilde{\mathcal{K}}\xi_i\}|^2 \quad (4.36)$$

Again, the error is well defined, and is given by:

$$E(x, y) = \sum_{\xi_i \in E_I} \lambda_i |\mathcal{F}^{-1}\{\tilde{\mathcal{K}}\xi_i\}|^2 \quad (4.37)$$

4.7.3 Comparison of MAST and ROCS computation times

The time required to compute an various numbers of aerial images is shown in Table 4.2. For this comparison, the mask and source are defined as they were above in the aerial image comparison. Because ROCS requires the one-time additional step of computing the singular value decomposition on the generating matrix M , ROCS is slower than MAST overall for small numbers of aerial images. However, for computing large series of images, ROCS quickly surpasses MAST in computation times.

Table 4.2: A comparison of aerial image computation times.

# of aerial images	MAST	ROCS	speed ratio
0 (overhead time)	0s	7s	-
1	1.1s	7.1s	0.2
10	11.2s	8.3s	1.3
100	112.4s	20.1s	5.6
1,000	1,123.6s	137.1s	8.2
10,000	11,242s	1,311s	8.57

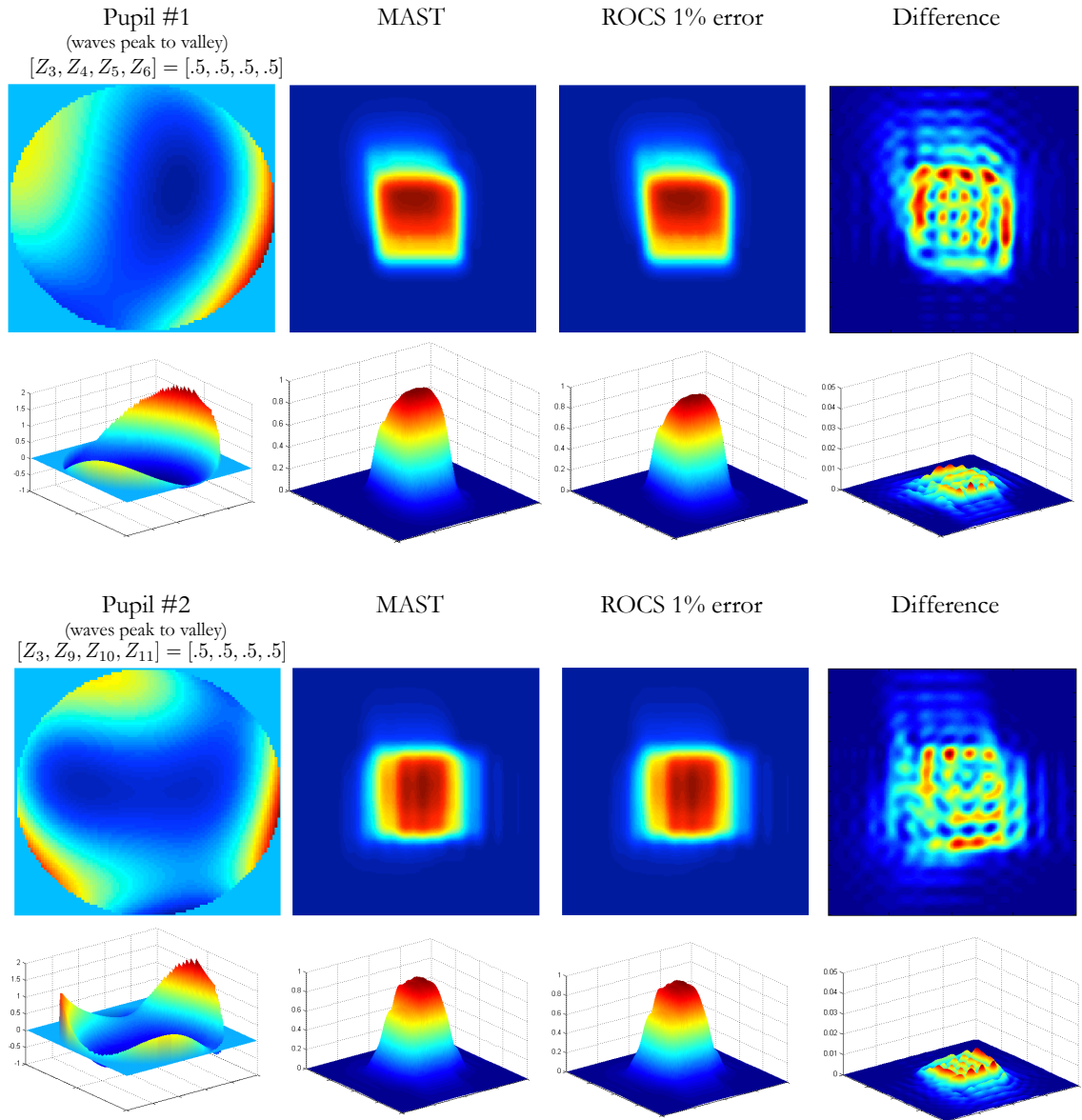


Figure 4.14: A comparison of MAST and ROCS aerial images using two sets of aberrations.

4.8 Summary

In this chapter we have demonstrated a novel decomposition of the Hopkins equation called ROCS. ROCS performs an eigenvalue decomposition of the system cross-coefficients matrix, which allows the rapid computation of aerial images in systems where the source coherence and the mask are fixed from image to image. In this regard, ROCS plays a parallel role to the SOCS decomposition for optical proximity correction.

When ROCS is used to approximate the aerial image by truncating its eigenvector sum, significant speed improvements can be realized. The error associated with this tradeoff is well-characterized and highly controllable, allowing the user to decide regions and the extent to which the error manifests.

The next chapter will deal with applications of ROCS in wavefront characterization for partially coherent optical systems.

Chapter 5

ROCS-based reconstruction of wavefront aberrations

Contents

5.1 Basic idea and experimental setup	90
5.2 Programming a defocus as an aberration	92
5.2.1 Linear transformation from the binomial basis to Zernike basis . . .	94
5.2.2 Approximation error	95
5.3 Search algorithms	96
5.3.1 Pattern search	98
5.3.2 Simulated annealing	99
5.4 Experiment: The SEMATECH Berkeley AIT	104
5.4.1 Setup	104
5.4.2 Search algorithm	105
5.4.3 Results	106
5.5 Discussion	110

In this chapter we apply the ROCS decomposition in an iterative procedure to solve for the aberrations of an optical system with arbitrary source coherence. In this procedure, a through-focus aerial image data series is compared with a ROCS-generated aerial image series with

an initial guess of pupil aberrations. A merit function describing the difference between the two series is used to guide a new guess of aberrations and the process is repeated iteratively until a desired tolerance is met. This iterative approach is applicable to nearly all systems with a well characterized source coherence where an aerial image is measured, including microscopes, actinic mask inspection tools, and lithography tools with aerial image monitors. Since there is no explicit dependence on numerical aperture, it scales well to next-generation high numerical aperture EUV optical systems.

Due to the ability of the ROCS decomposition to rapidly compute aerial images, this procedure is well-suited to analyze a large number of data points, which is ideal for optical systems that have strong field-dependent aberrations such as the SEMATECH Berkeley Actinic Inspection Toolⁱ [34]. Experimental data from the AIT will be presented at the end of the chapter.

5.1 Basic idea and experimental setup

A schematic representation of a typical EUV optical system is shown in Figure 5.1. Light radiates from an extended incoherent source placed in the rear focal plane of a condenser optic, creating a Kohler illuminator. The illuminated test pattern is imaged by the test optic onto the detector.

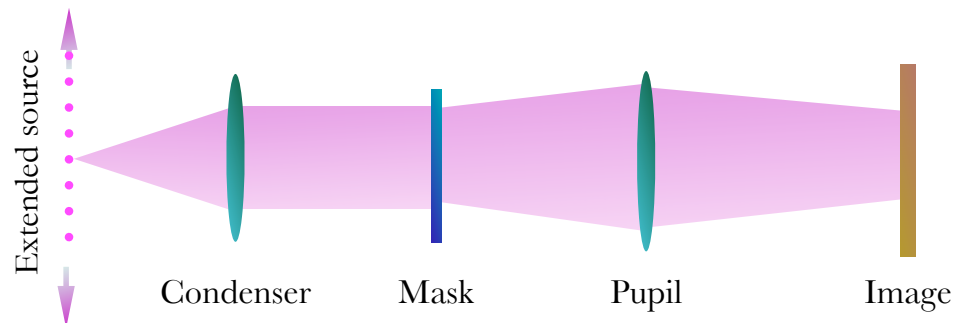


Figure 5.1: A model of a typical optical system.

Experimentally, a through-focus series of images can be obtained in a number of ways.

ⁱThe SEMATECH Berkeley AIT is operated by the Center for X-ray Optics at beamline 11.3.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory

In a microscope, the detector stage may be fixed and focus is adjusted by translating the mask position in z . In the case of the SEMATECH Berkeley AIT, through-focus images are obtained by making small adjustments to the wavelength of the light; since the imaging optic is a zone plate lens—which is highly chromatic—small wavelength changes manifest as focus shifts [34]. In a lithography setup, the detector stage might be translated to create the effect of a focus change.

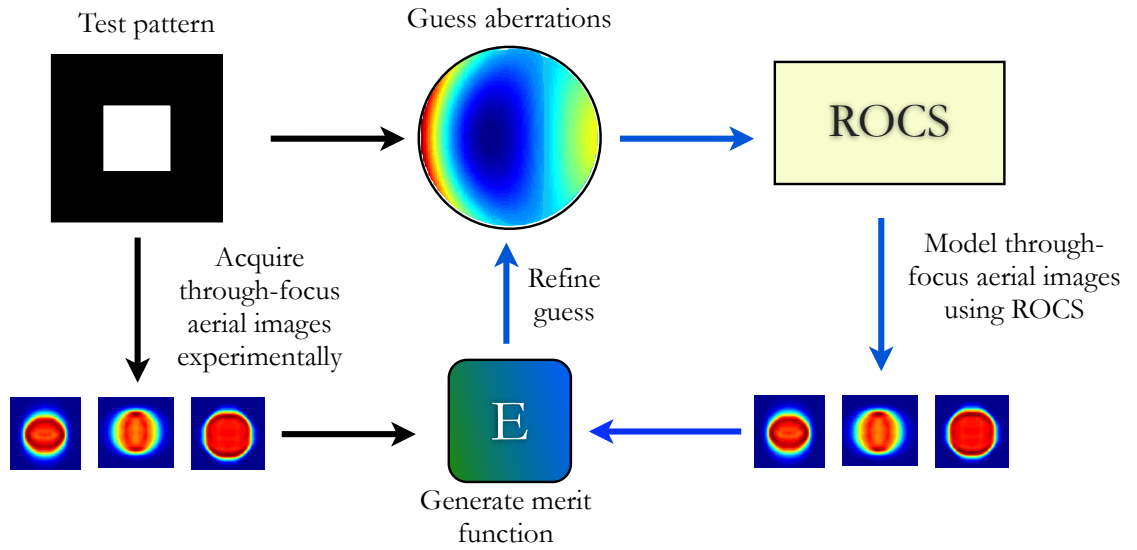


Figure 5.2: A ROCS-based iterative procedure for wavefront reconstruction. Each iteration uses a new pupil aberration to compute a through-focus aerial image series, which is compared with experimental data.

The mask pattern is modeled in the computer and fed into the iterative procedure outlined in Figure 5.2. In this procedure, a trial vector of pupil aberrations is initially chosen. These aberrations determine the pupil coherent transmission function $\mathcal{K}(\nu, \nu)$ which is used by the ROCS algorithm to produce a through-focus set of modeled aerial images. ROCS computes defocused images by adding or subtracting the Zernike decomposition of the phase that corresponds to the desired focus shift. The details of this calculation will be discussed in section 5.2.

Using a pupil aberration vector \vec{v} , the ROCS-modeled aerial images $\{I_{R_n}(x, y, \vec{v})\}$ are compared with the experimentally obtained aerial images $\{I_{E_n}(x, y)\}$, and a merit function is generated, which represents the sum of the integrated squared error between corresponding

images at each defocus step n .

$$M(\vec{v}) = \sum_n \sum_{x,y} [I_{R_n}(x, y, \vec{v}) - I_{E_n}(x, y)]^2 \quad (5.1)$$

The merit function is a measure of how close the modeled aerial images are to the experimental ones. The value $M(\vec{v}) = 0$ represents an exact match, meaning that insofar as the pupil aberrations uniquely specify the through focus series, the trial aberration vector \vec{v} is the desired solution. In practice, the presence of noise will prevent the merit function from ever reaching 0, in which case the goal is to find \vec{v} that gives the global minimum for M .

The flow diagram in Figure 5.2 illustrates the basic iterative procedure. The process is not a feedback loop in a strict sense because no information about the differences in the aerial image series is used to make specific changes in the trial vector of pupil aberrations. Therefore, unlike many other iterative procedures where each iteration converges toward the solution, there is no definitive way to know whether a given step will reduce or increase the merit function. This poses an interesting problem as to how to use the values of the merit function to form a strategy for finding the global minimum of M . A discussion of this strategy is handled in section 5.3.

5.2 Programming a defocus as an aberration

When translating the mask or the detector stage to create defocus experimentally, it is necessary to express the translation as an effective aberration. In the case of a detector translation, we mimic the displacement δ as a shift in the focused image plane by $-\delta$. In other words, if a spherical wave nominally converges to a point on a detector a distance z away and the detector is moved to a the position $z - \delta$, the image would be the same as if the detector remained fixed and the spherical wave was made to converge to a point $z + \delta$ as shown in Figure 5.3

This approximation is valid as long as $z \gg \delta$ which is generally the case in practice. To proceed, we need to find an *aberration* that turns a spherical wave converging to a point z into a new wave that converges to a point $z + \delta$. Let $S_i(\rho)$ and $S_f(\rho)$ be monochromatic, spherical waves with wavelength λ that converge to points z and $z + \delta$ respectively.

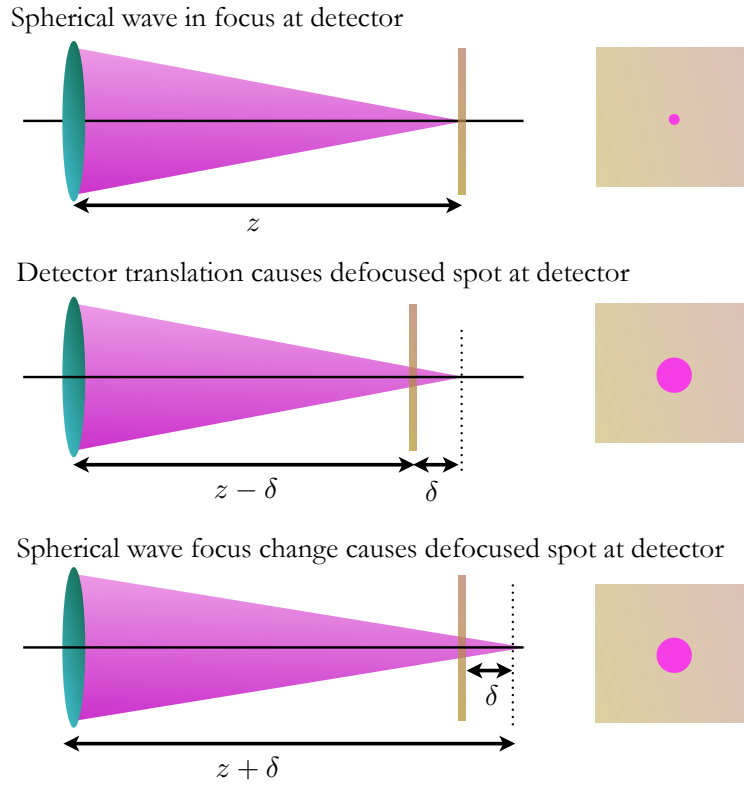


Figure 5.3: Equating detector translation to a shift in the focus of a converging spherical wave.

$$S_i(\rho) = \exp \left\{ \frac{-2\pi i}{\lambda} \sqrt{z^2 + \rho^2} \right\} \quad (5.2)$$

$$S_f(\rho) = \exp \left\{ \frac{-2\pi i}{\lambda} \sqrt{(z + \delta)^2 + \rho^2} \right\} \quad (5.3)$$

Here $\rho = \sqrt{x^2 + y^2}$ is the polar radial coordinate defined in the plane normal to the optical axis. We would like to use an aberration $\exp[2\pi i\Phi(\rho)]$ to change a spherical wave converging to a distance z into a spherical wave converging to a distance $z' = z + \delta$.

$$\exp \left\{ \frac{2\pi i}{\lambda} (z^2 + \rho^2)^{1/2} \right\} = \exp \left\{ \frac{2\pi i}{\lambda} [(z + \delta)^2 + \rho^2]^{1/2} \right\} \times \exp[2\pi i\Phi(\rho)] \quad (5.4)$$

Solving for $\Phi(\rho)$ and ignoring the bulk phase term, we have,

$$\Phi(\rho) = \frac{1}{\lambda} [((z + \delta)^2 + \rho^2)^{1/2} - (z^2 + \rho^2)^{1/2}] \quad (5.5)$$

We write (5.5) as a Taylor series in ρ ,

$$\begin{aligned}\Phi(\rho) &= \sum_{n=1}^{\infty} \frac{\rho^n}{n!} \frac{d^n}{d\rho^n} \Phi(\rho) \Big|_{\rho=0} \\ &= \frac{1}{\lambda} \sum_{n=1}^{\infty} a_n \rho^{2n} \left[\frac{1}{(z+\delta)^{2n-1}} - \frac{1}{z^{2n-1}} \right]\end{aligned}\quad (5.6)$$

where $a_1 = 1/2$, and a_n for $n > 1$ is given by:

$$a_n = -\frac{1 \cdot 3 \cdot 5 \cdots (2n-3)}{(-2)^n \cdot n!} \quad (5.7)$$

In (5.6), using that $z \gg \delta$, the bracketed expression approximates the derivative of $f(z) = z^{-(2n-1)}$:

$$\begin{aligned}\left[\frac{1}{(z+\delta)^{2n-1}} - \frac{1}{z^{2n-1}} \right] &= f(z+\delta) - f(z) \approx \delta f'(z) \\ &= \frac{-(2n-1)}{z^{2n}} \cdot \delta\end{aligned}\quad (5.8)$$

Introducing the dimensionless coordinate $\rho = \rho/(NA \cdot z)$, and defining modified coefficients:

$$b_n = -a_n(2n-1) = \frac{(2n-3)!!}{(-2)^n \cdot n!} (2n-1) \quad (5.9)$$

(5.6) can be rewritten concisely as:

$$\boxed{\Phi(\rho) = \frac{\delta}{\lambda} \cdot \sum_{n=1}^{\infty} b_n (\rho \cdot NA)^{2n}} \quad (5.10)$$

As expected, there are only spherical terms in the expansion. Note that each of the coefficients of ρ^2 scale linearly with the focus step δ . The first few terms in the difference of spheres expansion are:

$$\Phi(\rho) = \frac{\delta}{\lambda} \left[-\frac{1}{2} \rho^2 NA^2 + \frac{3}{8} \rho^4 NA^4 - \frac{5}{16} \rho^6 NA^6 + \dots \right] \quad (5.11)$$

5.2.1 Linear transformation from the binomial basis to Zernike basis

Let $\beta = \{\rho^2, \rho^4, \rho^6, \dots\}$ be the power basis for spherical decomposition, and let $\beta' = \{Z_3(\rho), Z_8(\rho), Z_{15}(\rho), \dots\}$ be the Zernike basis for spherical decomposition. Then by the

definition of the Zernike polynomials, we can write down the change of basis linear transformation:

$$[T]_{\beta}^{\beta'} = \begin{pmatrix} 2 & -6 & 12 & -20 & \dots \\ 0 & 6 & -30 & 90 & \dots \\ 0 & 0 & 20 & -140 & \dots \\ 0 & 0 & 0 & 70 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (5.12)$$

To convert the vector of coefficients $[\vec{\mathbf{b}}]_{\beta}$ into coefficients of Zernike polynomials $[\vec{\mathbf{b}}]_{\beta'}$, we apply the inverse map:

$$\begin{aligned} [\vec{\mathbf{b}}]_{\beta'} &= [T]_{\beta'}^{\beta} [\vec{\mathbf{b}}]_{\beta} \\ &= \{[T]_{\beta}^{\beta'}\}^{-1} [\vec{\mathbf{b}}]_{\beta} \end{aligned} \quad (5.13)$$

An important consequence of this result is that in the Zernike basis, the aberration coefficients that best approximate the difference of spheres in (5.5) depend on number of terms that are considered. As an illustration, take $\delta/\lambda = 1$ and $NA = .4$. Approximating $\Phi(\rho)$ with a single term would give:

$$\Phi(\rho) \approx \{[T]_{\beta}^{\beta'}\}^{-1} \cdot \frac{\delta}{\lambda} \left[\frac{1}{2} NA^2 \vec{\beta}_1 \right] = -.40 Z_3(\rho) \quad (5.14)$$

However, using two terms to approximate $\Phi(\rho)$

$$\Phi(\rho) \approx \{[T]_{\beta}^{\beta'}\}^{-1} \cdot \frac{\delta}{\lambda} \left[\frac{1}{2} NA^2 \vec{\beta}_1 - \frac{3}{8} NA^4 \vec{\beta}_2 \right] = -.35 Z_3(\rho) + .02 Z_8(\rho) \quad (5.15)$$

5.2.2 Approximation error

To analyze the error in the low order approximations, we plot the rms deviation of the approximated wavefront from a true difference of spheres as a fraction of the rms curvature as a function of NA . We take $\delta/\lambda = 10$ and let NA vary from 0 to 0.6.

From the graph we see that using only the defocus term will give a 7% error at $NA = 0.3$. At $NA = 0.5$, 4 terms are needed to bring the error under 1%.

In practice, we can implement ROCS to take in an arbitrarily defined pupil function, so creating the defocus is as simple as programming in equation (5.5) as an added aberration.

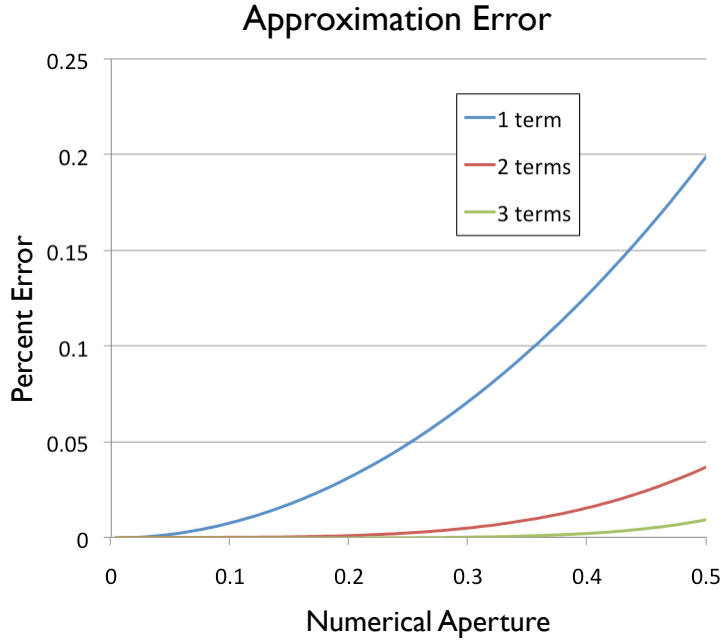


Figure 5.4: Error associated with truncating the defocus sum versus numerical aperture.

However, in applications where Zernike coefficients are desired, up to $NA = 0.5$ the coefficients are given by:

$$\{[T]_{\beta}^{\beta'}\}^{-1} \cdot \frac{\delta}{\lambda} \left[\frac{1}{2} NA^2 \vec{\beta}_1 - \frac{3}{8} NA^4 \vec{\beta}_2 - \frac{5}{16} \vec{\beta}_3 NA^6 + \frac{35}{128} \vec{\beta}_4 NA^8 \right] \quad (5.16)$$

5.3 Search algorithms

Once the merit function (5.1) is computed by comparing the experimental aerial image through-focus series to the images computed by ROCS, we require an optimization strategy for how to choose the next trial aberration vector for each subsequent iteration. The sample space of the pupil aberrations in this optimization consists of the set of N -dimensional vectors where N is the number of Zernike terms in consideration.

$$S = \{\vec{v} \mid \vec{v} \in \mathbb{R}^N\} \quad (5.17)$$

The dimensionality of the sample space can dramatically affect the speed of the search

algorithms. In the pattern search optimization algorithm which will be discussed in the next section, each iteration requires 2^N calculations. Therefore, when choosing N it is important to consider the nature of problem and how many Zernike terms are required to make a meaningful characterization.

When an optical system is misaligned, it is generally the lowest order Zernike terms that are affected the most. A lens tilt, for example, causes both astigmatism ($n = 4$) and coma ($n = 6$). A relative displacement between two lenses in a multiple-lens system can create spherical aberration ($n = 8$). Higher order aberrations such as tetrafoil ($n = 16$) or ashtray astigmatism ($n = 20$) may be indicative of an aberration in the optic itself, and cannot be corrected by an alignment procedure. Although these higher order aberrations will affect the image quality, when characterizing an optical system for the purposes of alignment, it is often the *correctable* Zernike terms are the primary focus of the measurementⁱⁱ.

While technically the optimization is unconstrained, we can safely restrict our search to aberrations that correspond to reasonable limits of an optical system. Extremely large aberrations would cause severe departures in image fidelity from the mask pattern that could be inspected visually. Assuming that this is not the case, we can set a limit on the absolute magnitude of the aberration vector v_{max} . We then can specify the sample space to be the N -dimensional ball of radius v_{max}

$$S = \{\vec{v} \mid \vec{v} \in \mathbb{R}^N, |\vec{v}| < v_{max}\} \quad (5.18)$$

The contour of the sample space depends heavily on the source coherence function, the mask pattern, and the number of defocus steps taken. To illustrate this, assume a disk illumination, and take the case where the mask consists of a rectangularly symmetric feature such as an isolated square contact, and only a single step through focus is considered. In this case an x - y astigmatism error is indistinguishable from a focus error, as both aberrations would create a equivalent broadening of the image in the x - and y -directions. For this configuration, there would be two global minima in the sample space, one at $\vec{v} = Z_3$ and one at $\vec{v} = Z_4$. However if we now consider 3-defocus steps, astigmatism and defocus become distinguishable since the astigmatism will create a focus shift between the horizontal and vertical

ⁱⁱThe SEMATECH Berkeley Microfield Exposure Tool (MET) uses a 2-bounce Schwartzchild optic where only Zernike terms 1-8 can be corrected

edges, whereas in the defocus will broaden horizontal and vertical edges simultaneously.

It is clear from the above example that there is the possibility of the sample space containing local minima. In general, the more defocus steps that are taken, the higher the degree of distinguishability between aberration vectors. Additionally, we can design more complex mask features that breaks the symmetry of some aberrations, although as mentioned in the previous chapter, too complex a feature will hinder the efficiency of the ROCS decomposition. Nonetheless, we cannot consider the sample space to be convex, which means that many algorithms designed for convex optimization cannot be implemented [2].

In this chapter we will consider two search routines. The first routine, known as *pattern search* optimization, is a direct search method that attempts to find the solution by examining the vertices of an N -dimensional cube where the volume gets progressively smaller. Pattern search is a good way to get a coarse mapping of the sample space when no prior information is available about the general location of the solution.

The second, *simulated annealing*, is a type of genetic algorithm that both resistant local minima and highly successful if the solution is within close proximity to the starting vector, as is the case in analyzing field-dependent aberrations where the aberrations of two adjacent cells are similar.

5.3.1 Pattern search

Pattern search optimization is a simple search algorithm that can be effective when relatively little information is available about the sample space. It works by probing several states simultaneously from an initial starting guess \vec{v}_0 where each candidate state is chosen by taking a step of length A_i in the positive and negative direction defined by a basis of orthogonal unit vectors that span the sample space. If the dimensionality of the sample space is N , each iteration requires the computation of $2N$ points since there is a positive and negative direction for each dimension. The first iteration of candidate states will be located at:

$$\vec{v}_{1,j} = \vec{v}_0 + (-1)^j A_0 \hat{x}_k, \quad 0 < j < 2N \quad (5.19)$$

where,

$$k = \left\{ \begin{array}{ll} j/2 & j \text{ even} \\ (j-1)/2 & j \text{ odd} \end{array} \right\}$$

If a state with a lower merit function is found, the state is accepted and the vector length doubles $A_{i+1} = 2A_i$.

$$\vec{v}_i = \{\vec{v}_{i,j} : M(\vec{v}_{i,j}) = \min_k M(\vec{v}_{i,k})\} \quad (5.20)$$

If no state is accepted, a *contraction* occurs in which the search is repeated using vectors with half the previous length $A_{i+1} = A_i/2$. The search ends when the vector length contracts beyond some specified tolerance A_{min} .

Pattern search does not guarantee convergence to the global minimum and the final value will depend heavily on the initial guess and the size of the starting vector length. In our implementation of pattern search, many trials are run in parallel with different initial guesses and the best outcome is chosen after they have all converged. This implementation can be extended to a genetic algorithm where the best result from a series of independent trials acts as a parent for a subsequent generation of trials. Each successive generation uses its parent solution vector as a seed for its starting guesses.

The pattern search path is visualized in 2 dimensions in Figure 5.5 where the candidate points are the vertices of a square. In N dimensions, the candidate points are the 2^N vertices of an N -dimensional cube.

5.3.2 Simulated annealing

Simulated annealing is a search algorithm designed to mimic metallurgic annealing, a process involving the heating and cooling of a crystalline material to reduce its defects or to relocate its impurities to a lower energy configuration. In an annealing process, the material containing impurities is heated to a temperature that significantly increases the mobility of the impurity atoms. The assumption is that from an energy perspective, each atom resides in a local minimum rather than the global minimum. The random thermal motion allows the impurities to *pop* out of their local minima and eventually migrate to a position that is more energetically favorable.

In simulated annealing, the trial vector \vec{v} represents the state of an impurity atom in the crystal. The *energy* of the state is defined as the merit function evaluated at the location of \vec{v} : $E(\vec{v}) = M(\vec{v})$. On each iteration, a candidate state \vec{v}' is chosen at random and the energy difference

$$\Delta E = E(\vec{v}') - E(\vec{v}) \quad (5.21)$$

is computed. The probability that the new state will be accepted is computed based on the acceptance probability function $P(\Delta E, T)$, where T is known as the *temperature*, and $P(\Delta E, T)$ as the following properties:

1. $P(\Delta E < 0, T) = 1$

A candidate state with a lower energy than the current state is always accepted.

2. $T_1 > T_2 \Rightarrow P(\Delta E, T_1) > P(\Delta E, T_2)$

For a given ΔE , the probability of moving to a state with higher energy decreases as the temperature decreases.

3. $\lim_{T \rightarrow 0} P(\Delta E > 0, T) = 0$

As the temperature approaches 0, the probability of leaving the current state for a candidate state of higher energy goes to 0.

The temperature starts at some initial value T_0 and is slowly decreased every few iterations until it eventually reaches 0. The basic idea is that the initial high temperature makes the state resistant to local minima, since a candidate state can be accepted even it resides at a higher energy state. The state bounces around until it approaches the global minima, and as the temperature is lowered it becomes less and less likely to leave, the same way that a real impurity atom finds its final state as it is annealed.

There are many choices of acceptance probability functions that satisfy the requirements above. The function used in the work presented in this chapter is:

$$P(\Delta E, T) = \begin{cases} \exp\{-\Delta E/T\} & : \Delta E > 0 \\ 1 & : \Delta E \leq 0 \end{cases}$$

Figure 5.6 demonstrates a sample simulated annealing path. The initial state is represented by the dark purple ball in a). In b), the candidate state has a lower energy, so it is accepted. The state in b) represents a local minimum, where many other search algorithms would get stuck; in simulated annealing however, a nonzero temperature T gives a finite probability of moving to a state of higher energy. In c), the candidate state is at a higher energy but is accepted anyway due to the nonzero acceptance probability. Finally, in d), the state descends to the global minimum.

One interesting property of simulated annealing is that each time the state changes, it is not necessarily closer to the solution, especially in the beginning when the temperature is high. For this reason, if first few steps are in the wrong direction, it is easy for the simulated annealing algorithm to get lost. This can be alleviated in certain cases such as in the analysis of field-dependent aberrations where neighboring field locations have similar solution vectors. Using the solution to one location as the initial guess for a neighboring location can help the simulated annealing algorithm to start off in the right direction.

Another way to make the problem easier is to simultaneously compute N paths instead of one, and choose the best answer that each path finds. A mechanical analogy illustrated in Figure 5.7 is to consider the sample space to be a vibrating surface with several bouncy balls distributed in a number of initial states. As the vibration (representing the thermal motion) goes to zero, there is a higher likelihood of finding the global minimum if there are many balls at work.

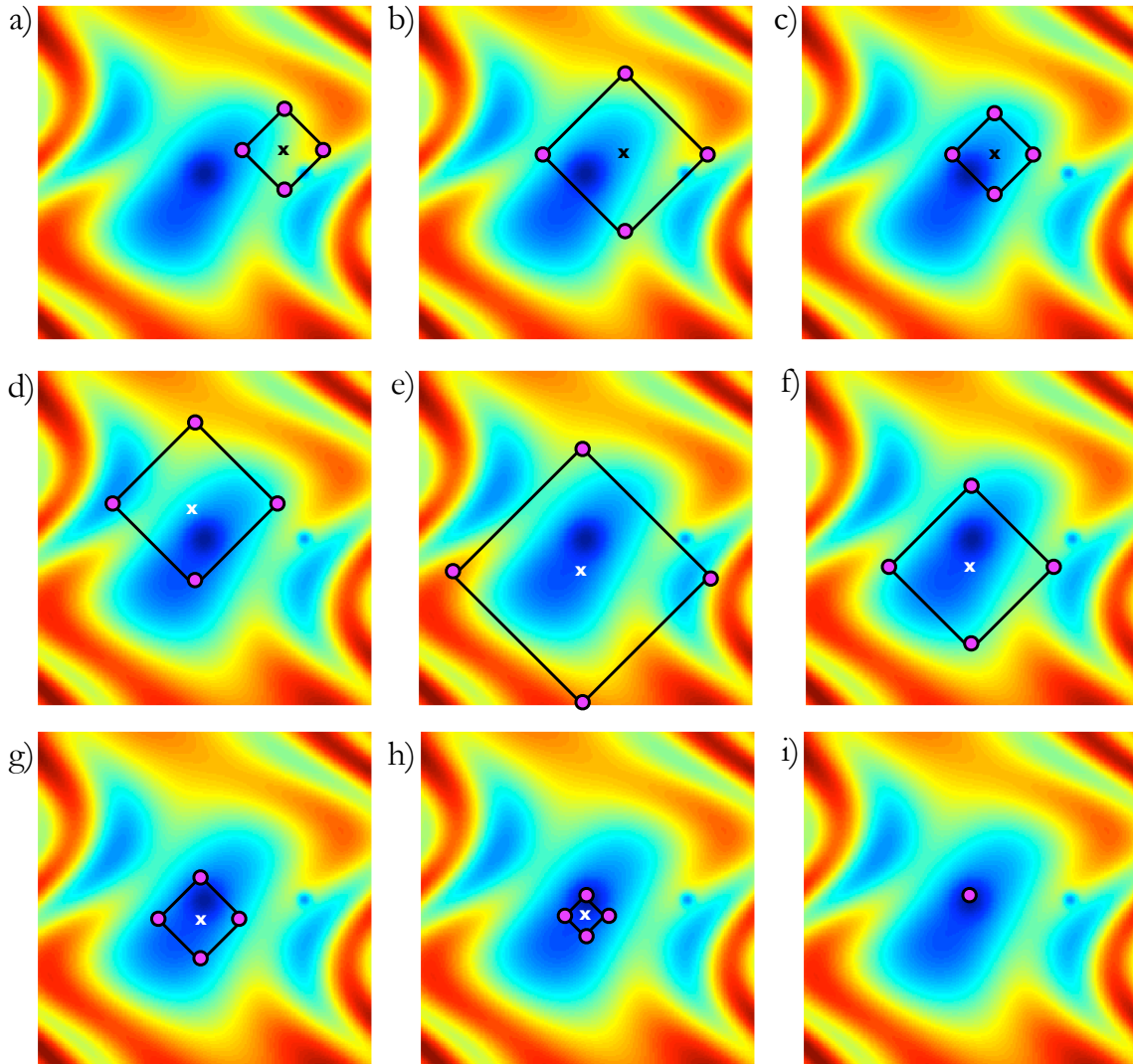


Figure 5.5: A sample pattern search path: a) initial guess, b) accept left, c) contract, d) accept left, e) accept down, f) contract, g) contract, h) contract, i) accept up, state is within tolerance.

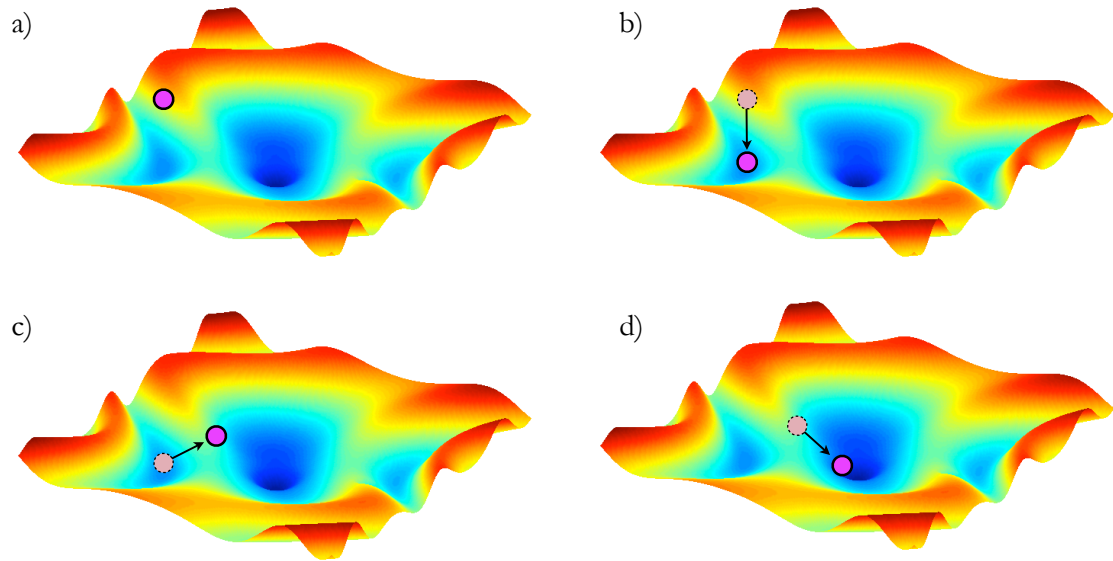


Figure 5.6: A sample simulated annealing state acceptance path.

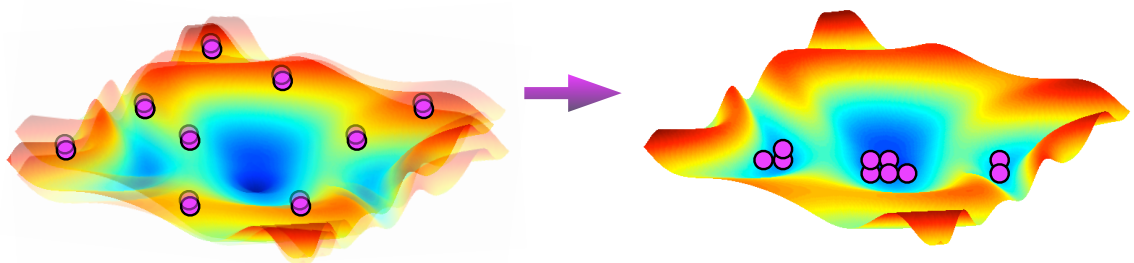


Figure 5.7: A mechanical analogy of simulated annealing. A vibrating surface helps balls find the global minimum.

5.4 Experiment: The SEMATECH Berkeley AIT

5.4.1 Setup

The SEMATECH Berkeley Actinic Inspection Tool (AIT) is an EUV mask inspection microscope that operates at beam line 11.3.2 at the Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory (LBNL). The imaging system consists of a single zone plate optic with a magnification of roughly 900x operating at a center wavelength of $\lambda = 13.4$ nm. Because of the synchrotron source, the illumination is highly coherent with a horizontal sigma of $\sigma_x = 0.2$ and a vertical sigma of $\sigma_y = 0.1$. The system numerical aperture is variable as zone plates can be swapped in and out, however all data presented in this section is taken at $NA = 0.0675$.

A field of 5:1 contactsⁱⁱⁱ with critical dimension 150 nm is imaged at 5 steps through focus where the focal length of the zone plate is adjusted by making fine adjustments to the wavelength of the synchrotron beam. Each focus step is separated by a defocus magnitude of 0.14 waves peak to valley. Due to the field-dependent nature of the wavefront, the aberrations at each contact will be different which allows us to make multiple localized aberration measurements with a single set of data. Since the variation of the aberrations across the field can be determined geometrically, the recovered wavefront at each location can be compared with a numerical model.

The region of interest of the measured field is subdivided into a grid with one contact per grid cell as shown in Figure 5.8. This grid specifies the object functions on which ROCS will operate.

The background noise in the images caused by dark current on the detector is subtracted to remove any intensity energy outside of the immediate contact profile. Failure to remove the noise causes errors in the ROCS decomposition since ROCS will mistake this noise for intensity blur caused by a point-spread function with large extent and attempt to recreate this blur with aberrations. Too much thresholding however hinders the ability to discern the tails in the intensity profile which reduces the sensitivity of the measurement. An example of a thresholded through-focus series of a single contact is shown in Figure 5.9. The data from

ⁱⁱⁱ $N:1$ denotes a periodic field of contacts where the period T is N times the critical dimension: $T = N \cdot CD$

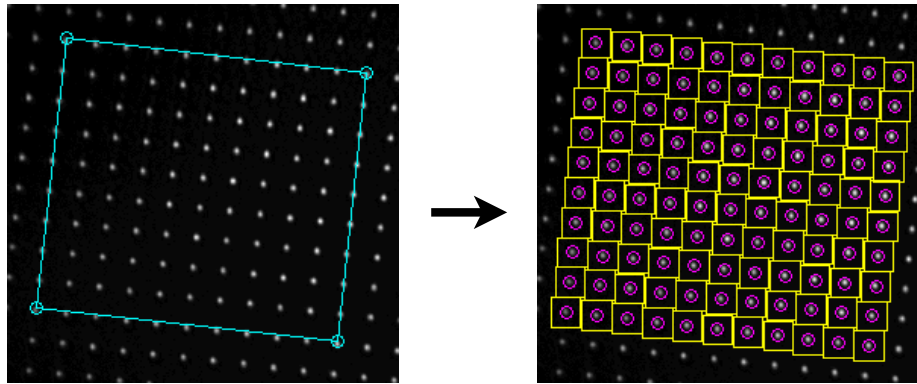


Figure 5.8: Left: The contact field region of interest. Right: The region of interest is divided into several cells that align to each contact location. Images were taken with $\lambda = 13.4$ nm, $NA = 0.0675$, $CD = 150$ nm, and $(\sigma_x, \sigma_y) = (0.2, 0.1)$

the SEMATECH Berkeley AIT is thresholded at 2% of the maximum intensity.

The imaging system is modeled in the computer and fed into ROCS to compute the aerial image through-focus series. In the ROCS setup, the mutual intensity spectrum is taken to be an ellipse with major and minor axes corresponding to the sigma values of the illumination. The SCC is computed in 6 seconds^{iv} with 76 source points, and 9 eigenvalues are retained using a 1.5% error bound on the aerial image intensity.

5.4.2 Search algorithm

The search algorithm consists of a hybrid approach that combines both the pattern search and the simulated annealing algorithms. Utilizing both methods gives the combined benefit of each algorithm; an initial pattern search gives a rough sense of which part of the sample space the solutions lie in, and the subsequent simulated annealing takes advantage of the proximity of the solutions between adjacent cells.

In the first phase, shown in Figure 5.10, a pattern search is performed on each of the contacts using 10 different initial starting guesses and vector lengths. The best result for each contact is chosen as preliminary solution for that contact. In the second phase, we form a set A of aberration vectors corresponding to the lowest 10% of the merit functions from the

^{iv}On a 2.4 GHz dual core machine

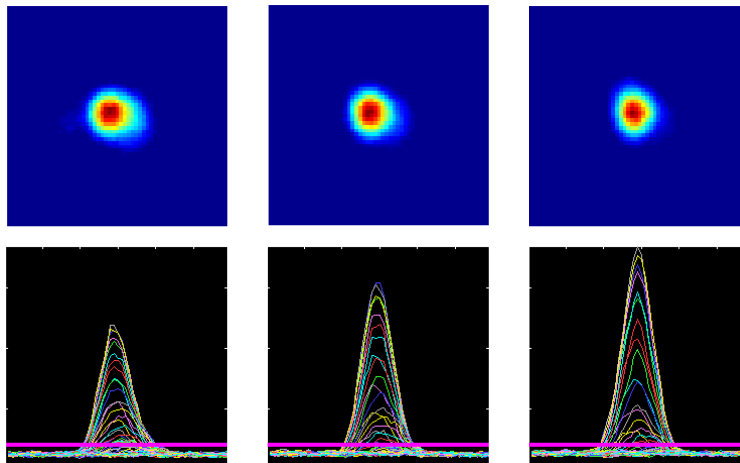


Figure 5.9: A through-focus series of contacts thresholded to remove noise. Top: Aerial images at three focus steps separated by 0.14 waves peak to valley with arbitrary origin. Bottom: Line-out profiles of the through-focus series. The pink line denotes where the threshold is taking place.

pattern search:

$$A = \{\vec{v} \mid M(\vec{v}) \in \text{Lowest 10\% of merit function values}\} \quad (5.22)$$

The vectors in A are used as seed vectors for the simulated annealing step shown in Figure 5.11a. Each of these seed vectors serves as the initial guess for adjacent cells (Figure 5.11b) and the solutions act recursively as initial guesses for the neighboring cells until the all of the contacts in the field have been processed. This step repeats for all vectors in A , and the final solution is the aggregate of all the best solutions for each contact location.

Due to the fact that square contacts have poor sensitivity to Zernike polynomials Z_6 and above, only polynomials Z_1 through Z_5 are reconstructed. To achieve a better sensitivity to higher order aberrations, more complexity is required in the source and object profiles.

5.4.3 Results

Figure 5.12 shows a comparison between the measured field-dependent aberrations and the theoretical model computed geometrically^v. In this particular set of data, there was a small tilt

^vThe geometric model is based on a ray tracing algorithm and was implemented in Zeemax

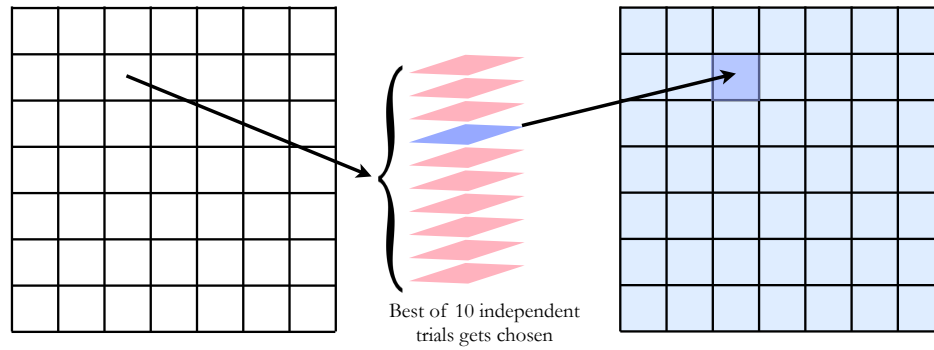


Figure 5.10: Pattern search is applied to each contact 10 times with randomized starting vectors. The best result is chosen as the preliminary solution.

offset due to a rotation in the mask along the optical axis. This offset is determined by fitting the measured contact center of masses to a grid and measuring the angular displacement of the grid axes from the x - y axes. Once this is corrected, the rms error between the theoretical model and the measured aberrations is $E_{rms} = 0.07 \pm 0.04$ waves.

One source of error may be real aberrations in the optical system itself, since the theoretical model is assuming ideal conditions. These system aberrations can be caused by aberrations in the zone plate optic itself, but are more likely to be caused by misalignment of the system. Another source of error may be the exclusion of higher-order Zernike polynomials in the reconstruction algorithm, since higher order aberrations will couple into the measurement. Bandwidth can also play a role, especially given the highly chromatic nature of zone plate lenses, although the spectrum of the SEMATECH Berkeley AIT is highly monochromatic with $\Delta\lambda/\lambda = < 1/1400$ [34].

In order to create an error metric for the aberration reconstruction that is independent of system aberrations, we measure the primary astigmatism of the contacts using a separate technique based on the observed ellipticity of the contacts through focus. In this technique, the best focus plane is found for both x - and y - directions, and the resulting focus shift is converted into astigmatism waves.

The two independent measurements are in good agreement, with an error of $E = 0.05 \pm 0.03$ waves. As before, higher order aberrations remain a possible source of error since they manifest in different ways. In the ROCS-based reconstruction, capping the ROCS decompo-

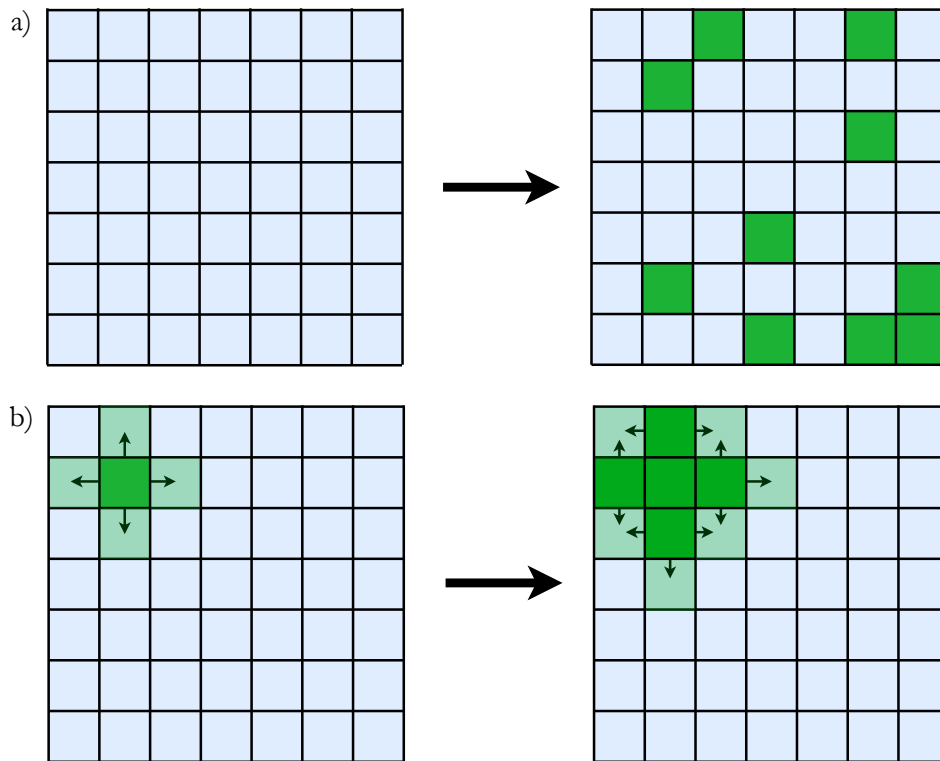


Figure 5.11: a) The best 10% of merit function values from the pattern search are used as seed vectors to the simulated annealing step. b) Simulated annealing is applied iteratively to cells adjacent to the seed vectors in *A*.

sition at 5 polynomials forces it to create the higher order blur using lower order aberrations. In the ellipticity-based reconstruction, higher order aberrations may shift the perceived focus locations which cause errors in the measured astigmatism values. An additional consideration is the lack of focus resolution (since only 5 focus steps were taken) which can also create uncertainty in the position of the foci in the ellipticity-based technique.

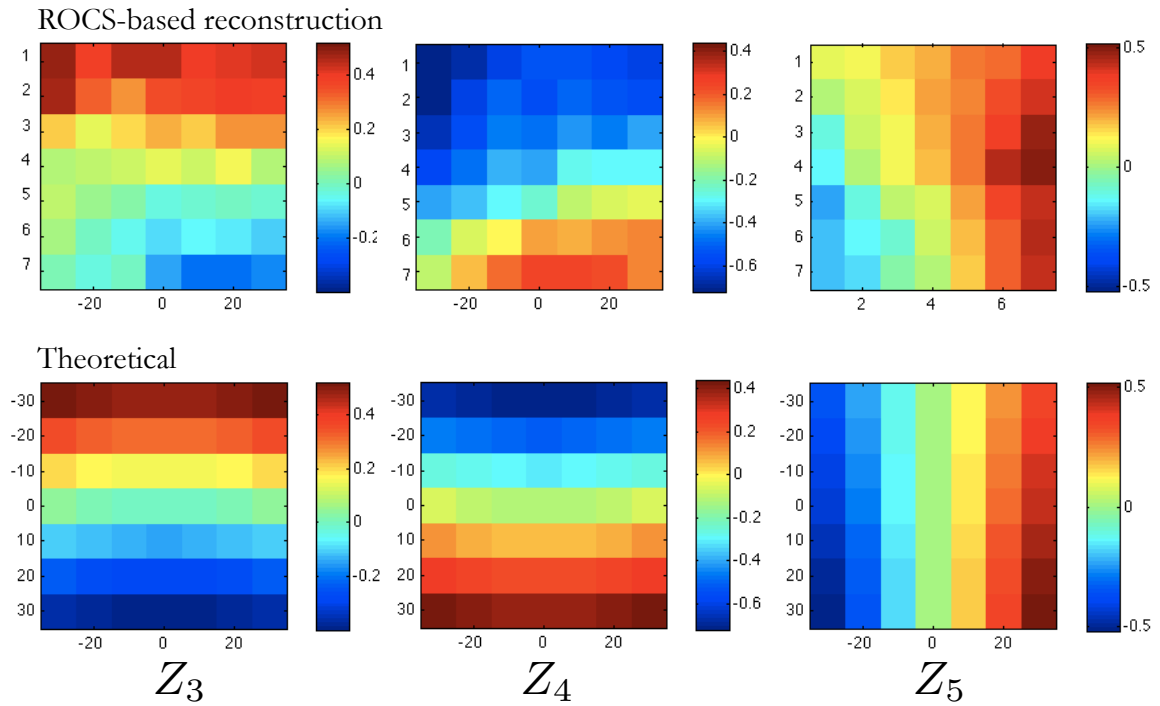


Figure 5.12: Top: Aberrations reconstructed using the ROCS-based iterative algorithm at each contact location. Bottom: Theoretical field-dependent aberrations based on a geometric model. The x and y axes represent the field location of the contacts (1 pixel = 750 nm). All colorbar scales are in waves peak to valley.

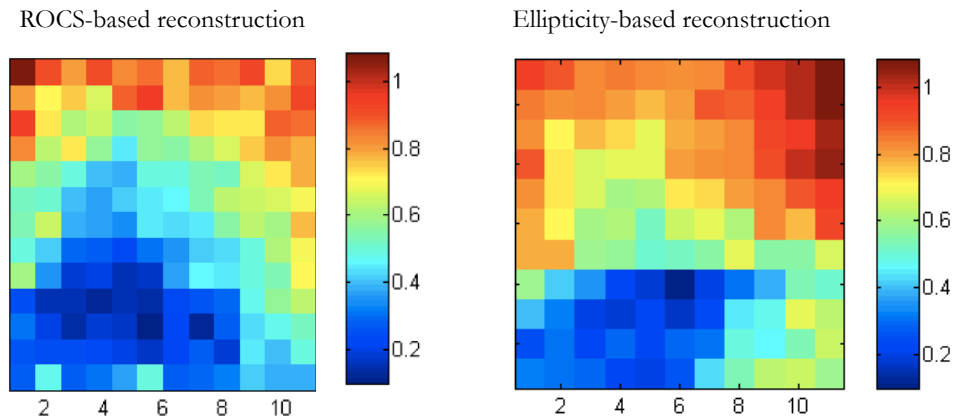


Figure 5.13: A comparison of the total astigmatism in waves peak to valley between the ROCS-based reconstruction and the ellipticity-based reconstruction. The x and y axes represent the field location of the contacts (1 pixel = 750 nm).

5.5 Discussion

The iterative, ROCS-based reconstruction has demonstrated high accuracy wavefront measurement $\lambda_{EUV}/20$ when compared with an independent astigmatism measurement. The ROCS-based reconstruction has the advantage of being compatible with nearly any optical system that produces a measurable aerial image intensity and scales well to the higher numerical apertures that are a key part of next generation EUV lithography. Since the reconstruction is based on images, there is little or no additional setup required to integrate it into the normal imaging workflow.

As with many iterative procedures, convergence of the search algorithms can be susceptible to long computation times depending on the number of Zernike terms and the complexity of the mask and source. Since non-convex searches like simulated annealing tend to converge much more slowly than convex algorithms, the computation of many generations may be required before a desirable tolerance is achieved, although these algorithms are easily parallelized to work on multi-threaded workstations. Finally, there is no guarantee that the final solution is the global minimum of the sample space.

These drawbacks notwithstanding, our preliminary simulations verify the viability of an iterative image-based approach to optical testing. As we continue to move toward higher resolution optical systems, iterative image-based optical testing is a promising alternative to interferometry and may play an important role in the metrology of next-generation optical systems.

Chapter 6

Conclusion

Contents

6.1 Lateral shearing interferometry	111
6.2 ROCS-based wavefront reconstruction	112
6.3 Discussion	113

In this dissertation, two viable methods have been presented for the optical testing of high resolution optical systems. These methods may play an important role in the characterization of optical systems as the lithography community turns to higher numerical apertures to support the continuing shrinkage of device critical dimensions.

6.1 Lateral shearing interferometry

Lateral shearing interferometry was shown to be an attractive alternative to reference wave-based interferometry for a number of reasons. Experimentally, LSI requires the addition of just a single optical element—the diffraction grating—into the existing optical system, which provides a simpler setup and alignment than interferometric tests requiring the multiple additional optical elements. Since the interference in an LSI test is only over the shear distance s , the coherence requirements are more relaxed than in a PS/PDI test, where the entire test optic must be illuminated coherently. Finally, LSI does not require an object-side spatial filter, which can be as small as 12 nm for next generation EUV lithography tools. This allows LSI

to scale to higher numerical apertures without severely affecting the difficulty of fabrication processes.

One of the key challenges in implementing LSI with a diffraction grating is the nonlinear angular response of the grating, which distorts the shape of incident wave. This effect was determined to be negligible at low numerical apertures ($NA < 0.15$), but at medium ($0.15 < NA < 0.35$) and high ($NA > 0.35$) numerical apertures, the null interferogram must be removed from the experimentally measured interferogram to prevent systematic aberrations from corrupting the reconstructed wavefront.

In order to pursue a quantitative investigation of the null interferogram, two mathematical approaches were considered. In the medium numerical aperture regime, a holographic construction derives an analytic form to the systematic aberrations. This analysis allows us to specify tolerances on the tilt of the grating and the detector to be less than 0.15 degrees to ensure a wavefront accuracy of $\lambda_{EUV}/100$ at $NA = 0.35$. In the high numerical aperture regime, a geometrical analysis provides a solution to the null interferogram in the form of a system of transcendental equations that can be solved numerically much more quickly than computing the Huygens-Fresnel diffraction integral.

6.2 ROCS-based wavefront reconstruction

An image-based wavefront reconstruction algorithm was also presented that can characterize the aberrations of an optical system with arbitrary numerical aperture and illumination coherence. This algorithm relies on ROCS, a novel decomposition of the Hopkins equation developed as part of this dissertation, that can rapidly compute aerial images from an imaging system with any illumination coherence, wavelength and NA. This flexibility is one of the key strengths of ROCS, as it allows ROCS to apply to nearly every imaging system with a well-defined coherence function including full-field microscopes, scanning microscopes, mask inspection tools, and lithography tools with an aerial image monitor.

ROCS describes the imaging process by a single operator S representing the system cross-coefficients matrix and has a customizable error bound allowing one to tradeoff aerial image fidelity for significant speed improvements. For aerial image errors of 1 – 3%, the ROCS algorithm computes aerial images up to 15 times faster than the Hopkins integration.

The versatility of ROCS allows the ROCS-based wavefront reconstruction to be compatible with an equally diverse set of optical systems and requires little or no experimental modifications. This reconstruction was used to characterize the field-dependent aberrations of the SEMATECH Berkeley Actinic Inspection Tool (AIT), matching an independent analysis of the astigmatism aberrations to within $\lambda_{EUV}/20$ rms.

6.3 Discussion

The choice of whether to use LSI versus the ROCS-based reconstruction will vary on a case by case basis. LSI is generally the preferred method from a reliability perspective, since it is a direct measurement. The ROCS-based reconstruction does not guarantee convergence to the global minimum of the sample space, and thus may give the wrong answer. LSI has advantages from a speed perspective as well, since the computation of the wavefront can be obtained immediately after the data is taken. The ROCS-based reconstruction can take thousands of iterations before finding the solution, depending on the parameters of the search algorithm. Finally, LSI reconstructs the entire wavefront, whereas the ROCS-based reconstruction can only reconstruct a finite set of Zernike polynomials.

Another key advantage of LSI over the ROCS-based reconstruction is its ability to work in systems where no aerial image is available, such as lithography tools. In principle, the ROCS-based method could be adapted to work with printed resist profiles rather than aerial images. In this modified experiment, the computed through-focus series would be convolved with a resist blur function and thresholded to simulate printing. The resulting line edges would be compared to the printed data to generate the merit function. Further work must be done to assess the viability of this technique, as it presents a more difficult reconstruction than the aerial image-based method.

A major drawback of LSI is that the optical system must be able to accommodate the experimental setup. In the ROCS-based method, there are no restrictions on numerical aperture or illumination coherence, making it extremely versatile. LSI by contrast requires coherence over the shear distance, requiring systems using incoherent sources to pre-filter the light before it reaches the test optic. Although LSI scales well to high NA in theory, actually implementing LSI at high NA can sometimes be difficult. Since the beam is rapidly expand-

ing, finite detector sizes force the diffraction grating to fit in a tight space between the test optic and the detector, which can be experimentally challenging, especially given the strict tolerances on the grating tilt specifications.

In summary, each of these methods has its strengths and their implementation will depend on the system in question. As the lithography community moves to higher resolution optical systems, these techniques provide two more promising ways of pushing technology forward.

Bibliography

- [1] A. MacDowell, O. Wood, and J. Bjorkholm. Interferometric testing of euv lithography cameras. *Proc. of SPIE*, 3152:202, 1997.
- [2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [3] C. Wagner and N. Harned. EUV lithography: Lithography gets extreme. *Nature Photonics*, 4:24–26, 2010.
- [4] O. Pornillos, B. K. Ganser-Pornillos, and M. Yeager. Atomic-level modeling of the HIV capsid. *Nature*, 469:424–427, 2011.
- [5] Photo: Nutfield Technology, Inc. 49 Range Road, Windham NH, 03087.
- [6] R. K. Maloney. The zyoptix xp: An advance in mechanical microkeratome technology. *Refractive Eyecare*, 9(10):10–13, 2005.
- [7] Image: <http://www.ophmanagement.com/article.aspx?article=102055>.
- [8] Image: http://www.journalofoptometry.org/10.3921/img/01_article_figure_01.jpg.
- [9] Figure: http://www.ieee.org/portal/cms_docs/sscs/sscs/07winter/rakeshfig2rn.jpg.
- [10] Photo: <http://micro.magnet.fsu.edu/micro/gallery/computers/computer2.html>.
- [11] Photo: <http://www.dvhardware.net/article25023.html>.
- [12] Figure: <http://electrons.wikidot.com>.

- [13] M. J. Madou. *Fundamentals of microfabrication: the science of miniaturization*. CRC Press, 2002.
- [14] S. Okazaki. Resolution limits of optical lithography. *JVST B9*, 9(6):2829–2833, 2001.
- [15] V. Ronchi. Forty years of history of a grating interferometer. *Applied Optics*, 3(4):437–451, 1964.
- [16] R. Miyakawa and P Naulleau. Lateral shearing interferometry for high-resolution euv optical testing. *Proc. of SPIE*, 7969, 2011.
- [17] P. Naulleau, K. A. Goldberg, and J. Bokor et al. Extreme-ultraviolet phase-shifting point-diffraction interferometer: a wave-front metrology tool with subangstrom reference-wave accuracy. *Applied Optics*, 38(35):7252–7263, 1999.
- [18] H. Meddecki, E. Tejnil, K. A. Goldberg, and J. Bokor. Phase-shifting point diffraction interferometer. *Optics Letters*, 21(19):1526–1528, 1996.
- [19] D. Attwood. *Soft X-rays and Extreme Ultraviolet Radiation*. Cambridge University Press, 1999.
- [20] J. W. Goodman. *Introduction to Fourier Optics, 2nd Edition*. McGraw-Hill Companies, 1996.
- [21] X-ray database at cxro: http://henke.lbl.gov/optical_constants/.
- [22] M. P. Rimmer. Method for evaluating lateral shearing interferograms. *Applied Optics*, 13(3):623–629, 1974.
- [23] D. Malacara, M. Servin, and Z. Malacara. *Interferogram Analysis for Optical Testing*. CRC Press, 2005.
- [24] K. Freischlad and C. Koliopoulos. Fourier description of digital phase-measuring interferometry. *JOSA*, 7(4):542–551, 1989.
- [25] K. G. Larkin and B. F. Oreb. Design and assessment of symmetrical phase-shifting algorithms. *JOSA*, 9(10):1740–1748, 1992.

- [26] <http://www.netlib.org/lapack/>.
- [27] D. Ghiglia and L. Romero. Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods. *JOSA*, 11(1):107–117, 1993.
- [28] N. Cobb. *Fast Optical and Process Proximity Correction Algorithms for Integrated Circuit Manufacturing*. PhD thesis, University of California, Berkeley, 1998.
- [29] N. Cobb, A. Zakhor, M. Reihani, F. Jahansooz, and V. Raghavan. Experimental results on optical proximity correction with variable threshold resist model. *Proc. of SPIE*, 3051:458–468, 1997.
- [30] K. Yamazoe. Computation theory of partially coherent imaging by stacked pupil shift matrix. *JOSA*, 25:3111–3119, 2008.
- [31] J. W. Goodman. *Statistical Optics*. Wiley Classics Library, 2000.
- [32] S. Friedberg, A. Insel, and L. Spence. *Linear Algebra*. Pearson Education, 2003.
- [33] J. Ye and Q. Li et al. Idr/qr: An incremental dimension reduction algorithm via qr decomposition. *IEEE transactions on knowledge and data engineering*, 17(8):1208–1222, 2005.
- [34] K. A. Goldberg et al. Actinic extreme ultraviolet mask inspection beyond 0.25 numerical aperture. *JVST B*, 26(6):2220–2224, 2008.