Modeling, Designing, and Measuring EUV Photomasks

Stuart Sherwin

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2023-37
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2023/EECS-2023-37.html

May 1, 2023

Acknowledgement

This work was performed in part at Lawrence Berkeley National Laboratory which is operated under the auspices of the Director, Office of Science, of the U.S. Department of
Energy under Contract No. DE-AC02-05CH11231.
This research was partially sponsored by Intel Corporation.
This research was partially sponsored by EUV Tech.
This research was partially sponsored by C-DEN (Center for design-enable nanofabrication). Member companies – ARM, ASML, Cadence, Carl Zeiss Group, Intel, KLA, Mentor
Graphics, and Qualcomm.
This research was carried out under the guidance of Laura Waller, Andy Neureuther, and Patrick
Naulleau. Pivotal CXRO staff: Ryan Miyakawa, Isvar Cordova, Markus Benk, and Eric Gullikson. Pivotal graduate students: Gautam Gunjala, Luke

Long, Jonathan Ma, and Wenhua Zhu.

Modeling, Designing, and Measuring EUV Photomasks

by

Stuart Larrick Sherwin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrial Engineering and Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Laura Waller, Co-chair
Professor Andrew Neureuther, Co-chair
Professor Roger Falcone

Fall 2021

Modeling, Designing, and Measuring EUV Photomasks

Abstract

Modeling, Designing, and Measuring EUV Photomasks

by

Stuart Larrick Sherwin

Doctor of Philosophy in Engineering – Electrial Engineering and Computer Science

University of California, Berkeley

Professor Laura Waller, Co-chair

Professor Andrew Neureuther, Co-chair

We present a selection of topics relating to modeling, designing, and measuring EUV (Extreme Ultraviolet) photomasks, with implications for high-volume nanofabrication of integrated circuits. These EUV photomasks must be accurately designed, but rigorously modeling large domains is extremely computationally intensive; we introduce an approximate Fresnel Double Scattering model which is 10,000x faster. This approximation can predict the trend of phase vs pitch, which is critical to designing EUV phase shift masks (PSMs). We also explore novel mask architectures to improve efficiency and contrast, such as an etched multilayer PSM (up to 6x throughput but restrictive applicability), aperiodic multilayers (up to +22% throughput and more general applicability), and multilayers with minimal propagation distance at certain angles (lower throughput but higher contrast with minimized 3D effects). Finally we explore computational metrology with EUV reflectometry, scatterometry, and imaging for probing the phase and amplitude response of an EUV mask, with experimental demonstrations at the Advanced Light Source synchrotron. We perform reflectometry experiments on 3 masks with different architectures to infer approximately 25 physical film parameters each. Another reflectometry application to contamination monitoring achieved single-picometer precision for thickness ($3\sigma < 6$pm) and sub-degree precision for phase ($3\sigma < 0.2^o$). We compare two implementations of phase scatterometry, either applying nonlinear optimization with approximate scattering, or linearizing the rigorous scattering relationship between intensity and phase; linearization is shown to generally be more accurate, but both methods have similar precision. We apply novel software and hardware for phase imaging, using PhaseLift convex phase retrieval, combined with a set of custom Zernike Phase Contrast (ZPC) zone plates. We perform hyperspectral ZPC phase imaging on 3 masks, where we see promising agreement with reflectometry in the trend of phase vs wavelength.

To my loving family who made this all possible: my father Mark, my mother Cathy, and my brother Evan.

Contents

List of Figures

List of Tables

Acknowledgments

This work was performed in part at Lawrence Berkeley National Laboratory which is operated under the auspices of the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

This research was partially sponsored by Intel Corporation.

This research was partially sponsored by EUV Tech.

This research was partially sponsored by C-DEN (Center for design-enable nanofabrication). Member companies – ARM, ASML, Cadence, Carl Zeiss Group, Intel, KLA, Mentor Graphics, and Qualcomm.

This research was carried out under the guidance of my two advisors, Laura Waller and Andy Neureuther, as well as the director of CXRO (Center for X-Ray Optics), Patrick Naulleau. Laura, Andy, and Patrick all contributed immensely over the years in directing the research both in the broader goals and the specific findings. Several staff members at CXRO were especially pivotal in carrying out the experiments: Ryan Miyakawa, who assisted with zone plate designs and imaging experiments; Isvar Cordova, who assisted with reflectivity and scattering experiments; Markus Benk, who assisted and advised on imaging experiments; and Eric Gullikson, who assisted and advised on reflectivity and scattering experiments. Several graduate students also bear mentioning: Gautam Gunjala, who led projects that were unfortunately not included in this thesis related to wavefront metrology based on random scattering; as well as the other graduate students under the guidance of Andy and Patrick: Luke Long, Jonathan Ma, and Wenhua Zhu, with whom I shared many illuminating conversations about research, and also with whom I attended myriad SPIE conferences.

Chapter 1

Inroduction

1.1 Moore's Law

In 1965, Gordon Moore published his famous article [32] "Cramming More Components onto Integrated Circuits," where he noticed a now famous trend. It was becoming more and more economical to manufacture integrated circuits (ICs) with more components. Although at any given point in time a larger IC would be more expensive to manufacture than a smaller one, nonetheless up to a certain point the larger chip could be manufactured at a lower cost per component. Chips with fewer components or more components could always be produced, but would ultimately be at an economic disadvantage compared to chips with the optimal number of components.

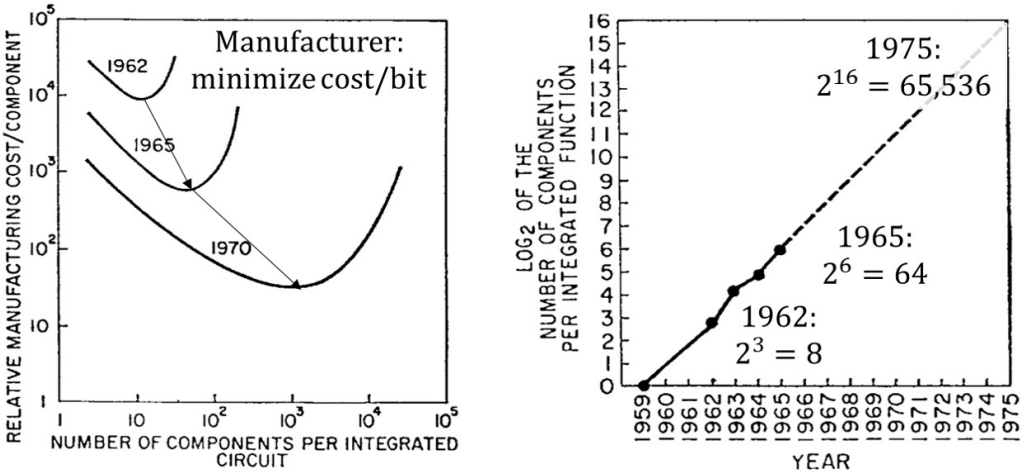

Figure 1.1: Moore's law, published 1965 [32] (annotations mine).

In 1962 that optimal number was about 8 components per IC; just three years later in

1965 when the article was published, it was 64. The entire cost curve had shifted down and to the right—meaning that a chip of any size was uniformly cheaper to manufacture, and also that the reduction in cost had been far greater for larger ICs than smaller ones. The projected curve for 1970 shows the process repeating itself, leading to a 1024 component chip becoming economical 5 years from the date of publication. But how far into the future could the trend really be expected persist? In retrospect Moore's prediction may seem somewhat modest, but in the context in which it was written it was anything but. From only 5 data points starting from one single component per circuit in 1959 to 64 in 1965, he projected out 10 years into the future to 1975, suggesting the optimal number of components per IC would grow a thousandfold to 65,536.

Now with the benefit of hindsight nearly six decades later, we can see the remarkable persistence of this trend. While the rate of doubling the number of components per chip every year was not maintained, in the 50 years from 1970 to 2020, a rate of doubling every two years has consistently been achieved. The rate of doubling for component density has been slightly slower at 2.5 years, and the additional scaling in number of components has been achieved by a more gradual increase in chip size, doubling roughly every 12 years.[52]

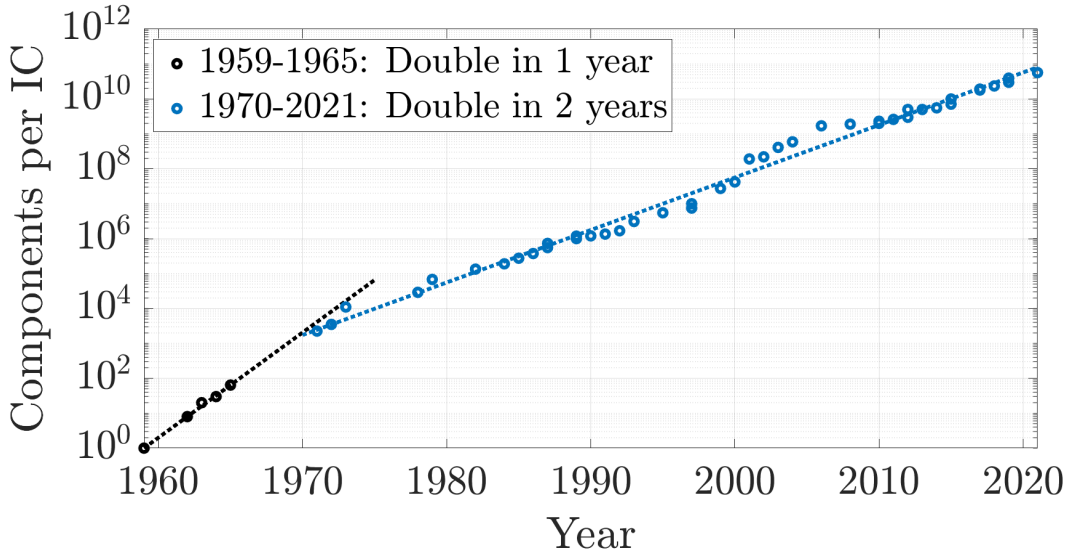

Figure 1.2: Historical progression of Moore's Law with original data from 1959-1965, and additional data from 1970-2021.[32, 52]

But Moore's Law is obviously not a law of nature. Doubling the number of components from one generation to the next requires countless engineers simultaneously to take on Herculean tasks in many disparate fields in order for IC technology to advance. In recent years the pace of improvement has seemed to slow slightly—famously Intel switched from a 2-year to 3-year cycle between nodes—and this rate may well continue to slow down in the years

to come. But the fundamental dynamic defining Moore's original observation will remain true in the sense that every year the cost landscape for IC manufacturing will shift due to technological innovations. As long as the cost curve keeps shifting down, ICs will continue to get cheaper; and as long as the curve keeps shifting right, ICs will continue to get more powerful.

1.2 Photolithography

In order for a newer, more advanced circuit design to be manufactured, there must be constant innovation in manufacturing technologies. Chief among the technologies which historically has dictated the progression of Moore's Law is photolithography, schematically shown in Figure 1.3.

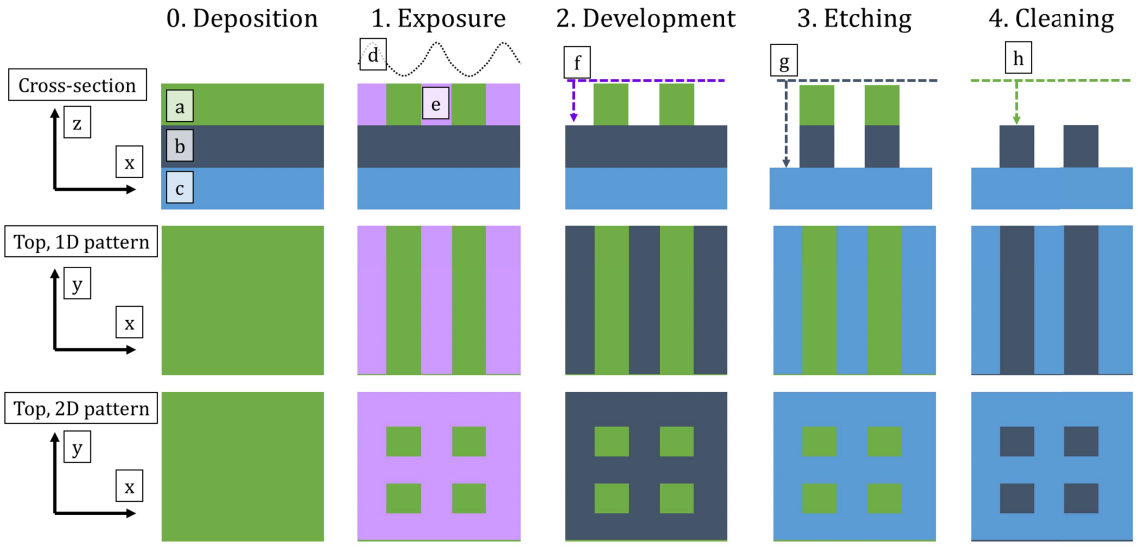

Figure 1.3: Basic steps in photolithography, in cross-section (top), top-down for a 1D line-space pattern (center), and top-down for a 2D contact array pattern (bottom). 0: A sample is prepared with a uniform film of photoresist (a) on top of a material to be patterned (b) which is supported by a flat substrate such as a Silicon wafer (c). 1: An image (d) is projected onto the photoresist, selectively exposing only certain areas to define a pattern (e). 2: The photoresist is developed (f) selectively removing only the exposed areas for positive tone resist, or only the unexposed for negative tone. 3: A vertical etch (g) selectively removes the material only in the areas where the photoresist has been removed by developement, transferring the pattern into the material. 4: The remaining photoresist is cleaned (h), leaving only the desired pattern in the material.

The term photolithography is the combination of three Greek roots: photo (light), litho (stone), and graph (write); literally the term means to write in stone with light, a very apt description. After preparing a sample with a photosensitive film called photoresist on top of a flat layer of material, an image is projected onto the photoresist triggering molecular exposure events which locally alter the photoresist; then a chemical process called development selectively removes either the exposed resist (positive tone) or unexposed resist (negative tone); then a vertical etch transfers the developed pattern to the material; and finally the residual photoresist is removed leaving only the pattern etched into the material.

Photolithography has been crucial to Moore's Law, where the patterns could be transistors, wires, or other electronic components. Whatever function the pattern serves in the final circuit, manufacturing devices with higher density will inevitably require printing smaller features. Therefore, Moore's Law creates a constant need for improvements in lithographic resolution.

1.3 Single-exposure resolution

Lithographic resolution is fundamentally set by the diffraction limit. The maximum spatial frequency for a coherent imaging system is $f_{max} = 1/p_{min} = (NA_{obj} + NA_{ill})/\lambda$, where f_{max} is the maximum spatial frequency, p_{min} is the corresponding minimum pitch, NA_{obj} is the numerical aperture (NA) of the objective lens, NA_{ill} is the NA of the illumination, and λ is the wavelength of the illumination. The resolution is proportional to the minimum pitch; most often in lithography resolution is defined as half of the minimum pitch.

Illumination is partially coherent, composed of many mutually incoherent illumination angles, and typically the maximum illumination angle corresponds to that of the NA of the objective. A version of the diffraction-limited resolution formula specific to lithography is commonly written as $p_{min}/2 = k_1\lambda/NA$ [29]. Here we specify only the NA of the objective ($NA = NA_{obj}$), and include resolution enhancement from off-axis illumination in the k_1 factor.

For on-axis illumination ($NA_{ill} = 0$) the minimum possible k_1 for a 1D pattern is 0.5; when illuminating at the edge of the pupil ($NA_{ill} = NA_{obj}$) the minimum possible k_1 shrinks by a factor of 2 to 0.25. For 2D patterns the resolution limit is somewhat larger; for example, the minimum pitch for a square contact array using on-axis illumination is $k_1 = 1/\sqrt{2} = 0.71$, and using off-axis illumination is $k_1 = 1/2\sqrt{2} = 0.35$. A factor on the order of 10% must also be added for a partially coherent imaging system to prevent diffraction orders being clipped by the collection aperture.

Improving the single-exposure resolution can be done in three ways: increasing NA (larger lenses), decreasing λ (different light source), or decreasing k_1 (off-axis illumination and process control). Up until 1998 with the introduction of $\lambda = 193$nm DUV (Deep Ultraviolet), lithography systems with successively shorter wavelengths of illumination were introduced[22]. At the same time, for each new illumination wavelength, tools with increasingly larger NA were introduced. From 1998 until 2009, although no new wavelength was

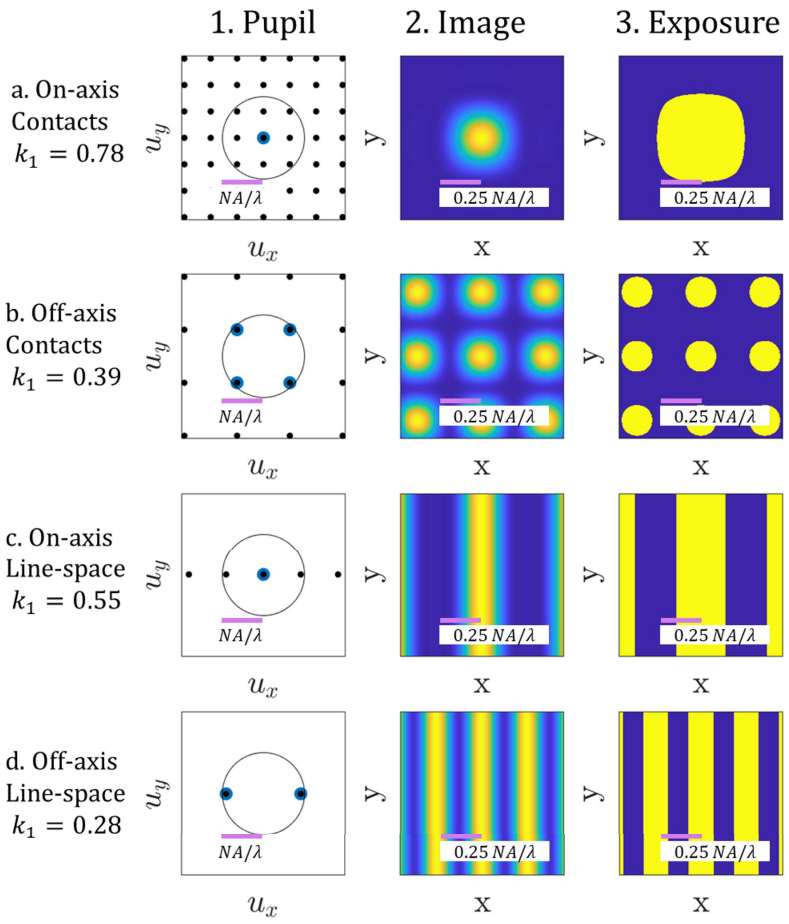

Figure 1.4: Single patterning resolution limit for 4 cases: on-axis contacts (a), off-axis contacts (b), on-axis line-space (c), and off-axis line-space (d). Column 1 shows pupil-space with normalized frequency/angle coordinates, marking illumination coordinates in blue and scattering in black. Column 2 shows the intensity images which expose the photoresist. Finally, column 3 shows the binary image after exposure, assuming a simple intensity threshold model.

introduced, NA scaling continued as subsequent 193nm systems were introduced with NA increasing from 0.6 to 1.35. The final push to an NA greater than 1 was achieved by water immersion due to water's higher refractive index than air; therefore the 1.35 NA tool is often referred to as the 193i (193-immersion).

At the same time as the DUV NA increased, computational lithography became increasingly utilized to enable printing at a smaller and smaller k_1 with optimized off-axis illumination sources. However, following the introduction of the 193i in 2009 there was no further improvement to wavelength or NA. In the following decade, the continued demand

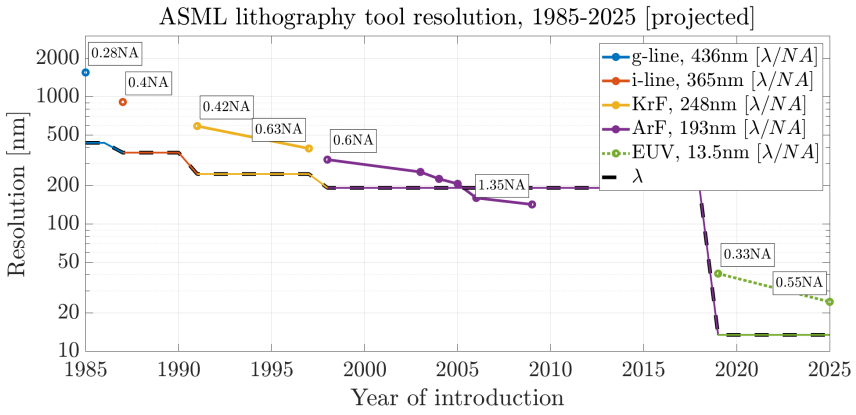

Figure 1.5: Historical scaling of lithographic single-exposure resolution for ASML systems since 1985, with trends of shrinking wavelength and growing NA[22].

Year	**Model**	**Source**	λ **[nm]**	**NA**	**Immersion**	λ **/ NA [nm]**
1985	PAS 2000/10	g-line	436	0.28	1	1557
1987	PAS 2500/40	i-line	365	0.40	1	913
1991	PAS 5000/70	KrF	248	0.42	1	590
1997	PAS 5500/500	KrF	248	0.63	1	394
1998	PAS 5500/900	ArF	193	0.60	1	322
2003	AT:1150i	ArF	193	0.75	1	257
2004	XT:1250i	ArF	193	0.85	1.33	227
2005	XT:1400i	ArF	193	0.93	1.33	208
2006	XT:1700i	ArF	193	1.20	1.33	161
2009	XT:1950i	ArF	193	1.35	1.33	143
2019	NXE:3400C	EUV	13.5	0.33	1	40.9
2025*	EXE:5000	EUV	13.5	0.55	1	24.5

Table 1.1: Single-exposure resolution for ASML tools since 1985[22]. *Projected.

for improved resolution necessitated reductions in k_1 far below 0.25. In other words, the 193i had to be pushed beyond the diffraction limit for Moore's Law to continue.

1.4 Multiple patterning

Reducing k_1 below 0.25 cannot be achieved in a single image due to the diffraction limit, but it is possible to go beyond this limit via multiple patterning, where multiple exposures are performed sequentially. To improve resolution by a factor of $1/k$ in both directions, k^2

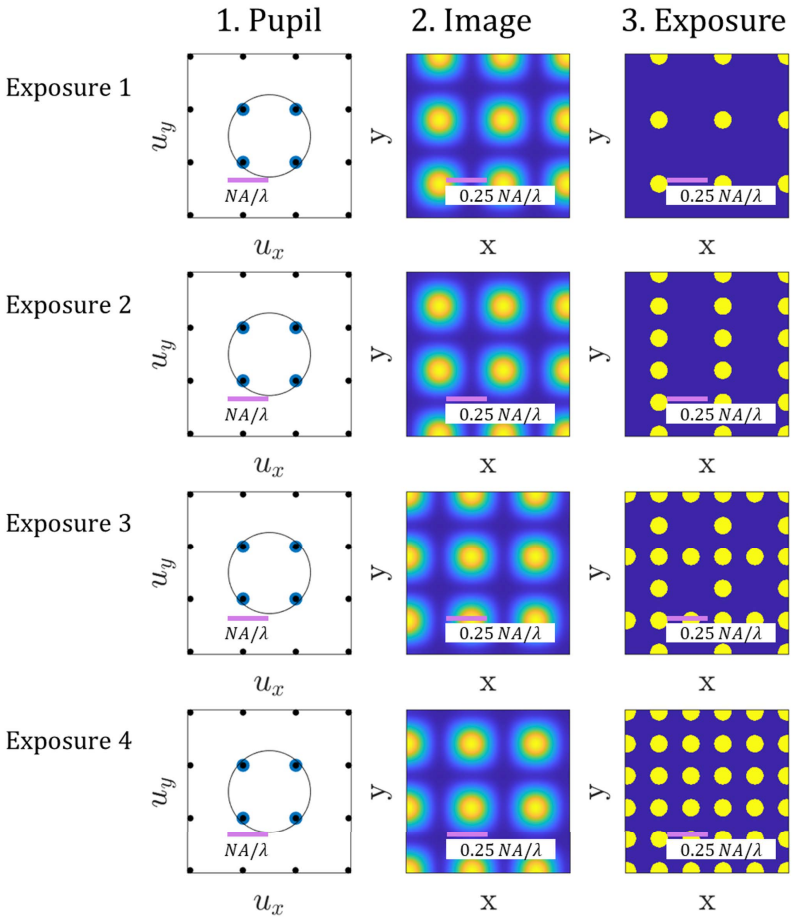

Figure 1.6: Doubling resolution in each direction with 4×-patterning for square contacts. As in Figure 1.4, the 3 columns represent pupil space, the intensity image, and the exposed pattern, while the 4 rows now represent the 4 exposures. Pupil space and intensity images still represent a single exposure each, but now the exposed pattern shows the cumulative effect of the multiple exposures, which defines the final pattern etched into the material. The illumination and pitch are the same as for single-patterning off-axis, however a bias is applied to the mask pattern and/or the intensity threshold to make the printed feature half the width. Then this same pattern is printed 4 times, each time shifting by half a pitch.

exposures would be required. In addition to the shear complexity introduced by multiple patterning (for instance indirect alignment between layers[28]), this approach has two serious drawbacks: a factor of at least k^2 increase in lithographic cost, while at the same time compounding random errors from each exposure. If each exposure has uncorrelated errors (e.g. overlay) of magnitude σ, then the combined pattern of k^2 exposures would have an

RMS of $k\sigma$; but at the same time as errors from each exposure compound, the requirements for printing at a lower k_1 become more stringent.

In other words, as you increase the number of exposures to print smaller patterns, the errors increase due to the accumulation of variance from uncorrelated random processes, while the error tolerances decrease simply due to the smaller dimensions[33]. This implies that there is a practical limit to the resolution enhancement that multiple patterning can achieve for a given level of overlay and critical dimension variability (the main two contributors to edge placement errors, which limit device yield and performance[33]). So ultimately a new lithography tool with better single exposure resolution would be required to continue reducing lateral feature sizes beyond the capabilities of 193i multi-patterning.

1.5 EUV lithography

This is where EUV lithography enters our story. In 2019, after decades of research and development, an EUV lithography tool with a wavelength of $\lambda = 13.5$nm and an NA of 0.33 went into high-volume manufacturing (HVM). A second generation with 0.55 NA is also under development and expected to go into production within approximately the next 5 years. The 13.5nm wavelength is 14.3 times smaller than 193nm, but at the same time the NA of 0.33 is 4.1 times smaller than 1.35. This leads to a reduction in λ/NA by a factor of 3.5 for 0.33 NA, or factor of 5.8 for 0.55 NA.

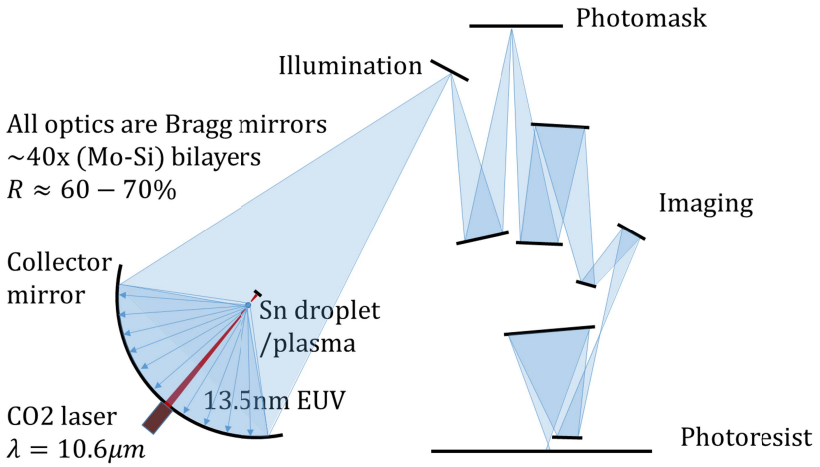

Figure 1.7: Schematic of an EUV scanner, consisting of a CO2 laser which irradiates a molten tin droplet, producing a plasma which emits 13.5nm EUV. The light is manipulated by multilayer mirrors, illuminating a photomask which reflects a predefined pattern, until finally an image of this reflection is projected onto the photoresist coating the device being manufactured.

But while EUV has already greatly improved single-exposure resolution in HVM, the full factors of 3.5 and 5.8 improvement from 193i will not be trivial to achieve in practice because 193i also is capable of printing at a lower single-exposure k_1 than EUV. Difficulties in bringing EUV to a comparable k_1 include illumination coherence, mask 3D effects (M3D), throughput, photoresist performance, and process control.

1.6 EUV photomasks

While there are many hardware components critical to EUV lithography, the focus of this work is the photomask, which consists of a multilayer mirror Bragg reflector substrate and an absorbing layer selectively etched with the mask pattern. The vertical structures of the multilayer and absorber both impact scattering and introduce so-called Mask 3D effects (M3D). M3D are explored in the following chapters, including some of the key issues caused by M3D: computational complexity, throughput, contrast, pattern phase, position errors, and focus shifts. Generally M3D are most severe for: features with smaller pitch, illumination angles further from normal incidence, thicker absorbers, and multilayers with deeper effective reflection planes.

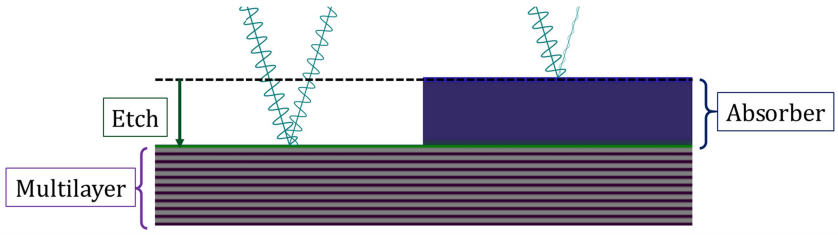

Figure 1.8: Schematic of an EUV photomask, consisting of a multilayer mirror Bragg reflector substrate and an absorbing layer selectively etched with the mask pattern.

Chapter 2

Modeling EUV Photomasks

In this section we discuss modeling the intensity image of an EUV mask which is projected onto the photoresist during exposure. We first introduce the idealized coherent imaging formulation in Section 2.1, and move on to more practical partially coherent imaging formulations in Section 2.2. We describe 3 approaches to partially coherent imaging, in particular the case of Kohler illumination, commonly used in lithography, where all illumination modes are mutually incoherent plane waves [29]. The first method, which is most intuitive and most general, is Abbe's method, Section 2.2.1, in which coherent images from each illumination mode are added together incoherently (due to the mutual incoherence of all illumination modes)[14]. Abbe's method can be applied to any scattering object, whether it is thick or thin. However, if the object is thin enough to apply a thin object approximation, one can apply accelerated calculation schemes such as Hopkins' method[19], Section 2.2.2, or Sum of Coherent Systems (SoCs)[9], Section 2.2.3, which is an approximation of Hopkins' method. Both Hopkins' and SoCs require precomputing interactions between the illumination source and the imaging pupil, so are only suited to applications where the source function is not changing such as optical proximity correction (OPC); and again these methods also are only applicable to thin objects and cannot directly accommodate thick objects which display significant mask 3D effects (M3D), as discussed in Section 2.2.4.

EUV masks are not thin, as discussed in Section 2.3, so we cannot ignore M3D and still expect to accurately model the partially coherent image. This forces us to adopt a rigorous solver such as Rigorous Coupled-Wave Analysis (RCWA) or Finite Difference Time Domain (FDTD) (Section 2.3.1) which is slow but accurate; or to make a thin-mask approximation which is very fast but inaccurate (Section 2.3.2); or to make an approximation that is more accurate than the thin-mask but still faster than a rigorous solver, which we call Fresnel Double Scattering (DblSc), introduced in Section 2.3.3.

We quantify the performance of DblSc relative to rigorous solvers in terms of speed (Section 2.3.4) and accuracy (Section 2.3.5). We find that the same power law applies to both DblSc and FDTD (quadratic for 2D domains), but DblSc is faster by a factor of 10,000 in the limit of large domains. Even despite the impressive speedup, DblSc is still fairly accurate, modeling the far-field intensity to within 3% and predicting the trends for

translation and phase shift vs pitch (Section 2.3.6). The trend of phase shift vs pitch is particularly important for designing phase shift masks, which famously require a phase shift of 1.2π rather than π[11, 49, 38]. Our results suggest that this view is not incorrect but somewhat simplistic; the desired phase shift for a given feature is still π, but one must account for the pitch-dependent phase shift and therefore increase the phase shift of the Fresnel coefficient (phase shift at infinite pitch) in order to hit the target phase at a relevant feature size. This analysis suggests that the exact phase shift required for optimal patterning with an EUV aPSM will be a function of the target feature size, as well as potentially other factors such as the multilayer mirror architecture.

Finally in Section 2.4 we present an example application of this DblSc scattering approximation to source-mask optimization (SMO). Unlike OPC, in SMO the source is constantly updated, making it ill-suited for Hopkins' or SoCs even for a thin object. Furthermore, given our thick EUV mask object we know we must use Abbe's method. We adopt the DblSc approximation rather than a rigorous solver for computational performance. In this way we demonstrate an example of co-optimizing the source and mask while being sensitive to EUV 3D effects, in order to maximize the process window (PW) with respect to errors in focus and dose.

2.1 Coherent imaging

We adopt the scalar wave optics approximation where we neglect polarization, but generally we do not make the thin-mask approximation that the scattering function for a given wavelength (and polarization) is shift-invariant in reciprocal space (defined by a linear shift-invariant transfer function). The theory presented in this section is rigorous for a polarized light source, and for a mixed polarization state one can treat each orthogonal polarization component separately using the theory presented, and combine the images from each component after the fact.

Consider a sample illuminated with a coherent plane-wave with lateral spatial frequency $f_{x,in}, f_{y,in}$; the illumination wave at reference plane $z = 0$ is described mathematically as a complex exponential, $E_0(x, y) = \exp 2\pi i(xf_{x,in} + yf_{y,in})$. If we assume the object being imaged is thin and has a complex-valued transmission function $t(x, y)$ with corresponding Fourier transform $T(f_x, f_y) = \mathcal{F}[t(x, y)]$, then the electric field after scattering would be $E_{NF}(x, y; f_{x,in}, f_{y,in}) = e^{2\pi i(xf_x+yf_y)}t(x, y) = \mathcal{F}^{-1}[T(f_x - f_{x,in}, f_y - f_{y,in})]$. This is called the near-field, which is the electric field immediately after scattering; it is this field which is then imaged by the lens. The optical system will only capture angles within the numerical aperture $NA = n \sin \theta_{max}$, where n is the refractive index of the medium in which the optics are immersed; for the remainder of this work we will assume that $n = 1$ because EUV lithography systems operate in vacuum, so the formula simplifies to $NA = \sin \theta_{max}$. This defines a cutoff frequency $f_c = NA/\lambda$ based on the diffraction relationship between lateral spatial frequency, wavelength, and angle: $\sin \theta - \sin \theta_0 = f_\perp \lambda$. This creates a frequency-space transfer function:

$$H(f_x, f_y) = \begin{cases} 1, f_x^2 + f_y^2 \leq f_c^2 \\ 0, \text{else} \end{cases} \tag{2.1}$$

The transfer function can also include phase due to defocus or other aberrations, and this phase is called the wavefront error function $\text{WEF}(f_x, f_y) = \arg[H(f_x, f_y)]$; so including this term the transfer function becomes:

$$H(f_x, f_y) = \begin{cases} e^{i\text{WEF}(f_x, f_y)}, f_x^2 + f_y^2 \leq f_c^2 \\ 0, \text{else} \end{cases} \tag{2.2}$$

This defines the transfer function at a nominal focal plane; a distance z from focus, the WEF is modified by zk_z, where $k_z = \frac{2\pi}{\lambda}\cos\theta$. So if we call $\text{WEF}_0(f_x, f_y) = \text{WEF}(f_x, f_y; z = 0)$ the wavefront error function at the focal plane, at plane z we write:

$$\text{WEF}(f_x, f_y; z) = \text{WEF}_0(f_x, f_y) + \frac{2\pi}{\lambda} z \cos\theta \tag{2.3}$$

Note that here the θ value is not constant, but rather corresponds to the angle of diffraction for each spatial frequency.

Now we can convolve the scattered field with the impulse response, i.e. the inverse Fourier transform of the transfer function, $h(x, y; z) = \mathcal{F}^{-1}[H(f_x, f_y; z)]$, to compute the far-field in the image plane:

$$E_{FF}(x, y; f_{x,in}, f_{y,in}; z) = [e^{2\pi i(xf_{x,in} + yf_{y,in})} t(x, y)] * h(x, y; z) \tag{2.4}$$

Finally, the only property of the image we must accurately predict for lithography is total energy deposited, which is proportional to intensity and independent of phase. So the final quantity of interest is the magnitude-squared of the far-field:

$$\begin{aligned} I(x, y; f_{x,in}, f_{y,in}; z) &= |\mathcal{F}^{-1}[T(f_x - f_{x,in}, f_y - f_{y,in}) H(f_x, f_y; z)]|^2 \\ &= |\mathcal{F}^{-1}[T(f_x, f_y) H(f_x + f_{x,in}, f_y + f_{y,in}; z)]|^2 \end{aligned} \tag{2.5}$$

For a thick mask the expression is the same, as long as we further include illumination dependence in the transmission function, $T(f_x, f_y) \to T(f_x, f_y; f_{x,in}, f_{y,in})$. For describing Abbe's method, Hopkins' method, and SoCs, we will omit z dependence for simplicity of notation; in each case the imaging computation can be repeated for a list of output z-planes with different pupil functions to compute a focus stack.

We also introduce discretized matrix notation for ease of use in computational applications. A periodic function can be exactly decomposed into Fourier coefficients discretely spaced by the fundamental frequency, which also defines a corresponding grid in real-space with the same number of pixels evenly spaced across one pitch. In matrix notation we write the field, for source point u (representing one illumination plane-wave):

$$\text{E}_u = \text{F}^*[\text{H}_u \cdot \text{T}] = \text{F}^* \text{diag}[\text{H}_u] \text{T} \tag{2.6}$$

where $\mathrm{E}_u \in \mathbb{C}^N$ is the field in the image plane from source-point u, $\mathrm{H}_u \in \mathbb{C}^N$ is the pupil function (transfer function) for the source point in question, and $\mathrm{T} \in \mathbb{C}^N$ is a vector of the Fourier coefficients of the mask transmission function. The image is calculated by taking the pointwise intensity in real-space:

$$\mathrm{I}_u = |\mathrm{E}_u|^2 = \mathrm{diag}[\mathrm{E}_u^*]\mathrm{E}_u = \mathrm{diag}[\mathrm{E}_u\mathrm{E}_u^*] \tag{2.7}$$

2.2 Partially coherent imaging

Neglecting the polarization state of the illumination, we describe the source as an intensity distribution in 3D reciprocal-space, $S(k_x, k_y, k_z)$. Instead of (k_x, k_y, k_z), other parametrizations for illumination are also possible, such as (f_x, f_y, λ) or (u_x, u_y, λ), where $u_{x,y} = f_{x,y}\frac{\lambda}{NA}$. From we here omit the λ variation for simplicity of notation and describe the monochromatric source with wavelength λ as $S(f_x, f_y)$.

Generally speaking, depending on the geometric configuration of the mask, each incident plane wave could map to any arbitrary linear combination of plane waves; this would be described by the 4D scattering function $T(f_{x,out}, f_{y,out}; f_{x,in}, f_{y,in})$. For a thin object the function is linear shift-invariant, so we can write it as $T(f_x, f_y)$ where $f_{x/y} = f_{x/y,out} - f_{x/y,in}$. If we take the 2D inverse Fourier transform this reduces to the more familiar real-space transmission function $t(x, y) = \mathcal{F}^{-1}[T(f_x, f_y)]$. In the thick-mask case we can still define a real-space transmission function, but must maintain the dependence on illumination spatial frequency (angle): $t(x, y; f_{x,in}, f_{y,in})$. In this context any deviations from a thin-mask transmission function are referred to as M3D (Mask 3D effects).

2.2.1 Abbe's method

A very straightforward approach to calculating the partially coherent intensity image with Kohler illumination is Abbe's method, in which we simply compute the image for each source point and then combine their intensities in real-space with a weighted integral over the source[14, 29]. We write out the formulation for a thick mask, understanding that the illumination dependence in T can simply be removed for a thin mask.

$$I(x, y) = \int S(f_{x,in}, f_{x,in})|\mathcal{F}^{-1}[T(f_x, f_y; f_{x,in}, f_{y,in})H(f_x + f_{x,in}, f_y + f_{y,in})]|^2 d^2 f_{x,in} f_{x,in} \tag{2.8}$$

Now writing the same integral out in matrix notation, it becomes a sum over discrete source-points contained in the vector $\mathrm{S} \in \mathbb{R}^{N_s}$. For element u, we have source weight $S_u \in \mathbb{R}$ and intensity image $\mathrm{I}_u \in \mathbb{R}^N$; the final image is calculated with a weighted sum of coherent intensity images:

$$\mathrm{I} = \sum_{u=1}^{N_s} S_u \mathrm{I}_u = \mathrm{I}_{PC}\mathrm{S} \tag{2.9}$$

where $\mathrm{I}_{PC} = [\mathrm{I}_u]_{u=1}^{N_S} \in \mathbb{R}^{N \times N_s}$ is a matrix, where each column is the coherent image from one source point. For a thick object we write $\mathrm{I}_u = |\mathrm{F}^* \mathrm{diag}[\mathrm{T}_u]\mathrm{H}_u|^2$, where T_u is the thick-mask transmission from source point u. For a thin mask we can drop the u dependence on T and write $\mathrm{I}_u = |\mathrm{F}^* \mathrm{diag}[\mathrm{T}]\mathrm{H}_u|^2$. This allows us to compute the whole matrix as:

$$\mathrm{I}_{PC} = |\mathrm{F}^* \mathrm{diag}[\mathrm{T}]\mathrm{H}|^2 \tag{2.10}$$

where $\mathrm{H} = [\mathrm{H}_u]_{u=1}^{N_S} \in \mathbb{C}^{N \times N_s}$ is a matrix, whose columns are the transfer functions from each source point.

2.2.2 Hopkins' method

In Hopkins' method we split the computation to first compute TCCs (transmission cross-coefficients), which combine the impact of the source and pupil, and then secondarily integrate across the mask[19, 29]. The TCC is a 4D function of scattering frequencies f_x, f_y and illumination frequencies f'_x, f'_y. We compute the TCC with an integral over the source frequencies:

$$TCC(f_x, f_y, f'_x, f'_y) = \int H(f_x + f'_x, f_y + f'_y) H^*(f_x + f'_x, f_y + f'_y) S(f'_x, f'_y) d^2 f'_x f'_y \tag{2.11}$$

Image calculation now requires a 4D integral, which we write as a 2D integral + 2D IFT:

$$I(x, y) = \mathcal{F}^{-1}[\int TCC(f_x, f_y, f'_x, f'_y) T(f_x, f_y) T^*(f'_x, f'_y) d^2 f'_x f'_y] \tag{2.12}$$

A physical interpretation of the TCC is summing up the contribution of each source-point according to its weight, for each frequency-pair in the auto-correlation of the shifted transfer function. This is leveraging the linearity of the partially coherent image in intensity; while the image is quadratic in electric-field, it is linear in the inner product of electric field with itself (intensity)—and it is also linear in the outer product of the field with its itself, which is the field autocorrelation matrix. In this way one precomputes the interactions of all pairs of frequencies, $TCC(f_x, f_y, f'_x, f'_y)$, and then applies this as a weight to the corresponding outer-product of mask-transmission $T(f_x, f_y) T^*(f'_x, f'_y)$. In matrix form, we start by rewriting the summation over the source:

$$\mathrm{I} = \sum_{u=1}^{N_s} S_u \mathrm{I}_u = \sum_{u=1}^{N_s} S_u |\mathrm{F}^* \mathrm{diag}[\mathrm{T}]\mathrm{H}_u|^2 = \sum_{u=1}^{N_s} S_u \mathrm{diag}[\mathrm{F}^* \mathrm{diag}[\mathrm{T}]\mathrm{H}_u \mathrm{H}_u^* \mathrm{diag}[\mathrm{T}]^* \mathrm{F}] \tag{2.13}$$

$$= \mathrm{diag}[\mathrm{F}^* \mathrm{diag}[\mathrm{T}]\mathrm{H} \mathrm{diag}[\mathrm{S}] \mathrm{H}^* \mathrm{diag}[\mathrm{T}]^* \mathrm{F}]$$

We then define the TCC matrix $\mathrm{M} = \mathrm{H}\mathrm{diag}[\mathrm{S}]\mathrm{H}^* \in \mathbb{C}^{N \times N}$. This yields the final expression:

$$\mathrm{I} = \mathrm{diag}[\mathrm{F}^*\mathrm{diag}[\mathrm{T}]\mathrm{M}\mathrm{diag}[\mathrm{T}]^*\mathrm{F}] \tag{2.14}$$

Just as in the continuous version, the TCC is precomputed based on the source and pupil, and then the TCC is combined with the mask transmission function in a quadratic form to compute the image.

2.2.3 Sum of coherent systems (SoCs)

In the SoCs approximate Hopkins' method, we first compute a source-pupil term just as in Hopkins' method, then we perform a truncated eigen decomposition to get a finite set of eigenfunctions approximating the matrix, and finally we compute the image as a weighted sum of the images from each eigenfunction[9, 29]. We define the imaging impulse function as $h(x, y) = \mathcal{F}^{-1}[H(f_x, f_y)]$ and the illumination mutual coherence as $J_0(x, y) = \mathcal{F}^{-1}[S(f_x, f_y)]$. We now define a function W that captures source-pupil interactions across real-space:

$$W(x, y; x', y') = J_0(x - x', y - y')h(x, y)h^*(x', y') \tag{2.15}$$

We decompose this function into eigenfunctions ψ_n with eigenvalues λ_n, truncating the decomposition to only retain the largest N eigenvalues (note that all eigenvalues of W must be real nonnegative numbers because this is a positive-semidefinite quadratic form making the truncated EVD equivalent to a truncated SVD in the discrete case, which is the best low-rank approximation to a matrix according to the Frobenius norm[53]). We write the truncated decomposition as:

$$W(x, y; x', y') \approx \sum_{n=1}^{N} \lambda_n \psi_n(x, y)\psi_n^*(x', y') \tag{2.16}$$

Finally, we take a weighted sum over the coherent image from each eigenfunction to obtain the SoCs approximation to Hopkins' method:

$$I(x, y) \approx \sum_{n=1}^{N} \lambda_n |\psi_n(x, y) * t(x, y)|^2 \tag{2.17}$$

In matrix notation, the final formula can be written:

$$\mathrm{I} \approx \sum_{n=1}^{N} \lambda_n |\mathrm{F}^*\mathrm{diag}[\mathrm{T}]\mathrm{F}\psi_n|^2 \tag{2.18}$$

We can understand this expression by expanding it similar to the TCC derivation. We define W by applying a Fourier similarity-transform to the TCC matrix M (equivalent to taking an IFT on input and output dimensions):

$$\mathrm{W} = \mathrm{F}^*\mathrm{MF} \iff \mathrm{M} = \mathrm{FWF}^* \tag{2.19}$$

We define the truncated EVD of W:

$$\mathrm{W} \approx \sum_{n}^{N} \lambda_n \psi_n \psi_n^* = \mathbf{\Psi \Lambda \Psi}^* \tag{2.20}$$

From the matrix formula for Hopkins' method:

$$\mathrm{I} = \mathrm{diag}[\mathrm{F}^* \mathrm{diag}[\mathrm{T}] \mathrm{M} \mathrm{diag}[\mathrm{T}]^* \mathrm{F}] = \mathrm{diag}[\mathrm{F}^* \mathrm{diag}[\mathrm{T}] \mathrm{F} \mathrm{W} \mathrm{F}^* \mathrm{diag}[\mathrm{T}]^* \mathrm{F}] \tag{2.21}$$

Then we substitute the truncated EVD:

$$\mathrm{I} \approx \mathrm{diag}[\mathrm{F}^* \mathrm{diag}[\mathrm{T}] \mathrm{F} \mathbf{\Psi \Lambda \Psi}^* \mathrm{F}^* \mathrm{diag}[\mathrm{T}]^* \mathrm{F}] \tag{2.22}$$

$$= \sum_{n}^{N} \lambda_n \mathrm{diag}[\mathrm{F}^* \mathrm{diag}[\mathrm{T}] \mathrm{F} \psi_n \psi_n^* \mathrm{F}^* \mathrm{diag}[\mathrm{T}]^* \mathrm{F}] = \sum_{n=1}^{N} \lambda_n |\mathrm{F}^* \mathrm{diag}[\mathrm{T}] \mathrm{F} \psi_n|^2 \tag{2.23}$$

The final expression is the desired matrix equation for SoCs, where we convolve the mask transmission function with each eigenfunction and take a weighted sum according to the eigenvalues.

2.2.4 Incorporating M3D

Although the SoCs (approximate Hopkins') method is by far the fastest in the case of a fixed source and pupil, nonetheless it is less theoretically general than Abbe's method because it requires the assumption of shift invariance in reciprocal space, an assumption which only holds true for a thin object but not for a thick object such as an EUV mask.

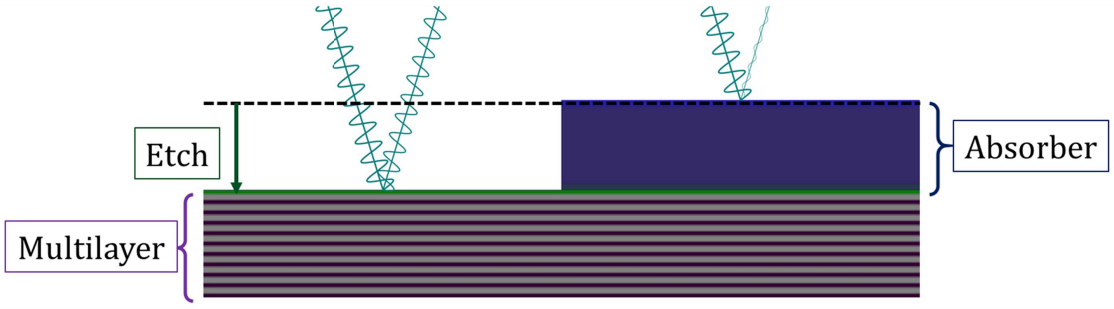

Figure 2.1: Schematic of photomask architecture, consisting of a multilayer mirror Bragg reflector and an absorber that has been selectively etched with the mask pattern. The vertical structure of the photomask makes the scattering no longer follow a thin-mask model, which necessitates including M3D in imaging calculations.

For a thick object it is more natural to apply Abbe's method, because it does not require this assumption of shift invariance. This is because in Abbe's method we separately compute the image from each source point and then sum them in the image plane, so there is no requirement that each source point produce the same scattering amplitudes. We simply include a dependence of T on source-point u, and write:

$$\mathrm{I} = \mathrm{I}_{PC}\mathrm{S}; \mathrm{I}_{PC} = [|\mathrm{F}^{*}\mathrm{diag}[\mathrm{T}_u]\mathrm{H}_u|^2]_u \tag{2.24}$$

Because we wish to model M3D in several projects in this work, we primarily consider Abbe's method for partially coherent imaging, incorporating M3D with either an approximate solver such as Fresnel Double Scattering (DblSc) or a rigorous solver such as RCWA. All results presented in the remainder of this section assume TE polarization (equivalent to s-polarization).

2.3 Scattering and M3D

2.3.1 Rigorous solvers

While not always practical, it is always possible to numerically solve Maxwell's equations to find the complex near-field scattered from an EUV mask for any given coherent illumination. This would be done with a method such as Rigorous Coupled-Wave Analysis (RCWA) or Finite Difference Time-Domain (FDTD). One needs only to specify the 3D geometric distribution of complex refractive index on the mask and the 3D illumination wave. To use with Abbe's method, we compute the complex near field scattered for the defined mask geometry from each source point (monochromatic plane waves), and then the coherent modes are combined incoherently with a sum over the source in the image plane. Both RCWA and FDTD provide an exact solution to Maxwell's equations (up to numerical limits), however taking very different approaches. In FDTD the 3D space is separated into a grid of voxels containing the complex refractive index within each location. The solution is found by propagating the incident electromagnetic fields in the time domain until the system converges to its time-harmonic (steady state) dynamics. RCWA uses an entirely different mathematical formalism, separating the space into slabs of arbitrary thickness rather than a uniform grid of voxels. The RCWA solution is not iterative but direct, solving a series of matrix equations dominated by eigen decompositions and matrix inversions.

For smaller domains RCWA is much faster because it does not require an iterative solution and the solutions to matrix equations can be computed very fast for small matrix sizes; furthermore, sensitive thin-film effects can be more easily captured with RCWA without a fine discretization in z, because the space is discretized into slabs with arbitrary thickness as opposed to voxels with a fixed height. However, RCWA scales with a power law of 3 or 6 for 1D or 2D respectively, arising from the cubic complexity of both eigen decomposition and matrix inversion, making RCWA a very impractical choice in the limit of large domain size. FDTD on the other hand has a power law of 1 and 2 for 1D and 2D respectively, meaning

it is relatively easy to scale the algorithm to larger domains, although the runtime for a small domain is much slower than RCWA. Both large and small simulation domains have important applications: larger domains are required to design the mask pattern at full-chip scale, at the same time as smaller domains are required to design the unit cell of the critical features on the mask (which often have the smallest pitch on the mask and suffer from the most severe M3D).

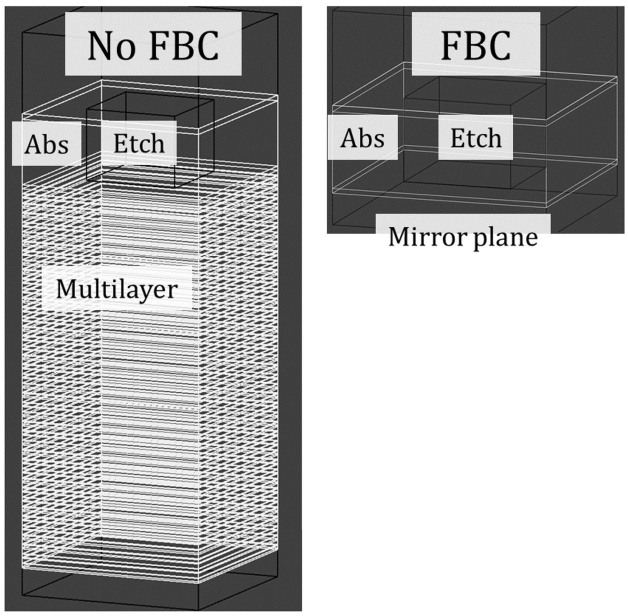

Figure 2.2: The Fourier boundary condition (FBC) replaces the multilayer mirror substrate with an effective reflection function to accelerate FDTD calculation.

FDTD can also be greatly accelerated by replacing the multilayer mirror with a 1D reflection function, using the so-called Fourier Boundary Condition (FBC). In the FBC, the sequence of flat layers in the multilayer substrate is replaced with a complex reflection function that simply reflects back each incident plane wave with amplitude and phase determined by the Fresnel reflectance of the multilayer. This greatly accelerates FDTD for an EUV mask, because more than half of the voxels can be removed. Using the FBC has very little impact on the converged solution, so it is generally recommended, and makes FDTD (FBC) the natural choice for rigorous modeling in the limit of large domain sizes as opposed to FDTD with no FBC or RCWA.

2.3.2 Thin-mask approximation

Instead of solving Maxwell's equations, a much faster alternative for scattering (which also is compatible with Hopkins'/SoCs) is to use a thin-mask approximation. Here the scattered

field is simply a binary pattern function with a complex amplitude determined by the Fresnel coefficients of multilayer and absorber at the nominal illumination conditions. The amplitude and phase of multilayer and absorber reflection cannot change as a function of illumination angle or as a function of pitch, because either of these would violate the shift-invariance assumption in Hopkins' method. So this method fails to capture even simple 3D effects like the angular variation of the absorber transmission coefficient, let alone the more complex pattern-dependent M3D captured in rigorous methods.

$$E(x,y) = r_{ML}P_{ML}(x,y) + r_{Abs}[1 - P_{ML}(x,y)] \tag{2.25}$$

2.3.3 Double-scattering approximation (DblSc)

We now describe an intermediate approach between a rigorous computation and a thin-mask approximation. The scattering is described now by a binary pattern and a set of 5 Fresnel transfer functions describing transmission through the absorber and etch, and reflections from the multilayer, absorber, and etch. The Fresnel transfer functions are evaluated by sampling the reflectance/transmittance coefficient at angles corresponding to the reciprocal-space grid for the given illumination and pitch. The grid is defined as follows: for illumination wave vector $k_{x,0}, k_{y,0}, k_{z,0}$ (with wavenumber $k_0 = \sqrt{k_{x,0}^2 + k_{y,0}^2 + k_{z,0}^2} = 2\pi/\lambda$), for pitch p_x, p_y in x and y respectively. The grid spacing in k-space will be $\Delta k_x = 2\pi/p_x$,$\Delta k_y = 2\pi/p_y$. If we index x with m and y with n, the discretized reciprocal space coordinate $[m,n]$ corresponds to: $k_{x,m} = k_{x,0} + m\Delta k_x$,$k_{y,n} = k_{y,0} + n\Delta k_y$,$k_{z,mn} = \sqrt{k_0^2 - k_{x,m}^2 - k_{y,n}^2}$. By evaluating a Fresnel coefficient on this grid in reciprocal space, we obtain a transfer function in frequency space, which is the Fourier transform of the discrete impulse response function (response to a discrete delta function at the origin). Therefore, we evaluate the Fresnel coefficient to get the transfer function, inverse Fourier transform to get the impulse response, and at several steps convolve the impulse response with a 2D pattern to get the resulting reflected or transmitted field.

There are seven steps in the approximate Fresnel Double Scattering computation, shown for an example case of $k_1 = 2$ [0.33 NA,p=655nm] square contacts, 50% duty cycle in x and y (25% multilayer by area), 6 degrees 13.5nm monochromatic illumination:

1. For given illumination and pitch, calculate 5 Fresnel reflection and transmission kernels for multilayer (r only), absorber (r, t), and etch (r, t)

2. Calculate the Fourier transform of the pattern (and inverse pattern)

3. Compute the pattern reflection (minor term), i.e. reflection that did not interact with multilayer

4. Compute pattern transmission (major term)—includes etch transmission times pattern plus absorber transmission times inverse pattern

5. Transmitted pattern reflected by the multilayer
6. Upward transmission of the reflected pattern
7. Combine terms for initial reflection and multilayer reflection with double transmission

2.3.4 Speed of DblSc

We quantify the improvement in speed for DblSc vs 3 rigorous methods: RCWA, FDTD, and FDTD with the FBC (Fourier Boundary Condition). DblSc was calculated in MATLAB; the three rigorous methods were calculated with the Panoramic EM-Suite API through MATLAB. By switching from a rigorous to approximate scattering model, computation time for one illumination condition is substantially reduced for lines and spaces (5-500x), and dramatically reduced for contacts (10-10,000x). Among the rigorous methods, FDTD is strictly faster with the FBC than without it, because it replaces a 3D grid of cells with a 1D z-structure and thereby reduces the simulation domain by well over half, without a significant change in the output field. But between RCWA and FDTD (FBC), neither is strictly faster than the other. For small feature sizes RCWA is much faster, but RCWA has a higher power-law scaling in both 1D and 2D, and particularly in the limit of a large 2D domains, FDTD (FBC) becomes orders of magnitude faster than RCWA. We model the scaling of the methods with a power-law plus constant model of runtime vs k_1, which is:

$$t(k_1) = t_0 + ak_1^{\alpha} \tag{2.26}$$

For 1D features α was nearly 0 for DblSc implying almost constant runtime at these feature sizes; α was around 1 for both FDTD and FDTD (FBC), and between 2 and 3 for RCWA. The theoretical scaling for the rigorous methods should be 1 for FDTD and 3 for RCWA. In 2D the theoretical scaling factors become 2 and 6, which we roughly see in the results; DblSc has a scaling of 2 as well. So FDTD (FBC) and DblSc (in 2D) have the same scaling law, but differ by a constant factor of 10,000×; in the small feature limit RCWA is faster than FDTD (FBC), but in all cases DblSc is faster than any rigorous method.

2.3.5 Accuracy of DblSc

In addition to being much faster to compute than RCWA or FDTD, the DblSc algorithm still manages to capture M3D effects with surprising fidelity. This suggests that M3D in EUV masks differ substantially from M3D in DUV masks, and are in some sense much simpler. There are intuitive physical reasons why M3D should be "simpler" to compute in EUV masks, because the refractive index of all materials used in EUV is within about 0.1 of unity, in stark contrast to DUV where much stronger absorbing and phase shifting materials are used. In addition to the stronger index contrast, the dimensions on the mask in DUV are laterally and axially much smaller relative to the wavelength, meaning that there is a very strong optical

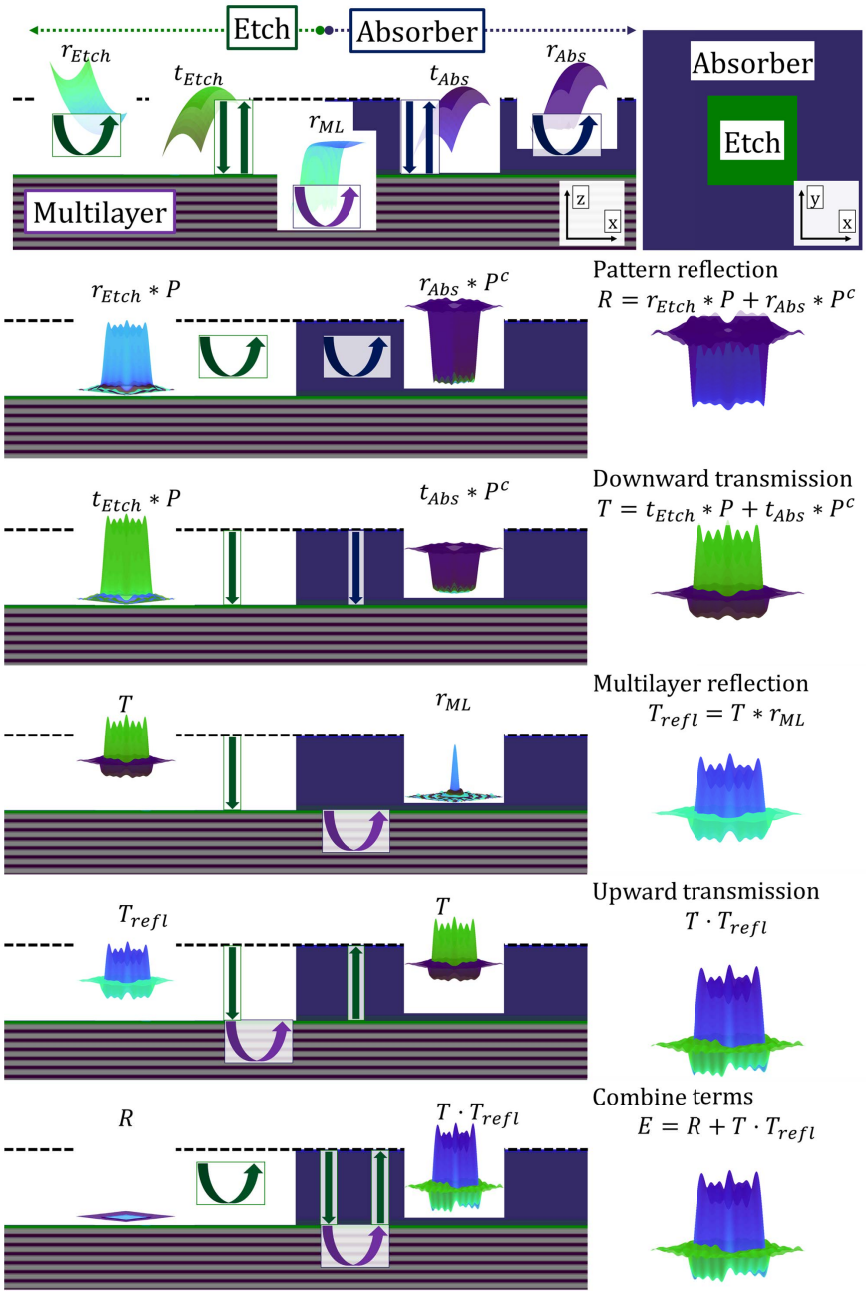

Figure 2.3: Double Scattering (DblSc) flowchart.

interaction—strongly absorbing conductors or strong phase shifters—over a small fraction of a wavelength, which leads to more complex scattering phenomena like electromagnetic edge and corner effects at the boundaries of conducting materials. On the other hand, EUV masks

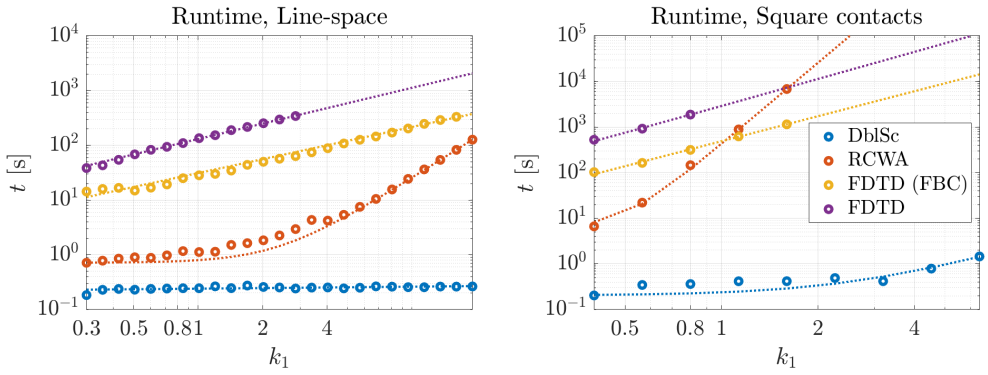

Figure 2.4: Runtime for lines and spaces (dx=1nm, dz=0.1nm) vs square contacts (dx=5nm, dz=0.1nm). Runtime fit using power-law plus constant. For each method the pitch is increased up to a maximum k_1 of 19.2 (6.4) or until the runtime surpasses 300s (1000s).

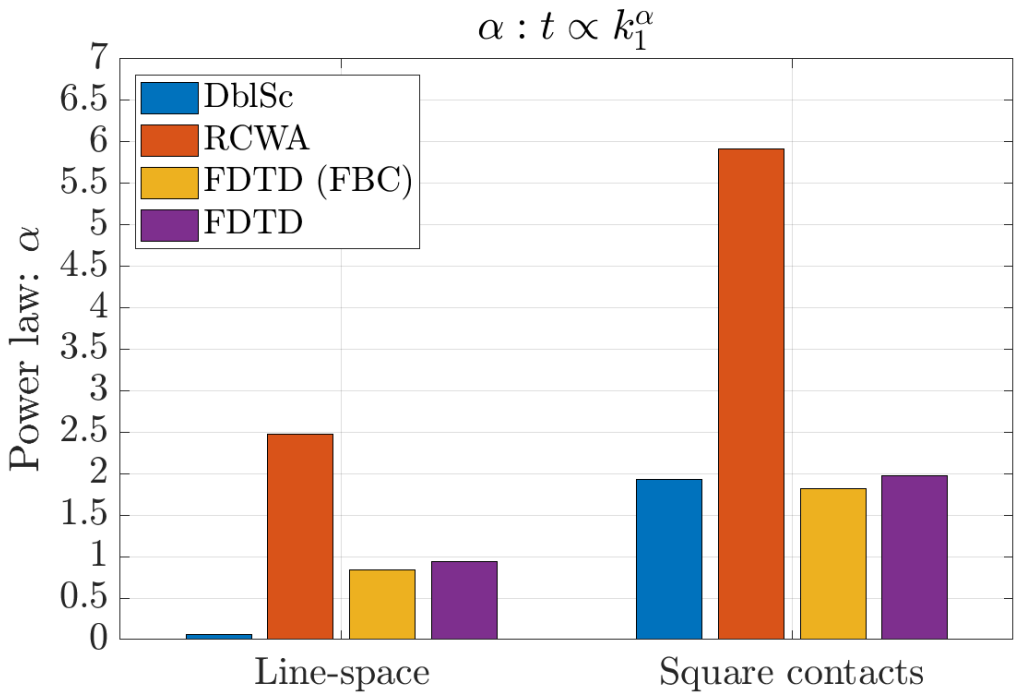

Figure 2.5: Fitted power law α, where $t = t_0 + ak_1^\alpha$.

have in some ways "worse" M3D in terms of imaging because the absorber is much thicker relative to the wavelength (providing space for wave propagation), as well as illumination-dependent distortions arising from the rapidly varying phase and amplitude response of the

multilayer. Furthermore, EUV illumination is non-telecentric meaning that M3D effects are asymmetric across the source. So EUV M3D are in some ways "worse" than DUV for the reasons provided, but at the same time it appears that computationally they may be more easily modeled by simply combining a sequence of scattering events using Fresnel coefficients to capture angular variation in amplitude, phase, and propagation.

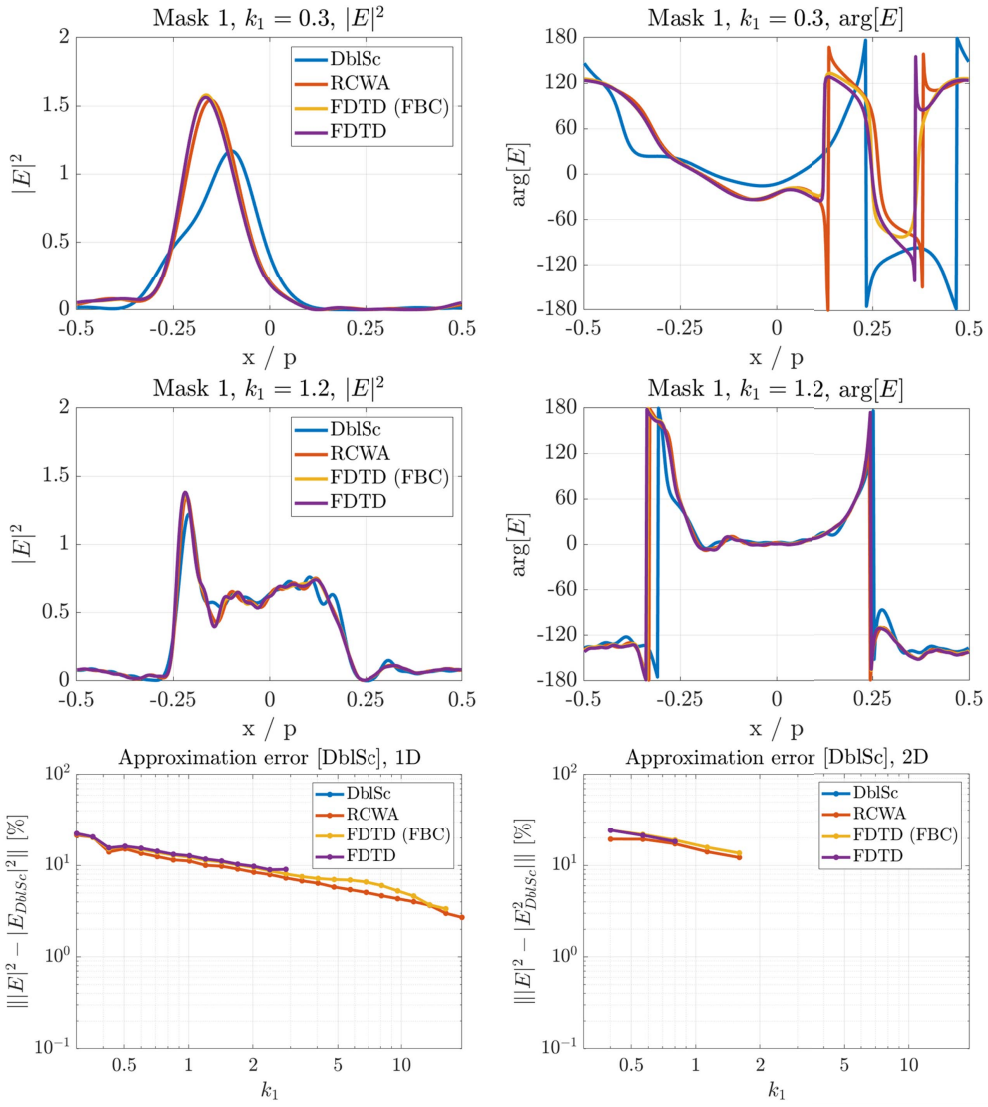

Figure 2.6: Accuracy of DblSc in near field.

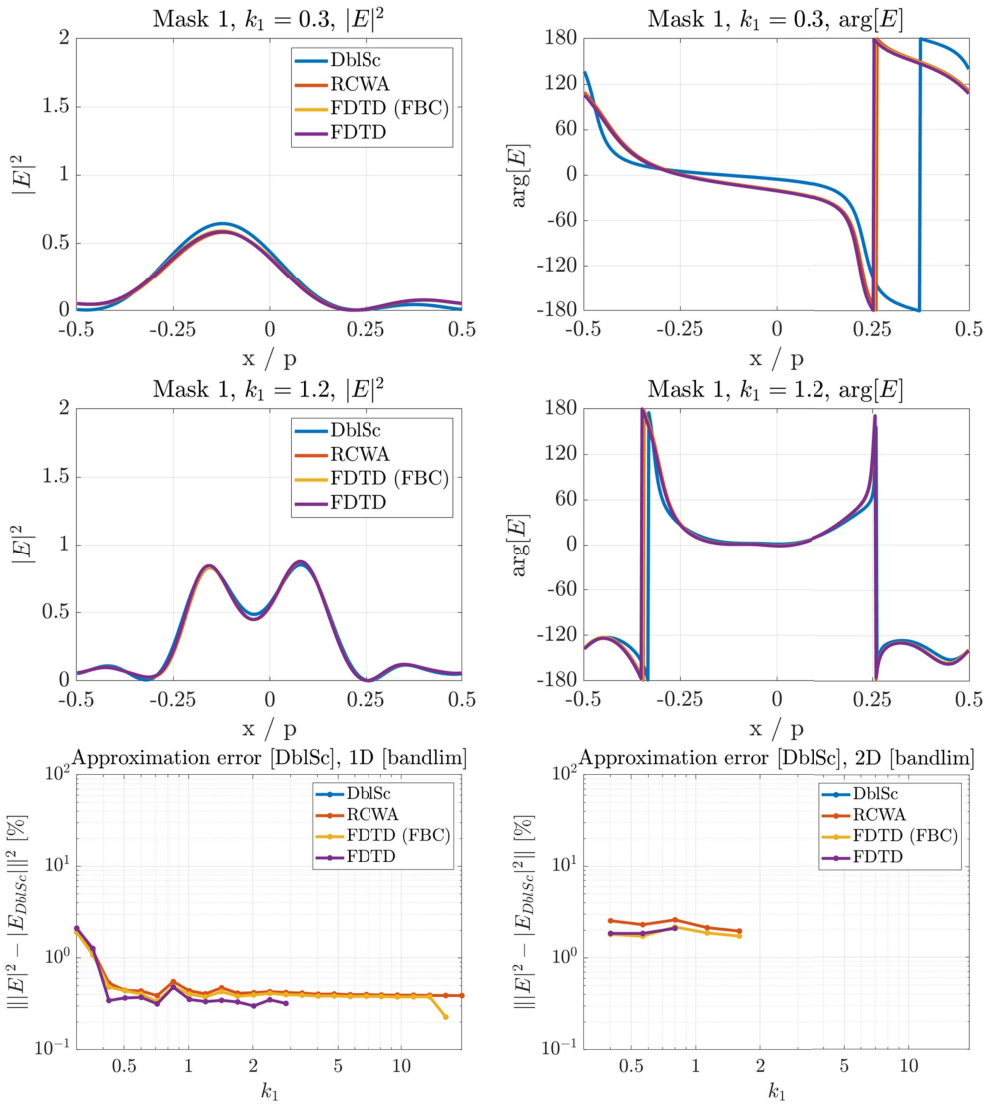

Figure 2.7: Accuracy of DblSc in far field.

2.3.6 M3D: phase

One of the most well-known EUV M3D effects is that the optimal phase shift for an aPSM is not π but roughly 1.2π. This number has been arrived at by numerical optimization using rigorous solvers such as RCWA; however, it appears that this effect can already be captured with an approximate DblSc solver, which can give important physical insights into the origins of such M3D effects in addition to accelerating computations. We show for 3 mask architectures (2 aPSMs and 1 standard TaN), the phase shift between absorber and multilayer is strongly dependent on pitch, generally decreasing with decreasing pitch. Trends

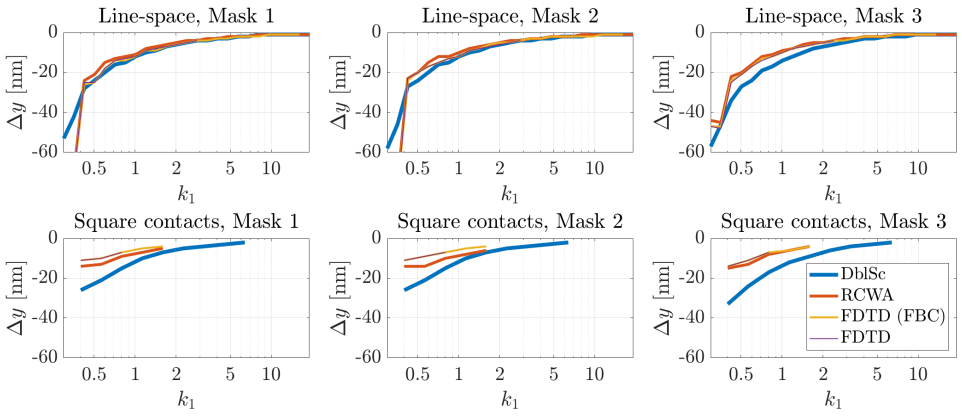

Figure 2.8: Pattern translation vs pitch.

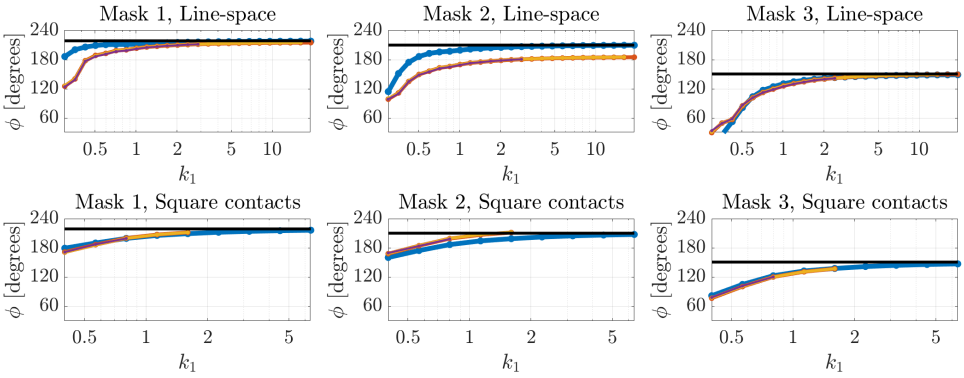

Figure 2.9: Pattern phase vs pitch.

are similar for contacts and lines and spaces. There is good agreement between DblSc and rigorous methods, except for Mask 2 (lines and spaces); it is not clear what causes this discrepancy, but it may be the more complicated 3-layer absorber architecture employed in Mask 2, and perhaps more than two scattering events must be computed for a more complicated absorber architecture such as this. Even in this case the trend of phase vs pitch is very similar for DblSc as the rigorous methods, but interestingly the rigorous methods do not converge to the Fresnel reflectance phase in the large pitch limit in this one case. It is somewhat perplexing how the phase could not converge to the Fresnel phase, but the behavior is remarkably consistent across the 3 rigorous solvers (RCWA, FDTD, and FDTD w/FBC) so is likely a real physical effect rather than a numerical artefact.

Based on the trend for contacts, both aPSMs have 180^o phase at a k_1 around 0.5, a

difference of about 36^o degrees (0.2π) from the large-pitch limit. A similar reduction in phase is observed in the TaN mask. The trend of the phase is in the same direction for all 3 masks, suggesting that this additional phase shift is largely independent of the absorber architecture and much more determined by other factors such as the multilayer.

Because of this pitch-dependent phase shift, a Fresnel phase of 0.8π is not equivalent to 1.2π; although these will be complex conjugates at the large-pitch limit, for a given small-pitch feature if the phase is reduced for each by 0.2π, they would become 0.6π and π respectively. A phase shift of π in the electric field will still yield the best in-image contrast, so the case of $\phi[k_1 \to \infty] = 1.2\pi$ is preferable, because this will result in the desired phase shift at the critical feature size of $\phi[k_1 = 0.5] \approx \pi$.

Another consequence of this theory is that the optimal amount of phase offset would not be universally 0.2π, but would instead be pitch dependent. If the minimum printable feature size were $k_1 = 1$, the optimal phase should be very close to π; if the minimum feature size were further decreased on the other hand, the optimal phase shift should be greater than 1.2π. Furthermore, because the effect appears to be related to the multilayer (since it is very similar for all 3 mask architectures), which would imply that future attempts to engineer the multilayer mirror reflection function would be likely to alter the optimal phase shift of the absorber. This relationship between multilayer and absorber phase is explored further in Section 3.3.4, where we indeed see that they are intimately linked.

2.4 Accelerated M3D-aware SMO

2.4.1 Partially coherent FEM algorithm

In this section we present a proof of concept for computational lithography using Abbe's method for partially coherent imaging and DblSc for scattering. Inputs to the algorithm such as the mask architecture, pattern, and source weights can be adjusted iteratively to minimize the defined cost metric.

We first compute approximate near-field scattering for each source point, which depends on the mask architecture and pattern, and the source point coordinates (but not their weights). Then for each source point we apply a set of transfer functions through focus, to obtain far-field intensities through focus for each source point. We compute the partially coherent focal stack by applying Abbe's method with the input source weights for each focal plane; note that we assume the source has an adjustable set of weights vs angle as well as a non-adjustable source spectrum vs wavelength (which in theory could be different for each source point but here is assumed to be the same for all source points). We then convert the partially coherent focal stack into a binary focus-exposure matrix (FEM) by applying a set of threshold values; this binary FEM is a computational analog to the common practice of exposing the same pattern with a range of focus and exposure values (an experimental FEM), then developing the pattern and inspecting the developed resist in an SEM (Scanning Electron Microscope). We quantify the accuracy of the pattern at each focus and exposure

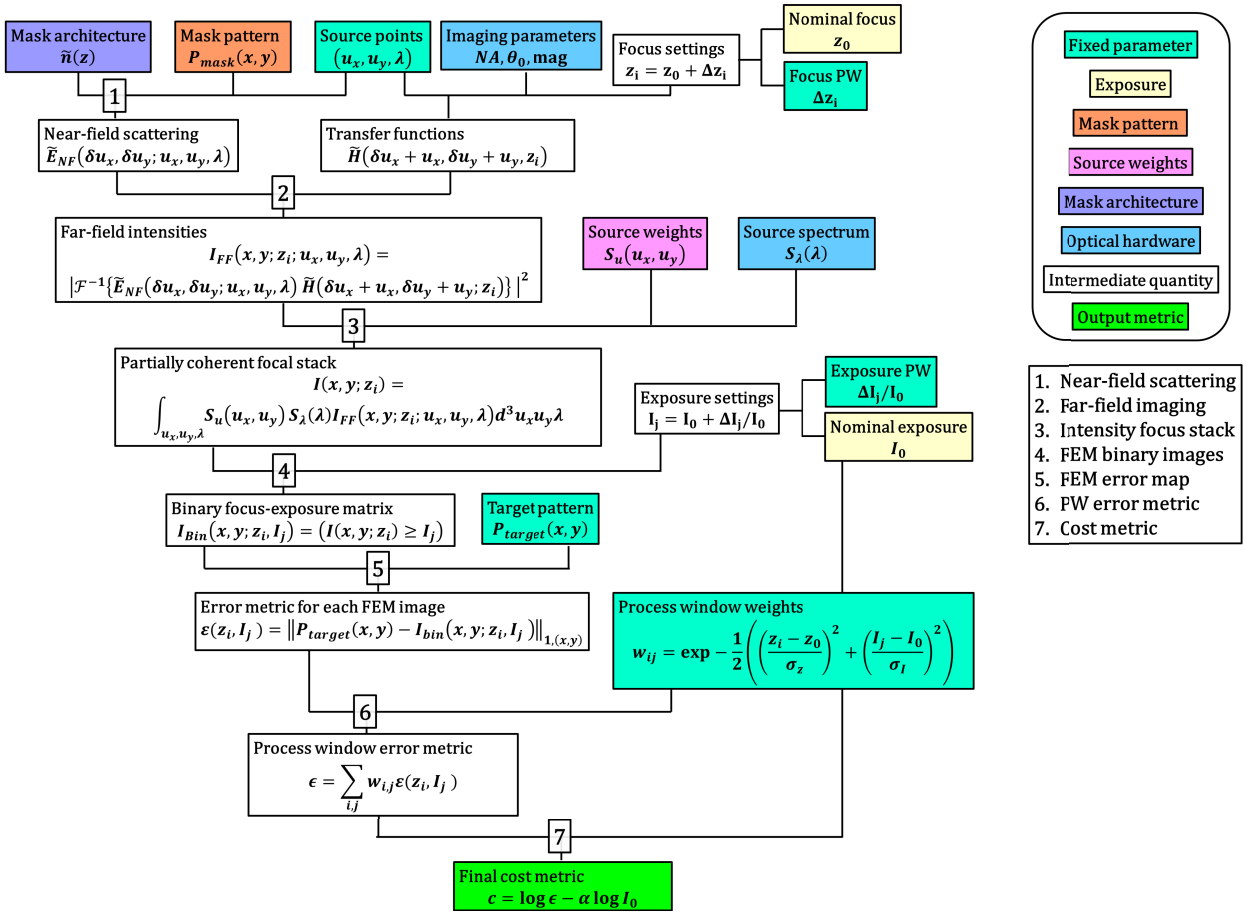

Figure 2.10: Flow chart for calculating FEM error metric using Abbe's method for imaging and DblSc for scattering.

level by subtracting off the binary target pattern and computing an error metric for each image, such as mean-absolute error of the FEM image minus the target pattern. We convert this grid of error values into a single error metric by taking a weighted sum over the process window (PW), using Gaussian decaying weights such that errors in the center of the PW are more significant than errors at the edge. This process window error metric could be interpreted as computing the expectation value of the error metric assuming focus and exposure errors are sampled independently from normal distributions with zero mean and known standard deviation. We finally compute a cost metric which combines both patterning error and throughput. For PW patterning error ϵ and nominal exposure threshold I_0, we define the cost as $c = \log\epsilon - \alpha\log I_0$ for some predefined $\alpha > 0$. To put this metric into words: try to print with the lowest expected error at the highest possible throughput. The trade-off between pattern fidelity and throughput can be controlled by making α smaller

(high fidelity) or larger (high throughput).

2.4.2 SMO for H lines

We now present an example of using this framework of computing a process-window cost metric, based on partially coherent imaging with Abbe's method + DblSc for M3D. We show a case of horizontal lines and spaces, with optimized dipole illumination. We perform the same source + mask optimization on 3 mask architectures (2 aPSMs + 1 TaN), for H lines and spaces with k_1 ranging from 0.3-0.6. Results are summarized in Figure 2.11.

The optimized sources in Figure 2.11 (a) are in some cases extremely asymmetric—particularly for $k_1 \approx 0.4$ for the TaN mask (fourth and fifth from the left, bottom row). Interestingly, these cases where the source is maximally asymmetric are precisely where the TaN mask is most clearly outperformed by the aPSMs. Exposure latitude (EL) in Figure 2.11 (e) is similar for the 2 aPSMs, and for both there is a substantial improvement over TaN for k_1 from 0.35-0.5.

At the same time as they have greater EL, the aPSMs also generally have higher intensity (higher throughput) than TaN, shown in Figure 2.11 (d). We define throughput as $1/\tau_0 = I_0 k_1^2$; the factor of I_0 represents the optical efficiency, whereas the factor of k_1^2 represents the resolution-sensitivity tradeoff, i.e. to maintain the same number of photons per feature we must increase the dose per area proportional to k_1^2 (the area of the unit cell).

We can combine EL and throughput into a single metric, $EL^2/\tau_0 = EL^2 I_0 k_1^2$, shown in Figure 2.11 (f). This single metric can be thought of as throughput at a fixed level of patterning fidelity, considering random variations in photon absorption (dose) and sensitivity to dose fluctuations (EL). If EL degrades, the pattern by definition becomes more sensitive to fluctuations in dose, therefore requiring reduced photon stochastics to print at the same error rate, requiring a higher exposure dose. From this metric (EL^2/τ_0), we can again see the clear benefits of aPSMs, which now are apparent for $k_1 > 0.35$. Interestingly all 3 architectures show that it is "easier" to print at $k_1 = 0.3$ than at some larger k_1 values, possibly related to the different illumination conditions used.

For all 3 masks there is a decrease in EL^2/τ_0 by over an order of magnitude for just a factor of 2 decrease in k_1, giving an estimate for the increased exposure dose required to print at the same error rate. This factor can be broken down as follows: a factor of 4 from reducing unit area; a factor of approximately 4 by reducing EL roughly in half; and the remainder due to reduced optical efficiency. This is essentially an explanation of the famous RLS tradeoff; as you reduce the feature size you must split the photons between more features, while at the same time each feature becomes more sensitive to photon stochastics and is also less optically efficient. Therefore, printing smaller features at the same error rate generally requires a higher exposure dose, because one needs an increased number of photons to be absorbed in each of the increased number of features even after accounting for the reduced optical efficiency. If this increase in exposure dose is intolerable one can always use a faster resist, but at the cost of worsening photon stochastics, resulting in higher error rates that would at some point become an even bigger concern than reduced throughput. Of course

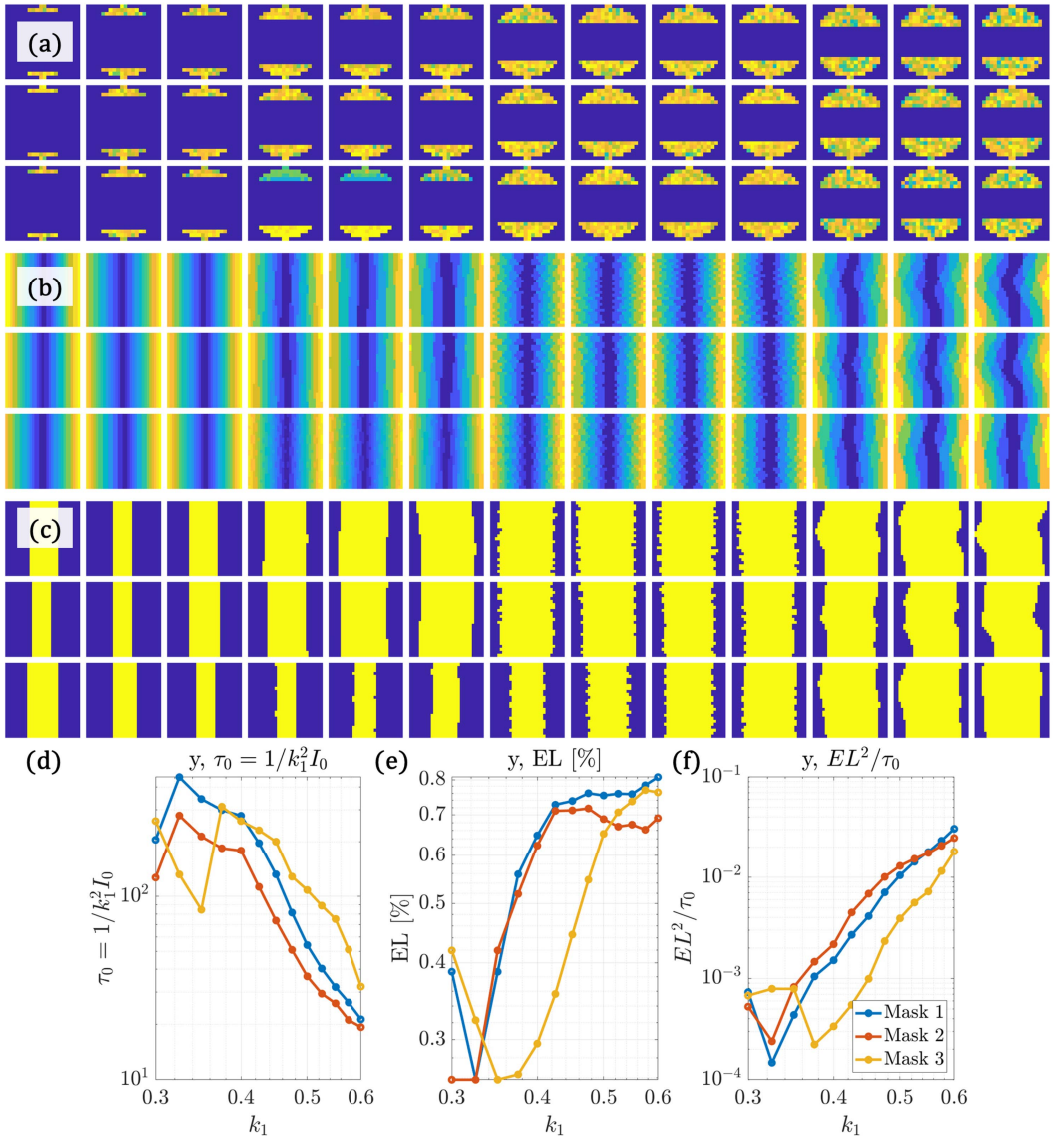

Figure 2.11: SMO for H lines and spaces with k_1 ranging from 0.3-0.6 for 3 photomasks. (a) Optimized sources. (b) Error through focus and dose. (c) Process window. (d) Throughput (efficiency and unit area). (e) Exposure latitude. (f) Throughput accounting for exposure latitude.

at any given feature size, manufacturers will look for ways to maximize throughput, but nonetheless according to these arguments, at every transition to a smaller design rule the minimum allowable exposure dose that a manufacturer can "get away with" will increase.

Chapter 3

Designing EUV Photomasks

In this chapter we explore 3 concepts for designing the architecture of multilayer and absorber on an EUV mask. These designs all provide benefits in terms of either throughput or image contrast.

In Section 3.1, which is based on the paper "Modeling high-efficiency extreme ultraviolet etched multilayer phase-shift masks"[42], we discuss the possibility of an etched multilayer phase shift mask (PSM), which would create phase-only modulation by selectively etching the multilayer. This type of mask architecture would be restrictive in terms of what kinds of patterns could be printed; only dense periodic patterns near the diffraction limit could easily be produced. However, high-resolution periodic patterns are often among the most difficult to print, so any improvement could potentially be quite valuable. And the gains reported are not insignificant: as much as 6x higher throughput could be obtained. This could enable printing with a much higher dose at the same throughput, which would allow for the higher photon density required to print smaller features, without increasing the exposure time.

In Section 3.2, which is based on the paper "Advanced multilayer mirror design to mitigate EUV shadowing"[43], we explore a new type of multilayer design. Instead of the traditional Bragg reflector which has a periodic unit cell of Mo and Si, we introduce an aperiodic multilayer in which the thickness of each layer has been perturbed to better print a specific feature under specific illumination. We focus on a case which suffers from some of the most severe M3D, which is horizontal dipole illumination at the edge of the pupil to print horizontal lines and spaces near the diffraction limit. We find significant improvements in throughput from an aperiodic multilayer design with oscillator layer thicknesses that introduces a secondary peak into the Bragg reflector. Even though the secondary peak is outside of the collection aperture, it can still improve the overall efficiency because some fraction of this reflected light will be re-scattered into the aperture after being reflected by the multilayer. Evaluating the designs with rigorous simulations, we find up to 22% and 14% increased throughput for the 0.33 and 0.55 NA systems respectively. Although much more modest than the improvements from an etched PSM, this optimized multilayer is much more general, and could be compatible with any absorber type. We also find that there is a large improvement for small features and relatively little change for larger features, implying

that this multilayer, which is optimized only for a single minimum-pitch feature, can offer substantial benefits close to this minimum feature size without sacrificing much performance for larger features.

Finally in Section 3.3 we consider co-optimizing the absorber and multilayer to print square contacts near the resolution limit with an attenuated phase shift mask (aPSM). We introduce an optimization metric which allows for an arbitrary tradeoff between contrast and throughput, and produce a range of designs for 3 feature sizes. We see that high power multilayer designs are very similar to traditional designs, but do have a pitch dependence where smaller features require extending the angular bandwidth slightly beyond the collection aperture, to enhance the reflection of diffracted light which can be re-scattered into the aperture (similar to the secondary peak in Section 3.2). The high contrast designs are probably impractical in that the throughput is orders of magnitude lower than high-power designs, however they are still worth studying because they exhibit interesting physical behavior that sheds light onto EUV M3D effects. The most interesting behavior, seen in the smallest two features, is that the effective reflection plane of the multilayer can be brought very close to the surface, and even slightly above the surface for a small range of angles. These high contrast designs align the illumination angles further from normal incidence with these nulls in the effective reflection plane. This means that light reflected at these angles will not propagate into the multilayer and therefore suffer far less from M3D. Interestingly, this causes the optimal phase shift for the aPSM absorber to be closer to π than 1.2π, which strongly suggests that the reason for the additional phase shift in the absorber is precisely due to the phase shift of the multilayer. In other words, if multilayer phase effects can be reduced in this way, the additional phase compensation required for the absorber will also decrease.

3.1 Etched multilayer PSM

This section explores an EUV mask architecture with no absorber and only multilayer. The contrast is created entirely by phase modulation, with part of the multilayer etched to create a phase shift between etched and un-etched multilayer. We will show that this can produce substantial improvements in throughput (up to 6x) for features near the resolution limit.

This section is heavily based on the paper "Modeling high-efficiency extreme ultraviolet etched multilayer phase-shift masks"[42], originally published in JM3 (Journal of Micro/Nanopatterning, Materials, and Metrology) in 2017.

Phase shift masks are well known to provide benefits in terms of resolution and process window [51, 27], and there is an existing body of work on modeling phase-shift masks for DUV lithography including fast algorithms [1] and thorough studies of electromagnetic effects [10, 13]. Furthermore, application to EUV masks has been shown for both relief masks [35] and etched multilayer masks [16]. Furthermore, Naulleau et. al. demonstrated the use of an etched multilayer phase shift mask (Fig. 3.1) for very high efficiency printing of dense line-space and contact array patterns [34]. The results for the contact array are particularly

encouraging, with a demonstrated 8x brighter image and 7x shorter exposure time [34]. In this paper we employ rigorous Finite Difference Time-Domain (FDTD) simulations carried out with Panoramic Hyperlith software, along with edge placement error (EPE) process window analysis to explore the feasibility of using etched multilayer phase-shift masks for EUV lithography.

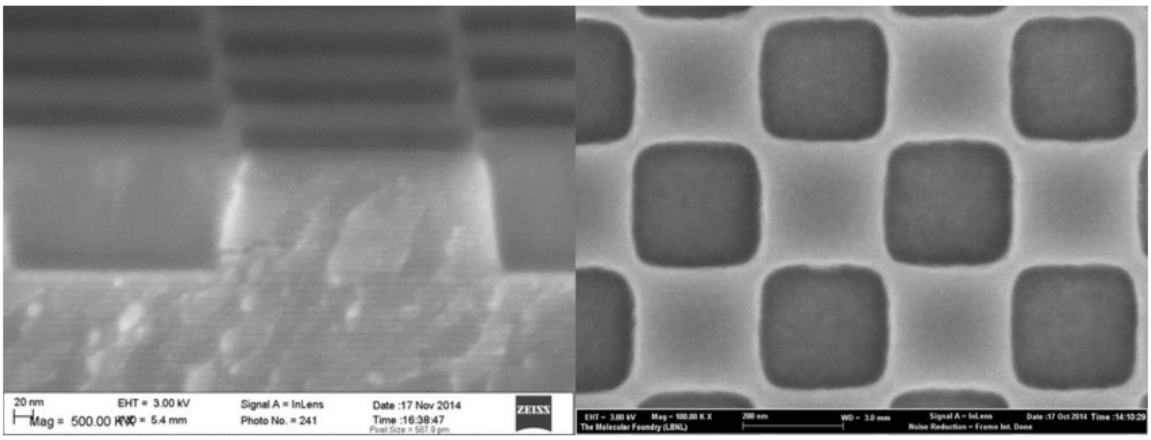

Figure 3.1: Cross-section and top down scanning electron micrograph of 50-nm pitch coded contact array; light areas are unetched multilayer and dark areas are etched multilayer[34].

3.1.1 Alternating PSM theory

An alternating phase shift mask (PSM) seeks to create image contrast by only modulating the phase with no attenuation. This is not a general purpose photomask, but for certain specific patterns it can provide substantial improvements in contrast and throughput. We explore the benefits of phase-shift masks for printing dense features near the resolution limit. The resolution limit, is given by:

$$R = k_1 \frac{\lambda}{NA}$$

where R is the smallest half-pitch that can be printed, λ is the wavelength, and NA is the image-side numerical aperture; the theoretical lower limit for k_1 is 0.25 [29]. However, achieving this resolution limit with an absorber mask requires off-axis illumination to shift diffraction orders with higher spatial frequencies into the imaging pupil; unfortunately, this illumination also has the effect of shifting other diffracted orders out of the pupil, reducing the incident power at the wafer. By contrast, phase-shift masks can achieve the theoretical lower limit for k_1 without resorting to off-axis illumination by exploiting frequency-doubling as shown in Fig. 3.2 for a line-space pattern with wafer pitch $p_{wf} = 25$nm, corresponding to $k_1 \approx 0.3$ for 0.33 NA. If the phase-shift mask has a dense, uniform array of alternating

reflectance 1 and -1, the 0-order or average electric field vanishes. For a feature near the resolution limit, where only the 0 and ±1 orders pass through pupil with on-axis illumination, this results in 2-beam imaging of the ±1 orders, which creates an electric field at the wafer that oscillates between positive and negative. This results in an intensity image with twice the spatial frequency of the electric field.

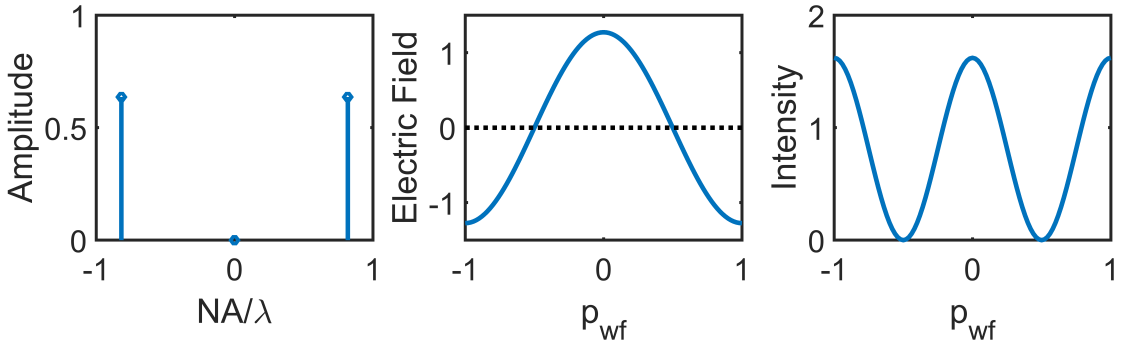

Figure 3.2: Left: Amplitudes of 0 and ±1 scattered orders for phase-shift mask. Center: Electric field at wafer. Right: Intensity image.

Since this process is based on 2-beam imaging of the ±1 orders, it cannot be used to print all pitches. Clearly the pitch must be large enough that the ±1 orders pass through the pupil. However, the pitch must be small enough that the ±3 orders miss the pupil; we will allow the ±2 orders into the pupil since the ideal phase-shift mask suppresses all even diffracted orders. Assuming conventional illumination with partial coherence radius σ, each diffraction order will be a shifted copy of the illumination pattern, centered about normalized spatial frequency $\frac{1}{4k_1}$, where the factor of 4 is due to frequency-doubling and the definition of k_1 in terms of half-pitch. By solving for k_1 where the ±1 orders start to leave the pupil and where the ±3 orders start to enter the pupil, we obtain the following condition on feature size:

$$\frac{1}{4(1-\sigma)} < k_1 < \frac{3}{4(1+\sigma)} \tag{3.1}$$

Evaluating this expression at the fully coherent case $\sigma = 0$, we obtain $0.25 < k_1 < 0.75$, while using the more realistic value of $\sigma = 0.2$, we obtain approximately $0.3 < k_1 < 0.6$.

3.1.1.1 Efficiency Gains for Lines and Spaces

To quantify the theoretical efficiency gains of phase-shift masks over absorber masks, we consider two idealized mask designs each composed of equal regions of alternating reflectance R_1 and R_2: $(R_1, R_2) = (1, 0)$ for the absorber mask, and $(R_1, R_2) = (1, -1)$ for the phase-shift mask. We will use each mask type to print a line-space pattern with wafer pitch $p_{wf} = 25$nm using $\lambda = 13.5$nm, $NA = 0.33$, magnification $m = 4$, meaning that this feature

has $k_1 = \frac{25\text{nm}}{2}\frac{0.33}{13.5\text{nm}} \approx 0.3$. For the absorber mask we use dipole illumination with a mask pitch $p_{mask} = mp_{wf}$, while due to frequency doubling for the phase-shift mask we use on-axis illumination with a mask pitch $p_{mask} = 2mp_{wf}$. Using Fourier diffraction methods as outlined in Smith [48], we may model the reflection function with arbitrary complex reflection coefficients R_1 and R_2 as:

$$R(x) = \left(R_1 \text{rect}\left[\frac{x}{p/2}\right] + R_2 \text{rect}\left[\frac{x - \frac{p}{2}}{p/2}\right]\right) * \text{comb}\left(\frac{x}{p}\right).$$

We then decompose the reflection function by its Fourier transform, representing decomposition into plane waves. Due to the periodic nature of the pattern, the Fourier transform only contains spatial frequencies that are integer multiples of $\frac{1}{p}$:

$$R(x) = \sum_j a_j e^{i2\pi x \frac{j}{p}},$$

where the a_j are called the diffraction efficiencies. Using the convention that the incident intensity on the mask is normalized to 1, these are given by:

$$a_j = \begin{cases} \frac{R_1+R_2}{2}, j = 0 \\ \frac{\sin\left(\frac{\pi j}{2}\right)}{\pi j}\left[R_1 + R_2 e^{-i\pi j}\right], j \neq 0 \end{cases} \tag{3.2}$$

From these formulas, we calculate the 0 and ± 1 order diffraction efficiencies for the two masks, denoted a_j^{Abs} and a_j^{PSM} for the absorber and phase-shift mask, respectively:

$$a_j^{Abs} = \begin{cases} \frac{1}{2}, j = 0 \\ \frac{1}{\pi}, j = \pm 1 \end{cases}, a_j^{PSM} = \begin{cases} 0, j = 0 \\ \frac{2}{\pi}, j = \pm 1 \end{cases}$$

Modeling the illumination poles as mutually incoherent delta-functions, we then compute the aerial images for the absorber and phase-shift masks, shown in Fig. 3.3. For the phase-shift mask, the on-axis illumination keeps the diffraction pattern centered in the pupil, allowing the 0 and ± 1 orders to pass through. In contrast, each of the two illumination poles for the absorber mask shifts the diffraction pattern, allowing only the 0 and 1 or 0 and -1 orders to pass through. Due to, symmetry we need only consider one illumination pole for the absorber mask, since both poles create identical images that add incoherently. Finally, note that since all the diffraction efficiencies are positive and real, all waves are in phase at $x = 0$, meaning that for either mask the maximum intensity is simply given by:

$$I_{max} = I(0) = |E(0)|^2 = \left|\sum_j a_j\right|^2,$$

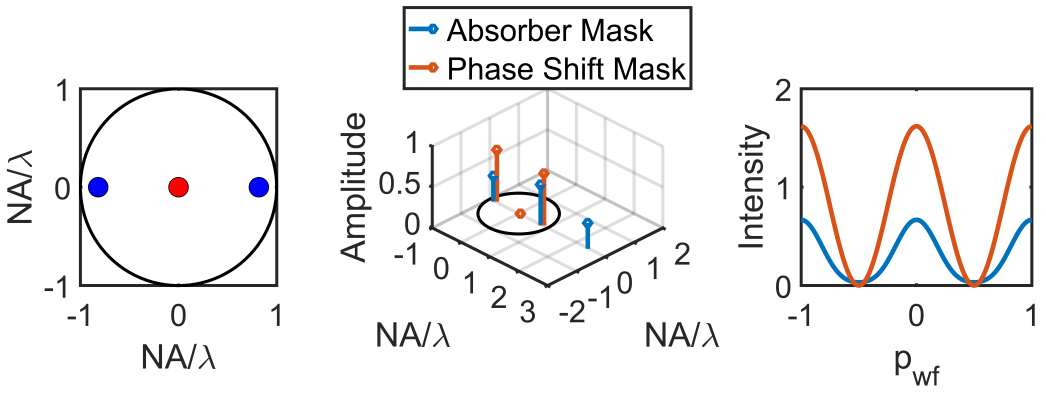

Figure 3.3: Left: Dipole illumination for absorber mask (blue) and conventional illumination for phase-shift mask (red). Center: Pupil plane plot of one illumination pole for each mask. Right: Aerial images from each mask.

where the summation is over orders j that fall within the pupil. This results in the phase-shift mask's peak intensity being higher by a factor of:

$$\frac{\left(\sum_{j=-1,0,1} a_j^{PSM}\right)^2}{\left(\sum_{j=0,1} a_j^{Abs}\right)^2} = \frac{\left(\frac{2}{\pi}+\frac{2}{\pi}\right)^2}{\left(\frac{1}{2}+\frac{1}{\pi}\right)^2} \approx 2.42$$

3.1.1.2 Efficiency Gains for Contacts

The contact array pattern is the two-dimensional version of the line-space pattern. For these special cases of $(R_1, R_2) = (1, 0)$ and $(R_1, R_2) = (1, -1)$ for the two masks, we obtain a separable form of the 2D reflection function:

$$R(x, y) = R(x)R(y)$$

This yields the formulas for the contact array diffraction efficiencies in terms of line-space diffraction efficiencies:

$$a_{j,k}^{Abs} = a_j^{Abs} a_k^{Abs}, a_{j,k}^{PSM} = a_j^{PSM} a_k^{PSM}$$

Similarly to the line-space pattern, the on-axis illumination for the phase-shift mask keeps the diffraction pattern centered in the pupil, now allowing a total of 9 orders to pass through. Again due to symmetry, we need only consider a single pole of the absorber mask, which shifts the diffraction pattern, allowing only 4 orders to pass through. Since again all diffraction

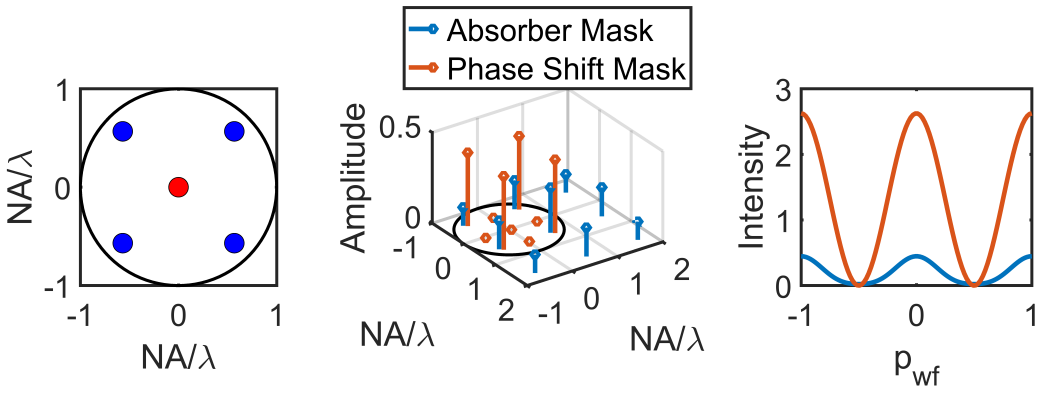

Figure 3.4: Left: Quadrupole illumination for absorber mask (blue) and conventional illumination for phase-shift mask (red). Center: Pupil plane plot of one illumination pole for each mask. Right: Aerial image cross-sections from each mask.

efficiencies are positive and real, we apply the same calculation for peak intensity, resulting in the phase-shift mask's peak being brighter by a factor of:

$$\frac{\left(\sum_{j,k=-1,0,1} a_j^{PSM} a_k^{PSM}\right)^2}{\left(\sum_{j,k=0,1} a_j^{Abs} a_k^{Abs}\right)^2} = \frac{\left[(\frac{2}{\pi})(\frac{2}{\pi}) + (\frac{2}{\pi})(\frac{2}{\pi}) + (\frac{2}{\pi})(\frac{2}{\pi}) + (\frac{2}{\pi})(\frac{2}{\pi})\right]^2}{\left[(\frac{1}{2})(\frac{1}{2}) + (\frac{1}{2})(\frac{1}{\pi}) + (\frac{1}{\pi})(\frac{1}{2}) + (\frac{1}{\pi})(\frac{1}{\pi})\right]^2} \approx 5.86$$

The differences between imaging with the two masks are outlined in Fig. 3.4.

3.1.1.3 Summary of Efficiency Gains

The efficiency gains for the idealized phase-shift mask are shown below in Table 3.1. Note that these gains are due both to the larger reflective area on the phase-shift mask as well as to the inefficiency of dipole and quadrupole illumination patterns, which shift significant scattered orders outside of the pupil reducing the power incident at the wafer.

3.1.2 Etched multilayer PSM for EUV

3.1.2.1 Multilayer mirror structure

Now that the theoretical advantages of phase-shift masks have been quantified, we turn our attention to one possible realization of phase-shift masks for EUV lithography: etched multilayer phase-shift masks. All EUV masks employ a multilayer mirror substrate due to

Table 3.1: Comparison of printing near minimum pitch using absorber and phase-shift masks. Analysis based on thin-mask model. $\lambda = 13.5$nm, $NA = 0.33$, $m = 4$. Dipole illumination poles placed at $(\frac{\pm 1}{2p_{wf}}, 0) = (\frac{\pm 1}{50nm}, 0) = (\pm 0.82, 0)\frac{NA}{\lambda}$. Quadrupole illumination poles placed at $(\frac{\pm 1}{2p_{wf}}, \frac{\pm 1}{2p_{wf}}) = (\frac{\pm 1}{72nm}, \frac{\pm 1}{72nm}) = (\pm 0.57, \pm 0.57)\frac{NA}{\lambda}$.

Pattern	Mask Type	Illumination	p_{wf} (nm)	p_{mask} (nm)	Peak Intensity	Peak Ratio
Line Space	Absorber	Dipole	25	100	0.67	1
Line Space	Phase Shift	Conventional	25	200	1.62	2.42
Contact Array	Absorber	Quadrupole	36	144	0.45	1
Contact Array	Phase Shift	Conventional	36	288	2.62	5.86

the need for high reflectivity at EUV. All simulations in this section are based on a multilayer design of Mo-Si bilayers with a d-spacing of $d = 6.95$nm and a Mo duty-cycle of $\gamma = 0.4$, with a $t_{Cap} = 2$nm-thick Ru capping layer, operating at a wavelength of $\lambda = 13.5$nm with off-axis illumination 6^o from normal. More complicated multilayer stacks accounting for such effects as inter-diffusion layers or interface roughness are not considered, but may be a subject of future work. For this multilayer mask structure, over the angular range 2^o-10^o (roughly corresponding to the mask-side acceptance angles), the reflectivity is fairly uniform and always over 0.7. The phase response on the other hand has significant non-uniformity, with a phase shift of over 0.1 waves over this range of angles. Fig. 3.5 shows the multilayer angular response, calculated using n and k values from the CXRO database[18], using the transfer-matrix method [17].

3.1.2.2 Phase Shift from Etching

As discussed in the previous section, the ideal phase-shift mask design has two regions with equal reflectivity and a relative phase shift of π. We denote the duty-cycle of the unetched pattern D and the etch depth in bilayers N_{Etch}. In dimensions of length, the width of the unetched region is $w = Dp_{mask}$ and the etch depth is $z_{Etch} = t_{Cap} + dN_{Etch}$. Again using the transfer-matrix method, we compute the reflectivity and phase-shift for each integer value of N_{Etch} from 0-20 bilayers, as shown in Fig. 3.6. Note that the reflectivity is nearly constant over this range, while the phase obeys a linear relationship with N_{Etch}. After an etch depth of 20 bilayers, a π phase-shift is achieved. However, we will later see that this large etch depth, ($z_{Etch} = 141$nm), is far from the regime of small vertical and large horizontal dimensions where the thin-mask and transfer-matrix methods are accurate. This introduces significant electromagnetic edge effects, which we will later compensate by adjusting N_{Etch} and D.

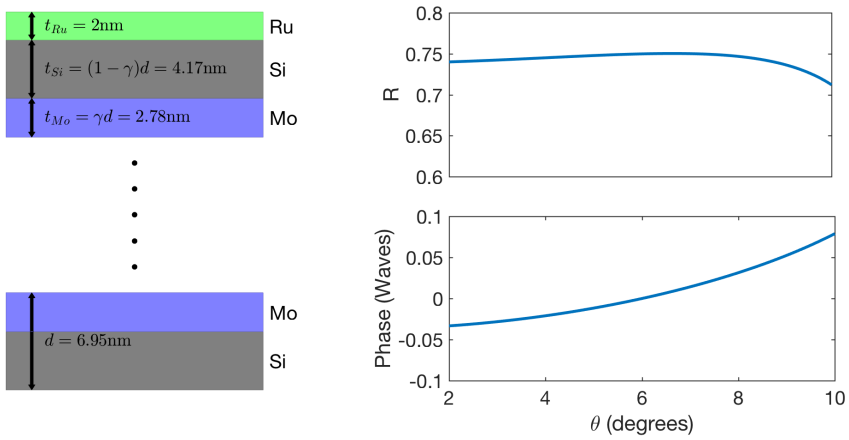

Figure 3.5: Left: multilayer stack design. Right: reflected intensity and phase vs angle. n and k values drawn from CXRO database for $\lambda = 13.5$nm. Values used: Si: $n = 0.99900154$, $k = 0.0018265$; Mo: $n = 0.923791$, $k = 0.0064358$; Ru: $n = 0.8864$, $k = 0.017066$.

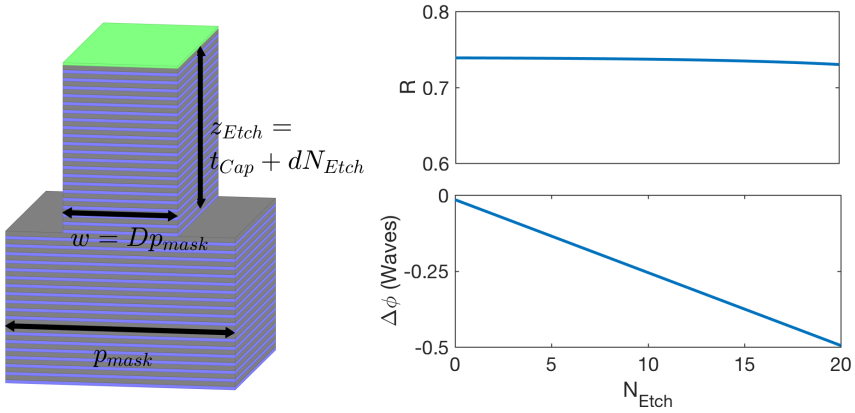

Figure 3.6: Left: schematic of pattern dimensions. Right: reflected intensity of etched region and phase shift between regions vs N_{Etch}.

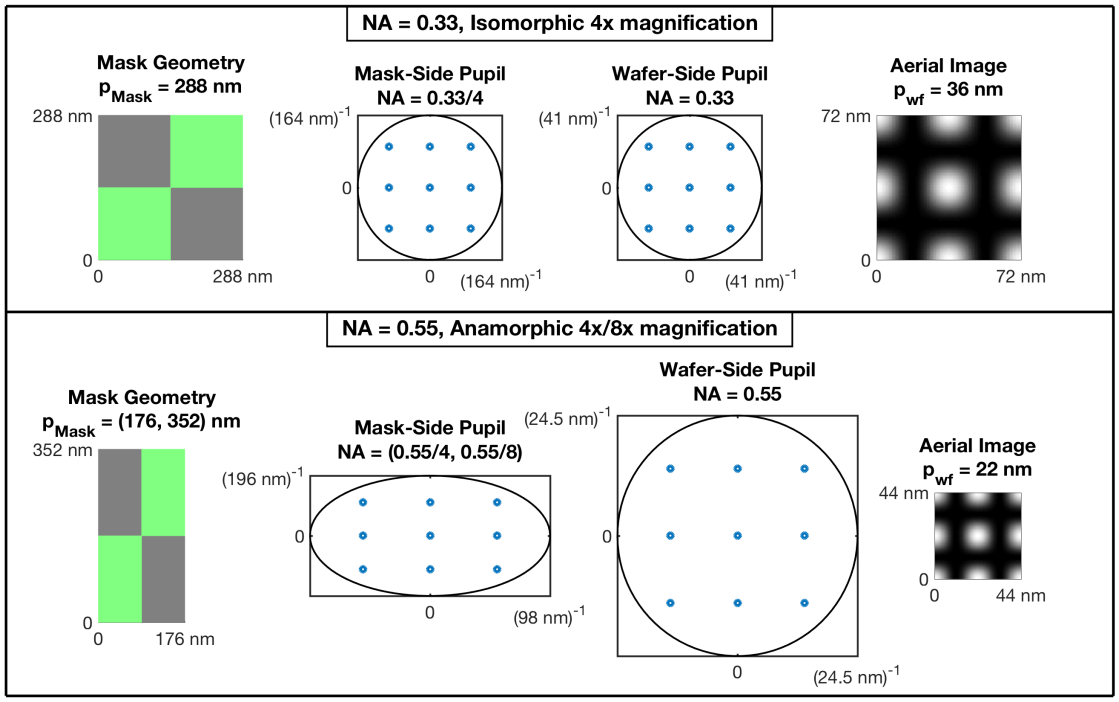

Figure 3.7: Top: 0.33 NA isomorphic 4x magnification to print p_{wf} = 36nm contacts. Bottom: 0.55 NA anamorphic 4x/8x magnification to print p_{wf} = 22nm contacts. Left to right: Mask geometry, mask-side pupil, wafer-side pupil, aerial image.

3.1.2.3 Anamorphic vs Isomorphic Magnification

Unlike traditional lithography tools, the next-generation of EUV lithography tools will use anamorphic magnification, which applies a different magnification to the two directions to minimize mask shadowing effects [36, 20]. We will analyze the performance of etched multilayer phase-shift masks using both current-generation 0.33 NA isomorphic 4x magnification, as well as next-generation generation 0.55 NA anamorphic 4x/8x magnification. An example of printing isomorphic contacts using both mask technologies is presented in Fig. 3.7.

3.1.3 M3D in etched multilayer PSM

Due to the off-axis illumination in EUV lithography, there is an asymmetry between features oriented with the direction of periodic variation perpendicular to the plane of incidence formed by the chief ray and the mask surface normal (non-shadowing) or in the plane of incidence (shadowing).

3.1.3.1 Multilayer Dispersion of Diffracted Waves

One source of asymmetry is ray-optical shadowing, whereby one sidewall is illuminated by incident light rays, and the other sidewall casts a shadow, illustrated in Fig. 3.8 (top). The angular dispersion of the multilayer is another source of asymmetry. As shown in Fig. 3.5, plane waves with different angles of incidence relative to normal acquire a different amplitude and phase upon reflection from the multilayer mirror. This applies to diffraction orders, which propagate through the multilayer at different angles, particularly affecting the relative phase of the orders. This angular dispersion is a source of asymmetry between the shadowing and non-shadowing orientations because the angle of the j^{th} diffracted order relative to the multilayer surface normal (θ_j) is different for the two orientations, given by:

$$\sin\theta_j = \begin{cases} j\frac{\lambda}{p_{mask}} + \sin\theta, & \text{Shadowing} \\ \sqrt{(j\frac{\lambda}{p_{mask}})^2 + \sin^2\theta}, & \text{Non-shadowing} \end{cases} \tag{3.3}$$

where θ is the angle of incidence (6^o). These angles are shown for $j = 0, \pm 1$ in Fig. 3.8, where we see that in the non-shadowing orientation the ± 1 orders propagate at the same angle relative to normal, causing them to receive the same phase shift. However, in the shadowing orientation the ± 1 orders propagate above and below the angle of incidence respectively, leading to a relative phase-shift between the orders which can degrade image quality. Also note that in the 0.33 NA isomorphic system, the angular distance between the 0 and 1 orders is much larger in the shadowing direction than in the non-shadowing direction. This can be understood by Eq. 3.3, due to squaring the $\frac{\lambda}{p_{mask}}$ term (≈ 0.0675), which is significantly smaller than the $\sin\theta$ term (≈ 0.1045). The anamorphic system mitigates this effect by doubling the size of all shadowing orientation features on the mask, which leads to a similar angular distance between the 0 and 1 orders in both orientations. Overall, compared to the 0.33 NA isomorphic system, the 0.55 NA anamorphic system simultaneously produces larger diffracted angles and stronger edge effects in the non-shadowing direction, and conversely smaller diffracted angles, reduced edge effects, and reduced geometric shadowing in the shadowing direction. This is illustrated in Fig. 3.8 by the increased non-shadowing dispersion and decreased shadowing dispersion in the 0.55 NA system.

3.1.3.2 Pitch-Dependence

Thin-mask and transfer-matrix methods are only accurate in the regime of small vertical and large horizontal dimensions (approximately vertical dimensions $< \frac{\lambda}{2}$ and horizontal dimensions $> 2\lambda$)[29], which are not valid assumptions for these features as the etch depth is on the order of 10λ. Using rigorous Finite-Difference Time-Domain (FDTD) analysis, we may quantify the deviation of the amplitudes of the diffracted electric fields as a function of etch depth from the thin-mask transfer-matrix prediction, which is calculated by computing the R_1 and R_2 values for the etched and unetched multilayer stacks as in Fig. 3.6 then calculating diffraction efficiencies from Eq. 3.2. This is presented below in Fig. 3.9, where

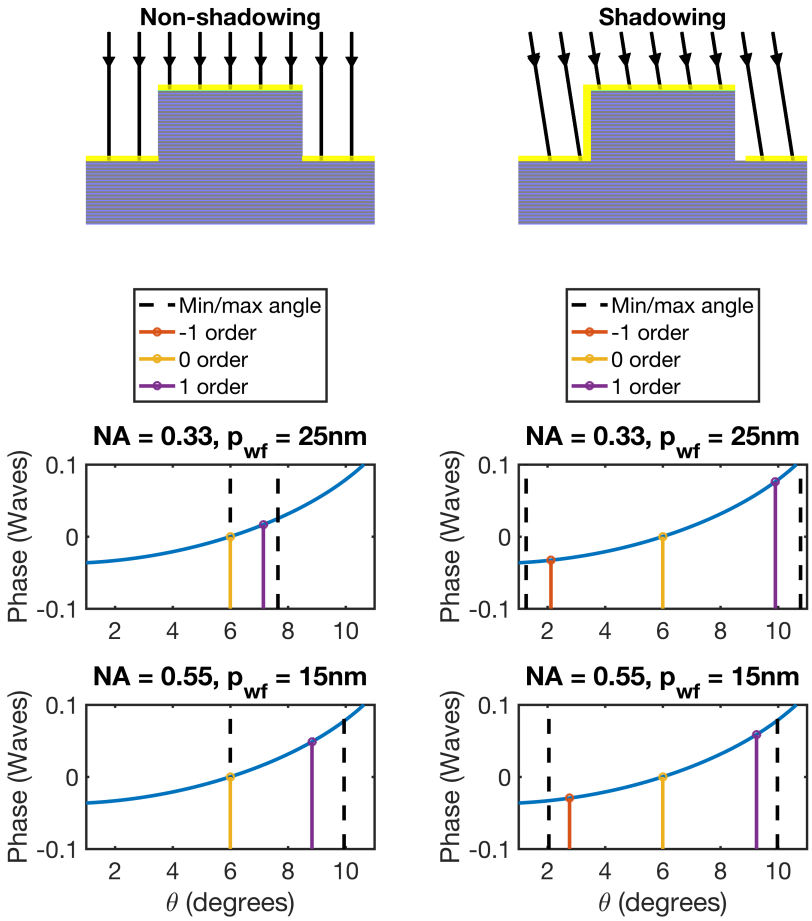

Figure 3.8: Left: Non-shadowing orientation. Right: shadowing orientation. Top to bottom: Incident ray diagram; diffracted angles at 0.33 NA isomorphic 4x magnification, $p_{wf} = 25$nm; diffracted angles at 0.55 NA anamorphic 4x/8x magnification, $p_{wf} = 15$nm.

we see that decreasing the pitch causes further deviation from the thin-mask model. Again, we use the convention that the incident intensity on the mask is normalized to 1. Whereas the thin-mask model predicts that the 0 order is minimized at $N_{Etch} = 20$ bilayers (when the π phase-shift occurs), the FDTD simulation shows that as the pitch decreases, a deeper etch is required, as shown in Fig. 3.10. Note that this effect of a pitch-dependent phase shift has also been observed in DUV phase-shift masks [13, 10]. Additionally, as the pitch decreases, the amplitudes of the ± 1 orders become attenuated, likely due to a combination of edge effects and a lower multilayer reflectivity at higher angles.

The different etch depths required to print different pitches may limit what features can be printed in a single exposure, because it is not feasible to include multiple etch depths on

the same mask. This effect may be overcome at 0.33 NA, since the difference in optimal etch depth between 25nm ($k_1 \approx 0.3$) and 50nm ($k_1 \approx 0.6$) features is only about 1 bilayer. However, the effect becomes much more concerning at 0.55 NA, where the difference in optimal etch depth is roughly 4 bilayers. These feature sizes were chosen based on $\sigma = 0.2$ and the criterion $\frac{1}{4(1-\sigma)} < k_1 < \frac{3}{4(1+\sigma)} \Rightarrow 0.3 < k_1 < 0.6$. Therefore, the two features cover the entire printable range at $\sigma = 0.2$ and include both the cases of 3-beam imaging of the 0, ± 1 orders ($k_1 \approx 0.3$) and 5-beam imaging of the 0, ± 1, ± 2 orders ($k_1 \approx 0.6$).

3.1.3.3 Orientation-Dependence

Carrying out similar analysis in the shadowing orientation, in Fig. 3.11 we can see that, just as in the non-shadowing orientation, smaller features on the mask require a deeper etch to minimize the 0 order. In this orientation the ± 1 orders become asymmetric, an effect which becomes more severe at smaller mask pitches. However, due to the anamorphic magnification of the 0.55 NA system, features at the same k_1 in the shadowing orientation are larger by a factor of 1.2 on the mask than at 0.33 NA, causing them to experience somewhat less deviation from the thin-mask prediction. This stands in contrast to the non-shadowing orientation (Fig. 3.9), where at 0.55 NA, features at the same k_1 are smaller by a factor of 0.6 on the mask at than at 0.33 NA, causing them to experience significantly more deviation from the thin-mask prediction.

3.1.4 Process window optimization

To quantify the patterning performance of etched multilayer phase-shift masks, we introduce a metric referred to as the maximum uncorrectable Edge Placement Error (EPE), or simply EPE_{max}, which tracks the maximum deviation of all edges from their nominal locations after correcting for an average shift of the entire pattern at best focus and exposure threshold. Note that for the line-space pattern, due to frequency-doubling 2 lines are printed in each electric-field period; since 2 edges are considered in each line, a total of 4 edge points are used to compute EPE_{max}. By contrast for the contact-array pattern, frequency-doubling in both horizontal and vertical directions introduces 4 contacts in each electric-field period; since 2 horizontal and 2 vertical edges are considered for each contact, a total of 16 edge points are used to compute EPE_{max}.

As shown in Fig. 3.12, to calculate this metric we (a) calculate image edge location, $x_i(d,t)$ for each (i,d,t) (edge index, defocus, intensity threshold); (b) subtract off the nominal edge locations x_i^{nom} to obtain: $EPE_i(d,t) \triangleq x_i(d,t) - x_i^{nom}$; (c) calculate the average pattern shift at each (d,t): $\Delta x(d,t) \triangleq \text{mean}_i[EPE_i(d,t)]$; (d) find the best focus, intensity threshold, and pattern shift by minimizing: $(d^*,t^*) \triangleq \text{argmin}_{(d,t)} \max_i |EPE_i(d,t) - \Delta x(d,t)|$, then defining $\Delta x^* \triangleq \Delta x(d^*,t^*)$; (e) calculate uncorrectable EPE for each edge: $|EPE_i(d,t) - \Delta x^*|$; (f) calculate maximum uncorrectable EPE: $EPE_{max}(d,t) \triangleq \max_i |EPE_i(d,t) - \Delta x^*|$.

By tracking the positions of all edges, this metric accounts for many types of patterning errors, including CD errors, placement errors, and telecentricity errors. For any given mask

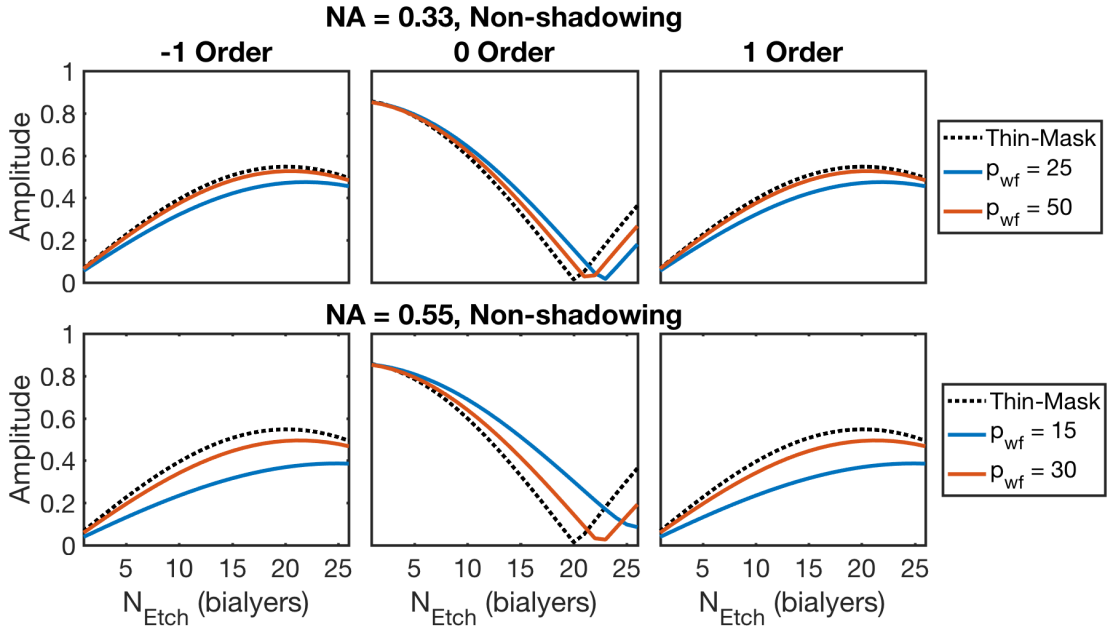

Figure 3.9: Amplitude of 0, ± 1 orders vs N_{Etch} for 0.33 NA (top) and 0.55 NA (bottom) non-shadowing features.

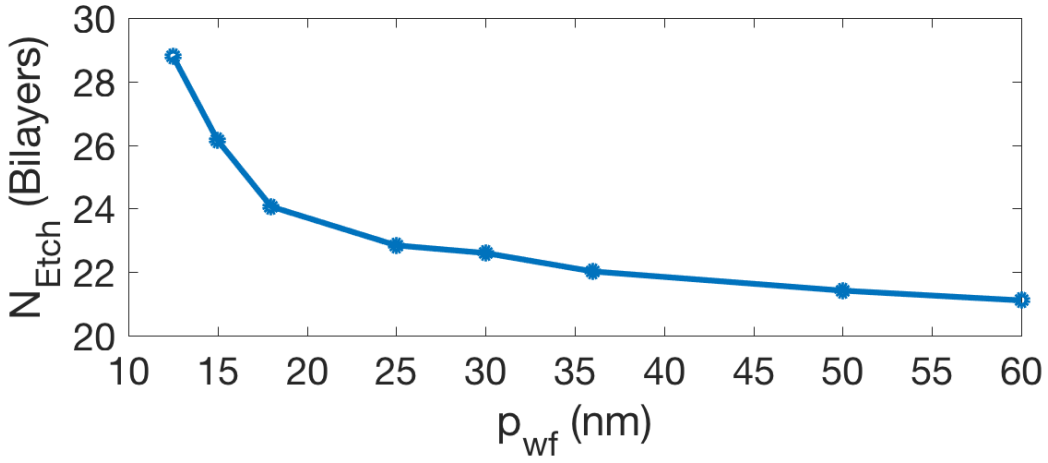

Figure 3.10: Etch depth that minimizes 0 order amplitude vs wafer pitch, non-shadowing features.

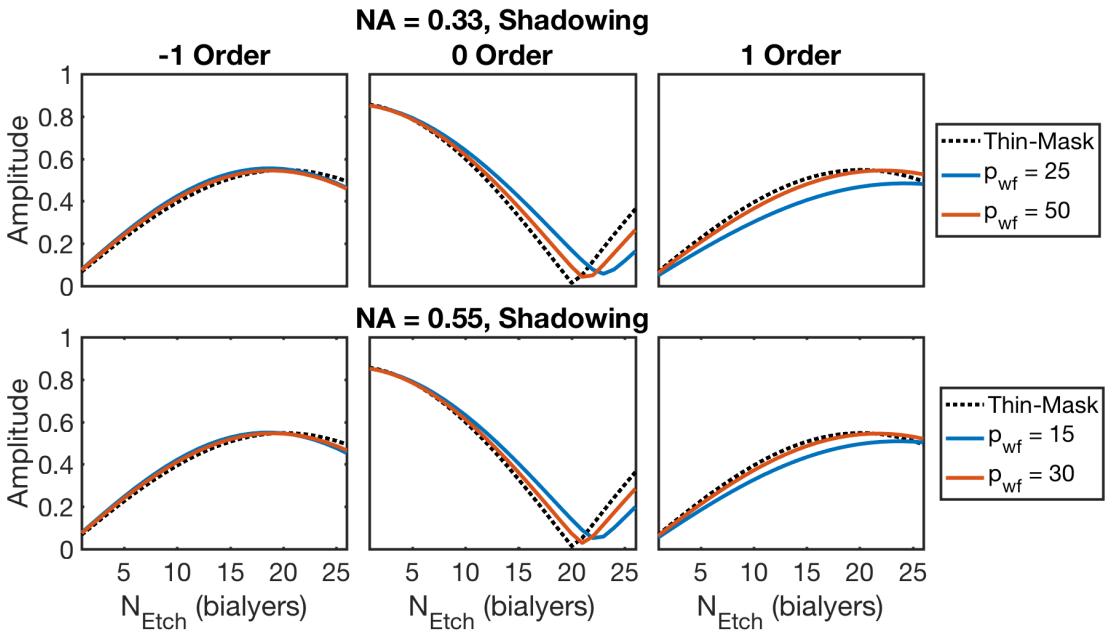

Figure 3.11: Amplitude of 0, ± 1 orders vs N_{Etch} for 0.33 NA (top) and 0.55 NA (bottom) shadowing features.

design, we compute EPE_{max} through focus and intensity threshold, set a specification limit EPE_{spec}, and calculate the process window, or range of acceptable focus and dose errors. We can then optimize the mask design to maximize the area of the process window. Furthermore, we can co-optimize the design for printing multiple features on the same mask by maximizing their process-window overlap.

3.1.4.1 Optimized designs

In this section, we apply the EPE_{max} methodology to optimize mask designs for both the line-space and contact-array patterns, for both the current-generation 0.33 NA isomorphic and the next-generation 0.55 NA anamorphic systems. For each system, we consider the same pitches used previously ($k_1 \approx 0.3, 0.6$). The designs are optimized to maximize the area of the EPE_{max} process window as a function of defocus and intensity threshold, by adjusting the etch depth N_{Etch} and the duty cycle D as defined in Fig. 3.6 for line-space patterns, and by adjusting the etch depth N_{Etch} for contact arrays. All images were calculated using FDTD in Panoramic Hyperlith software, using a single (fully coherent) illumination pole at the center of the pupil, with incident intensity on the mask normalized to 1.

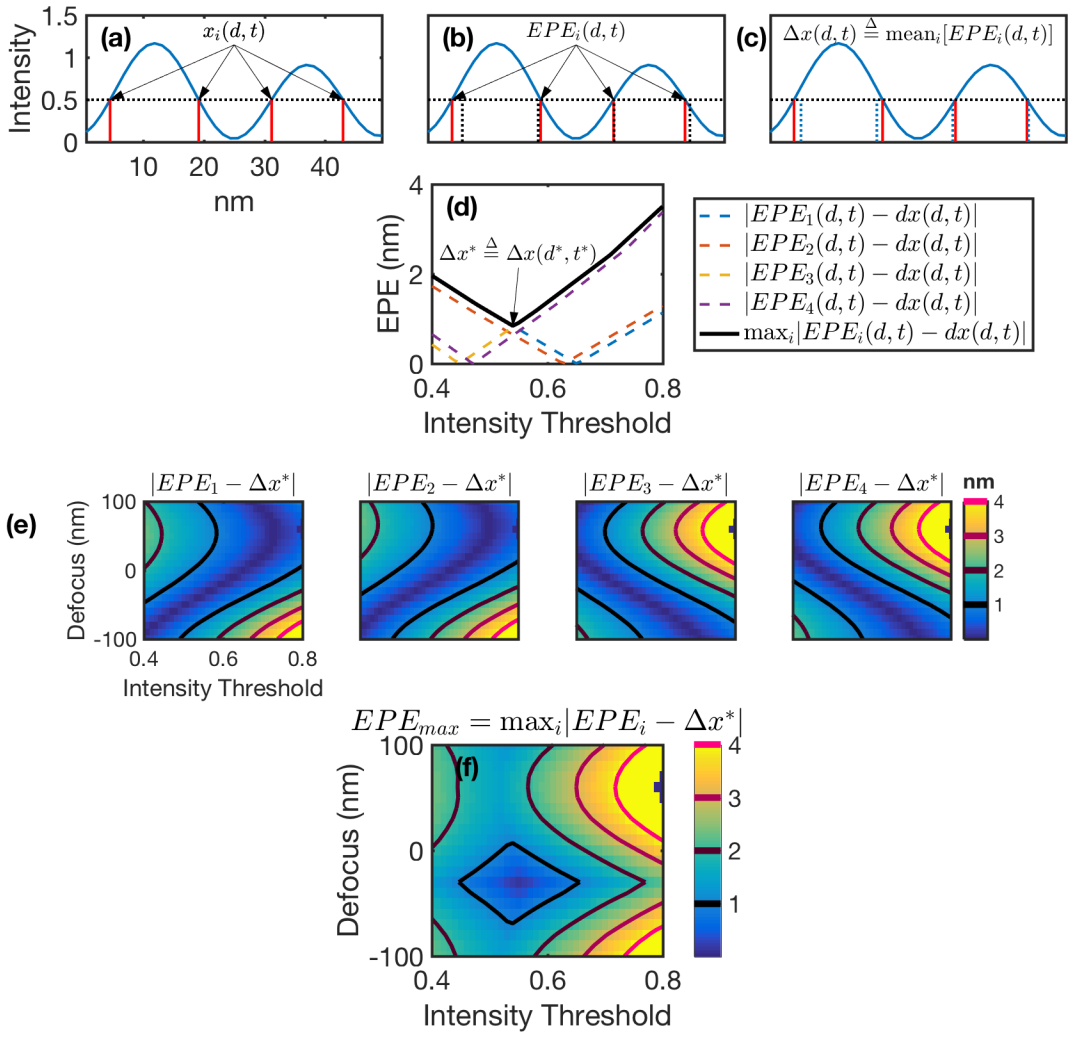

Figure 3.12: Calculation of EPE_{max}

3.1.4.2 Line-space, NA = 0.33 Isomorphic

Below in Fig. 3.13, we show the process windows and optimized aerial images through focus for the 0.33 NA isomorphic system. Each row shows one p_{wf} and orientation, and each column shows one N_{Etch}; the process window shown for each case uses the optimized D. The aerial images through focus use the optimized N_{Etch} and D. Note that for the 3-beam imaging cases ($p_{wf} = 25$nm), an extended depth of focus is achieved at the optimized N_{Etch}. As might be expected, these N_{Etch} values found by process window optimization turn out to be the values that minimize the 0 order amplitude (Fig. 3.10). The process window is narrower for the larger features, due to the presence of the ± 2 orders in the pupil, and is especially restrictive for the $p_{wf} = 50$nm shadowing feature. Another reason the process window is narrower for these features is that the same EPE_{spec} of 1nm is applied to all features despite the factor of 2 difference in pitch.

Whether these process windows are acceptable for manufacturing will depend on which features must be printed on the same mask, as well as the precise specifications for each feature. Printing multiple features in the same exposure would cause a loss in process window, both because there is no single etch depth that is optimal for all features, and because the exposure process windows do not align for all features. As we will later show, sensitivity to etch depth as well as overall depth of focus can be greatly improved with a central obscuration. As a topic of future work, it may also be possible to increase the process window overlap using sub-resolution assist features (SRAFs).

3.1.4.3 Contacts, NA = 0.33 Isomorphic

At an N_{Etch} value that minimizes the 0 order amplitude (24 bilayers) and using an equal checkerboard design, the $p_{wf} = 36$nm contact array pattern enjoys a similar extended depth of focus and wide exposure latitude as the optimized line-space pattern. Furthermore, the nominal intensity threshold at 0.33 NA is 0.9, or 91% of the thin-mask transfer-matrix prediction of 0.99, which is based on a 20-bilayer etch depth using the transfer-matrix reflection coefficients and the thin-mask model to generate diffraction efficiencies.

3.1.4.4 Line-space, NA = 0.55 Anamorphic

Whereas for the 0.33 NA system we considered $p_{wf} = 25$nm, 50nm. In this section, we proportionally scale down these pitches for a wafer-side NA of 0.55, to $p_{wf} = 15$nm, 30nm, shown in Fig. 3.15. The most significant difference from 0.33 NA is that at 0.55 NA different features can achieve their widest process window at N_{Etch} values that differ by up to 4 bilayers. This behavior can also be observed in Figs 3.9, 3.11, and arises because the anamorphic magnification leads to much smaller mask pitches in the non-shadowing direction, which then require larger values of N_{Etch} to minimize the 0 order. These differences in optimal etch depth greatly increase the difficulty of printing these features simultaneously; however, this effect can be mitigated with a central obscuration.

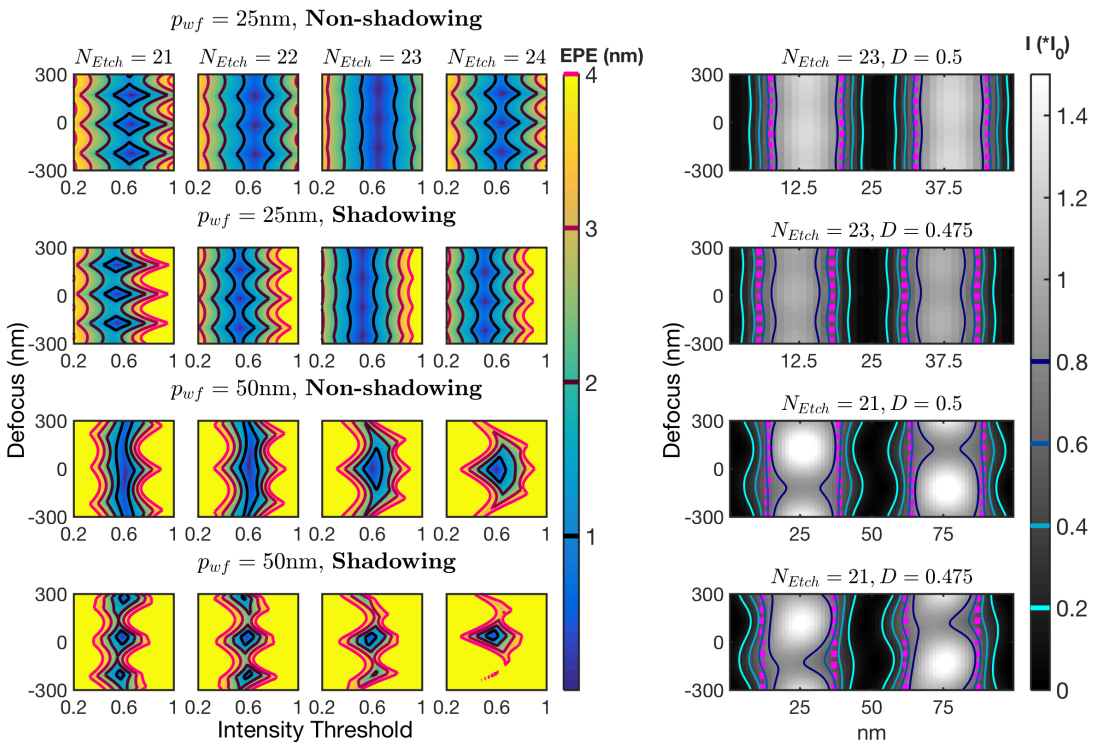

Figure 3.13: Left: EPE_{max} process windows for $N_{Etch} = 21 - 24$ bilayers, each pitch and orientation; each process window calculated with optimal D. Right: Aerial images through focus. Solid lines are contours at highlighted threshold values and dotted lines are nominal edge positions after correcting for Δx^*. Aerial images use (N_{Etch}, D) with largest process window at $EPE_{spec} = 1$nm.

3.1.4.5 Contacts, NA = 0.55 Anamorphic

We can achieve a similar extended depth of focus and wide exposure latitude to print 22nm contacts in the 0.55 NA anamorphic system, using a somewhat deeper etch depth (26 bilayers) again with the equal checkerboard design. Furthermore, the nominal intensity threshold at 0.55 NA is 0.98, or 99% of the thin-mask transfer-matrix prediction. The increase in brightness from the 0.33 NA to the 0.55 NA system is due to the reduced shadowing. Indeed, a similar effect is observed with $k_1 \approx 0.3$ shadowing-orientation lines and spaces, where the intensity at 0.55 NA is 15% higher than at 0.33 NA.

3.1.4.6 Summary of Optimized Mask Designs

Tables 3.2 and 3.3 summarize printing results for the lines-space and contacts-array patterns. The first row in each table is calculated using the thin-mask model (TMM), with reflection

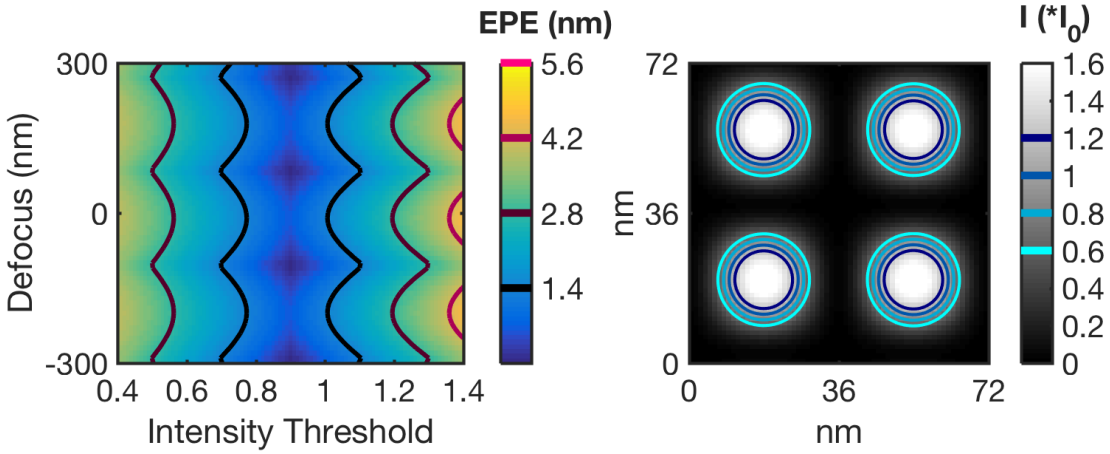

Figure 3.14: Left: EPE_{max} process window for $p_{wf} = 36$nm contacts, 0.33 NA, $N_{etch} = 24$ bilayers. Right: Aerial image at best focus, contours at exposure thresholds.

coefficients calculated using the transfer-matrix method with a 20-bilayer etch depth into 60 bilayers of multilayer. All other rows were calculated using FDTD in Panoramic Hyperlith software. Threshold refers to the intensity threshold at best printing conditions (minimum EPE_{max}); note that the threshold is based on normalizing the incident intensity on the mask to 1. Exposure latitude is calculated as the max minus min threshold as a fraction of nominal, at best focus. N_{Etch} and D values for each feature maximize the process window area at the listed EPE_{spec}.

3.1.4.7 Lines and spaces

All pitches and orientations are within 87%-113% the intensity of the reference thin-mask feature, with the brightest being the 0.55 NA 15nm non-shadowing feature, and the darkest being the 0.33 NA 25nm shadowing feature. Shadowing features tend to be darker than non-shadowing features, with smaller mask pitches corresponding to darker images due to more severe shadowing. This could potentially cause problems overlapping the process windows of shadowing and non-shadowing features, unless measures such as SRAFs are used to equalize the intensities. All shadowing features ultimately received the same optimized unetched duty cycle of $D = 0.475$, meaning that the optimized etched trench is slightly wider than nominal for these features to compensate for shadowing. The non-shadowing features all retained the nominal value of $D = 0.5$, except for the 0.33 NA 15nm feature, which received $D = 0.55$, meaning that a somewhat narrower trench than nominal can compensate for electromagnetic edge effects in this, the smallest mask pitch considered. The exposure latitude of all features with $k_1 \approx 0.3$ is between 37.6%-45.5%, while the exposure latitude for features with $k_1 \approx 0.6$ drops to between 14.8%-19.8%. This large drop can be accounted for both because the

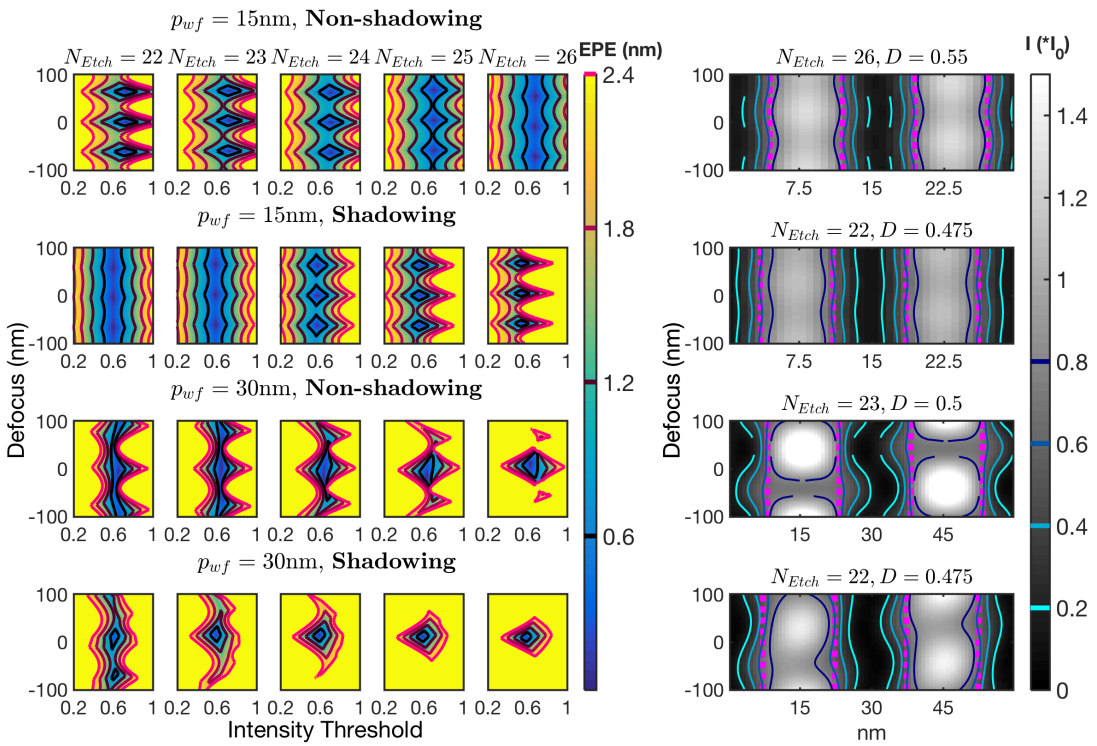

Figure 3.15: Left: EPE_{max} process windows for $N_{Etch} = 22 - 26$ bilayers, each pitch and orientation; each process window calculated with optimal D. Right: Aerial images through focus. Solid lines are contours at highlighted threshold values and dotted lines are nominal edge positions after correcting for Δx^*. Aerial images use (N_{Etch}, D) with largest process window at $EPE_{spec} = 0.6$nm.

± 2 orders are in the pupil for the larger pitch, and because the same EPE_{spec} is used for all features at one NA instead of scaling the specification with the pitch. The narrowest exposure latitude is found in the $k_1 \approx 0.6$ shadowing features, where the asymmetric ± 1 and ± 2 orders further degrade the image quality.

3.1.4.8 Contacts

The contact array achieves 91% of the throughput gains predicted by the thin-mask model at $p_{wf} = 36$nm, 0.33 NA and fully 99% of the predicted gains at $p_{wf} = 22$nm, 0.55 NA. The brighter image at 0.55 NA is due to the reduced shadowing, which is a result of the anamorphic design.

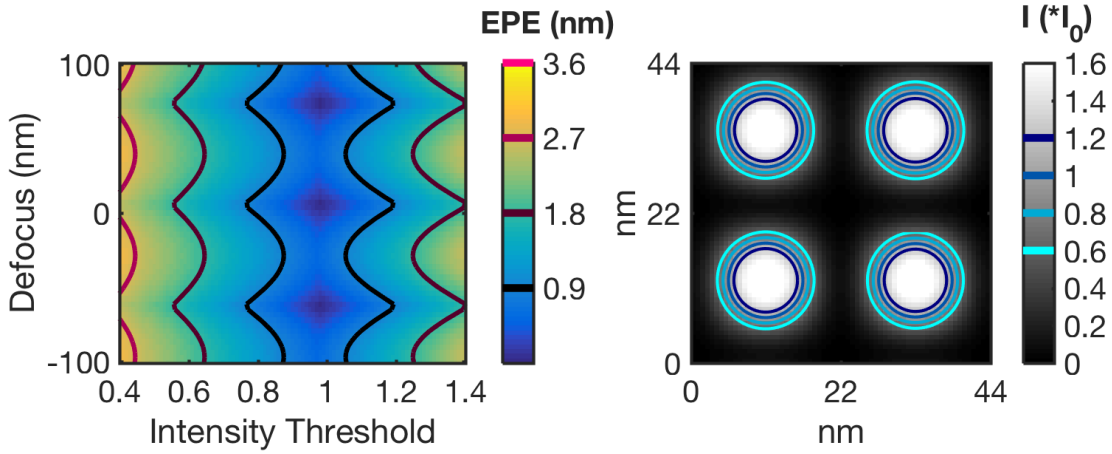

Figure 3.16: Left: EPE_{max} process window for $p_{wf} = 22$nm contacts, 0.55 NA, $N_{etch} = 26$ bilayers. Right: Aerial image at best focus, contours at exposure thresholds.

Table 3.2: Line-space summary for concurrent optimization of threshold, exposure latitude, etch depth and duty-cycle.

Mode	p_{wf} (nm)	Orientation	NA	EPE_{spec}	Threshold	Exposure Latitude	N_{Etch}	D
TMM	25	Non-Shadowing	0.33	1	0.6	25.1%	20	0.5
FDTD	25	Non-Shadowing	0.33	1	0.65	45.4%	23	0.5
FDTD	25	Shadowing	0.33	1	0.52	45.3%	23	0.475
FDTD	50	Non-Shadowing	0.33	1	0.62	19.8%	21	0.5
FDTD	50	Shadowing	0.33	1	0.59	17.6%	21	0.475
FDTD	15	Non-Shadowing	0.55	0.6	0.68	37.6%	26	0.55
FDTD	15	Shadowing	0.55	0.6	0.6	38.5%	22	0.475
FDTD	30	Non-Shadowing	0.55	0.6	0.65	21.7%	23	0.5
FDTD	30	Shadowing	0.55	0.6	0.6	14.8%	22	0.475

Table 3.3: Contact array summary for concurrent optimization of threshold, exposure latitude and etch depth..

Mode	p_{wf} (nm)	NA	EPE_{spec}	Threshold	Exposure Latitude	N_{Etch}
TMM	36	0.33	1.4	0.99	46.0%	20
FDTD	36	0.33	1.4	0.9	44.9%	24
FDTD	22	0.55	0.9	0.98	43.2%	26

3.1.5 Partial Coherence

Whereas partial coherence is often used to improve imaging performance and throughput, for phase-shift masks printing dense periodic patterns, deviations from coherent illumination tend to degrade imaging performance. This is due to reliance on frequency-doubling, which necessitates the precise annihilation of the 0 order. This annihilation cannot be precisely achieved for all angles of illumination simultaneously, therefore in this section we explore how much deviation from the ideal delta-function illumination is acceptable. Fig 3.17 shows the effect of increasing partial coherence σ on printing 25nm lines and spaces in the non-shadowing orientation at 0.33 NA. Increasing σ reduces the intensity threshold, exposure latitude, and depth of focus. This process window
degradation is due to several factors: first, for $\sigma > 0.18$, diffraction orders get partially clipped by the outer edge of the pupil, which greatly reduces the process window; note that the exposure latitude is almost constant until after this point, suggesting that this is the dominant effect at best focus. The partial coherence effectively removes the infinite depth of focus of a fully-coherent image due to the introduction of a continuous spectrum of radii in the pupil, which all change phase with depth at different rates. Other effects that degrade the process window as σ increases include different bulk phase-shifts in reflection between multilayer and air, different edge effects, and different dispersion within the unetched multilayer for different angles of illumination. It is not straightforward to untangle all of these effects; however, combining all these effects together with FDTD, at $\sigma = 0.2$ the nominal intensity drops by only 8%, the exposure latitude is 39.2%, and the depth of focus at 10% exposure error is 340nm. This illumination condition of $\sigma = 0.2$ corresponds to a pupil fill ratio (PFR) of 4%, which is more coherent than current sources but falls within the ASML road map for future improvements to the tool.

3.1.6 Central Obscuration

A central obscuration will be included in the high-NA EUV lithography systems[37]. In our analysis, this will provide substantial benefits for printing dense line-space and contact array patterns with phase-shift masks. Indeed, even using a 0.33 NA system to print these patterns, modification to include a central obscuration would most likely be indispensable. The reason is that a central obscuration greatly relaxes the requirement to completely annihilate the 0 order on the mask, because it will be blocked in the pupil. This means that an essentially infinite depth of focus can be achieved even with an imperfect etch depth, which not only reduces sensitivity to mask making errors but also may enable simultaneous printing of line-space features with different optimal etch depths (although the exposure windows would still need to be equalized by biasing, SRAFs or other means).

These benefits are visualized in Fig. 3.18, which shows how a central obscuration substantially improves process window and reduces requirements on etch uniformity for $p_{wf} = 36$nm, 0.33 NA contacts. Without the obscuration (left), variations in the etch depth significantly shift the best focus location, while incomplete suppression of the 0 order leads to undesirable

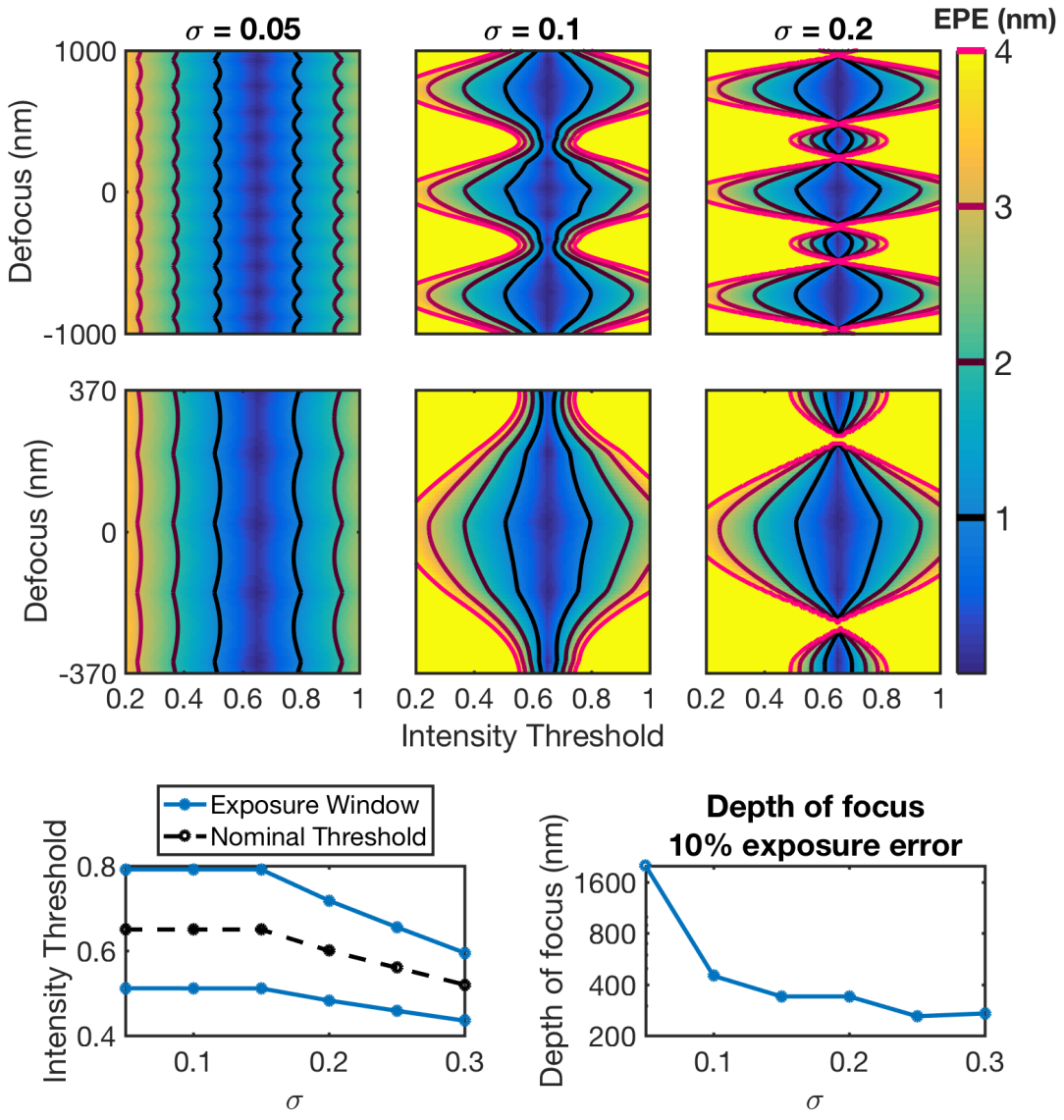

Figure 3.17: Top: Process windows with increasing σ at two defocus ranges. Bottom: Effect of σ on exposure window (at best focus) and depth of focus (at 10% exposure error). Based on FDTD simulation in Hyperlith of a $p_{wf} = 25$nm line-space pattern, $NA = 0.33$, $N_{Etch} = 23$ bilayers, $D = 0.5$, non-shadowing orientation.

fluctuations through focus even in the best case. By blocking the unwanted 0 order light with a central obscuration (right), through-focus variations are completely removed and errors in the etch depth result in only a slight shift in intensity. The complete removal of through-focus variation in this case is due in part to the suppression of the $(\pm 1, 0), (0, \pm 1)$ orders on the mask by the perfectly even checkerboard design. On a real mask, errors in the dimensions of checkerboards may also introduce nontrivial $(\pm 1, 0), (0, \pm 1)$ orders. If necessary, these orders could also be blocked by a pupil filter that blocks all 5 of the $(0, 0), (\pm 1, 0), (0, \pm 1)$ orders. This would require a custom pupil filter with obscurations at locations specific to the pattern on the mask, which, although introducing additional sources of complexity, has been demonstrated to work experimentally[30].

3.1.7 Sub-Resolution Assist Features

In this section, we optimized the process windows for different features by modifying only N_{Etch} and D on the mask. However, further improvements may be possible using sub-resolution assist features (SRAFs). These could take a variety of forms: either sub-resolution etched features, which would preserve the low cost and complexity of a single mask writing step, or sub-resolution absorber features, which would be patterned during a second mask writing step. Either type of SRAF could be used to suppress the ± 2 orders when printing larger pitches, or to equalize the exposure process windows for different line-space pitches and orientations. Absorber SRAFs would allow for much more flexibility in etched phase-shift mask design by enabling modulation of the amplitude; however, these benefits must be weighed against the substantial increase in cost and complexity from a second mask-writing step. It will likely still be beneficial to include a central obscuration even after designing a mask with SRAFs, due to the reduced sensitivity to mask manufacturing errors and the larger depth of focus.

3.1.8 Engineered Multilayer Mirror

Others [55, 25] have explored engineering broadband multilayer mirrors, which sacrifice some reflectivity in exchange for a more uniform angular response. A similar approach could be applied to engineer a multilayer mirror for use in etched phase-shift masks. The key to a good design would have a few components: First, the phase shift per nanometer should be as large as possible to minimize the total etch depth and hence minimize edge effects. Second, both the amplitude and phase of the angular response must be as uniform as possible for all angles within the mask-side NA, to minimize angular dispersion. Finally, the maximum reflectivity should be kept moderately high to maintain a high throughput; however, given the already substantial throughput gains of phase-shift masks, a slight decrease in reflectivity can likely be tolerated if the other two metrics can be substantially improved.

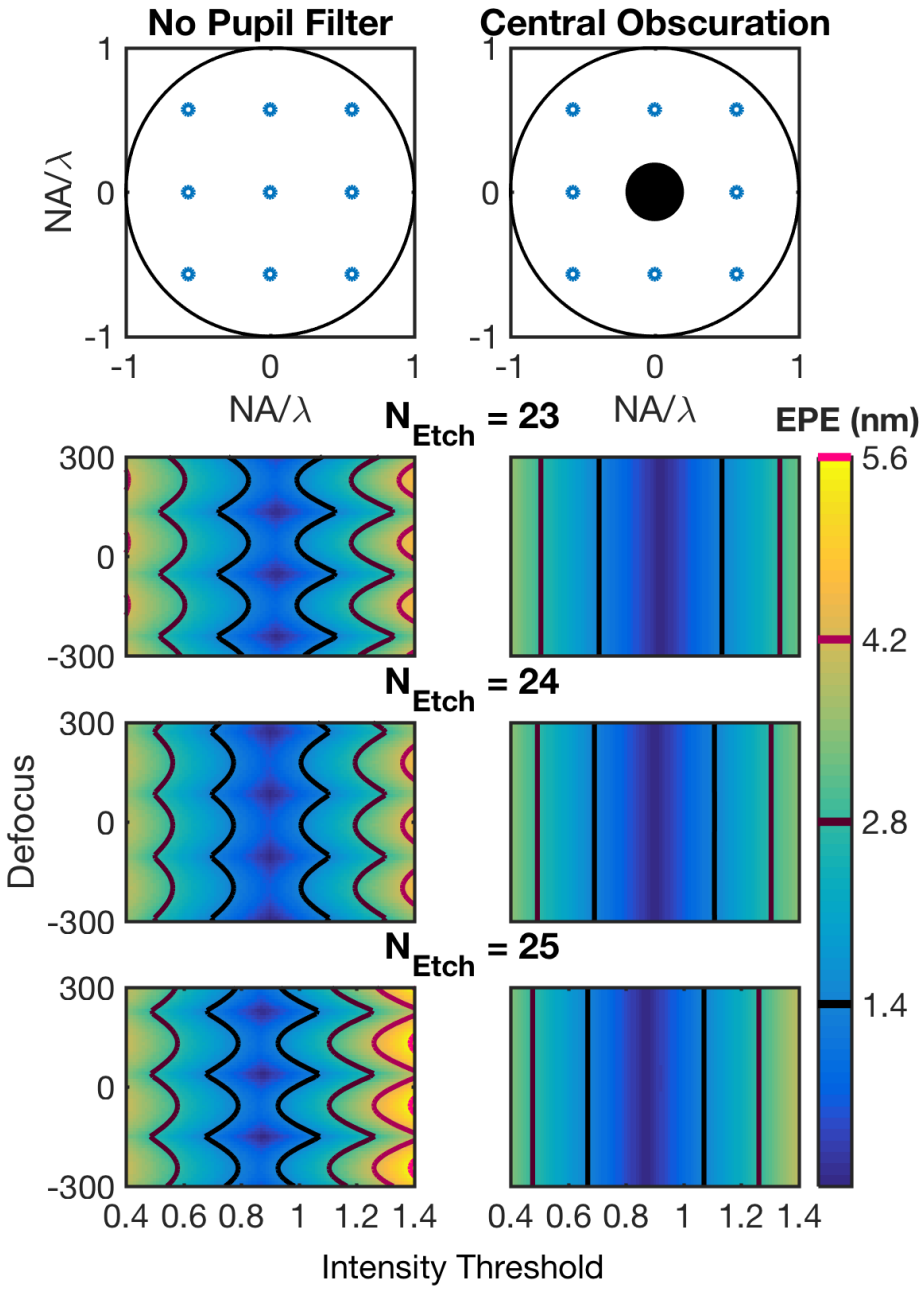

Figure 3.18: Left: EPE_{max} process windows without central obscuration. Right: EPE_{max} process windows with central obscuration. Based on FDTD simulation in Hyperlith of a $p_{wf} = 36$nm contact array pattern, $NA = 0.33$, $N_{Etch} = 24$ bilayers.

3.1.9 Conclusion

In conclusion, we have explored many aspects of the feasibility of etched multilayer phase-shift masks for EUV lithography. The primary advantage that phase-shift masks can offer is increased throughput when printing dense features near the resolution limit compared to traditional absorber masks. We derived analytic formulas for these throughput gains using the thin-mask model, and calculated them to be 2.42x for lines and spaces and 5.86x for contacts. We quantified variations in these gains by pitch and orientation through rigorous FDTD simulations, and found between 87%-113% of the thin-mask prediction for lines and spaces, and 91%-99% for contacts. Furthermore, we found that while larger partial coherence degrades the process window, an achievable[39] σ of 0.2 resulted in a 340nm depth of focus, a 39.2% exposure latitude, and only an 8% drop in intensity, suggesting that partial coherence will likely not be the limiting factor in the applicability of this technology.

To ensure printability of all patterns, we introduced a metric to account for the maximum uncorrectable edge placement error among all edges, EPE_{max}, which accounts for not only CD errors but also telecentricity errors and relative feature motion. We used this metric to both optimize mask designs for maximal process window area, and also to assess whether different features could feasibly be printed on the same mask. We found two major factors limiting what line-space pitches and orientations can be printed on a single mask: differences in the etch depth to achieve an extended depth of focus, and differences in exposure windows. The first problem can be mitigated using a central obscuration to block the 0 order, which allows for features to be printed with an extended depth of focus even with a non-optimal etch depth; the second problem remains to be addressed, but one possible solution may be to use biasing or SRAFs to equalize the exposure windows by modifying the mask pattern.

Many questions remain unanswered, and future directions of work may include exploring whether SRAFs can improve patterning with etched multilayer phase-shift masks, determining the impact of more complicated pupil filters than a central obscuration, and optimizing the multilayer stack for this application.

3.2 Optimized aperiodic multilayer

In this work, we demonstrate a method to design the Mo-Si multilayer stack of an EUV photomask to increase the optical efficiency of shadowing-orientation equal lines and spaces imaged under dipole illumination. We achieve this using a computational framework written in the PyTorch machine learning library, which is capable of optimizing the multilayer for partially-coherent imaging rather than specular reflectivity. After computing optimal multilayer designs for both 0.33 and 0.55 NA EUV systems, we verify the improvements via RCWA simulation. We demonstrate optical efficiency gains of up to 22%/14% for the 0.33/0.55 NA systems, respectively.

This section is heavily based on the conference paper “Advanced multilayer mirror design to mitigate EUV shadowing”[43], originally published at SPIE Advanced Lithography 2019.

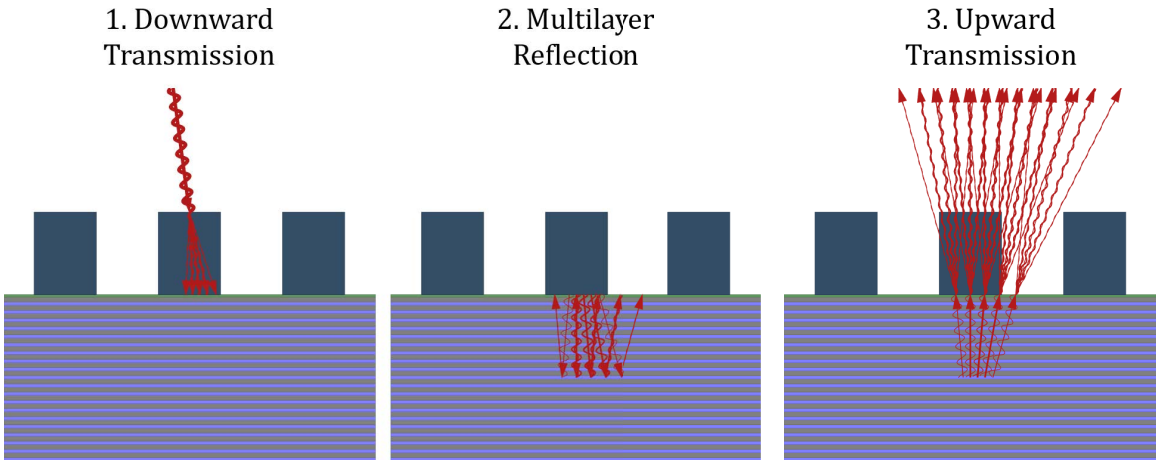

Figure 3.19: 3-step transmission-reflection-transmission model to calculate the reflected near-field. (1) calculate the downward transmission through the absorber. (2) modulate each diffraction order by the multilayer reflection coefficient. (3) coherently sum the upward transmission of all reflected diffraction orders.

3.2.1 Multilayer 3D effects

3.2.1.1 Mathematical Model

The role of the multilayer mirror in EUV lithography can be approximately modeled using a 3-step transmission-reflection-transmission model [8, 41]. In this model, we compute the reflected near-field for a periodic pattern as follows: (1) calculate the downward transmission of the incident light through the absorber, (2) modulate each diffraction order by the multilayer reflection coefficient for the relevant angle, and (3) coherently sum the upward transmission of all reflected diffraction orders. The process is visualized in Fig 3.19, and is essentially a simplified version of the Fresnel Double-Scattering (DblSc) model introduced in [section 1].

Consider the special case of a grating of mask pitch p illuminated from angle θ_0, both in the shadowing orientation. Let $\sin\theta_i = \sin\theta_0 + f_i\lambda$; $f_i = i/p$ denote the angles and spatial frequencies of diffraction orders, respectively. Denote the mask reflection function as $r(x;\theta)$ and its Fourier-transform as $\tilde{r}(f;\theta) = \mathcal{F}[r(x;\theta)]$. Throughout this section we will use $\tilde{}$ to represent a Fourier-transform over the spatial dimension. Denote the absorber transmission function as $t(x;\theta)$. Finally, denote the multilayer Fresnel reflection coefficient as $r_{ML}(\theta)$. We compute the diffraction orders reflected off the mask using the following formula:

$$\tilde{r}(f,\theta_0) \approx \sum_i \tilde{t}(f - f_i;\theta_i) r_{ML}(\theta_i) \tilde{t}(f_i;\theta_i) \tag{3.4}$$

If we additionally assume an ideal thin absorber, the transmission function is independent

of angle, yielding:

$$\tilde{r}(f,\theta_0) \approx \sum_i \tilde{t}(f-f_i) r_{ML}(\theta_i) \tilde{t}(f_i) \tag{3.5}$$

In vector notation:

$$\tilde{\mathbf{r}} \approx \tilde{\mathbf{t}} * \left(\tilde{\mathbf{t}} \circ \mathbf{r_{ML}}\right) \tag{3.6}$$

Finally, in real-space vector notation:

$$\mathbf{r} \approx \mathbf{t} \circ \mathcal{F}^{-1}\left\{\tilde{\mathbf{t}} \circ \mathbf{r_{ML}}\right\} \tag{3.7}$$

Where $\circ$ denotes element-wise multiplication, $*$ denotes convolution, and bolded quantities are vectors containing the function evaluation on the appropriate real-space or Fourier-space grid. To put the formula into words: the final reflection is the transmission pattern multiplied by the reflected transmission pattern.

3.2.1.2 Physical Interpretation

Fig. 3.20 shows the reflectivity and phase versus angle for a traditional multilayer mirror. From $0-10^o$, the reflectivity is approximately constant and the phase is approximately a scaled cosine. In this regime we can approximate the multilayer reflection function as a scaled Fresnel propagation kernel, such that the multilayer is modeled as a reflection plane at some depth h below the surface. By fitting the reflection function phase from $0-10^o$ to the function $2h(k_0-k_z) = h\frac{4\pi}{\lambda}(1-\cos\theta)$, we estimate the effective reflection plane to be at depth $\hat{h} = 44.6$nm. Therefore, an intuitive geometric model of shadowing at low angles is that there is a 70% reflective mirror 44.6nm (6-7 bilayers) below the top surface of the multilayer. This gives the light $2\hat{h} = 89.2$nm to translate and defocus between the downward and upward passes through the absorber. From geometric optics, the lateral translation of the pattern would then be $\Delta x = \Delta z \tan\theta = 2\hat{h}\tan\theta$. For the 0.33 NA system, the maximum illumination angle is $\sin^{-1}(\sin 6^o + 0.33/4) = 10.8^o$, with a corresponding translation of 17.0nm. Mask-side pitch can be as small as $0.5\frac{13.5\text{nm}}{0.33/4} = 81.8$nm, implying that the reflected pattern is misaligned by nearly a quarter of the pitch.

To qualitatively validate our geometric interpretation, we rigorously simulate three geometries: (a) absorber only, which models a mirror at the top surface of the multilayer; (b) absorber + propagation, which models a mirror at the effective reflection plane; (c) full-model, which includes both the absorber and multilayer. Fig. 3.21 schematically defines these geometries in the top row, and the center row shows how the geometries were actually simulated in RCWA. Note that to represent an ideal reflector, in scenarios (a) and (b) we have simulated an equivalent transmission system by adding a mirror copy of the absorber and measuring the transmitted rather than reflected near-field.

All three scenarios were rigorously simulated using RCWA in Panoramic EM-Suite API for Matlab. We consider a mask with $t = 60$nm thick TaN absorber, illuminating a grating

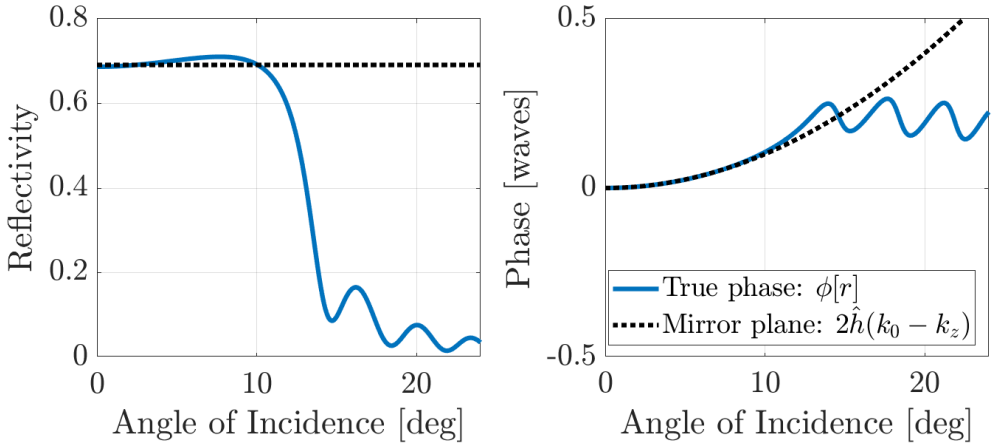

Figure 3.20: Multilayer reflection coefficient (blue) and scaled Fresnel propagation kernel (black dotted). The Fresnel propagation approximation is valid up to approximately 10^o. The depth of the effective reflection plane was determined to be $\hat{h} = 44.6$nm by fitting the phase over this range.

with mask pitch 600nm from $0 - 10^o$. Fig. 3.21 (bottom row) plots the near-field for each scenario, where cases (a) and (b) have also been scaled by the reflectivity of the multilayer at the angle of incidence to account for overall attenuation. Both the absorber only and absorber + propagation models roughly match the full-model at normal incidence, but at 10^o from normal the absorber + propagation model much more accurately predicts the position of the left edge. This means that our physical interpretation of the multilayer as a reflector 44.6nm below the surface is a better model than absorber only. Furthermore, this example clearly shows that the thickness of the absorber, while not insignificant, is not the whole story when it comes to shadowing.

3.2.2 Horizontal dipole illumination

Define the normalized spatial frequency coordinate $u := \frac{\sin\theta - \sin\theta_0}{NA}$, where θ is the illumination angle, θ_0 is the center of the optical axis, and NA is the mask-side numerical aperture. Also define the normalized half-pitch $k_1 := \frac{p}{2}\frac{NA}{\lambda}$. All features and illumination in this section are 1D and in the horizontal (shadowing) orientation. Consider equal lines and spaces under horizontal dipole illumination from $u_0 = \pm 0.95$ in the 0.33 NA system. Due to the off-normal optical axis, these two illumination poles are at different angles. We will refer to $u_0 = -0.95$ as the low angle pole ($\theta = 1.5^o$) $u_0 = 0.95$ as the high angle pole ($\theta = 10.5^o$). Throughout the remainder of this section we will consider this illumination condition. Fig 3.22 shows the imaging performance for $k_1 = 0.26$. We plot the intensity of each pole at the wafer, along with the averaged partially-coherent image to see the effects of each. Shadowing reduces the

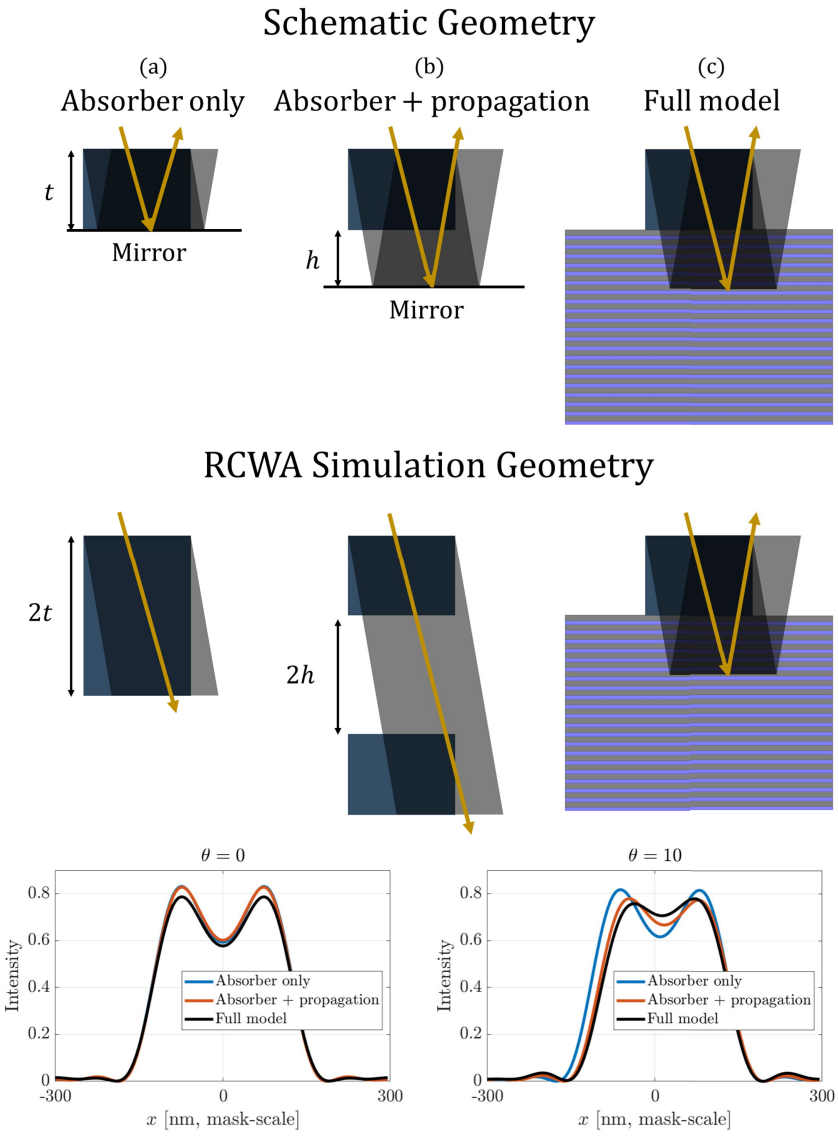

Figure 3.21: Comparing 3 models of EUV mask reflection for $t = 60$nm TaN absorber. The top row shows the geometries schematically, the center row shows the equivalent geometries used for RCWA, and the bottom row shows the computed near-field intensity. (a) absorber only simulates an ideal reflector on the top plane of the multilayer. (b) absorber + propagation simulates an ideal reflector at the fitted depth $\hat{h} = 44.6$nm underneath the surface of the multilayer. (c) full model rigorously simulates reflection from an EUV photomask. All 3 models are approximately equivalent at normal incidence, but at 10^o the absorber-only model grossly underpredicts the translation of the left edge, whereas the absorber + propagation model much more accurately models the edge position.

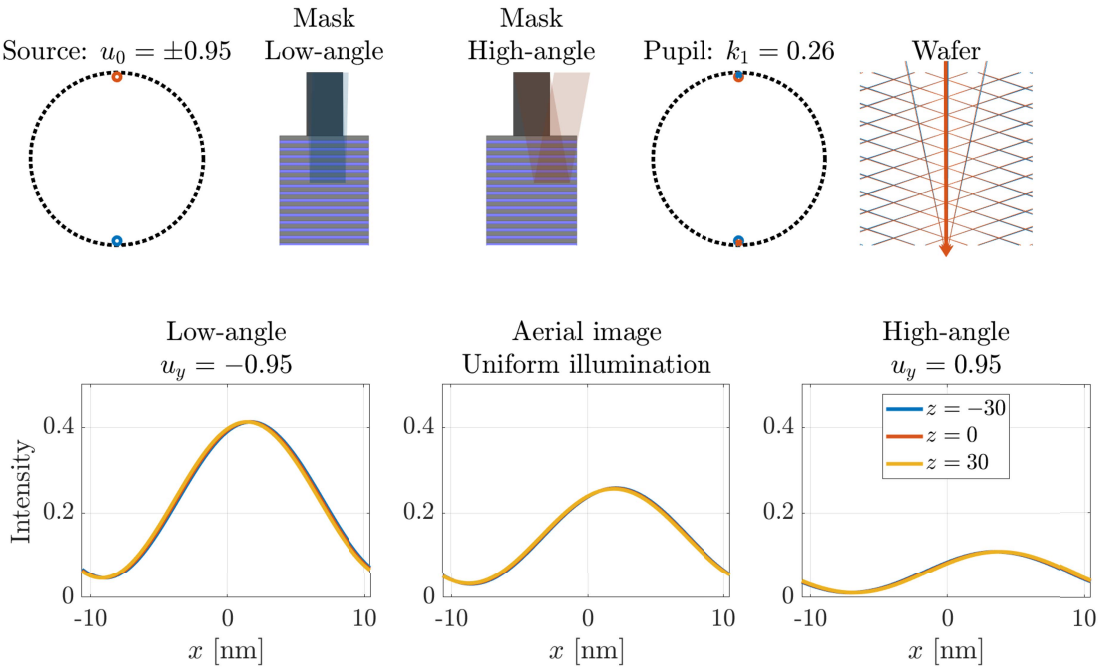

Figure 3.22: EUV dipole imaging, illumination matched to pitch. $k_1 = 0.26$, $u_0 = \pm 0.95$, $NA = 0.33$. Top row, left to right: source points, mask schematic for low and high angle, pupil plane diffraction, wafer 2-beam image. Blue represents illumination from $u_0 = -0.95$ (low-angle) and red represents illumination from $u_0 = 0.95$ (high angle). Bottom row, left to right: intensity at wafer for $u_0 = -0.95$ (low-angle), incoherent sum of both poles (uniform illumination), $u_0 = 0.95$ (high angle). The large disparity in intensity of the two illumination angles arises due to shadowing.

average intensity of the image from $u_0 = 0.95$ (high-angle) by a factor of 3.8. However, the partially-coherent aerial image still has an extended depth of focus due to the matching of the illumination angles to the pitch such that the two diffraction orders of each monopole image propagate with approximately the same k_z.

For comparison, Fig 3.23 shows the same plots imaging a larger feature with $k_1 = 0.5$ under the same $u_0 = \pm 0.95$ dipole illumination. As illustrated in the middle row, now that the illumination is not matched to the pitch, the asymmetry in intensity between the two illumination poles introduces telecentricity error (TCE) or translation through focus. The bottom row illustrates how to correct telecentricity error for this feature by adjusting the source weights to equalize power at the wafer. Balancing the illumination necessitates taking power away from the more optically efficient source point (low-angle), and transferring it to the less efficient point (high-angle). So, while this correction is necessary to avoid loss of process window, it inevitably comes at the cost of optical efficiency.

Denote the average intensities of the low-angle and high-angle poles as I_0, I_1 respectively.

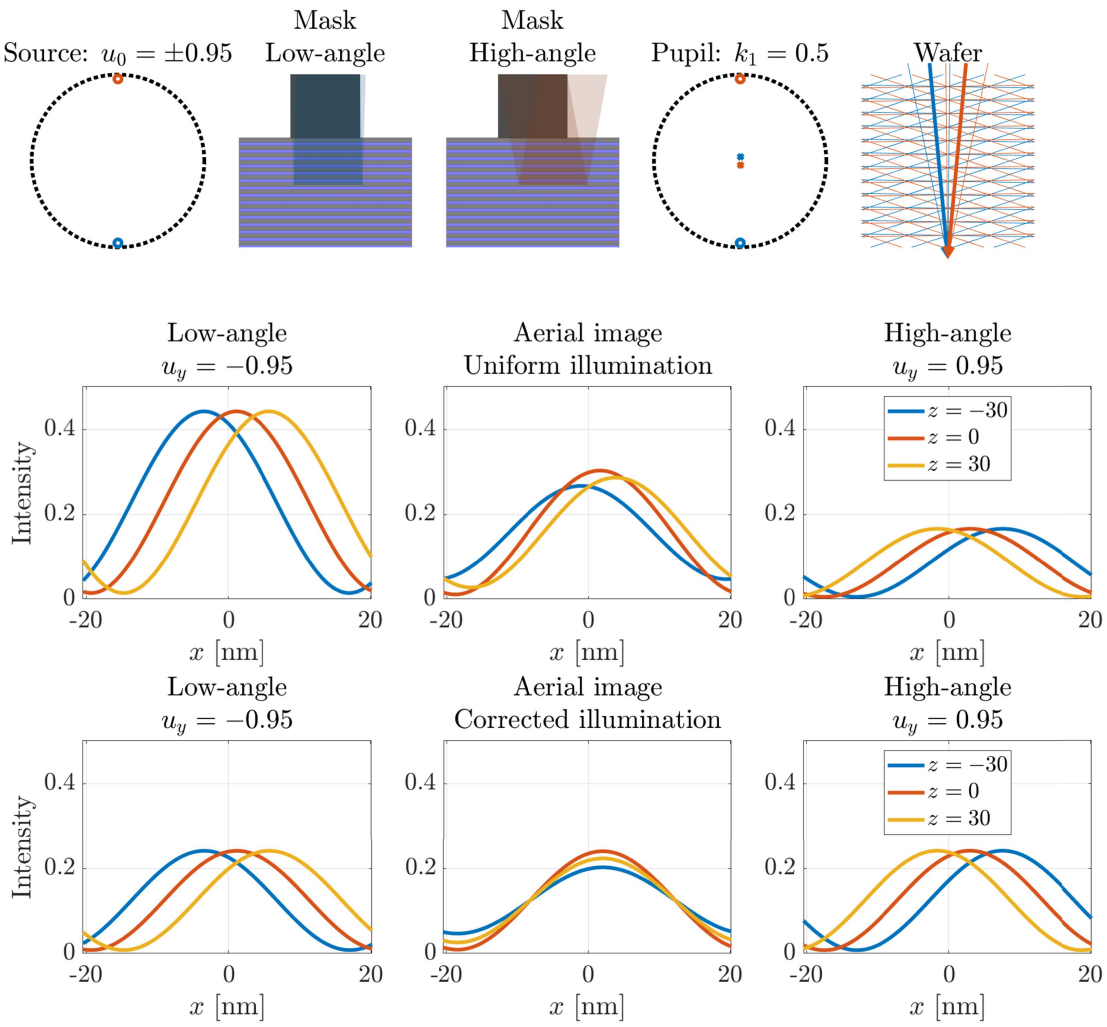

Figure 3.23: EUV dipole imaging, illumination not matched to pitch. $k_1 = 0.5$, $u_0 = 0.95$, $NA = 0.33$. Top row, left to right: source points, mask schematic for low and high angle, pupil plane diffraction, wafer 2-beam image. The middle row plots wafer intensity for each pole alone and with uniform illumination. Because the illumination is not matched to the pitch, shadowing leads to lateral translation with defocus. The bottom row uses corrected illumination, which entails adjusting source weights to equalize the power of each pole at the wafer. Although necessary, this illumination correction reduces optical efficiency.

Similarly denote the source weights as w_0, w_1. We normalize the source such that $w_0+w_1 = 1$. Then in the case of a single feature under dipole illumination, the optimal source weights to minimize telecentricity error are given by:

$$w_0 = \frac{I_1}{I_0 + I_1}; w_1 = \frac{I_0}{I_0 + I_1} \tag{3.8}$$

Therefore, the average intensity at the wafer is:

$$I_0 w_0 + I_1 w_1 = \frac{2I_0 I_1}{I_0 + I_1} \tag{3.9}$$

We then define this as a figure-of-merit for a given feature under given dipole illumination:

$$I_{TCE} := \frac{2I_0 I_1}{I_0 + I_1} \tag{3.10}$$

We can conceptualize of I_{TCE} as the average intensity at the wafer, after applying a telecentricity error correction to the source. This metric drives the optimization to improve the efficiency of both poles, focusing more on improving the high angle.

3.2.3 Multilayer Optimization

So far, we have established that: (1) the reflection function of the multilayer mirror plays an important role in EUV shadowing, (2) shadowing severely attenuates images from high angles of illumination, (3) this leads to imbalanced optical efficiency across the shadowing axis, (4) correcting the illumination to mitigate telecentricity error improves the process window but reduces optical efficiency. We now turn our attention to engineering the multilayer mirror to mitigate this problem. To achieve this, we have developed a framework to optimize the sequence of layers in a multilayer mirror for imaging a given set of features under specified illumination. The framework is written in PyTorch, which allows for automatic differentiation with respect to the multilayer design. The optimized design is computed by gradient descent on the multilayer design with respect to the cost function, $-I_{TCE}$.

3.2.3.1 Computational Framework

To compute I_{TCE}, we need to compute the average intensity from each illumination pole in the far-field. We achieve this by using Eq. 3.7 to calculate the reflected near-field, and then apply the optical bandlimit to get the far-field. For near-field calculation we need the absorber transmission and the multilayer reflection. For simplicity and to make this an absorber-agnostic multilayer optimization, we assume an idealized thin absorber with 100% attenuation. To fully model multilayer effects, we calculate the Fresnel reflection coefficient rigorously using the transfer-matrix method[17]. We assume the nominal index of refraction from the CXRO database for all materials [18]. We include $MoSi_2$ interdiffusion layers based on the model proposed by Aquila [3]. In our optimization, interdiffusion layers are

assumed to be known but not controllable; in practice the same interdiffusion thicknesses of 0.95nm for Mo-on-Si and 0.6nm for Si-on-Mo are used for all multilayers in this section. All multilayers have $N = 60$ bilayers, each composed of 4 layers, from top to bottom: Si-$MoSi_2$-Mo-$MoSi_2$. There is an additional Ru capping layer assumed to be 2.5 nm thick, with an Ru_2Si_3 interdiffusion layer of 1 nm. Thus, in total our reflection coefficient is the result of interference from 242 layers; since we only optimize the Mo and Si thicknesses, our optimization is over a 120 dimensional vector space.

The code is implemented in Python using the PyTorch nn.Module class. This is the base class for neural network models, here repurposed for physics-based computation. We developed two Modules for our optimization: one to calculate the Fresnel reflection coefficient of the multilayer, and a second to calculate the near-field coming off the EUV mask for each feature and source-point.

We optimize the multilayer design by differentiating $-I_{TCE}$ with respect to the vector of Mo-Si thicknesses in the multilayer mirror and performing gradient descent. The analytic gradient is computed by backpropagation, leveraging PyTorch's auto-differentiation feature, thus obviating the need for hand-computation of analytic derivatives.

3.2.4 Optimized multilayer designs

The optimization was carried out for $k_1 = 0.27$, $u_0 = \pm 0.95$ for both the 0.33 and 0.55 NA systems. In both cases, the multilayer consisted of $N = 60$ Mo-Si bilayers, and was first optimized for maximum average reflectivity over the mask-side acceptance angles forcing a periodic design (i.e. a traditional design), then optimized for I_{TCE} forcing a periodic design, and finally optimized for I_{TCE} allowing an aperiodic sequence of layer thicknesses (holding interdiffusion and capping layers constant). All optimizations were carried out using gradient descent with backtracking line-search. Following the optimization, we performed rigorous RCWA simulations for imaging shadowing-orientation features from $0.26 \leq k_1 \leq 0.5$ under the same illumination conditions, to validate our designs.

3.2.5 Optimized Multilayer for 0.33 NA

Fig 3.24 shows the sequence of Mo and Si layer thicknesses in the optimized multilayer mirror for the 0.33 NA system. The optimized thickness curves appear to be approximately sinusoidal, with a mean thickness close to the traditional design and very strong correlation between the Mo and Si thicknesses. These facts conform with our understanding of the traditional Bragg reflector design, where each layer's optical path length is a quarter wavelength. Detuning from this condition rapidly decreases the reflectivity, so it is somewhat expected to only see small deviations from the traditional design in order to maintain high reflectivity.

Fig 3.25 shows the reflectivity vs angle and normalized spatial frequency for the 0.33 NA system. There are several key differences to note: the optimized curve has a slightly wider angular bandwidth, slightly lower reflectivity at normal incidence, and most strikingly a new secondary peak at approximately 19^o.

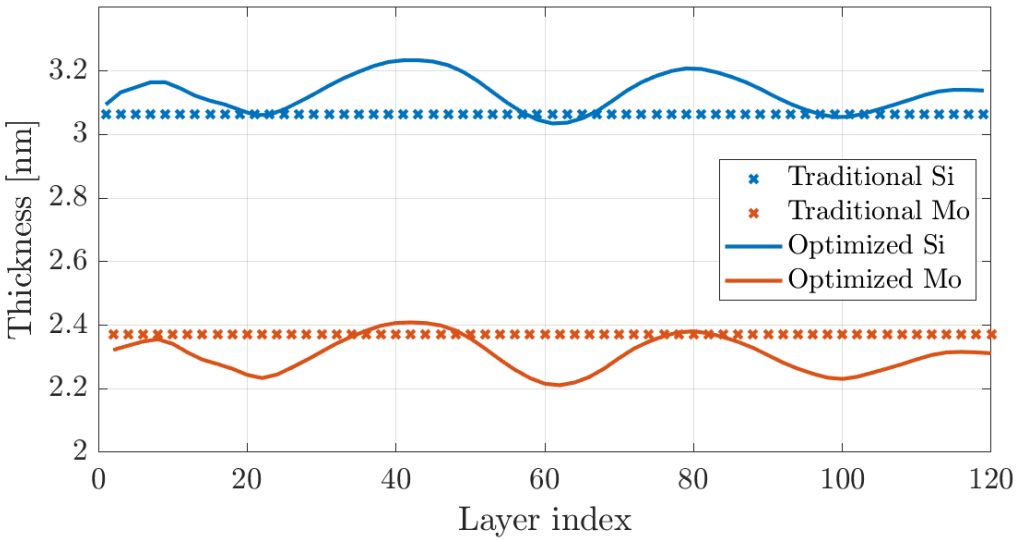

Figure 3.24: Layer thickness for Si (blue) and Mo (red) of a traditional (x) and optimized (solid line) multilayer for the 0.33 NA system.

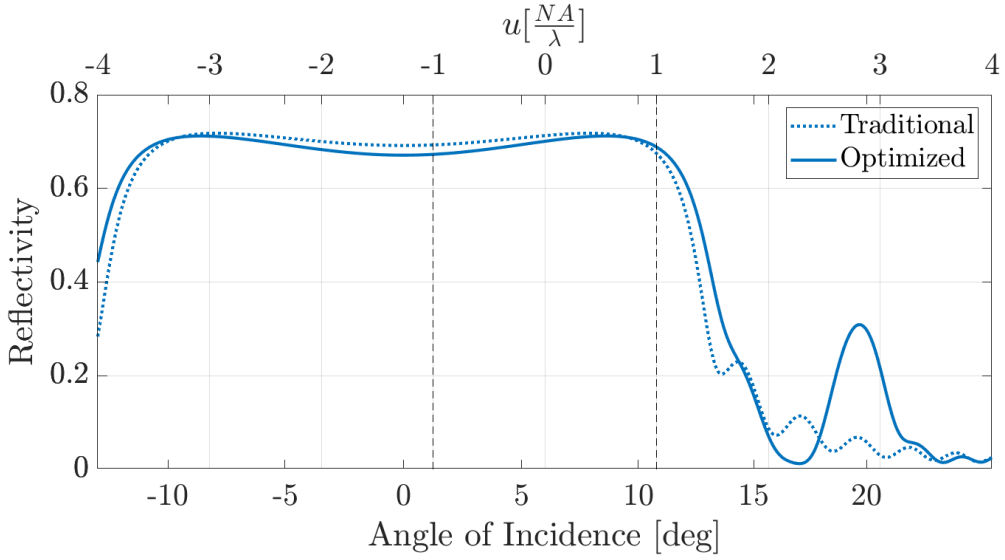

Figure 3.25: Reflectivity vs angle and normalized spatial frequency for the 0.33 NA system. The optimized curve has a slightly wider angular bandwidth, slightly lower reflectivity at normal incidence, and a new secondary peak at approximately 19^o.

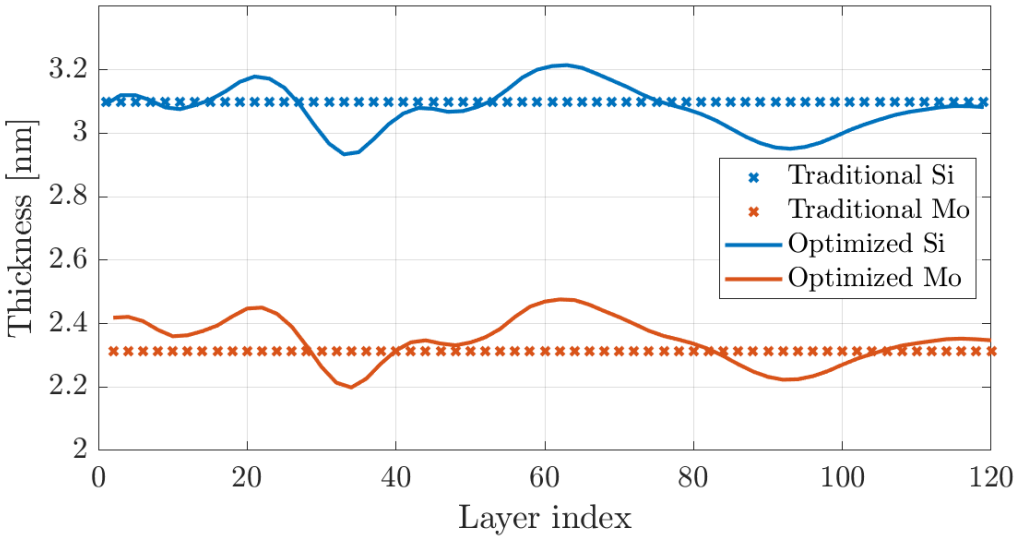

Figure 3.26: Layer thickness for Si (blue) and Mo (red) of a traditional and optimized multilayer for the 0.55 NA system.

3.2.6 Optimized Multilayer for 0.55 NA

Fig 3.26 shows the sequence of Mo and Si layer thicknesses in the optimized multilayer mirror for the 0.55 NA system. The optimized thickness curves are no longer sinusoidal, but still maintain a mean thickness close to the traditional design and very strong correlation between the Mo and Si thicknesses. Again, this suggests the optimization tries to maintain the constructive interference of the traditional Bragg reflector design, while introducing small deviations to boost imaging performance for certain features.

Fig 3.27 shows the reflectivity vs angle and normalized spatial frequency for the 0.55 NA system. Changes in the reflectivity curve are broadly similar to the 0.33 NA system: the angular bandwidth increases, the reflectivity at normal incidence decreases, and again we have a secondary peak at about the same normalized spatial frequency, around 16^o.

3.2.7 Rigorous evaluation of multilayer designs

Fig 3.28 summarizes the improvements in optical efficiency achieved for the two systems, based on rigorous simulation with RCWA. We see that the low-angle efficiency is essentially unchanged, but the high-angle efficiency is substantially increased, particularly for $k_1 < 0.35$. This leads to an overall improvement in the optimization metric of TCE-corrected optical efficiency. We achieve improvements up to 22% for the 0.33 NA system and 14% for the 0.55 NA system. For the 0.33 NA system, all features considered experienced at least some improvement in optical efficiency. However, for the 0.55 NA system there was a slight

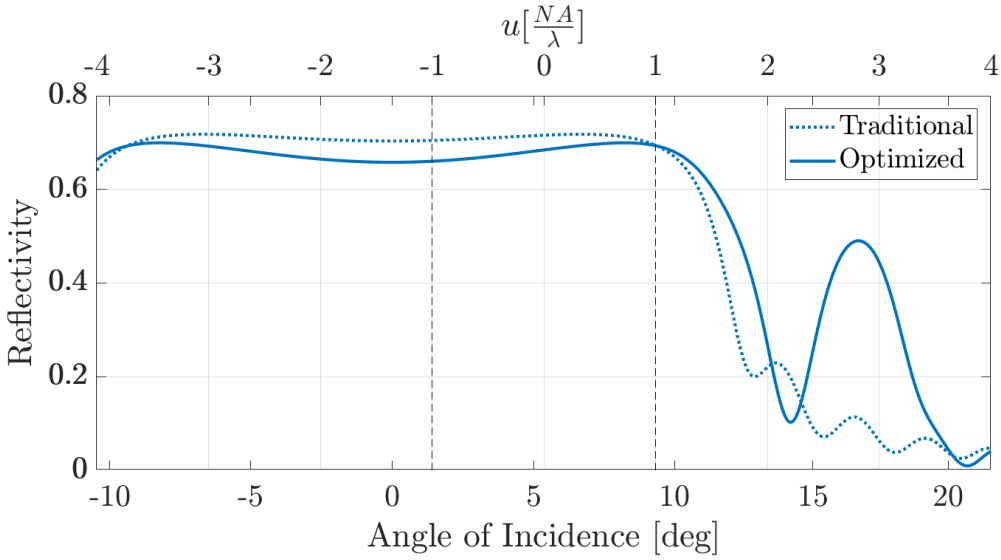

Figure 3.27: Reflectivity vs angle and normalized spatial frequency for the 0.55 NA system. The optimized curve has a slightly wider angular bandwidth, slightly lower reflectivity at normal incidence, and a new secondary peak at approximately 16^o.

decrease in efficiency for $k_1 \geq 0.38$. Indeed, this behavior is somewhat expected since $k_1 = 0.27$ was the only feature used in the optimization. This suggests that the optimization is working as expected as a way to design a multilayer specific to the pattern and illumination, validating both our relatively simple optimization algorithm as well as our approximate physical model. Refining both these aspects of the code may lead to further improvements, and should be a study of future work.

3.2.8 Physical interpretation of multilayer designs

We believe that the improved optical efficiency is primarily due to a counter-intuitive physical mechanism: the secondary peak slightly below $u = 3$ in each optimized curve, where $u = \frac{\sin\theta - \sin\theta_0}{NA}$ is the normalized spatial frequency. The +1 diffraction order from the $u_0 = 0.95$ is reflected at $u = 0.95 + \frac{1}{2(0.27)} \approx 2.8$, roughly coinciding with the secondary peaks. Some fraction of this scattered +1 order light will re-scatter to the -1 order on the upward pass, ultimately contributing to the 0 order of the reflected near-field and therefore increasing the power at the wafer. This scatter-scatter mechanism underscores a counter-intuitive realization: the multilayer reflection coefficient at angles well outside of the imaging system's NA can still impact the image. Indeed, for illumination at $\pm u_0$ and pitch $p = 2k_1\frac{\lambda}{NA}$ we see that the relevant range of spatial frequencies is:

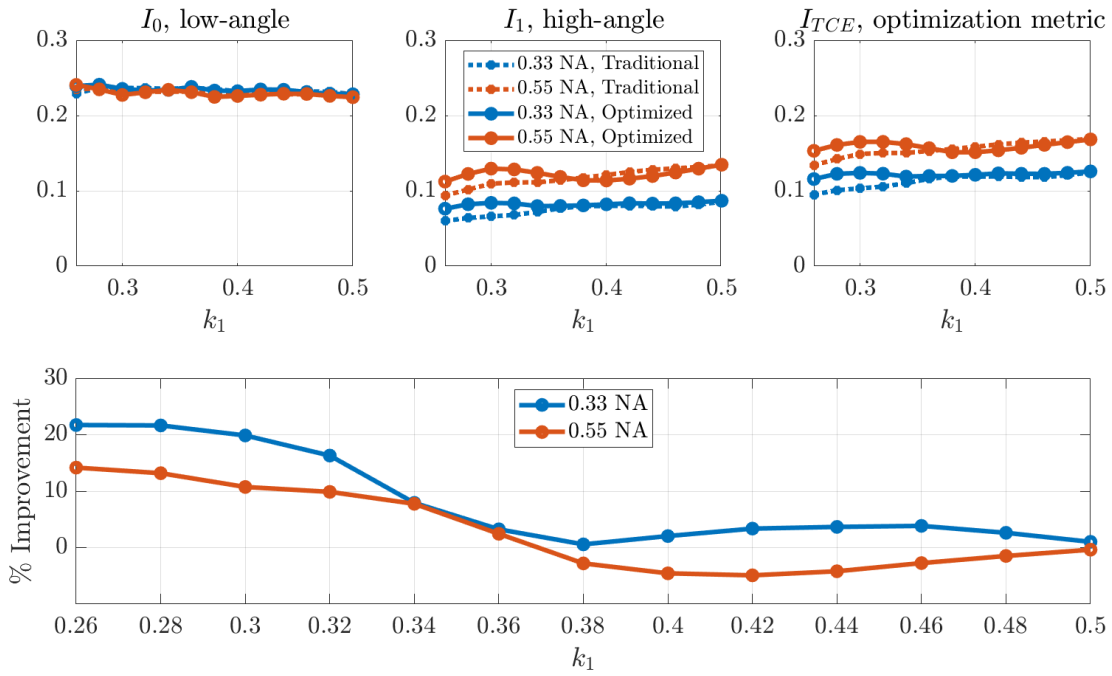

Figure 3.28: Summary of efficiency gains. Top row, left to right: Average intensity for low-angle only (I_0), high-angle only (I_1), and partially-coherent image under TCE-corrected illumination (I_{TCE}). Bottom row shows the relative improvement in I_{TCE} for the optimized design.

$$-u_0 - \frac{1}{2k_1} \leq u \leq u_0 + \frac{1}{2k_1} \tag{3.11}$$

Plugging in the most extreme values of $u_0 = 1$, $k_1 = 0.25$, we obtain the maximum range of spatial frequencies to consider for the multilayer reflection function:

$$-3 \leq u \leq 3, \tag{3.12}$$

in terms of angles:

$$\sin\theta_0 - 3NA \leq \sin\theta \leq \sin\theta_0 + 3NA, \tag{3.13}$$

where NA is the mask-side NA and θ_0 is the center of the optical axis. This range of angles is $-8.2^o \leq \theta \leq 20.6^o$ and $-6.5^o \leq \theta \leq 17.4^o$ for the 0.33 and 0.55 NA systems, respectively.

3.2.9 Conclusion

We have shown rigorous simulations suggesting substantial improvements in optical efficiency of up to 22%/14% for the 0.33/0.55 NA EUV systems by optimizing the multilayer mirror substrate of the photomask for $0.27k_1$ shadowing lines and spaces under $\pm 0.95u_0$ shadowing dipole illumination. The computational framework developed for multilayer optimization is quite general and can be applied to different features or imaging heuristics. One clear extension for future work is to optimize the multilayer for printing 2D patterns such as square or staggered contact-arrays, or indeed any arbitrary pattern. Another possible extension would be to optimize the multilayer for patterning metrics such as edge placement error or image log-slope rather than simply intensity. Ultimately, the optimization of the multilayer design could be integrated into the overall inverse-lithography optimization, along with the source, mask, and pupil. The improved optical efficiency demonstrated in this section suggests a promising future for multilayer optimization to help achieve the ultimate resolution limit of EUV lithography.

3.3 Optimized aPSM architecture

In this section we explore co-optimizing the absorber and multilayer architecture for an EUV attenuated phase shift mask (aPSM) to print square contacts near the resolution limit. Through numerical optimization, for 3 feature sizes we produce a range of designs each with a different tradeoff between throughput and contrast. There are two primary findings: First, for highest throughput the multilayer angular bandwidth should be pitch dependent, with a larger angular bandwidth extending beyond the collection aperture necessary for smaller features (very similar to the secondary peak in the optimized multilayer designs in Section 3.2, extending the multilayer angular bandwidth beyond the aperture allows light diffracted out of the aperture to re-scatter back after being reflected by the multilayer, thereby increasing the overall efficiency). Second, for the highest contrast designs at the two smallest feature sizes we see the effective reflection plane of the multilayer is brought very close to or even slightly above the top surface for certain angles (aligned with the illumination angles further from normal incidence). While these designs are too inefficient to be practical, this interesting physical phenomenon suggests that engineering the multilayer phase response could be critical to improve image contrast at small feature sizes. Furthermore, in these high-contrast designs where the reflection plane is very close to the multilayer surface, the optimal phase shift of the absorber becomes closer to the theoretical value of π rather than the often cited 1.2π. This strongly suggests that the source of this absorber phase compensation is very closely tied to the phase response of the multilayer, and engineering the multilayer phase could greatly reduce the magnitude of this correction.

3.3.1 Attenuated PSM theory

Recent studies into next-generation EUV absorbers have found striking improvements from aPSM absorbers over traditional and even in many cases over thin high-k absorbers. The largest gains for aPSMs have been reported for dense 2D periodic patterns such as square contacts (similar to the patterns suited to an etched/alternating PSM)[50]. Benefits arise in both contrast and throughput. In a traditional binary absorber, the 0 order always has greater magnitude than any of the scattered orders; the 0 order can be lowered relative to scattered orders by applying a bias to shrink the open area on the mask, until in the limit of a delta-function all the orders are equal in amplitude.

On the other hand, an aPSM can equalize the amplitude of the 0 order with ± 1 orders without such an extreme bias, and can even reduce the 0-order amplitude to be lower than ± 1 via destructive interference, and in the extreme case of an alternating PSM even complete attenuation of the 0 order, which leads to frequency doubling. Maximum modulation contrast is achieved from the interference of waves with equal amplitude; a binary absorber balances the 0 and ± 1 amplitudes in the limit of an infinitesimally small opening, whereas an aPSM balances the orders with a finite opening.

The size of the opening to balance the 0 and ± 1 diffraction orders depends on the relative intensity (and phase) of the absorber compared to multilayer. We start with the following equations, with absorber reflectance (relative to multilayer) r:

$$s_0 = D + r(1 - D) \tag{3.14}$$

$$s_{m\neq 0} = (1 - r) \sin \frac{m\pi D}{m\pi} \tag{3.15}$$

If we assume that $r = -a, 0 < a < 1$ (i.e. the absorber has amplitude between 0 and 1 and is perfectly out of phase with the background), the expression for orders 0 and 1 become:

$$s_0 = D(1 + a) - a \tag{3.16}$$

$$s_1 = (1 + a) \sin \frac{\pi D}{\pi} \tag{3.17}$$

These are equalized when:

$$D - \sin \frac{\pi D}{\pi} = \frac{a}{1 + a} \tag{3.18}$$

For an ideal binary mask $a = 0$ so the expression becomes $D = \sin \frac{\pi D}{\pi}$, which is only satisfied in the limit $D \to 0$ (infinitesimal opening). For the example of Ru with 30% reflectivity $a = \sqrt{0.3} = 0.548$, we can solve the equation numerically and we find $s_0 = s_1$ when $D = 0.641$. For the example of Pd with 6% reflectivity $a = \sqrt{0.06} = 0.245$ and $s_0 = s_1$ when $D = 0.515$. The total scattered intensity is:

$$\|s\|_2^2 = D^2 + a^2(1-D)^2 \tag{3.19}$$

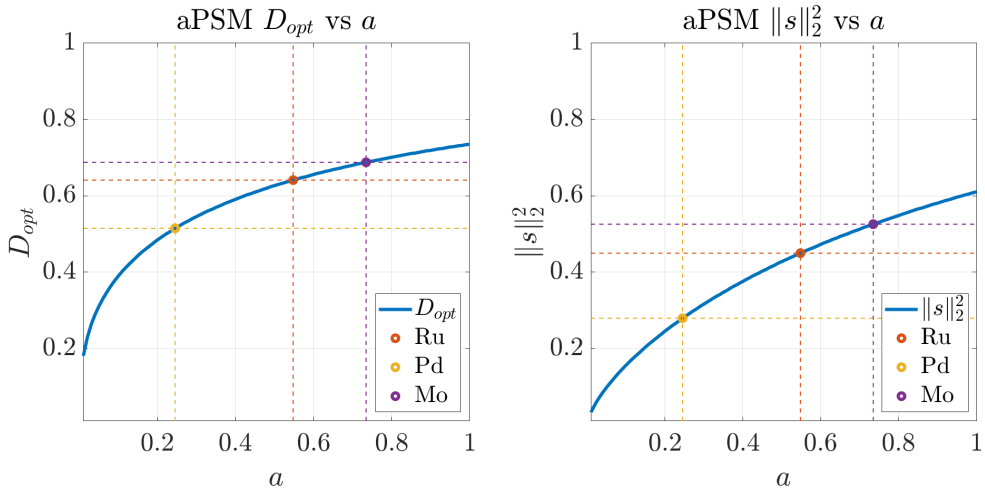

Figure 3.29: aPSM absorber strength a and relation to optimal bias D and total intensity $\|s\|_2^2$.

So, in the limit $a \to 0$, $s_0 = s_1$ can only be achieved in the limit $D \to 0$, implying that the intensity also tends toward 0. But as a takes on a larger value the orders are equalized at a higher value of D, increasing the power at which one achieves optimal image contrast. Finally in the limit $a = 1$ (alternating PSM) $s_0 = s_1$ when $D = 0.735$, while $D = 0.5$ would result in complete cancelation of the 0 order (frequency doubling), as in Section 3.1. A stronger aPSM (higher reflectivity) achieves $s_0 = s_1$ (maximum modulation) at a higher value of D, leading to a more open feature and ultimately higher power, both because more of the feature is multilayer/open and because the absorber/closed areas have higher intensity. So fundamentally an aPSM can print a 2-beam image with higher contrast and higher throughput than a binary absorber, making an aPSM a logical choice to operate at a better point in the RLS trade-off.

A key finding in past studies of EUV aPSMs has been a now famous "1.2π" optimal phase value rather than the expected value of π. This same value is suggested to be optimal for all 3 phase shifting materials in the ASML study[26], as shown in Figure 3.30; furthermore the same study found that aPSMs provide the best overall image quality for printing dense contacts, superior to both standard Ta absorbers and thin high-k absorbers. Results from DblSc in Section 2.3.6 and rigorous modeling in Section 3.3.4 suggest that the additional 0.2π phase shift is not entirely universal (even though the same value appears for different materials)—but a different value would be optimal depending on the multilayer substrate and the feature size. In this study we dig deeper into the design of an aPSM architecture,

Optimum for p28	Ru	Pd	Mo
Optimum NILS	3.6	3.5	3.3
Phase shift	1.2π	1.2π	1.2π
Reflectivity	30%	6%	54%

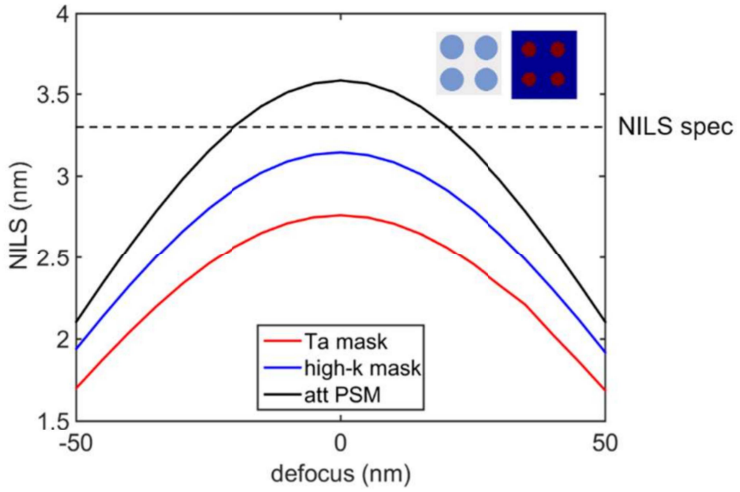

Figure 3.30: aPSM absorbers, numericallly optimized by ASML using RCWA.[26]

exploring the trade-offs between contrast and throughput, and the relationship between the multilayer phase and the optimal absorber phase.

3.3.2 Optimizing contrast and throughput

Because the advantage of an aPSM is both related to power and image quality (contrast/modulation/exposure latitude), when designing an aPSM architecture one needs to define an ultimately subjective trade-off. In this study we use an arbitrary parameter α ($\alpha > 0$) to define the trade-off, essentially equivalent to the α parameter introduced in the Section 2.4. We define the mask architecture and 2D pattern by a set of parameters contained in the vector v (parameters such as spatial dimensions and material properties), and we define a figure of merit in terms of the image contrast and average power:

$$\mathrm{FOM}\left(\mathrm{v}, \alpha\right) = \log\left[\mathrm{contrast}\left(\mathrm{v}\right)\right] + \alpha \log\left[\langle I \rangle\left(\mathrm{v}\right)\right] \tag{3.20}$$

Note that $\langle I \rangle$ is the average image intensity for fixed illumination intensity, and therefore is directly proportional to throughput (inversely proportional to exposure time). Also

note that contrast is defined to reflect the worst case among all rows and columns of the image, so we define contrast $= \frac{I_{max}-I_{min}}{I_{max}+I_{min}}$, where $I_{max} = \max_{x,y} I(x,y)$ and $I_{min} = \max\left[\min_x\left[\max_y\left[I(x,y)\right]\right], \min_y\left[\max_x\left[I(x,y)\right]\right]\right]$. We update the parameter values in v to maximize the figure of merit:

$$\max_{\mathrm{v}} \left(\log\left[\mathrm{contrast(v)}\right] + \alpha \log\left[\langle I\rangle(\mathrm{v})\right]\right) \tag{3.21}$$

We focus on square contacts of $k_1 = 0.4, 0.5, 0.6$ (0.33 NA, $p = 33 - 49$nm wafer-scale, labeled $p33$, $p41$, and $p49$ for pitches 33, 41, and 49nm respectively), printed with quadrupole illumination that is aligned with the feature such that the 4 primary waves (00,10,01,11) all have equal k_z in image space, which is $u_x = u_y = \pm 1/2\frac{\lambda}{NA}\frac{1}{p} = \frac{\pm 1}{4k_1}$ (corresponding to respectively 0.625, 0.5, 0.417 for the 3 feature sizes). We assume equal source weight in each illumination pole, and compute the partially coherent image using Abbe's method, with RCWA to compute near-field scattering.

We iteratively update a total of 7 parameters relating to the mask bias, the thickness and refractive index of the absorber, and the multilayer unit cell. For each of the 3 feature sizes we run the optimization with 10 values of α exponentially spaced from 0.01 (highest contrast) to 10 (highest throughput).

The optimization draws inspiration from genetic algorithms as well as gradient descent. For each feature we start with some nominal design $v_0 \in \mathbb{R}^d$ plus lower and upper bounds for all parameters $v_L, v_U \in \mathbb{R}^d$ ($d = 7$), as well as a set of α values $\{\alpha_j\}_{j=1}^{n_a}$ ($n_a = 10$). To start we generate a set of $2n_a$ random perturbations $\{\delta v_k\}_{k=1}^{2n_a}$, drawn from a normal distribution with standard deviation proportional to the difference between the upper and lower bounds in each dimension. Then for each α, we evaluate the merit function of each design and store the optimal design. This defines the value for each design at generation 0; for generation i and α_j we denote the solution $v_{i,j} \in \mathbb{R}^d$. Now we produce a series of iterative updates with $n_g = 6$ generations, inspired by a genetic algorithm where the "population" of solutions optimal at each value of α will "evolve" to have higher merit functions. In each generation (i), for each α (j) we first apply a random perturbation to each parameter, $w_j = v_{i-1,j} + \delta v_j$, again drawing perturbations from a multivariate normal distribution with standard deviation proportional to $v_U - v_L$. Then we take one gradient step with step size s, $w_j* = w_j - s\partial_v \mathrm{FOM}\left(w_j, \alpha_j\right)$. At the end of the generation, as in the initialization, we update the value of $v_{i,j}$ with the optimal design among the previous design ($v_{i-i,j}$), and the new candidate designs ($\left\{w_j^*\right\}_{j=1}^{n_a}$). Although the algorithm is of course not guaranteed to converge to a global optimum in this non-convex problem, we hope to mitigate this by adding random perturbations and cross-evaluating solutions found for different α values.

The results of these optimizations are shown in Figure 3.31, where contrast is plotted vs optical efficiency, and inset images display representative examples of the image quality for different cases. The bottom-right shows one extreme of low-contrast but high optical efficiency ($\alpha = 10$), and the top-left shows the other extreme of high-contrast but low efficiency ($\alpha = 0.01$). This spans an extreme range of designs, with optical efficiency varying by 4 orders of magnitude and the image quality going from very poor on the bottom right

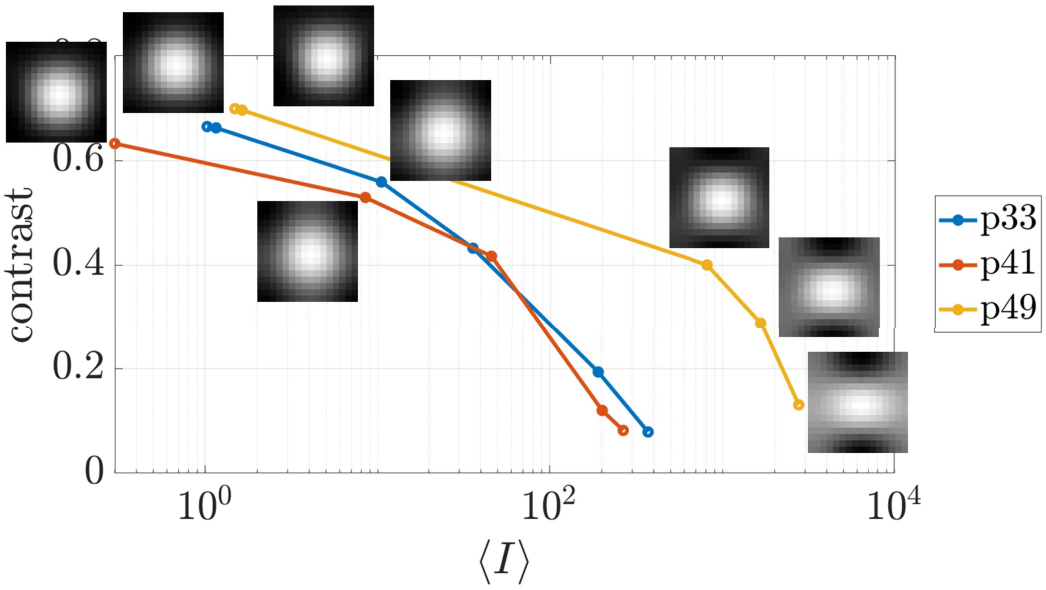

Figure 3.31: aPSM optimization for 3 features, trading off contrast and optical efficiency. Each point represents the optimal solution at one of the 10 α values, after 6 generations of optimization. Inset images display representative examples of the image quality.

to essentially ideal on the top left. It is also worth considering that both contrast and efficiency can have implications for throughput, as discussed in Section 2.4.2. Low efficiency decreases throughput because more time is required to deposit the same dose; on the other hand low contrast decreases throughput more indirectly, because the increased sensitivity to random dose fluctuations necessitates either accepting degraded printing quality (which is often unacceptable) or using a photoresist with a higher exposure dose that suffers less from photon stochastics but of course takes longer to expose.

A practical design would likely need to balance both metrics, such as the designs at roughly 40% contrast. For these intermediate designs, the two smaller features have comparable efficiency, both roughly an order of magnitude below the largest feature. The situation is quite different for the highest contrast designs in Figure 3.31, where the middle feature (p41) has much lower efficiency than the other two features. Indeed, whereas the smallest and largest features are getting closer together in the limit of high contrast, the middle feature appears to follow a very different trend with ultimately much lower throughput and somewhat lower contrast. These trends arise from the different multilayer and absorber architectures employed, and are explored in the following sections.

3.3.3 Optimized multilayer reflectance

Among the parameters optimized was the multilayer unit cell (d-spacing and gamma or equivalently Mo and Si thickness, with MoSi interdiffusion thickness held constant). Show in Figure 3.32 (a), the highest contrast designs for the two smaller features produce a multilayer with substantially lower reflectivity near normal incidence; these designs probably place too extreme a value on contrast and not enough on throughput—but even if impractical for manufacturing, they provide some very interesting physical insights.

Although the reflectivity is much lower in the highest contrast designs for the two smaller features, nevertheless the contrast is much higher. This improvement in contrast appears to be related to the multilayer phase, specifically the effective propagation distance (phase gradient vs k_z) for high-angle illumination (further off-axis). In the most extreme high contrast designs for p33 and 41, as shown in Figure 3.32 (d), the effective propagation is reduced close to 0 and even in some cases reversing the sign of the phase gradient (placing the effective reflection plane *above* the surface of the multilayer at certain angles).

The highest power designs in Figure 3.33 (a)-(b) contain fairly traditional multilayers, although the angular bandwidth appears to be a function of pitch—understandably a wider angular bandwidth is required for a smaller pitch due to the larger angles of diffraction, at the expense of a slight reduction of reflectivity at normal incidence. These designs are also visualized in Figure 3.34, where we can very clearly see the multilayer bandwidth extending far beyond the pupil (black circle) for the smallest pitch. This can be understood as promoting the efficiency of multiple scattering, where light can diffract on the downward transmission outside the pupil, then be reflected by the multilayer, and scatter a second time back into the pupil during upward transmission.

Next, we inspect designs for an intermediate level of contrast, 40%, in Figure 3.33 (c)-(d). Interestingly there are quite different trends in the largest feature and the smaller 2. For the largest feature the angular bandwidth appears very similar, while the smaller features extend the angular bandwidth even further, at the cost of a substantial reduction in power at normal incidence. In this way the power at low illumination angles is decreased and at high-angles it is increased, at least partially balancing the optical efficiency. As discussed in Section 3.2, improving the efficiency of high-angle source points will have a disproportionately positive impact on overall throughput, because it was shown that the least efficient points have the largest effect on overall efficiency if one applies a telecentricity correction to the source weights with positive and negative u_y. On the other hand if no correction is applied to the source weights, equalizing the power in the poles can improve contrast by mitigating TCE.

The highest contrast designs in Figure 3.33 (e)-(f) are (as mentioned previously) likely impractical in that they place too high a value on contrast relative to throughput; case in point the reflectivity is under 20% for the two smallest features at 6 degrees. These smaller features have Bragg reflectors where the peak is shifted out beyond the edge of the pupil. So clearly this is not a good design for throughput, but it does manage to achieve high contrast from its engineered phase profile even at a very small k_1 with a uniform source. Due to

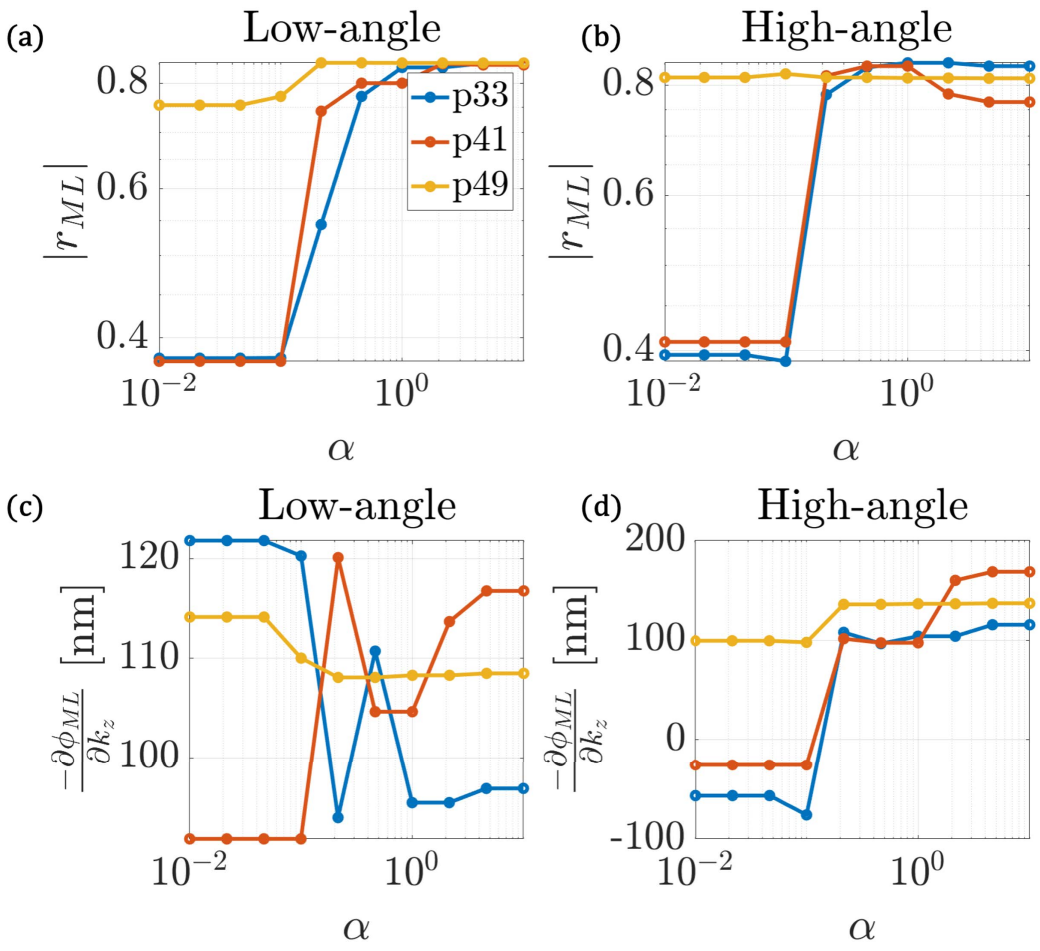

Figure 3.32: aPSM optimization for 3 features, ML amplitude vs α and effective propagation distance vs α (high α corresponds to high contrast). Highest contrast designs have much lower multilayer reflectivity (a)-(b), but the multilayer reflection for the high-angle illumination pole effectively travels a negative distance (d).

oscillations in the amplitude, there is a small range of angles around 8^o where the curvature of the phase switches sign. When the phase curvature vs angle is 0, so is its slope vs k_z, meaning that illumination in this narrow range of angles can be reflected much closer to the multilayer surface, or even above the surface when the phase curvature (briefly) switches sign. We see very clearly in Figure 3.35 (c) and (f) that these nulls in the effective propagation distance are aligned with the more oblique illumination angles. Physically speaking, the reflection plane is still deep inside the multilayer for illumination near normal incidence, but for the oblique illumination angles the reflection plane has been brought to the surface, which produces an image with much higher contrast.

3.3.4 Optimized absorber phase shift

So, we have seen that the multilayer design can be tuned to improve contrast and/or throughput; what about the absorber? The amplitude and phase of the optimized absorber designs is presented in Figure 3.36. First, shown in 3.36 (a)-(b), we note that the absorber amplitude displays a trade-off with α (contrast): generally higher contrast correlates with a darker absorber (relative amplitude reducing from roughly 30%-10%). There is also a trend with the phase, shown in 3.36 (c)-(d), particularly for the most unique designs, the high-contrast designs for k_1 0.4 and 0.5. In these designs, the phase shift of the absorber becomes closer to 180^o as the contrast increases, whereas for the higher power designs the phase is quite close to the notorious value of 1.2π (216^o). This might appear odd at first—after all the conclusion of recent studies has been that the absolute best patterning for small k_1 with an EUV aPSM occurs essentially universally 1.2π phase shift—why now can even better contrast be achieved with a shift closer to 1.1π?

A solution to this apparent mystery is suggested in Figure 3.36 (f), which shows the optimized phase shift in the absorber vs the effective propagation distance in the multilayer. The distance is always around 100nm, except for 3 points clustered on the bottom-right, which correspond to the high-contrast designs in question. And for precisely these designs again we see the relative phase shift significantly closer to 180^o than for the other designs. As earlier suggested in the DblSc study on M3D phase shift vs pitch, the source of the phase compensation is pitch dependent and similar for different absorber architectures on similar multilayers; now we see that if we change the multilayer to have a different phase response we also need to readjust the phase of the absorber, strongly suggesting that the multilayer phase is the primary reason previous studies have shown 1.2π phase shift to provide optimal imaging performance in rigorous simulations with a standard multilayer substrate.

In summary, it is possible to optimize the phase response of the multilayer to decrease the effective propagation distance for the high-angle illumination. This is effectively bringing the reflection plane of the multilayer closer to the surface (but only for a limited range of angles, aligned with source points where shadowing is most severe). Although there is a substantial throughput penalty, much better contrast may be achieved with a multilayer designed for its phase response. Bringing the reflection plane to the surface of the multilayer greatly reduces multilayer 3D effects and leads to an optimal phase shift which is much closer to 180^o than a traditional multilayer engineered solely for reflectivity. All this suggests that the multilayer phase is the source of the additional 0.2π phase in previous studies of aPSM design, and that if multilayer phase variation can be reduced then the magnitude of this phase compensation will also be reduced. A similar (but less extreme) effect would also likely be present using alternative multilayer stacks such as Ru-Si, which sacrifices some reflectivity to bring the effective propagation plane closer to the surface.

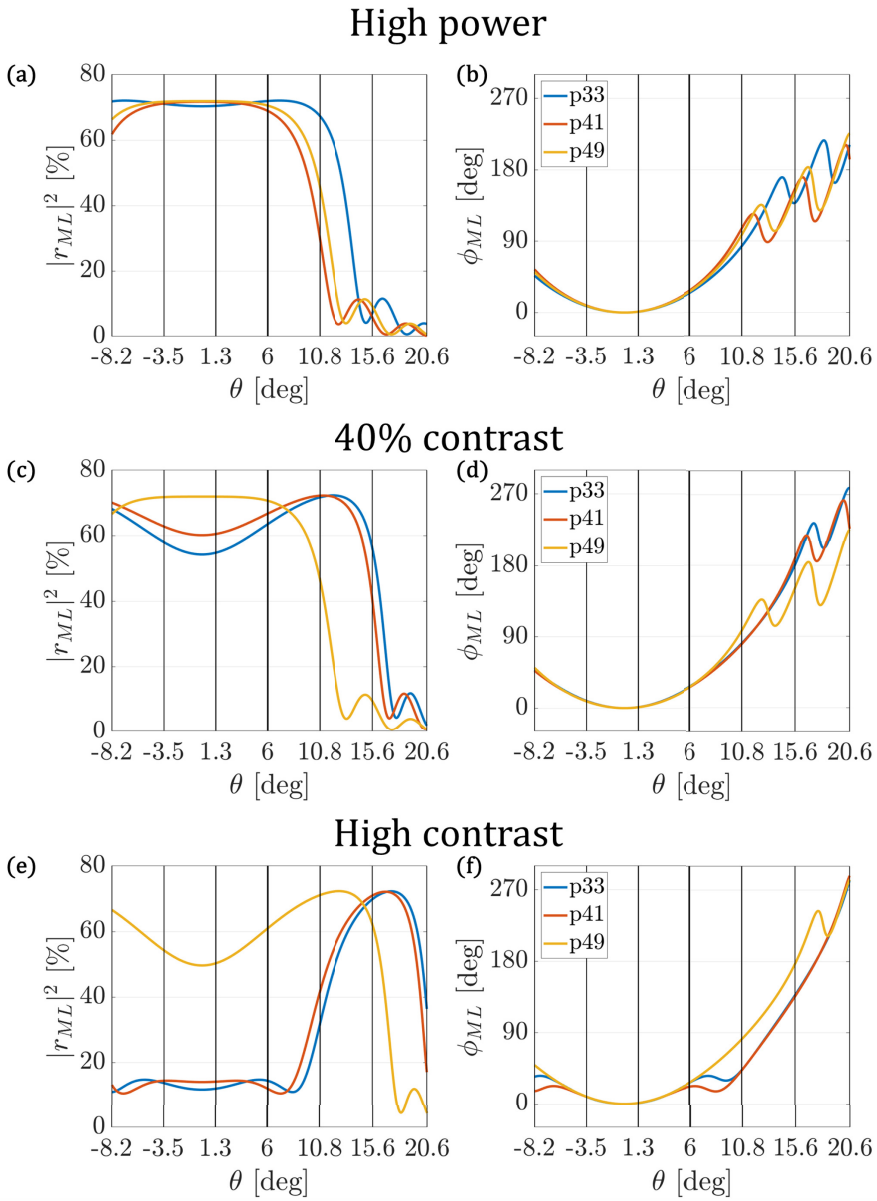

Figure 3.33: aPSM optimization for 3 features, optimized multilayer reflectance for cases of highest throughput (a)-(b), 40% contrast (c)-(d), and highest contrast (e)-(f). Throughput is increased in (a) for the smallest feature by extending the multilayer bandwidth to angles outside the pupil. Similar extension of the bandwidth is seen for the intermediate designs in (c), now for both p33 and p41. The highest contrast designs in (e)-(f) for p33 and p41 are completely unlike the traditional design, placing the Bragg reflector peak entirely outside the pupil in order to improve the phase response.

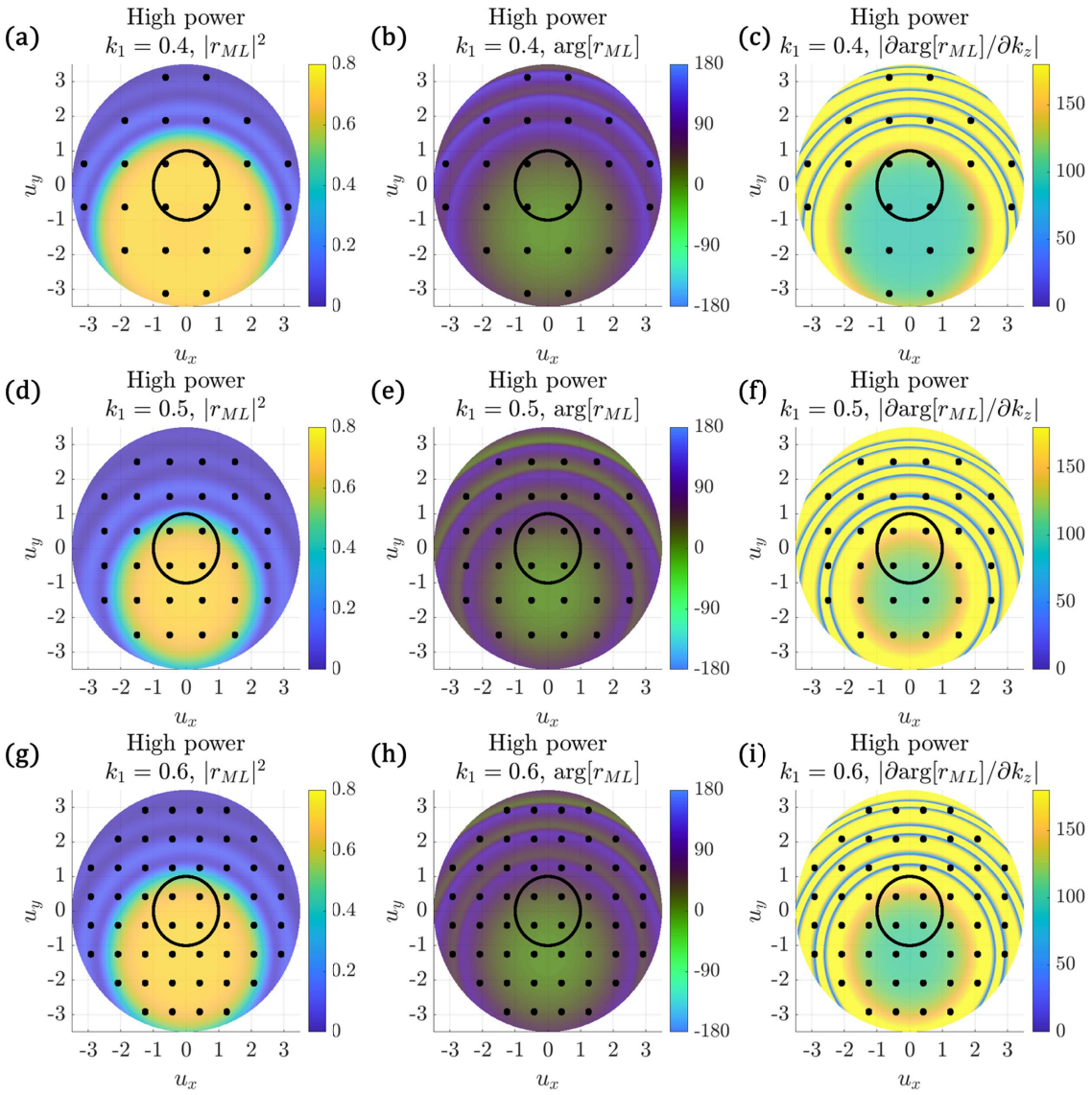

Figure 3.34: aPSM optimization for 3 features, optimized multilayer reflectance (amplitude, phase, and effective propagation distance) for highest power design. (a)-(c) show the smallest feature, $0.4k_1$ (p33), where we clearly see the extended multilayer bandwidth in (a), which increases the amplitude of reflections outside the pupil, which can partially scatter back into the pupil after reflection.

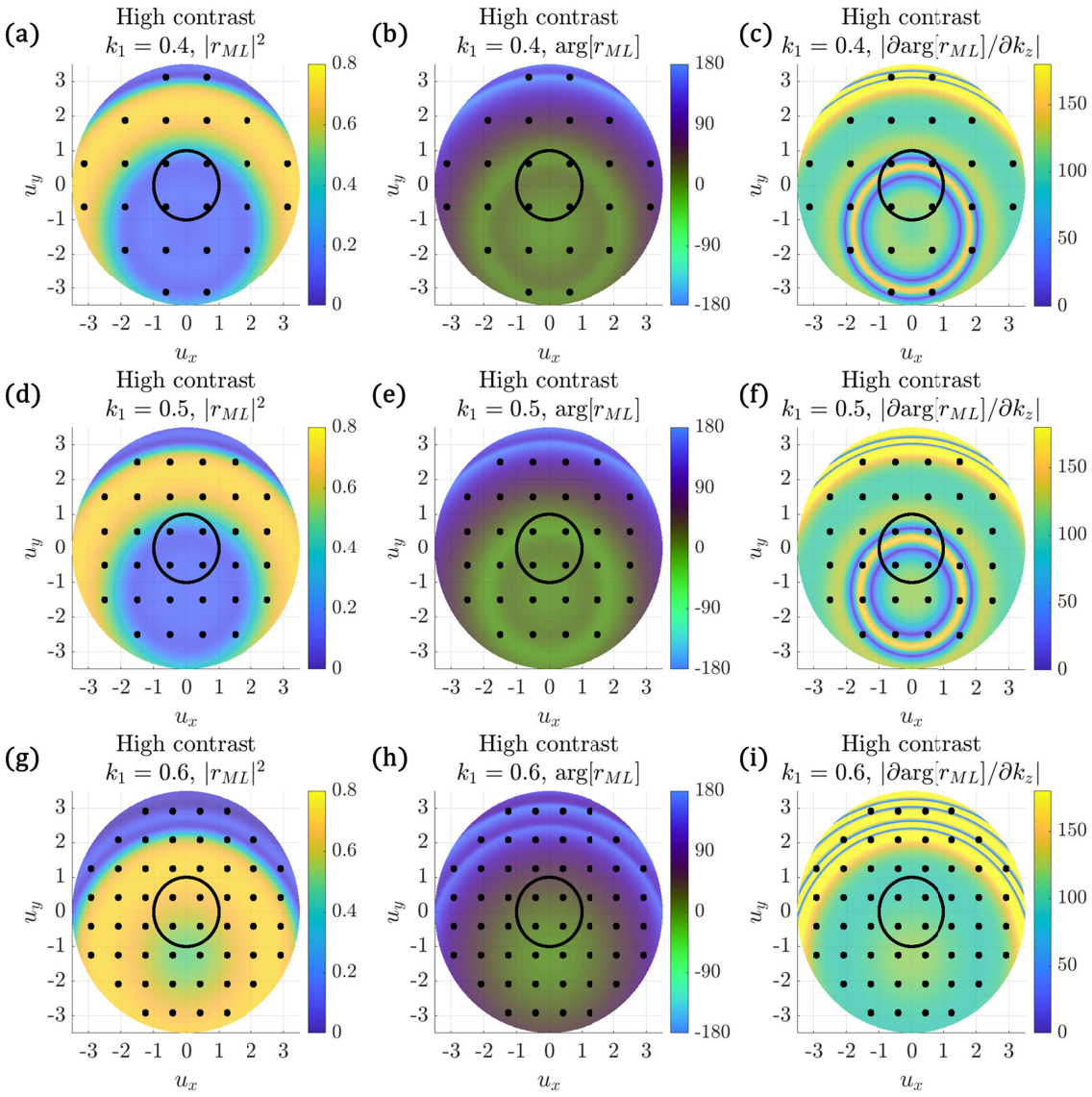

Figure 3.35: aPSM optimization for 3 features, optimized multilayer reflectance (amplitude, phase, and effective propagation distance) for highest contrast design. (a)-(f) show the unique designs for the two smaller features, where the reflectivity is dramatically reduced but the multilayer reflection plane is raised to the top surface for certain angles. These nulls are aligned with the two more oblique illumination angles, shown in (c) and (f).

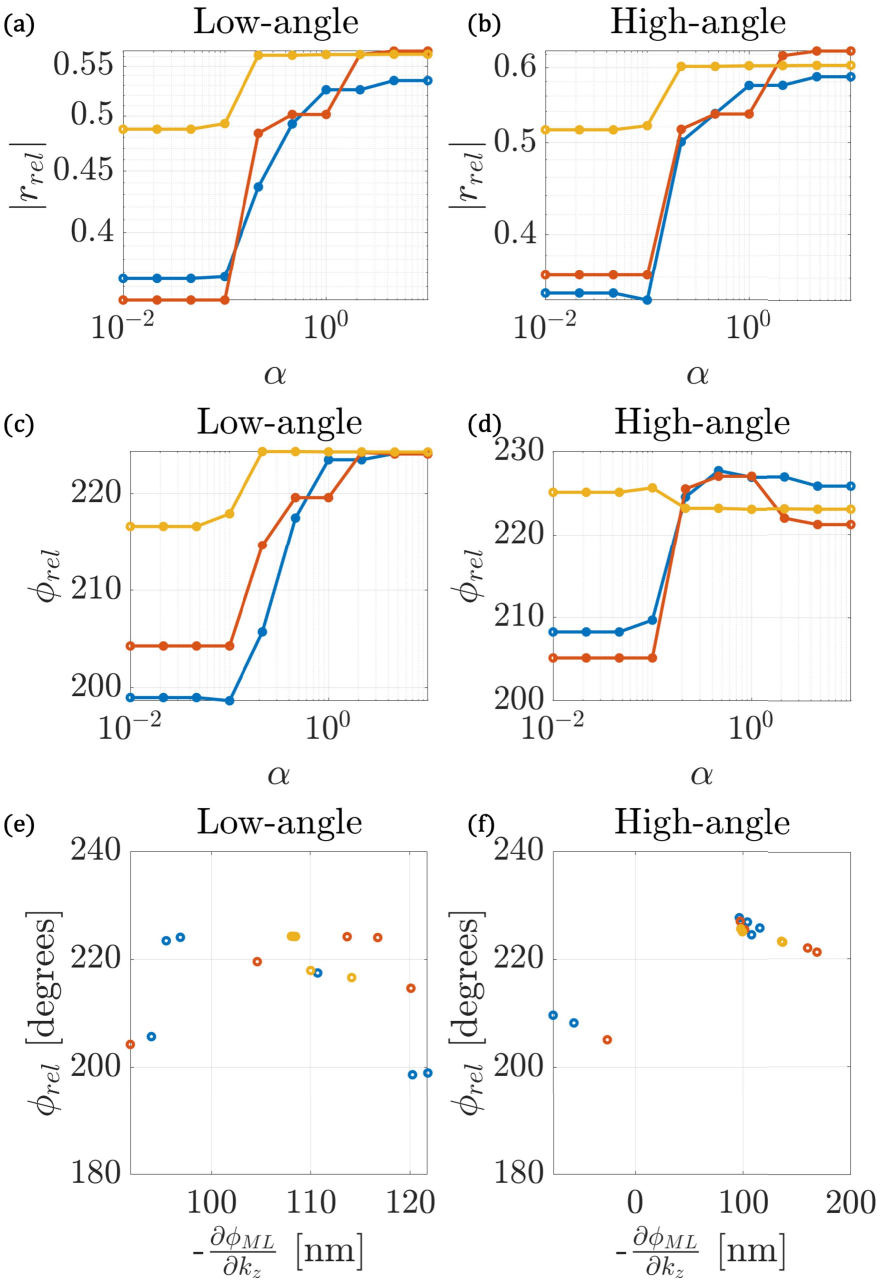

Figure 3.36: aPSM optimization for 3 features, optimized amplitude and phase of absorber over multilayer. (a)-(b) show amplitude vs α, (c)-(d) show phase vs α, and (e)-(f) shows the effective propagation distance in the multilayer vs the phase shift between absorber and multilayer. A shorter effective propagation distance into the multilayer for oblique illumination corresponds with a phase value closer to the traditional 180^o, demonstrating the link between multilayer and absorber phase in EUV mask design.

Chapter 4

Measuring EUV Photomasks

In this chapter we explore 3 methods for measuring EUV masks: reflectometry, scatterometry, and phase imaging. Particularly, we are interested in measuring the amplitude and phase response of the mask at wavelengths and angles relevant to EUV lithography. The amplitude response is quite straightforward to measure with the proper hardware; so all of these studies fundamentally are focused on computationally recovering the phase response. The phase can only be inferred computationally from measurements of intensity, because the temporal phase of 13.5nm EUV is far too fast to measure electronically. All the methods shown have been demonstrated at the Center for X-Ray Optics (CXRO) using the Advanced Light Source (ALS) synchrotron to produce high-power coherent EUV in the viscinity of 13.5nm wavelength. Reflectometry and scatterometry were conducted at the ALS Calibration, Optics Testing, and Spectroscopy Beamline 6.3.2, and imaging was conducted on the ALS EUV Lithography Photomask Imaging Beamline 11.3.2 on the SHARP full-field EUV microscope.

The first method we explore is reflectometry (Section 4.1), in which we illuminate a flat sample of multilayer or absorber on the photomask with a range of incident angles and wavelengths and measure the reflected intensity for each. For a given 1D film-stack geometry, we can also calculate the theoretical Fresnel reflectivity at the same wavelengths and angles to compare with the measurements. We perform nonlinear least-squares regression and iteratively update the physical parameters defining the film-stack, to computationally recover a model that matches the data better than the initial model. Then once this is complete we compute the Fresnel reflectance (including phase) for multilayer and absorber, and compute the relative phase shift between the two, which we call the phase of the mask. Following this multidimensional optimization, we also present a contamination monitoring application where we freeze all but one parameter (representing top-surface contamination) and perform single-dimensional nonlinear least-squares to fit the contamination thickness. With this latter contamination monitor method, we experimentally demonstrate single-picometer precision for thickness and sub-degree precision for phase.

The second method is scatterometry (Section 4.2), which is conducted on the same tool as reflectometry but with a different sample and an additional detector scan. The flat sample is replaced with a scattering sample such as a grating, and instead of only measuring the

specular reflection, the detector is scanned to measure light scattered to different angles. A very direct approach which we choose not to take would be to again perform nonlinear least-squares as in reflectometry, but instead of Fresnel reflectivity using a rigorous solver such as RCWA as the forward model; however due to the long runtime required we decide that this approach is not practical. Instead we compare two approximate strategies: first, performing again parametric nonlinear least-squares but using the approximate DblSc model rather than RCWA for computational efficiency; and second, linearizing the relationship between intensity and phase produced by RCWA in the neighborhood of the nominal 3D model. We find in simulation that linearized RCWA is often substantially more accurate, but both algorithms have similar precision. We also evaluate both algorithms at two levels of partial coherence, representing a more coherent synchrotron source and a less coherent plasma source. The performance is similar on either source, except that for larger features on the plasma source it is necessary to account for partial coherence when linearizing RCWA; if partial coherence is neglected during linearization the performance is almost identical for the synchrotron source and for smaller features on the plasma source, but the accuracy drops for larger features on the plasma source where the diffraction orders start to blur together. We also present the same algorithms applied to experimental scatterometry measurements of the TaN mask, where we observe a similar through-pitch trend and comparable levels of precision for all algorithms, but often significant differences in the trends of phase over time.

The third method is phase imaging (Section 4.3), in which we measure a set of images of the photomask under multiple imaging conditions and use these images to recover the amplitude and phase response. For the phase retrieval algorithm, we focus on PhaseLift convex phase retrieval, which has the attractive mathematical property of convexity, meaning that any initial guess should lead to a global optimum, as opposed to the more commonly used nonlinear least-squares algorithms which are nonconvex and therefore can fall into local minima. In addition to the algorithm, we present a novel hardware approach using Zernike Phase Contrast (ZPC), and an extension to hyperspectral imaging with hyperspectral ZPC (hZPC). In ZPC we image the same feature under the same illumination with multiple different zone plates. Each zone plate is encoded with a unique phase shift in a predefined region aligned with the illumination. This allows the 0 order forward scattered light to be phase shifted while the rest of the diffracted orders are imaged normally. We show how the ZPC method is more accurate than through-focus imaging because the quadratic defocus function is not sufficiently sensitive to the phase of low spatial frequencies (unless a very large focus range is collected). Because ZPC uses a discontinuous transfer function, the same level of phase diversity can be obtained for both low and high spatial frequencies, which greatly improves the accuracy. We also present hZPC reconstructions for horizontal and vertical lines on the 3 experimentally measured photomasks (a standard TaN absorber plus two aPSMs), where we see promising agreement with reflectometry, particularly for vertical lines and spaces.

4.1 Reflectometry

This section is heavily based on the paper "Picometer sensitivity metrology for EUV absorber phase"[45], originally published in JM3 (Journal of Micro/Nanopatterning, Materials, and Metrology) in 2021, with the addition of Section 4.1.6, in which we apply the multidimensional film parameter regression experimentally to 3 photomasks with different architectures.

With growing interest in EUV attenuated phase shift masks due to their superior image quality for applications such as dense contact and pillar arrays, it is becoming critical to model, measure, and monitor the intensity and relative phase of multilayer and absorber reflections. We present a solution based on physical modeling of reflectometry data, which can achieve single picometer phase precision and sensitivity to changes in average film thickness below one atomic monolayer. We measure absorber and multilayer reflectivity to determine thin-film parameters with a multi-dimensional optimization, and then acquire a new measurement of either multilayer or absorber to determine perturbations in surface contamination thickness. While it is difficult to assess the accuracy of the first step, the simplicity of the second step allows us to characterize our sensitivity to changes in contamination thickness. We apply this analysis using an initial set of measurements, and repeated measurements after a period of storage. For the multilayer the total contamination growth was 1068pm, which occurred almost exclusively during storage (1085pm) and decreased very slightly during repeated measurements (-17pm). For the absorber the behavior was quite different, with a total growth of 126pm, which occurred much less during storage (28pm) and primarily during repeated measurements (98pm). Ultimately the change in relative phase (absorber minus multilayer) was $-0.86°$ for the multilayer and $-1.12°$ for the absorber. We estimate the precision of the surface contamination measurement to be $3\sigma < 6$pm for measuring thickness and $3\sigma < 0.2°$ for measuring phase.

In recent years interest in EUV attenuated phase shift masks (aPSMs) has increased, particularly for small-pitch dense patterns[26]. A traditional absorber cannot create a high contrast image without substantial bias due to the relative imbalance of the 0-order and scattered waves. An aPSM on the other hand uses destructive interference between the pattern (absorber) and the background (multilayer) thus reducing the power of the 0-order and transferring that power into the scattered waves. Therefore, an aPSM offers both higher contrast and better throughput than a traditional absorber for these critical patterns[26]. And yet the introduction of EUV aPSMs requires addressing the challenges of precisely measuring and controlling the phase. Particularly, we are concerned with the relative phase between the absorber and multilayer reflections at a defined reference plane, which we in this work take to be the topmost surface of the photomask, including both the absorber and any contamination that might be present.

We present a reflectometry-based metrology solution which can measure the relative phase between multilayer and absorber with single picometer precision. The method requires measurements of absorber and multilayer reflectivity under varying illumination conditions. These measurements are then used in a multidimensional nonlinear least-squares regression to determine a possible physical model for the multilayer and absorber; we can then compute

the Fresnel reflectance from the physical model to extract the relative phase of multilayer and absorber. Subsequent measurements of the absorber and multilayer can then be used to determine changes in surface contamination thickness from a brute-force single-dimensional nonlinear least-squares regression.

Extracting phase from measurements of only amplitude can be explained through Kramer's-Kronig relations, which imply that there is a unique relationship between the phase and amplitude of an analytic function (such as the reflection coefficient)[2]. Therefore *any* analytic function which matches the measured amplitude over a sufficiently large wavelength range must match the phase of the true reflection coefficient. By construction our computed Fresnel coefficient will always be analytic because it is composed of analytic functions of the tabulated atomic scattering factors[18], which are themselves analytic. Furthermore our Fresnel coefficient cannot be any analytic function; it is greatly restricted in what values it can take on by prior knowledge of the mask. Prior knowledge comes in several forms: the known wavelength-dependence of each material's refractive index, the nominal thickness for each layer, and a measurement of top surface roughness conducted by AFM (Atomic Force Microscopy). This allows us to restrict the range of variables, which given strict enough bounds should mitigate remaining ambiguities so that any solution which matches the measured reflectivity for both multilayer and absorber to sufficient accuracy within the variable bounds should also provide an accurate prediction of the relative phase.

At the same time there is certainly still room for further studies into the absolute accuracy to assess the exact requirements on the range of illumination conditions and prior knowledge. Additional prior knowledge could also be obtained from supplemental measurements such as grazing incidence reflectometry to measure refractive index. So while we expect that the initial multi-dimensional optimization should be able to accurately measure the phase, given enough data and enough prior information, nonetheless to quell any remaining fears about the uniqueness of the solution to the multi-dimensional optimization, we restrict our major claims to the precision of the secondary single-dimensional nonlinear least-squares regression used to determine changes in contamination thickness on the top surfaces of the absorber and multilayer. In this way, we hope to demonstrate that at the very least our approach is highly repeatable and sensitive to small changes, while leaving the door open to future studies into absolute accuracy.

We validated our technique on a test mask which has a standard 40 bilayer Mo-Si multilayer mirror and a 60nm TaN absorber. We performed an initial reflectivity measurement for the multilayer, and 2 months later for the absorber. After an additional 3 months we repeated both multilayer and absorber measurements back-to-back. We observed a substantial change in the multilayer signal during storage, but during the repeated measurements the signal was relatively stable. In contrast the absorber changed systematically both during storage and measurement.

We hypothesized that the most likely explanation for contamination would be a thin carbon film growing on top of the multilayer and absorber, based on previous studies of carbon contamination on EUV multilayers and Ta-based absorbers depending on storage conditions and EUV exposure[56, 47, 21]. Therefore from this point forward we hold all other

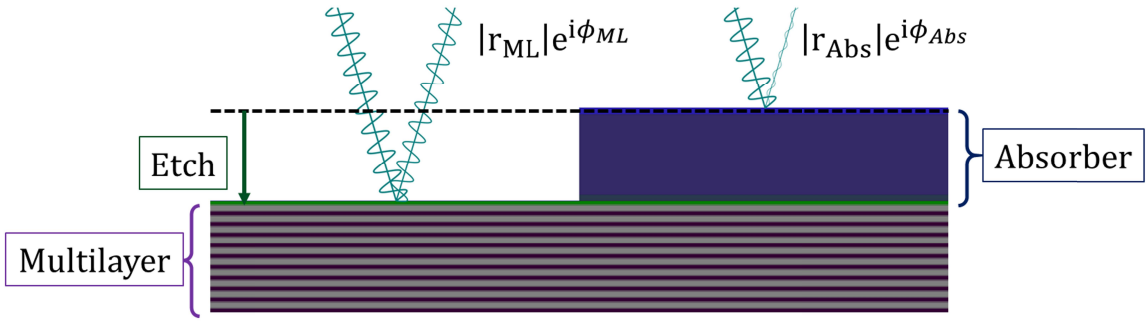

Figure 4.1: Photomask schematic Schematic depiction of reflections from absorber and multilayer. Absorber layers from top to bottom are: TaON-TaN-Ru-(Si-MoSi-Mo-MoSi)x40. Multilayer uses same film-stack, with an additional etch depth parameter to replace etched layers with vacuum. Additional C layers are added on top of the absorber and multilayer, representing contamination. The complex reflection coefficient for the multilayer is $r_{ML} = |r_{ML}|e^{i\phi_{ML}}$, and for the absorber is $r_{Abs} = |r_{Abs}|e^{i\phi_{Abs}}$.

parameters constant and adjust only the carbon layer thickness on either the multilayer or absorber when assessing repeatability of the technique; this assumption allows us to use a simple brute-force search which guarantees convergence to a unique optimal solution (given a fixed initial model and a set of reflectivity measurements). This allows us to quantify changes in the carbon film thickness and the relative phase. During storage, we estimate the change in carbon contamination thickness to be 1068pm for the multilayer (over 5 months) and 28pm for the absorber (over 3 months). During reflectivity measurement the contamination rates were approximately -1.7pm/measurement (11 measurements) for the multilayer and 12.3pm/measurement (9 measurements) for the absorber. We estimate an overall change in the relative phase of $-0.86° \pm 0.1°[3\sigma]$ for the multilayer, and $-1.12° \pm 0.16°[3\sigma]$ for the absorber.

Section 4.1.1 defines the mask geometry and experimental conditions. Section 4.1.2 defines the mathematical model and optimization approach. In section 4.1.2.1 we describe the Fresnel coefficient model and the mean-squared error (MSE) objective function used in our optimization. In section 4.1.2.2 we discuss our choice of optimization algorithm for the initial multidimensional optimization. We wrap up discussion of computational methods in 4.1.2.3 where we describe the brute-force search used for the carbon contamination measurements. Section 4.1.3 presents the results of both optimization steps. In section 4.1.3.1 we present the initial model, which demonstrates the ability of our technique to find at least *some* solution matching the measured reflectivity. Then in section 4.1.3.2 we show clear trends revealed by the brute force search for carbon contamination on the absorber and multilayer. In section 4.1.4 we analyze the phase recovered from the modeled Fresnel coefficient. In section 4.1.4.1 we define the effective propagation distance of the reflection coefficient for both multilayer and absorber, and discuss some of its implications on imaging. In section 4.1.4.2 we discuss

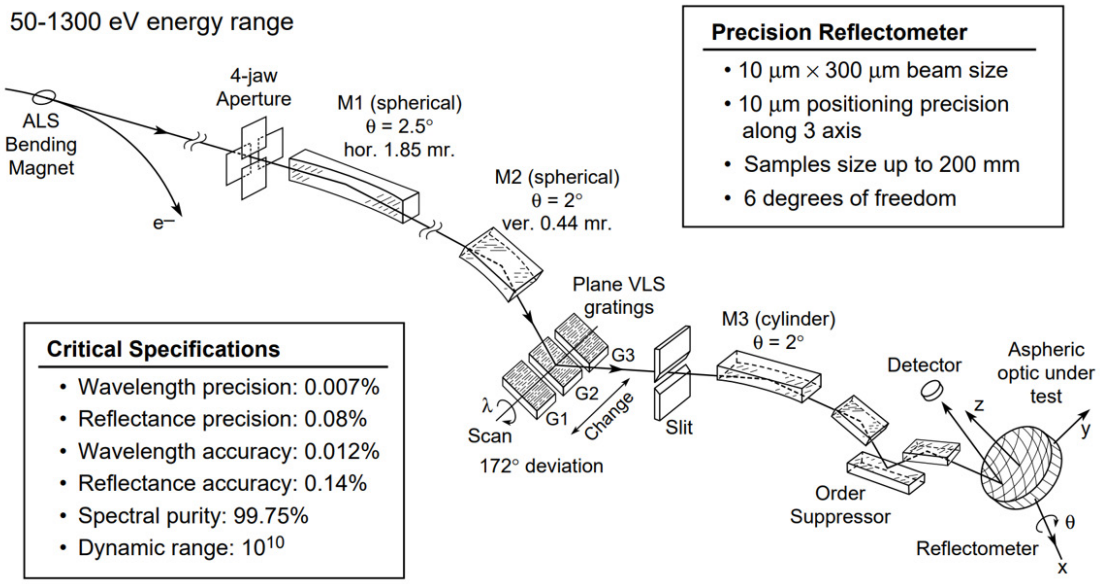

Figure 4.2: Advanced Light Source Calibration, Optics Testing and Spectroscopy Beamline 6.3.2.[15]. A beam of photons with energy 50-1300eV exits the bend magnet and is spatially filtered by the aperture, then projected with Kirkpatrick–Baez mirrors M1 and M2 onto a varied line-space grating monochromator. After the slit filters the +1 diffraction from the grating, a monochromatic illumination wave is projected by M3 onto the sample in the reflectometer, and the reflected intensity is captured by the detector.

how this single absorber has not one but a range of phase values for different illumination conditions. In section 4.1.4.3 we describe interference effects in multilayer and absorber reflections, particularly how transmission alone is insufficient to predict the relative phase or to model the impact of contamination. Finally, in section 4.1.5 we quantify the precision of our method, considering random and systematic error.

4.1.1 Experimental methods

We characterize an EUV photomask using measurements of reflectivity from pure multilayer and absorber regions on the mask, taken at the Advanced Light Source Calibration, Optics Testing and Spectroscopy Beamline 6.3.2.[15], depicted in 4.2. In each reflectivity measurement, we illuminate a flat spot on the photomask with a defined wavelength and angle of incidence, and measure the reflected intensity; we sequentially repeat this process over a range of wavelengths and angles to produce a scan of reflectivity.

Figure 4.1 depicts a cross-sectional view of measuring multilayer and absorber reflectivity. The reflectometer measures the amplitude-squared of the reflection coefficient of each region;

our goal is to extract the relative phase between these two reflections. In order to define the phase, we must specify a reference plane, which we define as the topmost surface of the photomask. Although this reference plane may not necessarily coincide with the plane of best focus, it is a very natural choice because it is the lowest possible reference plane which is completely above the surface. If we assume an absorber of thickness t and index of refraction $\tilde{n} = 1 + \delta - i\beta$, then the relative phase would be $2tk_z\delta$ (the factor of 2 arising due to reflection). However, as discussed in Section 4.1.4.3, it is not sufficient to only consider light transmitted through the absorber, because there is also a substantial effect due to interference with light reflected by the absorber, which is captured in the rigorous Fresnel reflection coefficient calculation. One counter-intuitive aspect of this approach to keep in mind is that the reference plane now depends on the thickness of contamination on the absorber; if the focal plane were not adjusted to compensate, then in the far-field image you would observe changes in both phase (scattering) and focus (imaging).

The multilayer nominally consists of a 2.5nm Ru cap as well as 40x Mo-Si bilayers; the "bilayers" actually contain 4 layers rather than 2 in our model, as an additional MoSi interdiffusion layer is added at both the Mo-Si and Si-Mo interfaces [3]. The absorber has an additional 58nm TaN absorber topped with a 2nm TaON anti-reflection coating (ARC) layer; note that the ARC is to reject out-of-band DUV light and not to suppress the reflection of EUV. The measured "multilayer" is actually absorber that has been etched, so we model this with an additional etch depth parameter, which can either represent residual TaN thickness (if positive) or reduced Ru thickness (if negative); the nominal value is -0.2nm, meaning that nominally the Ru would be 0.2nm thinner after etch than it is under the absorber. We further place a layer of pure carbon on top of the absorber and multilayer to represent contamination.

The refractive index of each layer is based on the atomic composition of the layer and the atomic scattering factors of the constituent elements [18, 44].

We use the atomic scattering factors, f_0, f_1, in the Center for X-Ray Optics database in order to capture the wavelength-dependent changes in absorption and refraction for each element[18].

For soft x-ray and EUV wavelengths, the refractive index can be approximated as $n = 1 - \delta + i\beta = 1 - \frac{n_a r_a \lambda^2}{2\pi}(f_0 + if_1)$, where the n_a is the atomic density, r_a is the Bohr radius, and λ is the wavelength; in the case of a multi-atomic material, one can then sum up the impact of each material on δ and β [4].

For each material, m, we define a nominal concentration $n_{a,m}$, which, together with the tabulated scattering factors $f_{0,m}, f_{1,m}$, determine a nominal refractive index: $n_m = 1 - \delta_m + i\beta_m = 1 - \frac{n_{a,m} r_a \lambda^2}{2\pi}(f_{0,m} + if_{1,m})$.

If a layer is composed of a set of materials, where material m has an atomic density of $w_m n_{a,m}$, we compute the refractive index with a weighted sum: $n = 1 - \delta + i\beta = 1 - (\sum_m w_m [\delta_m - i\beta_m])$.

Our model contains a total of $N = 101$ wavelengths, $M = 7$ materials, and $K = 10$ layers, meaning that we must predict a total of 1010 values for complex refractive index. The above

model suggests we may greatly reduce the dimension of the problem by controlling each layer's wavelength-dependent complex refractive index, $\mathbf{N} \in \mathbb{C}^{N\times K}$, via the much smaller matrix, of each layer's concentration of each element, $\mathbf{W} \in \mathbb{R}^{M\times K}$, and a matrix containing the nominal $\delta - i\beta$ for each material, $\mathbf{A} \in \mathbb{C}^{N\times M}$. $(\mathbf{N})_{n,k}$ is the complex refractive index in layer k at wavelength n; $(\mathbf{W})_{m,k}$ is the concentration of element m in layer k; and $(\mathbf{A})_{n,m}$ is $\delta - i\beta$ for element m at nominal atomic density for wavelength n. Finally we compute the wavelength-dependent refractive index matrix for all layers as $\mathbf{N} = \mathbf{1} - \mathbf{AW}$. Because most layers consist of only one or two materials, only 14 of the 70 concentration matrix elements are nonzero, allowing us to control a 1010-element complex valued matrix $\mathbf{N}$ via only 14 positive real numbers in $\mathbf{W}$. In Table 4.1 we report the resulting refractive index of each layer at $\lambda = 13.5$nm.

Nominal roughness values were obtained by AFM measurement of the top surface which found 134pm of RMS roughness on the multilayer and 382pm on the absorber. The 134pm of roughness on the multilayer was also used to apply a constant 1.95% reflectance loss to the multilayer (impacting both multilayer and absorber reflections), based on previous empirical measurements of power losses due to replicated surface roughness in EUV masks[31]. The final thickness, roughness, and refractive index of each layer is listed in Table 4.1.

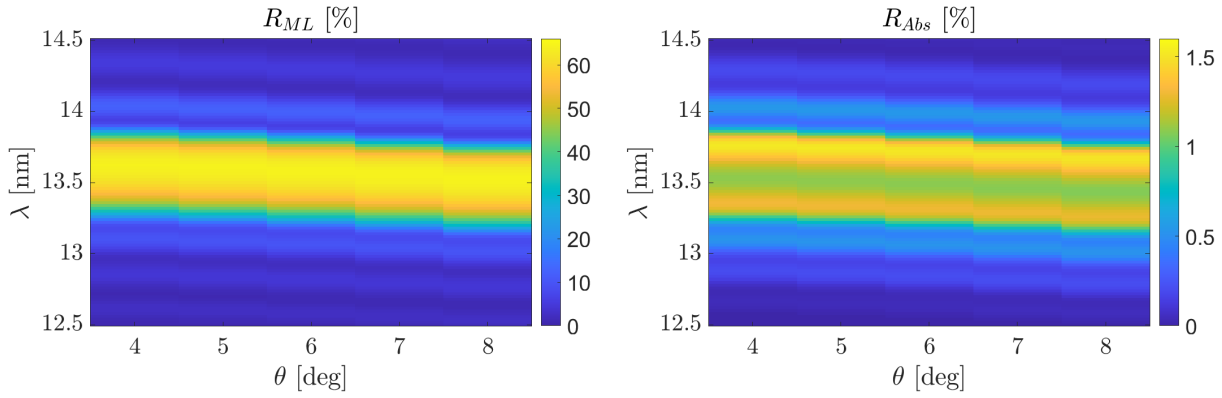

Figure 4.3: Reflectometry raw data scan of wavelength (λ) and angle (θ) for multilayer (left) and absorber (right). Note that the absorber reflectivity differs strikingly from the multilayer, with a drop in reflectivity near the multilayer's peak. This is due to destructive interference between a dominant component reflected by the multilayer and attenuated by the absorber transmission squared, and a much weaker modulation component reflected from the absorber only.

We measure reflectivity for both multilayer and absorber over a range of wavelengths (12.5-14.5nm) and angles (4-8°), depicted in Figure 4.3. We carried out our measurements multiple times to characterize changes over time as well as precision. Ultimately we acquired

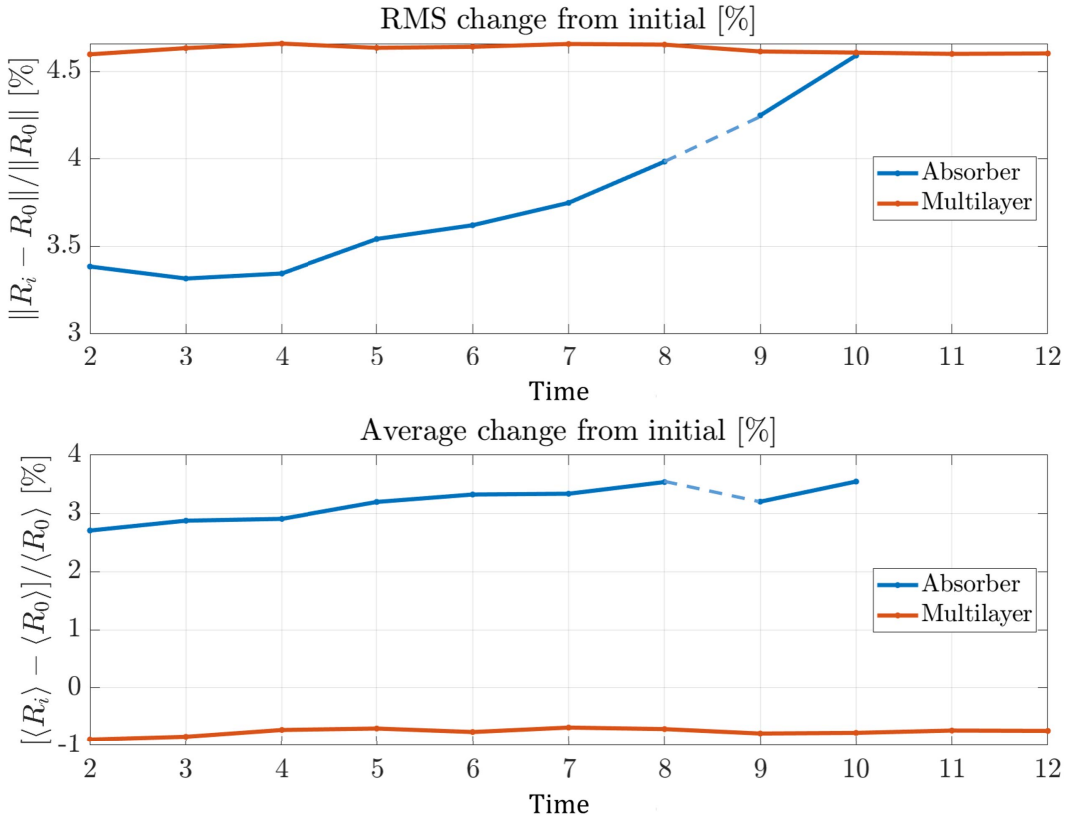

Figure 4.4: Changing reflectivity over time for multilayer and absorber. R_0 denotes the initial reflectivity for all wavelengths and angles, and R_i denotes the i^{th} subsequent measurement

. The multilayer experiences a change of 4.6% RMS (-0.89% average) during storage (5 months), but then relatively little change over the remaining measurements. On the other hand for the absorber there is a relatively smaller change of 3.4% RMS (+2.7% average) after the first measurement (3 months), followed by an increasing trend for both metrics. A dotted line is used to connect absorber measurements 8 and 9, when the light source was temporarily down.

an initial multilayer measurement, followed 2 months later by an initial absorber measurement, followed 3 months later by an additional 11 multilayer and 9 absorber measurements at the same mask locations collected back-to-back (all multilayer followed by all absorber measurements). Note that a dotted line is used to connect absorber measurements 8 and 9, when the light source was temporarily down.

We observe the reflectivity in both absorber and multilayer regions changing over time, shown in Fig. 4.4; here R_0 denotes the initial reflectivity measurement for all wavelengths and angles, and R_i denotes the i^{th} subsequent measurement. We plot the data both in terms

of the RMS change over all wavelengths and angles (top), as well as the average change again over all wavelengths and angles (bottom). The RMS change ($= \sqrt{\langle R_i^2 - R_0^2 \rangle}$) is a metric of the distance of R_i from R_0, adding up the magnitude of change across the entire signal and can only be positive; the average change ($= \langle R_i - R_0 \rangle$) sums up all changes to reflectivity regardless of where they occur in the signal, and it can be either positive (increased total power) or negative (decreased total power). There is a 4.6% RMS (-0.89% average) change in the multilayer signal after storage between the initial measurement and the subsequent measurements (5 months), but then relatively little change over the remaining measurements. The trend for the absorber on the other hand is quite different: there is a relatively smaller 3.4% RMS (+2.7% average) change after the first measurement (3 months), followed by a clear increasing trend for both metrics. This suggests that contamination of the multilayer occurred almost exclusively during storage, whereas contamination of the absorber occurred both during storage as well as during exposure. Also note that whereas the multilayer reflectivity decreases on average due to contamination, the absorber reflectivity actually increases, which can only be understood by considering interference between reflected and transmitted components in the absorber, further described in Section 4.1.4.3.

We hypothesize that the changes in both the multilayer and absorber are likely due to carbon contamination on the top surface, which has been observed growing in both storage[56, 21] and during exposure[56], as well as being cleaned during exposure in a scanner environment[21]. The differing contamination behaviors observed for multilayer and absorber are intriguing. Regarding the seemingly faster contamination rate of the multilayer during storage we refrain from speculating due to insufficient data. On the other hand there is a clear physical explanation for why we would observe the steady contamination during reflectivity measurement only in the absorber due to its much higher EUV absorption, which should create a steady stream of escaping secondary electrons that can "crack" hydrocarbon molecules in the chamber and allow freed carbon to bind to the surface.

4.1.2 Computational methods

4.1.2.1 Problem statement

We parametrize each layer of our mask by a thickness and a concentration of each elemental species, and parametrize each interface by its RMS surface roughness. Given all these physical parameters, we compute the Fresnel reflection coefficient[12, 23, 46] and compare its amplitude to a measurement of reflectivity vs wavelength (λ) and angle (θ). We then perform an iterative optimization to find a set of physical parameters that minimizes the mean-squared error. We write the error $\mathcal{E}$ as:

$$\mathcal{E}\left(x_1, x_2, ..., x_m\right) = \frac{1}{n} \sum_{i=1}^{n} \left[\left| r\left(x_1, x_2, ..., x_m; \lambda_i, \theta_i\right) \right| - \sqrt{R_{meas}\left(\lambda_i, \theta_i\right)} \right]^2,$$

where $x_1, x_2, ..., x_m$ are the unknown variables, r computes the Fresnel reflection coefficient, R_{meas} is the measured reflectivity, and $\{\lambda_i, \theta_i\}_{i=1}^{n}$ are the measured pairs of wavelengths and

angles. We express our nonlinear least squares optimization problem as:

$$\min_{x_1,...,x_m} \mathcal{E}\left(x_1, ..., x_m\right).$$

4.1.2.2 Initial multidimensional optimization

We choose a simple yet robust method for our optimization - cyclic coordinate descent[54], solving each 1D sub-problem with golden-section search[40, 7]. In this method we cycle through the unknown parameters (cyclic), optimizing the function with respect to only one parameter at a time (coordinate descent). We initialize with $\left\{x_j^0\right\}_{j=1}^m$, and obtain $\left\{x_j^k\right\}_{j=1}^m$ after k complete cycles. In each cycle we loop through all coordinates and update one at a time. The sub-problem for unknown x_j on iteration k is:

$$\min_{x_j} \mathcal{E}\left(x_1^k, ..., x_{j-1}^k, x_j, x_{j+1}^{k-1}, ..., x_m^{k-1}\right) = \min_{x_j} f_{j,k}\left(x_j\right).$$

We solve each 1D sub-problem with a golden-section search, which is a derivative-free optimization technique closely related to binary search [24, 40]. Whereas the well-known zero-finding algorithm binary search shrinks the distance between two points by a factor of 2 each iteration and converges to a zero, the golden-section search shrinks the distance by a factor of the golden ratio $\phi = \frac{1+\sqrt{5}}{2}$ each iteration and converges to a local minimum between the initial bounds. We denote the golden-section search algorithm as:

$$x^* = \mathrm{GSS}\left[f\left(\cdot\right), L, U, N\right],$$

where $f\left(\cdot\right)$ is the function to be minimized, L is the initial lower bound, U is the initial upper bound, and N is the number of iterations.

In each iteration we apply this sub-optimization to $f_{j,k}$, using bounds centered around the current iterate x_j^{k-1}, with a radius $\Delta_{j,k} = \Delta_j \gamma^{k-1}$ such that $L = x_j^{k-1} - \Delta_{j,k}$ and $U = x_j^{k-1} + \Delta_{j,k}$. Here Δ_j is the initial search radius, which is chosen to be a fraction β $(0 < \beta < 1)$ of the distance between the upper and lower variable bounds, UB_j and LB_j respectively: $\Delta_j = \beta\left(\mathrm{UB}_j - \mathrm{LB}_j\right)$. On each iteration the search radius shrinks by a factor of γ $(0 < \gamma < 1)$, making the search on each iteration more refined than the previous. We write this sub-optimization as:

$$x_j^* = \mathrm{GSS}\left[f_{j,k}\left(\cdot\right), x_j^{k-1} - \Delta_{j,k}, x_j^{k-1} + \Delta_{j,k}, N\right].$$

After completing the golden-section search we update x_j with a convex combination of the previous and new value using some α $(0 < \alpha < 1)$:

$$x_j^k = \alpha x_j^* + (1 - \alpha) x_j^{k-1}.$$

In practice we use an α value slightly below 1 (e.g. 0.9), which is equivalent to slightly reducing the step size for each update. Reducing the α value slows down convergence,

but generally helps improve stability. An advantage of this approach is that golden-section search does not require us to perform any cumbersome gradient or Jacobian computations for the Fresnel reflection coefficient (unlike other approaches like gradient descent or Gauss-Newton) [5].

4.1.2.3 Contamination measurement

After performing this multi-dimensional optimization based on the initial multilayer and absorber measurements only, we freeze all parameters except the thickness of carbon contamination. Then for each measurement we perform a brute-force search, evaluating a range of thicknesses from 0-8nm and choosing the thickness with the lowest MSE (Mean-Squared Error). For measuring the contamination on the absorber, we use the first multilayer measurement and change only the measurement of the absorber, and vice versa for contamination on the multilayer. The simple brute-force optimization algorithm used in this final step should help allay fears about the physical accuracy of the initial optimization because any remaining errors in the physical parameters are held constant for all measurements. All claims about sensitivity and precision to thickness and phase are based solely only on this secondary brute-force optimization (which is guaranteed to find the global optimum), rather than the initial multi-dimensional optimization (which is not). Furthermore, regardless of the physical correctness of the global optimum of this function, as long as there is a sufficient gap in MSE from the next lowest local minimum, then the global optimum will be chosen robustly and the precision will simply be due to the curvature of the minimum and the noise in the measurement, exactly like a linear regression. Still, we acknowledge that unless almost all physical parameters are known in advance, this reflectometry phase measurement may not be absolutely accurate, but should remain highly repeatable and sensitive to small changes. Future work is still needed to study the absolute accuracy of this method, and it is possible that accuracy can only be guaranteed with additional measurements, for example AFM to measure surface roughness (which is included in these results) or grazing incidence reflectometry to measure refractive index (which is not).

4.1.3 Results

4.1.3.1 Initial optimization

We perform our initial fit on the first reflectivity scan only (each consisting of 505 measurements) for multilayer and absorber, fitting 31 physical parameters using in total 1010 reflectivity measurements. We specify initial values as well as upper and lower bounds for each variable, which restricts the domain of our search to a 31-dimensional box. We then loop through variables updating one at a time iteratively to increase the fidelity of the model to the measured reflectivity as described in Section 4.1.2.2.The predicted amplitude of the initial fit is shown in Fig. 4.5 alongside the experimental measurement. Note that both experiment and fit are plotted as a function of k_z ($= 2\pi \cos\theta/\lambda$), whereas the actual experimental data

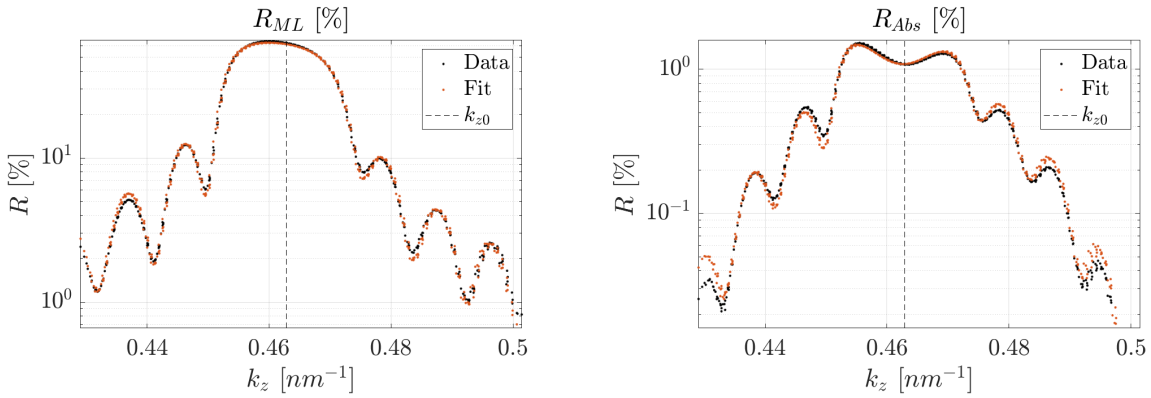

Figure 4.5: Modeled reflectivity vs k_z. Comparison of raw (black) and fitted (red) reflectivity based on initial 31 parameter fit, demonstrating the ability to find a solution consistent with the data.

	Thickness	**Roughness**	**delta**	**beta**
C (Absorber)	1.683	0.327	0.046	0.008
TaON (Absorber)	1.857	0.373	0.028	0.020
TaN (Absorber)	58.110	0.346	0.046	0.033
Vac (Etch)	58.105	0.000	0.000	0.000
C (Etch)	3.854	0.159	0.046	0.008
TaN (Etch)	-0.311	0.000	0.046	0.033
Ru (Multilayer)	2.151	0.141	0.120	0.018
Si (Multilayer)	3.625	0.136	0.001	0.002
MoSi2 (Multilayer)	0.568	0.140	0.025	0.004
Mo (Multilayer)	1.858	0.141	0.074	0.006
MoSi2 (Multilayer)	0.919	0.138	0.029	0.005

Table 4.1: Physical model for the film-stack. "Mutlilayer" layers (Ru, Si, MoSi, Mo, MoSi) are common to absorber and etch. On top of these are either the 3 "absorber" layers (C, TaON, TaN) or "etch" layers (Vac, C, TaN). If TaN (Etch) is positive, then some absorber remains un-etched; on the other hand if it is negative (as it is here) then some of the Ru has been etched. Roughness is RMS surface roughness. Index of refraction at $\lambda = 13.5$nm is $\tilde{n} = 1 - \delta + i\beta$.

is collected as a function of both λ and θ. This simplification is possible because over this measurement range the reflectance has very little dependence on the transverse wave vector due to planar symmetry and the relatively small range of measured angles. We perform this compression only for ease of visualization, which is to say that the actual Fresnel coefficient

calculation takes into account both the wavelength and the angle. This approximation does lead to a faint amount of apparent "noise" in Fig. 4.5, which is in fact not noise but rather the manifestation of slight reflectance variation orthogonal to k_z. While there is certainly no mathematical guarantee about the uniqueness of our solution, Fig. 4.5 demonstrates that our optimization procedure can at least find *some* physical solution consistent with the data. More likely than not some physical parameters may be "compensating" for others and possibly introducing some systematic errors in the prediction; however, this issue (if present) could be resolved given a sufficiently accurate initial guess and sufficiently tight bounds on each variable, which may require that additional supplementary measurements be performed. The final physical model for thickness, roughness, and refractive index is presented in Table 4.1, but again we stress that this is merely one possible physical model rather than the only possibility consistent with the data.

4.1.3.2 Contamination measurement

Following this initial optimization, we freeze the model and perform a brute-force search for the carbon contamination thickness corresponding to each multilayer and absorber measurement. This is depicted in Figure 4.6 for multilayer (top) and absorber (bottom). The leftmost plot shows the MSE vs carbon thickness for the each measurement, and the black dots show the optimal values. For the multilayer there is a clear separation between the first and later measurements, whereas all the absorber measurements are closely clustered. The middle plot shows the carbon contamination thickness for each measurement, which very clearly shows the sharp jump in contamination on the multilayer as opposed to the steady increase on the absorber. The rightmost plot shows how the contamination impacts the relative phase value (absorber phase minus multilayer phase). The multilayer contamination causes the phase to jump down by $-0.86°$ and then remain approximately constant, whereas the absorber contamination causes the phase to gradually decrease by $-1.12°$. Note that both are changing in the same direction, meaning that the combining the two effects would imply a $-1.98°$ change in relative phase.

Two somewhat counter-intuitive observations arrive from these results: first, a smaller change in the absorber contamination thickness (126 vs 1068 pm) causes a larger change in the phase ($-1.12°$ vs $-0.86°$); and second the growth of the same material on absorber and multilayer causes the same sign of change in the relative phase (absorber phase minus multilayer phase). Both these apparent mysteries are caused by interference effects, discussed in Section 4.1.4.3. If we accounted only for transmission, then the phase shift from a change in thickness would be $(\pm)2tk_z\delta$, corresponding to $+0.31°$ and $-2.59°$ for the absorber and multilayer respectively. As expected, the signs are opposite and the magnitude of the change in phase is proportional to the change in thickness. But interference impacts both the multilayer and absorber reflections, leading to a smaller value with the same sign for the multilayer ($-0.86°$ vs $-2.59°$) and a larger value with opposite sign for the absorber ($-1.12°$ vs $+0.31°$). These substantial deviations from the simplified transmission model can be attributed to interference effects in the reflection, which are explored further in Section

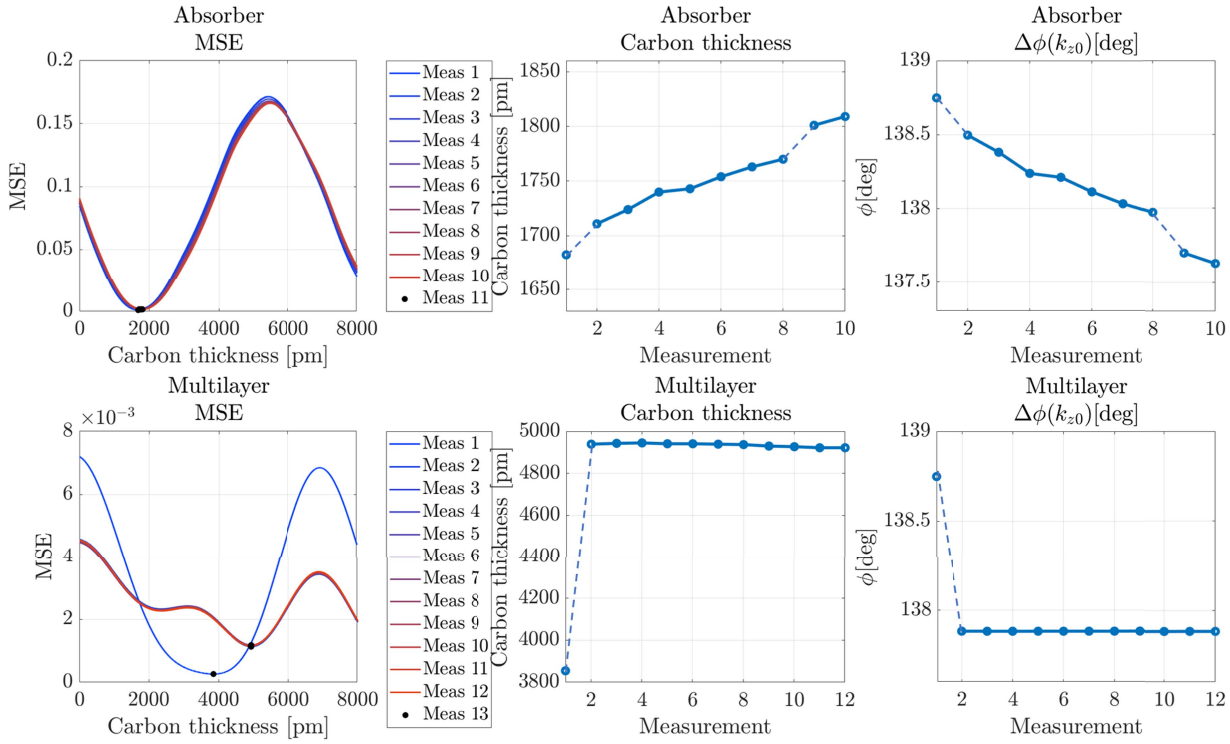

Figure 4.6: Carbon growth on multilayer (top) and absorber (bottom): single parameter fit to model contamination. Left: MSE vs carbon thickness; black dots show optimal value for each measurement. Center: Recovered carbon thickness: increases sharply for multilayer during storage but not measurement; increases systematically during both storage and measurement for absorber. A dotted line is used to connect data points before and after storage, and in the absorber between measurements 8 and 9 when the light source was temporarily down. Right: Relative phase between absorber and multilayer; total change of $-0.86°$ for multilayer and $-1.12°$ for absorber.

4.1.4.3.

As to the question of how we could possibly be sensitive to thickness changes smaller than an atomic monolayer, the answer is that we are measuring an average thickness across the area of the beam. While the contamination must consist of an integer number of atoms, due to nonuniformity it need not consist of an integer number of atomic monolayers. In this way we can detect changes averaging only a fraction of a monolayer across the beam spot.

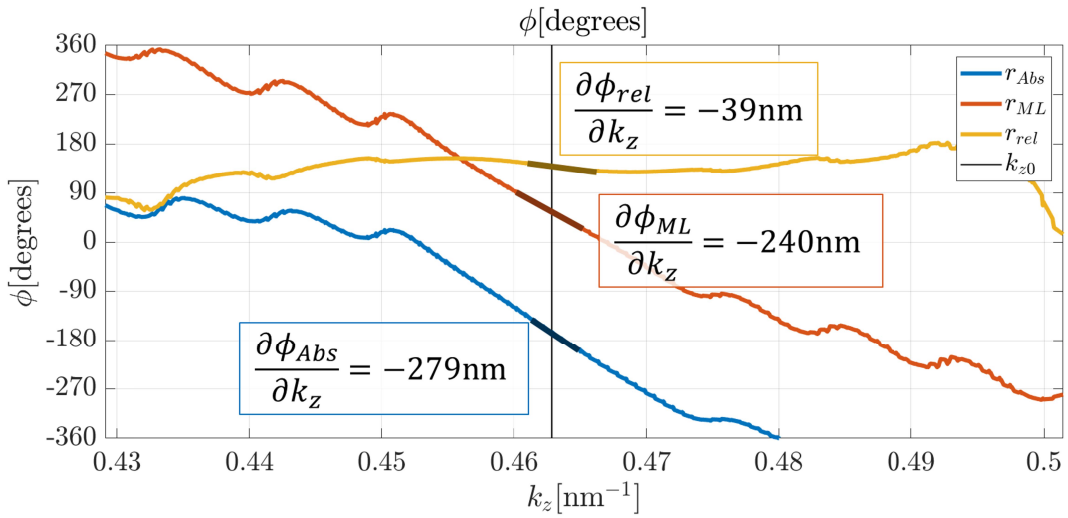

Figure 4.7: Recovered phase based on fitted Fresnel reflection coefficient from initial 31 parameter fit for absorber (blue), multilayer (red), and relative (yellow). The partial derivative $\partial\phi/\partial k_z$ represents an effective propagation distance. The absorber has a significantly larger effective propagation distance than the multilayer (285 vs 240nm), which will impact the Mask 3D (M3D) effects.

4.1.4 Analysis

4.1.4.1 Effective propagation distance

From our initial model, we can extract the phase as a function of k_z, shown in Figure 4.7. In addition to the relative phase shift between multilayer and absorber at the nominal operating condition $k_{z0} = 2\pi \cos 6°/(13.5\text{nm})$, we can also use our model to obtain useful information about mask 3D effects from the partial derivative $\partial\phi/\partial k_z$. This derivative represents an effective propagation distance for each reflection coefficient. The effective propagation distance for the multilayer is 240nm, including both reflection by the multilayer and transmission twice through vacuum of equal height to the absorber. The propagation distance for the absorber on the other hand is 279nm, an increase of 39nm (17%). Pattern translation as a function of illumination angle is proportional to this propagation distance ($\Delta x = \frac{\partial \phi}{\partial k_z} \tan\theta$), meaning that for this mask light reflected from the absorber will translate 17% further than light reflected from the multilayer, which should impact Mask 3D (M3D) effects. Bear in mind however that these numbers, as they are derived from the initial optimization which is not guaranteed to find a uniquely optimal solution, must still be taken with a grain of salt and should be further confirmed with other measurements. It may be the case that the actual numbers differ from those presented, however the same qualitative effects would be present albeit with possibly a different magnitude.

NA	$\Delta\phi_\lambda$ [2%]	$\Delta\phi_{\theta_y}$	$\Delta\phi_{\theta_x}$	m_y	m_x	θ_0	θ_y^{max}	θ_x^{max}
0.33	21.0°	18.3°	3.6°	4	4	6.0°	10.7°	7.7°
0.55	21.0°	13.5°	10.1°	8	4	5.355°	9.3°	9.6°

Table 4.2: Phase change across source for the 0.33 and 0.55 NA systems, based on recovered $\partial\phi_{rel}/\partial k_z = 42$nm. For both systems a 2% λ variation leads to a 21.0° variation in the phase. The 0.33 NA system also has substantial variation in the Y direction (18.3°) but much less in X (3.6°). In contrast, in the 0.55 NA system the variation is reduced in Y (13.5°) but greatly increased in X (10.1°).

4.1.4.2 Phase variation across the source

In addition to impacting M3D effects, the 39nm difference in effective propagation distance between multilayer and absorber will cause the pattern phase to depend on the illumination condition–that is to say the same absorber will have different phase from different source points. Variation in k_z across the source could arise from variation in either angle or wavelength. For this mask a 2% change in wavelength would cause a phase shift of $\Delta\phi_\lambda = (\partial\phi/\partial k_z)\, k_0\, (2\%) = 21.0°$. A change in angle from one edge of the pupil to the other (Y, shadowing orientation) would cause a phase shift of $\Delta\phi_{\theta_y} = (\partial\phi/\partial k_z)\, k_0\, [\cos\theta_{min} - \cos\theta_{max}] = 18.3°$ and 13.5° for the 0.33 and 0.55 NA systems respectively; the effect is somewhat mitigated in the higher NA system due to the increased magnification and decreased chief-ray angle. On the other hand we observe the opposite trend in the X direction due to increased mask-side NA, where the variation increases from 3.6° to 10.1°. These effects are summarized in Table 4.2.

The fact that images from different illumination angles (and wavelengths) can span such a wide range of phase shifts will surely have consequences for imaging and therefore should be considered in designing future EUV aPSMs. Quantifying the severity of these effects and exploring options for mitigation are open-ended problems that surely merit further study; here we propose just a few possible strategies for mitigation: In the case of low k_1 (where there is the most interest in EUV aPSMs to begin with), the required illumination will introduce substantial M3D effects in source points far from normal incidence (positive u_y); it may be possible to equalize the contrast for source points with positive and negative u_y by designing the absorber phase primarily for positive-u_y source points at the expense of the negative where M3D effects are less severe to begin with. Another strategy, which could perhaps best accommodate arbitrary sources, would be to design an absorber whose phase varies more slowly vs k_z, which would reduce the magnitude of this effect overall; in principle this should be possible, particularly if one were willing to use a more complicated absorber architecture. Finally, if the resolution permits, one could simply restrict the source to reduce k_z variation without redesigning the absorber.

4.1.4.3 Interference effects

In order to interpret the multilayer and absorber reflection coefficients, a simplified "transmission-only" analysis will not suffice. For both the multilayer and absorber, we must consider interference between different reflections, which is rigorously accounted for in the Fresnel coefficient calculation via the transfer matrix method [12] in which each layer (j) is described by a 2×2 transfer matrix M_j, and the overall reflection or transmission through a stack of layers is computed from the product of all transfer matrices: $M = \Pi_j M_j$. Due to the associativity of matrix multiplication we may group together any set of layers without modifying the final result, so we define three matrices M_{ML}, M_{Abs}, and M_{Etch}, which each contain the product of transfer matrices in the respective group of layers.

The multilayer consists of $N = 40$ repetitions of a 4-layer unit cell (Si-MoSi-Mo-MoSi): $M_{ML} = (\Pi_{j=1}^{4} M_{ML,j})^N$. The absorber consists of 4 layers (C-TaON-TaN-Ru): $M_{Abs} = \Pi_{j=1}^{4} M_{Abs,j}$. And the etch consists of 3 layers (Vac-C-Ru): $M_{Etch} = \Pi_{j=1}^{3} M_{Etch,j}$. The final transfer matrix in the absorber region of the mask is $M_{Abs} M_{ML}$, and in the multilayer region it is $M_{Etch} M_{ML}$. The reflection from the transfer matrix is then given by $r = M_{2,1}/M_{1,1}$, and the transmission is given by $t = 1/M_{1,1}$.

A simplified "transmission-only" analysis would be equivalent to considering only $t_{Abs0} = 1/M_{Abs1,1}$ and $t_{Etch0} = 1/M_{Etch1,1}$, then taking the ratio t_{Abs0}^2/t_{Etch0}^2. However, this approach does not capture the impact of interference between reflections from different layers, which can be substantial, such as the destructive interference visible in the absorber reflectivity. A slightly more advanced approach which can begin to explain these interference effects is to consider the final reflection as the sum of two components: a dominant component which transmits twice and is reflected by the multilayer, and a much weaker modulation component which is reflected before interacting with the multilayer. We write these components for the absorber as $t_{Abs0}^2 r_{ML0}$ and r_{Abs0}, and for the etch as $t_{Etch0}^2 r_{ML0}$ and r_{Etch0}, where r_{ML0}, r_{Abs0}, and r_{Etch0} are $M_{2,1}/M_{1,1}$ for the corresponding transmission matrix. This allows us to introduce the approximate reflection formulas $r_{abs} \approx t_{abs0}^2 r_{ML0} + r_{abs0}$ and $r_{ML} \approx t_{etch0}^2 r_{ML0} + r_{etch0}$.

This interference process is illustrated in Figures 4.8 and 4.9, which show the total reflection (blue) with its dominant (red) and modulation (yellow) components. Note that all three components are calculated from the product of multiple transfer matrices, so each component itself contains interference between multiple layers; furthermore the approximate decomposition into dominant and modulation components is purely for illustration and the approximation is not used in computing the final reflection.

In the case of the multilayer region, shown in Figure 4.8, the final reflection is almost entirely due to the dominant component, because the C and Ru have well over 90% transmittance but less than 10% reflectance. This leads to the relative imbalance of the dominant multilayer reflection ($t_{etch0}^2 r_{ML0}$) and the much smaller modulation (r_{etch0}). The greater slope $\partial\phi/\partial k_z$ of the dominant component implies that it has a larger effective propagation distance, which is expected due to the effective reflection plane of the multilayer.

As a result of contamination, both the dominant and modulation components are attenuated, but the overall reflection has increased amplitude for high k_z and decreased amplitude

for low k_z. This overall amplitude change is due to interference between the two components, specifically because the phase between the two components is decreased for high k_z and increased for low k_z. The dominant component phase increases by roughly a constant, but the modulation component phase decreases by a much larger amount in roughly a parabolic trend. This causes the overall change in phase at k_{z0} to be much smaller than the change in phase from the dominant component, hence why the simplified transmission-only analysis greatly over-predicts the change in phase. Also note that the phase change of the final reflection has an overall negative slope with k_z, implying that the overall propagation distance is increased, i.e. the surface contamination appears to push the reflection plane deeper into the multilayer.

In the case of the absorber, as shown in Figure 4.9 , interference plays an even greater role in forming the final reflection. Whereas the dominant absorber component is roughly just an attenuated version of the multilayer, the final reflection amplitude oscillates around this value, with a local minimum approximately where the multilayer attains its maximum. This local minimum is easily observed in the raw data, and is caused by destructive interference between the dominant ($t_{abs0}^2 r_{ML0}$) and modulation (r_{abs0}) components. The components interfere destructively because at k_{z0} (black vertical line) the phase of the dominant and modulation components are almost exactly 0° and 180° respectively. Also note that there is substantial k_z variation in the absorber modulation component itself; this would not be the case if it were simply a reflection from the very top interface, however k_z variation arises because the modulation component itself includes interference from the 4 layers comprising the absorber (C-TaON-TaN-Ru), captured in the transfer matrix calculation.

Contamination on the absorber results in a negligible change to the amplitude of the dominant component, but a noticeable phase shift due to the additional optical path length of transmission through carbon. On the other hand the total amplitude clearly shows the left peak decreasing and the right peak increasing (which is easily observable in the measured reflectivity data). A similar trend is also seen in the changes to the modulation component amplitude, but the final reflection has additional oscillations due to interference. The changes in the phase of the final reflection oscillate by several degrees around the changes to the dominant component, again demonstrating the importance of a rigorous Fresnel analysis rather than a simplified transmission-only analysis to assess phase effects.

4.1.5 Precision

In computing the precision, we exclude the initial measurements for both multilayer and absorber to remove the impact of storage. In addition, we compute the 3σ precision in two ways: using the raw data, and using the residuals of a linear fit. Before taking the linear residuals the 3σ for contamination thickness is 8.0 and 30.8pm for the multilayer and absorber respectively, but after removing the linear trend we obtain a better estimate of only the random component of the variation, 3.5 and 5.8pm. So the sensitivity to thickness on both multilayer and absorber surfaces appears to be roughly similar, and lower than 6pm. The corresponding precision for phase after accounting for the linear trend is 0.16° for the

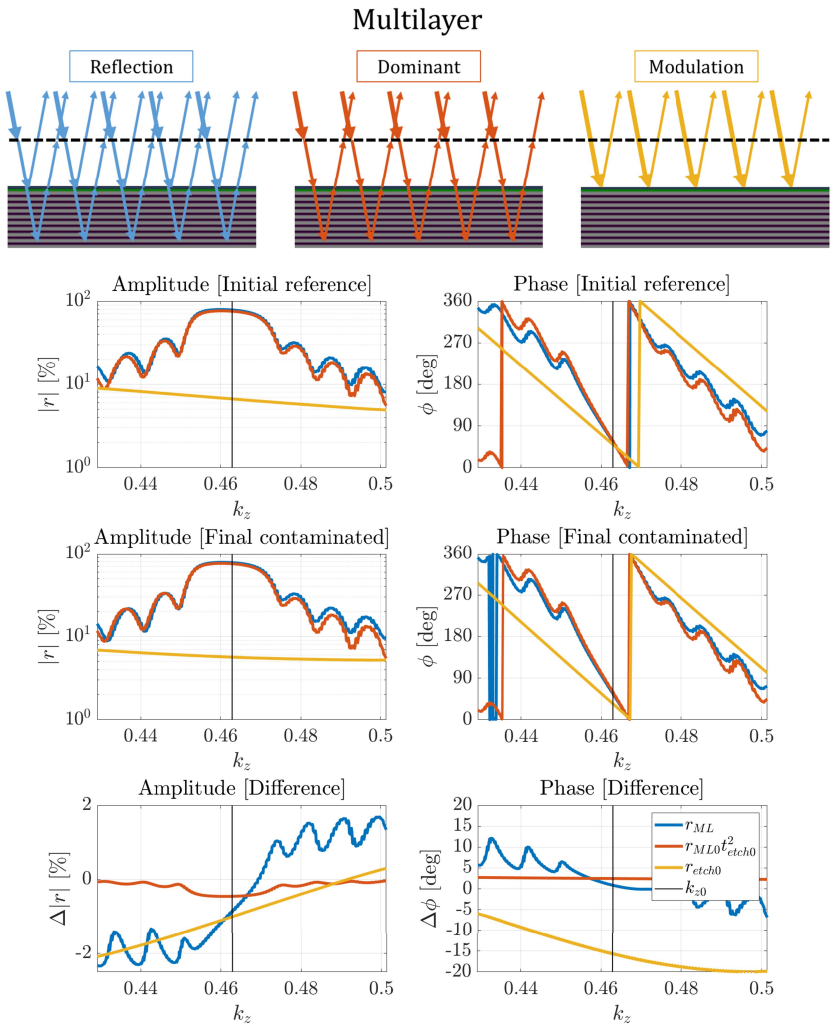

Figure 4.8: Schematic illustration of decomposing multilayer reflection into dominant and modulation components. Initial measurement: The amplitude of the dominant component is approximately the amplitude of the multilayer reflection coefficient, whereas the amplitude of the modulation component is approximately linear with k_z. Final measurement: After the growth of approximately 1nm of carbon, the amplitude of both components decreases uniformly. Difference: both the dominant and modulation components are attenuated, but the overall reflection has increased amplitude for high k_z and decreased amplitude for low k_z. Phase changes in opposite directions for the dominant and modulation components, reducing the final change in phase. The final phase change contains a slope due to the different effective propagation distances in the two components.

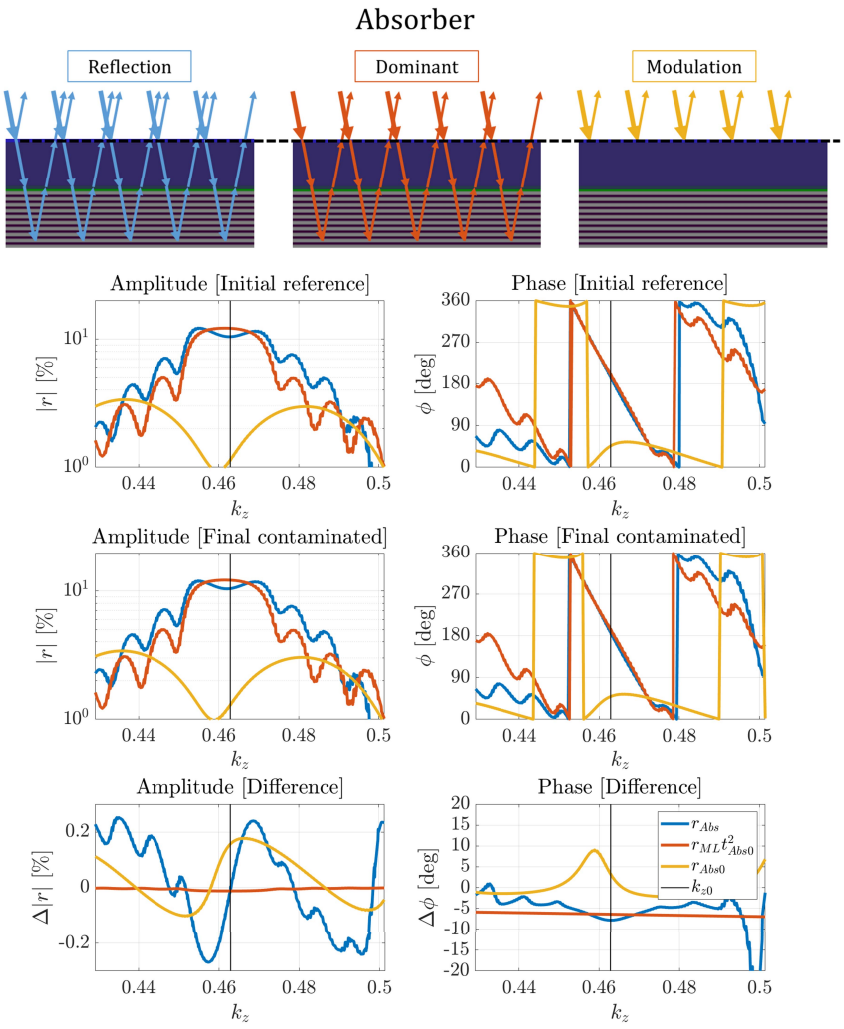

Figure 4.9: Schematic illustration of decomposing absorber reflection into dominant and modulation components. Initial measurement: The amplitude of the dominant component closely follows the amplitude of the multilayer reflection coefficient, whereas the amplitude of the modulation component has additional features, such as the local minimum at k_{z0} (due to interference between the 4 layers composing the absorber). Final measurement: After the growth of approximately 100pm of carbon, there is a slight change in the relative intensities of the two highest peaks. Difference: Observe the changes due to contamination from final minus initial model. The left peak decreases and the right peak increases by a similar amount. There are substantial changes in both modulation component amplitude and the phase of both components.

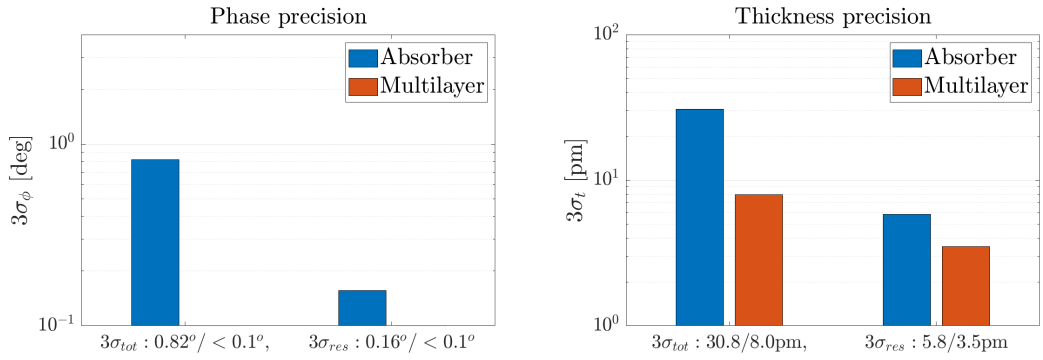

Figure 4.10: Precision: 3σ of total and linear residuals for multilayer and absorber relative phase (left) and contamination thickness (right). First data point has been removed to remove the impact of storage and only consider precision of repeated measurements. Taking the linear residuals is a rough way to separate a systematic trend from noise. The multilayer trend has $3\sigma < 0.1°$ whether or not we take the linear residuals. The absorber trend is highly systematic, so the precision is greatly reduced from the raw data ($3\sigma = 0.82°$) by taking the linear residuals (0.16°). Thickness precision considering the linear residuals is $3\sigma = 5.8$pm and 3.5pm for the multilayer and absorber, respectively.

absorber, and less than 0.1° for the multilayer. However the multilayer result being so low is likely not a general phenomenon, but rather seems to be specific to interference effects in the multilayer reflection, whereby increasing contamination thickness shifts the phase of the dominant and modulation components of the multilayer reflection in opposite directions leading to a reduced overall phase shift. So the absorber precision of 0.16° should be seen as the more representative estimate for the phase precision of the technique.

Conclusion

We have presented a computational reflectometry technique to characterize the phase of an EUV photomask via measurements of reflectivity from multilayer and absorber regions. A 31 parameter physical model was capable of producing a highly accurate match to the experimental reflectivity data, although further studies are required to understand the absolute accuracy of the model because the optimization is not guaranteed to find a uniquely optimal solution. We used the physical model to extract the relative phase of the absorber and multilayer reflections, as well as the phase partial derivative with respect to k_z, which represents an effective propagation distance for each reflection coefficient. According to our model, the effective propagation into the absorber (279nm) is 39nm further than the multilayer (240nm); among other things this will cause the phase to vary across the source. We then held our initial physical model fixed and used a brute-force search to measure carbon contamination on the multilayer and absorber; this simplified algorithm can guarantee convergence

to a unique optimum for each measurement. We found that the multilayer was contaminated substantially more during storage than the absorber (1068pm vs 28pm), whereas during repeated EUV reflectometry measurements the contamination actually slightly decreased on the multilayer but continued growing on the absorber (-1.7pm/measurement vs 12.3pm/measurement). Although changes in absorber thickness amounted to only a fraction of an atomic monolayer, we were able to measure the average thickness of a non-uniform layer across the beam-spot to sub-atomic precision. Finally we used repeated measurements to quantify the method's precision. The precision in measuring contamination thickness determined by the linear residuals was $3\sigma = 3.5$ and 5.8pm for the multilayer and absorber. The precision for measuring phase based on the contamination thickness was $3\sigma = 0.16^\circ$ for the absorber and better than 0.1° for the multilayer. Therefore we have demonstrated single picometer precision for measuring both contamination thickness ($3\sigma < 6$pm) and phase ($3\sigma < 0.2^\circ = 7.5$pm wavefront).

4.1.6 Experimentally recovering mask architecture, 3 masks

In addition to the standard TaN absorber, we have also carried out reflectometry measurements for multilayer and absorber of two aPSMs, one with 2 absorber layers, the other with 3 layers.

The attenuation of the absorber relative to multilayer is about 13% for Mask A, 36% for Mask B, and 39% for Mask C. The phase shift is about 151^o,218^o,206^o for the 3 masks respectively. A phase shift larger than 180^o is optimal for EUV aPSMs, due to M3D effects, and such a phase shift around $1.2\pi = 216^o$ is observed in the two aPSMs. The effective propagation distance for the absorber reflection was respectively 293, 259, and 261nm; for the multilayer it was 234, 235, and 243nm; and the difference was 58, 24, and 19nm.

Therefore, the aPSMs have higher reflectance, larger phase shift, and less of a discrepancy between the propagation distance in the absorber and multilayer reflections. The higher reflectance and shorter propagation distance are due to using a thinner absorber, while the higher phase shift (in a shorter absorber thickness) is due to refractive indices with greater phase contrast from vacuum.

Layer	Parameter	Type	Min	Nom	Fit	Max
C	Thickness	Absorber	0.00	2.00	1.07	4.00
TaON	Thickness	Absorber	1.00	2.00	1.75	3.00
TaN	Thickness	Absorber	56.00	58.00	58.20	60.00
Ru	Thickness	Absorber	1.50	2.50	2.86	3.50
Si	Thickness	Multilayer	3.14	3.24	3.29	3.34
MoSi2	Thickness	Multilayer	0.59	0.60	0.61	0.61
Mo	Thickness	Multilayer	2.06	2.16	2.16	2.26
MoSi2	Thickness	Multilayer	0.94	0.95	0.95	0.96
dEtch	Thickness	Etch	-0.40	-0.20	-0.40	0.00
C	Thickness	Etch	0.00	2.00	0.26	4.00
C - C	Concentration	Absorber	1.0000	1.0000	1.0000	1.0000
TaON - N	Concentration	Absorber	0.0000	0.3333	1.6364	2.0000
TaON - O	Concentration	Absorber	0.0000	0.3333	1.6364	2.0000
TaON - Ta	Concentration	Absorber	0.0000	0.3333	0.5005	2.0000
TaN - N	Concentration	Absorber	0.0000	0.8229	0.0015	2.0000
TaN - Ta	Concentration	Absorber	0.0000	0.8229	0.7983	2.0000
Ru - Ru	Concentration	Absorber	0.0000	1.0000	0.8403	2.0000
Si - Si	Concentration	Multilayer	0.0000	1.0000	1.4454	2.0000
MoSi2 - Mo	Concentration	Multilayer	0.0000	0.6667	0.5540	2.0000
MoSi2 - Si	Concentration	Multilayer	0.0000	0.3333	1.5669	2.0000
Mo - Mo	Concentration	Multilayer	0.0000	1.0000	0.8497	2.0000
MoSi2 - Mo	Concentration	Multilayer	0.0000	0.6667	0.6468	2.0000
MoSi2 - Si	Concentration	Multilayer	0.0000	0.3333	0.1390	2.0000
C - C	Concentration	Etch	1.0000	1.0000	1.0000	1.0000

Table 4.3: TaN mask architecture, nominal, fit, and optimization bounds for each variable.

Layer	Parameter	Type	Min	Nom	Fit	Max
C	Thickness	Absorber	0.00	2.00	0.17	4.00
TaB	Thickness	Absorber	2.00	4.00	3.98	6.00
MoRe	Thickness	Absorber	46.00	48.00	48.46	50.00
Ru	Thickness	Absorber	1.50	2.50	3.02	3.50
Si	Thickness	Multilayer	3.14	3.24	3.29	3.34
MoSi2	Thickness	Multilayer	0.59	0.60	0.61	0.61
Mo	Thickness	Multilayer	2.06	2.16	2.16	2.26
MoSi2	Thickness	Multilayer	0.94	0.95	0.95	0.96
dEtch	Thickness	Etch	-0.40	-0.20	-0.40	0.00
C	Thickness	Etch	0.00	2.00	0.91	4.00
C - C	Concentration	Absorber	1.0000	1.0000	1.0000	1.0000
TaB - B	Concentration	Absorber	0.3512	0.3863	0.4202	0.4249
TaB - Ta	Concentration	Absorber	0.5236	0.5760	0.5586	0.6336
MoRe - Mo	Concentration	Absorber	0.5020	0.5522	0.5642	0.6074
MoRe - Re	Concentration	Absorber	0.3935	0.4329	0.4273	0.4762
Ru - Ru	Concentration	Absorber	0.0000	1.0000	0.1459	2.0000
Si - Si	Concentration	Multilayer	0.0000	1.0000	1.7064	2.0000
MoSi2 - Mo	Concentration	Multilayer	0.0000	0.6667	0.4165	2.0000
MoSi2 - Si	Concentration	Multilayer	0.0000	0.3333	0.6069	2.0000
Mo - Mo	Concentration	Multilayer	0.0000	1.0000	0.8442	2.0000
MoSi2 - Mo	Concentration	Multilayer	0.0000	0.6667	0.7178	2.0000
MoSi2 - Si	Concentration	Multilayer	0.0000	0.3333	0.0006	2.0000
C - C	Concentration	Etch	1.0000	1.0000	1.0000	1.0000

Table 4.4: 2-layer aPSM mask architecture, nominal, fit, and optimization bounds for each variable.

Layer	Parameter	Type	Min	Nom	Fit	Max
C	Thickness	Absorber	0.00	2.00	0.00	4.00
TaB	Thickness	Absorber	15.20	17.20	17.70	19.20
MoRe	Thickness	Absorber	7.10	9.10	8.90	11.10
Ru	Thickness	Absorber	20.30	22.30	22.63	24.30
Ru	Thickness	Absorber	1.50	2.50	2.64	3.50
Si	Thickness	Multilayer	3.14	3.24	3.29	3.34
MoSi2	Thickness	Multilayer	0.59	0.60	0.60	0.61
Mo	Thickness	Multilayer	2.06	2.16	2.16	2.26
MoSi2	Thickness	Multilayer	0.94	0.95	0.95	0.96
dEtch	Thickness	Etch	-0.40	-0.20	-0.40	0.00
C	Thickness	Etch	0.00	2.00	3.51	4.00
C - C	Concentration	Absorber	1.0000	1.0000	1.0000	1.0000
TaB - B	Concentration	Absorber	0.3324	0.3656	0.3347	0.4022
TaB - Ta	Concentration	Absorber	0.5725	0.6297	0.5726	0.6927
MoRe - Mo	Concentration	Absorber	0.4697	0.5167	0.4934	0.5684
MoRe - Re	Concentration	Absorber	0.4010	0.4411	0.4270	0.4852
Ru - Ru	Concentration	Absorber	0.0000	1.0000	1.1006	2.0000
Si - Si	Concentration	Multilayer	0.0000	1.0000	1.3970	2.0000
MoSi2 - Mo	Concentration	Multilayer	0.0000	0.6667	0.3470	2.0000
MoSi2 - Si	Concentration	Multilayer	0.0000	0.3333	0.0006	2.0000
Mo - Mo	Concentration	Multilayer	0.0000	1.0000	0.8091	2.0000
MoSi2 - Mo	Concentration	Multilayer	0.0000	0.6667	0.7687	2.0000
MoSi2 - Si	Concentration	Multilayer	0.0000	0.3333	1.2116	2.0000
C - C	Concentration	Etch	1.0000	1.0000	1.0000	1.0000

Table 4.5: 3-layer aPSM mask architecture, nominal, fit, and optimization bounds for each variable.

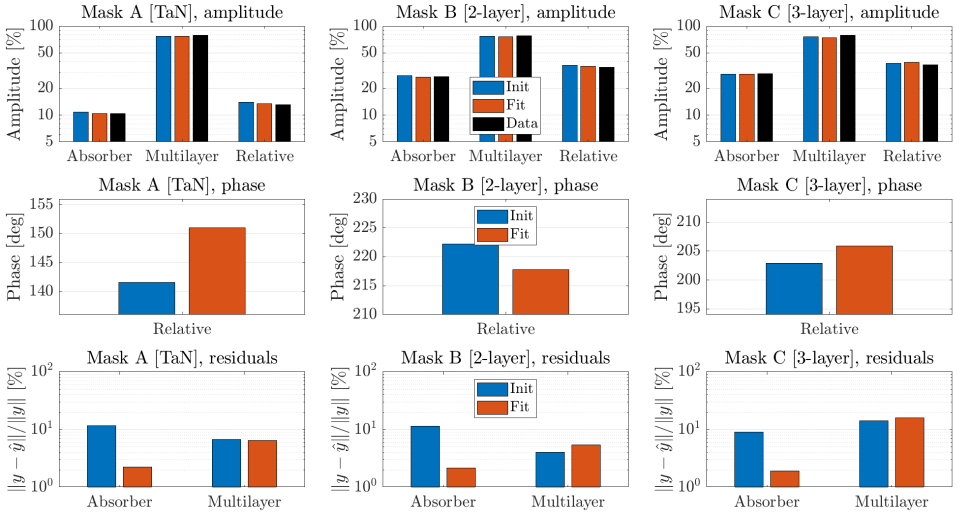

Figure 4.11: (a)-(c) show amplitude at $\lambda = 13.5$nm $\theta = 6^o$ for the 3 masks. Each subplot shows multilayer, absorber, and relative amplitude and compares the initial, fit, and measured values. (d)-(f) show the initial and fitted values for relative phase for the 3 masks; only relative phase is relevant to imaging so absolute phase of multilayer and absorber not shown. The true value is also omitted, because in these experiments it is unknown. (g)-(i) show the residuals for multilayer and absorber for the initial and fitted models. In all cases the absorber residuals decrease significantly after the fit, while multilayer residuals are constant or even slightly worse after the fit.

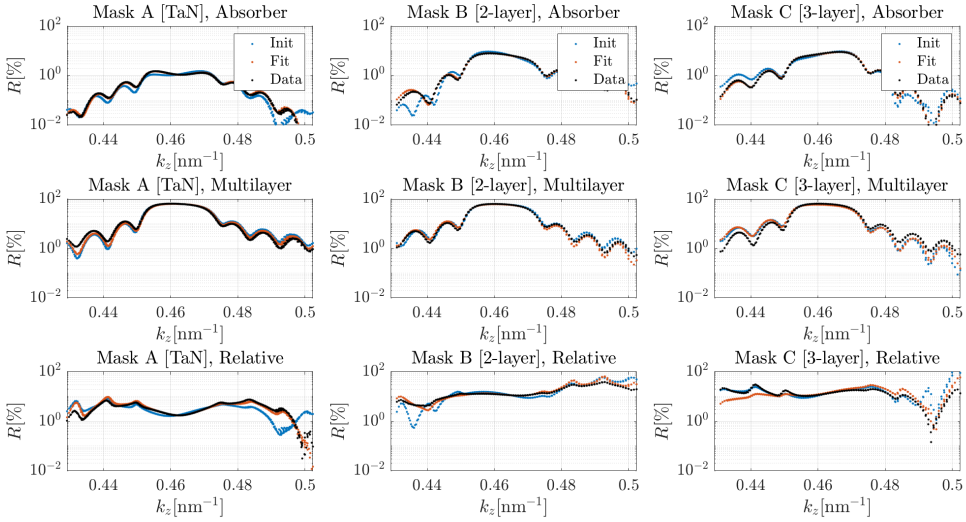

Figure 4.12: Initial, fitted, and measured reflectivity for the 3 masks; multilayer, absorber, and relative. (a)-(c) show absorber reflectivity vs k_z for the 3 masks. Note that in all cases and particularly for the two aPSMs the absorber reflectivity curve is agrees much better with the data after the fit. (d)-(f) show multilayer reflectivity vs k_z for the same cases. There is not a substantial change before and after the fit in any of the cases. (g)-(i) show the relative amplitude (absorber over multilayer), which defines the aPSM strength, i.e. the amplitude of the phase shifted absorber signal relative to the multilayer background. As with the absorber reflectivity, the fitted curves show clear improvement over the initial.

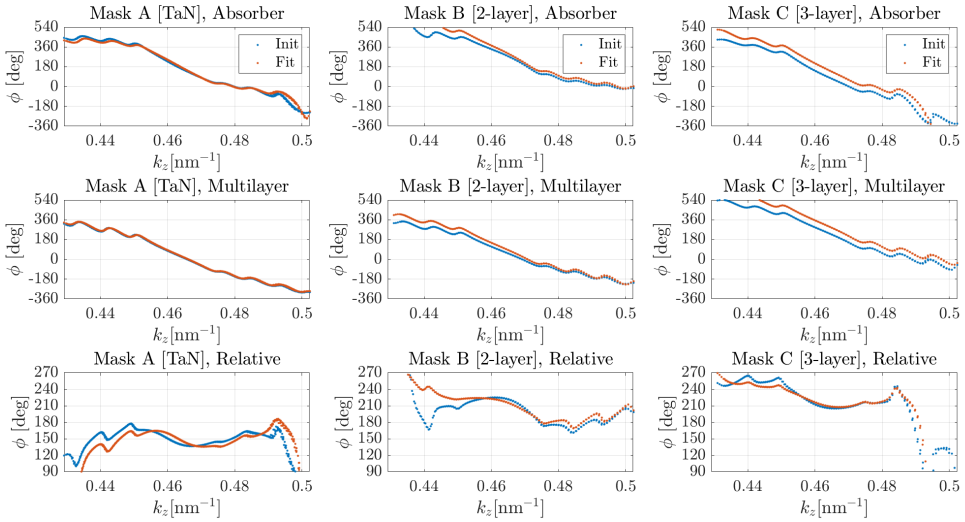

Figure 4.13: Initial and fitted phase for the 3 masks; multilayer, absorber, and relative. (a)-(c) show absorber phase vs k_z for the 3 masks. We see a constant phase shift and a change of slope (propagation distance) for the two aPSMs. (d)-(f) show multilayer phase vs k_z for the same cases. The change for the two aPSMs is visually quite similar, likely because a change in absorber thickness occurs, which changes the optical path length for both multilayer and absorber by the roughly same amount (determined by the real part of the abosrber's refractive index, which is within about 10% of unity). (g)-(i) show the relative phase (absorber minus multilayer), which defines the aPSM phase shift (for large-pitch features). Many changes in multilayer and absorber cancel out, but the relative phase still changes by $5 - 10^o$ at nominal k_z, due to a lateral shift in (g), a change in slope and curvature at low k_z in (h), and more subtle distortions in (i).

4.2 Scatterometry

In this section we explore EUV phase scatterometry, which is conducted on the same tool as reflectometry but with a scattering rather than flat sample. We scan the detector for each illumination condition to measure scattered light rather than only measuring the specular reflection. One potential intuitive strategy would be to exactly follow the approach of reflectometry and apply nonlinear least-squares using a rigorous solution to Maxwell's equations to max the measured intensity. But whereas a rigorous solution can be found very quickly with the transfer matrix method for reflectometry, in scatterometry such a rigorous solution requires a much more time-consuming model such as RCWA. There is no fundamental issue with this approach, but practically the long runtime of RCWA would require some combination of increasing the optimization time, reducing the number of parameters, or reducing the number of iterations. So because of these practical limitations to the most intuitive computational approach, instead we compare two approximate strategies presented in Section 4.2.2: first, nonlinear least-squares using the approximate DblSc model rather than RCWA; and second, linearizing the relationship between intensity and phase produced by RCWA in the neighborhood of the nominal 3D model. Generally we find in simulation that linearized RCWA is more accurate, but both algorithms have similar precision. We evaluate both algorithms in simulation at two levels of partial coherence in Section 4.2.3, representing a more coherent synchrotron source and a less coherent plasma source. The performance is similar for both, except that for larger features on the plasma source it is necessary to account for partial coherence when linearizing RCWA; if partial coherence is neglected during linearization the performance is almost identical for the synchrotron source and for smaller features on the plasma source, but the accuracy drops for larger features on the plasma source (while the precision is not particularly impacted). We also present the same algorithms applied to experimental scatterometry measurements of a TaN absorber mask in Section 4.2.5, where we observe a similar through-pitch trend and comparable levels of precision for all algorithms, but often significant differences in the trends of phase over time.

4.2.1 Measurements

Just as in reflectometry, measurements for scatterometry were carried out at the ALS Calibration, Optics Testing, and Spectroscopy Beamline 6.3.2 at CXRO, depicted in Figure 4.2. The only difference between the measurements are the sample and the positioning of the detector. The flat sample used in reflectometry is replaced with a scattering sample such as a grating, and instead of only measuring the specular reflection, the detector is scanned to also measure scattered or diffracted light.

Our scatterometry measurements consist of 41 wavelengths from 13-14 nm, illumination at 6^o, and detector angles from 2-12.5^o. Additional illumination angles were collected in the experiment, but in this analysis we mostly restrict ourselves to a single angle. Raw data for one illumination angle is shown in Figure 4.14 (f) for illustration. We extract diffraction

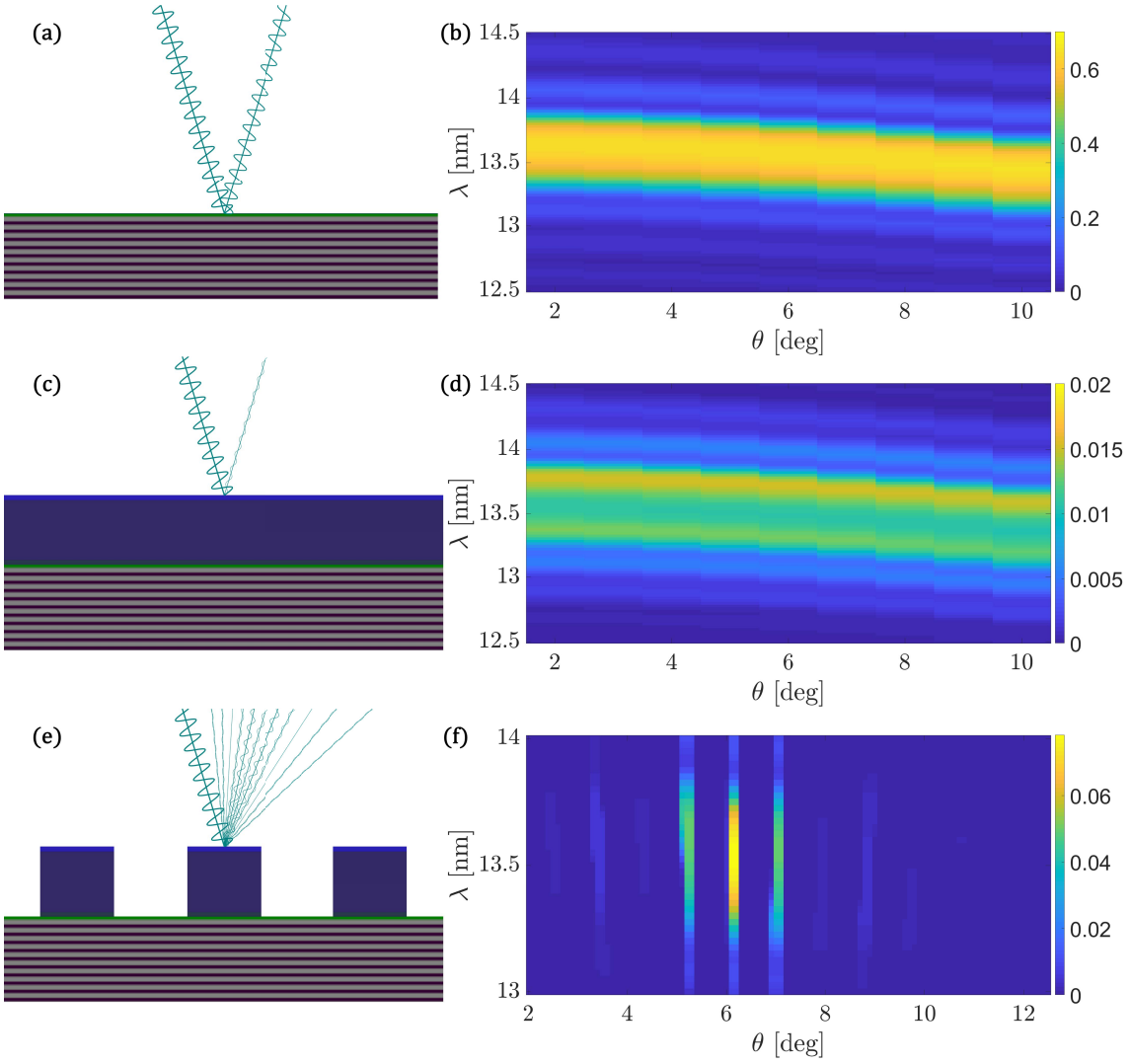

Figure 4.14: Sample schematics plus raw data for multilayer (a)-(b), absorber (c)-(d), and scatterometry (d)-(e).

efficiencies with a local maximum over a predefined range of angles corresponding to the angles of diffraction for the (known) pitch and illumination angle.

4.2.2 Phase scatterometry algorithms

Our task is to determine either the geometric structure that caused the observed scattering, or more specifically in our case to determine the phase shift between absorber and multilayer in the near-field at a defined reference plane, the top surface of the absorber. The most direct

approach to solving this might be to directly minimize the MSE between the measured signal and the output of a rigorous scattering solver such as RCWA. However, as discussed in Section 2, this approach suffers from very long runtimes, so we do not consider this to be a practical approach in this section. Instead, we compare two approximate approaches: first, using the Fresenel Double Scattering (DblSc) approximation and directly solving a nonlinear least squares (NLSQ) problem for the parameters describing a 3D model; and second, linearizing the relationship between measured intensity and near-field phase based on RCWA.

The two approaches to phase scatterometry are summarized in Figure 4.15. For NLSQ + DblSc Fig. 4.15 (1), we initialize the film model from the reflectometry fit in (a), which is itself NLSQ using Fresnel reflectivity as the forward model. Then we move to collecting a scatterometry measurement shown in (b), and use the reflectometry film model plus the nominal pitch and duty cycle to define a nominal 3D model for the scattering target. The diffraction efficiencies from the model are directly compared to the measured signal, and parameters are adjusted to minimize the MSE; the optimization algorithm is the same coordinate descent + golden section search used for reflectometry. Finally the phase is extracted from the updated 3D model using DblSc to generate the approximate near-field.

For linearized RCWA Fig. 4.15 (2), we again use the fitted film model from reflectometry for the vertical dimension, and the nominal pitch and duty cycle for lateral dimensions. But unlike (1), we now introduce an intermediate computational step before taking the scatterometry measurement shown in (b), where we run RCWA simulation for both the nominal 3D geometry as well as a set of perturbed geometries. We define the nominal scattering signal as y_0, and the nominal phase as ϕ_0. For the N randomly perturbed geometries we define a list of training pairs $[\delta y_i, \delta\phi_i]_{i=1}^{N}$ by taking the difference of the perturbed RCWA output from the nominal value. We transform the training dataset into a matrix $\delta Y = [\delta y_i]_i$ and a vector $\delta\Phi = [\delta\phi_i]_i$. We then solve for a set of weights w using regularized linear regression, defining two new quantities: a matrix Σ which is the noise covariance and a scalar α for numerical stability. We then solve for w using a regularized pseudo-inverse:

$$w = (\delta Y^* \delta Y + \Sigma + \alpha I)^{-1} \delta Y^* \delta\Phi \tag{4.1}$$

Finally in Fig. 4.15 2. (c) we take a scatterometry measurement, and we produce an estimate of the phase, $\hat{\phi}$, using the following formula:

$$\hat{\phi} = \phi_0 + w^*(y - y_0) \tag{4.2}$$

4.2.3 Partial coherence effects

In a practical metrology system, the light source may not be as coherent as the synchrotron used at Beamline 6.3.2. While any light source can be made arbitrarily coherent by filtering, this comes at the expense of throughput. Therefore it is important to determine the sensitivity of the method to partial coherence as well as noise statistics from readout and photon variance. We compare the performance of NLSQ + DblSc vs linearized scatterometry for

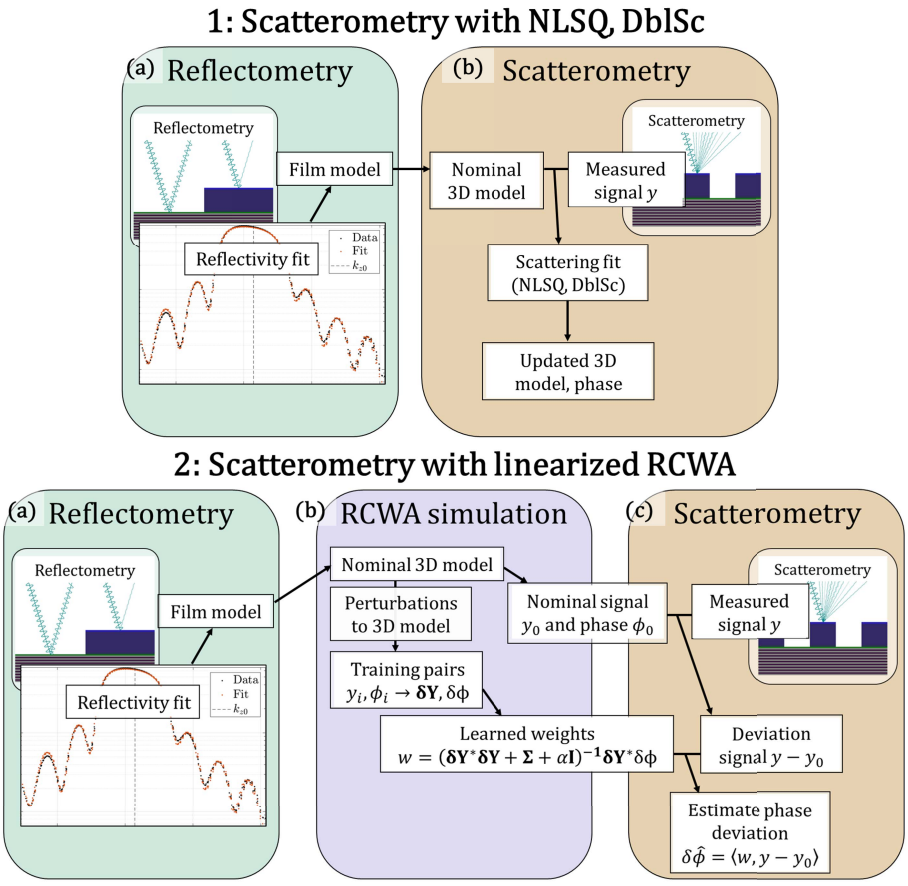

Figure 4.15: Flowcharts describing the two phase scatterometry algorithms considered. 1. NLSQ + DblSc (nonlinear least-squares with Fresnel Double Scattering). (a) Reflectometry fit to determine film model. (b) Scatterometry fit, initializing 3D model from fitted reflectometry model and the nominal pitch and duty cycle. Diffraction efficiencies from the model are approximated with DblSc and directly compared to the measured signal; parameters are adjusted to minimize the MSE; and finally the phase is extracted from the updated 3D model using DblSc. 2. Linearized RCWA. (a) Just as in 1, a fitted reflectometry model is required to initialize the 3D scattering model. (b) Unlike 1, we now have an intermediate step where we use RCWA to evaluate scattering from the nominal structure and from a set of perturbed structures to generate training pairs. The training data is used to learn a linear mapping from intensity to phase. (c) The measured scattering signal is fed into the learned function to reconstruct the phase.

two light sources: the sychrotron a Beamline 6.3.2 and a plasma source. For the synchrotron we assume $\sigma_\lambda = \lambda/400$, $\sigma_\theta = 0.25^o$[15], and for the plasma source we assume $\sigma_\lambda = \lambda/200$,

$\sigma_\theta = 0.7^o$. Examples for both levels of coherence at 3 feature sizes are shown in Figure 4.16.

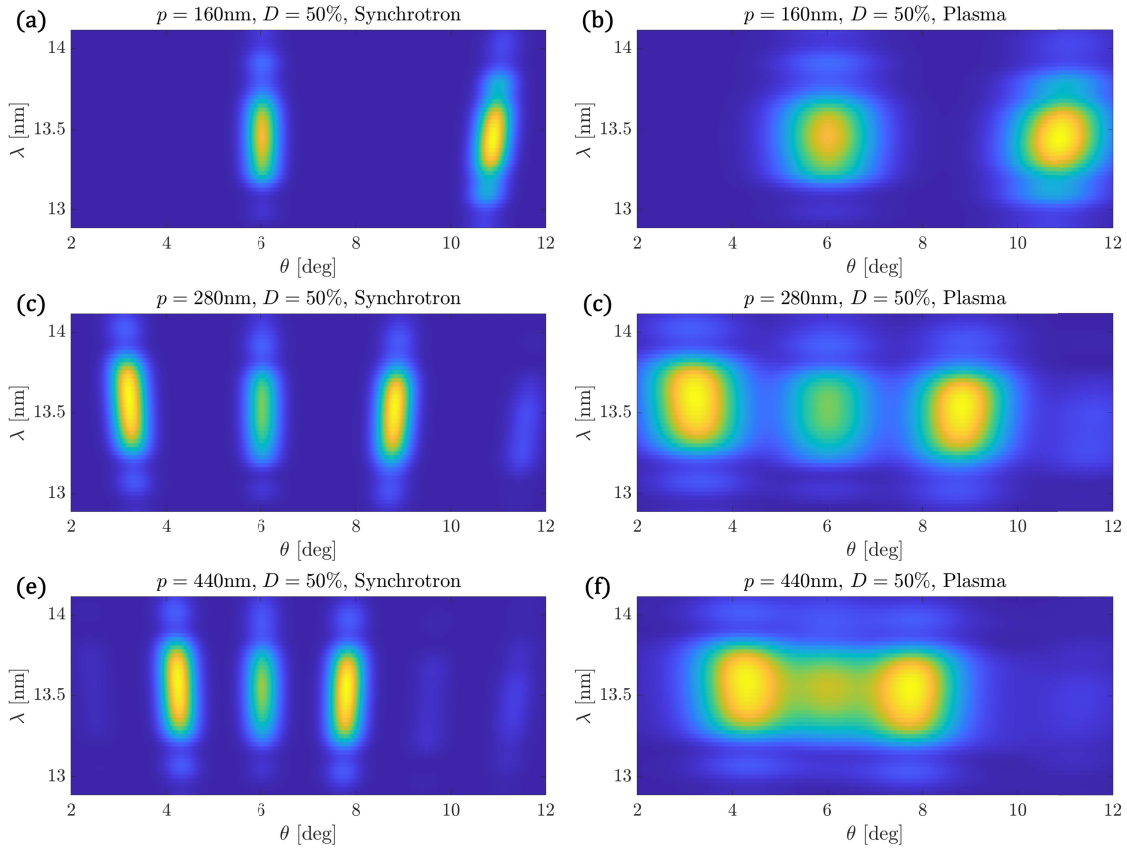

Figure 4.16: Examples of measuring 3 targets at 2 different levels of partial coherence, corresponding to a synchrotron source and a plasma source; x-axis is detector angle and y-axis is illumination wavelength. (a)-(b) show $p = 160$nm, $D = 50\%$ for the two sources. (c)-(d) show $p = 280$nm, $D = 50\%$ for the two sources. (e)-(f) show $p = 440$nm, $D = 50\%$ for the two sources. (a), (c), and (e) are simulated using a synchrotron source, $\sigma_\lambda = \lambda/400$ spectral bandwidth and $\sigma_\theta = 0.25^o$ angular bandwidth. (b), (d), and (f) are simulated using a plasma source, $\sigma_\lambda = \lambda/200$ spectral bandwidth and $\sigma_\theta = 0.7^o$ angular bandwidth.

4.2.4 Accuracy and precision

In this section we compare both the accuracy and precision of the two scatterometry algorithms (DblSc + NLSQ and linearized RCWA). Additionally, we consider two "flavors" of linearized RCWA, first using fully coherent training data, and then using partially coher-

ent training data. For both the synchrotron and plasma source we assume the same noise statistics, of 2×10^6 incident photons, 150 photons standard deviation readout noise.

For each case of 3 photomasks and 9 grating targets, we train linearized RCWA using 50 perturbations around the nominal design, drawing parameters from a normal distribution with standard deviation 10nm for mask width, 0.5nm for layer thickness outside of the multilayer, and 2.5% for material density for layers outside the multilayer; for layers inside of the multilayer the deviations in thickness and density are reduced by a factor of 40 to account for the greater sensitivity to these repeating layers. For all algorithms we use the same testing data, which consists of 10 randomly sampled geometries from the same distribution as the training set, with partial coherence applied based on the light source. Furthermore, for each scattering signal we generate 10 simulated measurements with realistic noise statistics (based on readout noise and photon counting), and test each algorithm on each signal in the training set.

Both accuracy and precision are determined using the training set. We compute the absolute accuracy ϵ_ϕ as the RMS error of the recovered phase minus the true phase. We compute the phase precision $3\sigma_\phi$ as 3 times the standard deviation of the phase output over the 10 unique noise instances.

Using the synchrotron source, Figure 4.17 (a) shows very similar accuracy for linearized RCWA whether or not coherence is considered in the training set, with linearized RCWA giving about a factor of 4 better accuracy than NLSQ + DblSc. The accuracy for a plasma source (b) is similar, except that for larger targets the accuracy gets much worse for linearized RCWA if coherence is not considered in training (because the diffraction orders of larger features are merging together in this less coherent source). In terms of precision with either source (c)-(d) all 3 algorithms give comparable performance, with NLSQ + DblSc often the most precise (but not significantly). This implies that DblSc is less accurate, but has similar or slightly better precision than linearized RCWA.

4.2.5 Experimental results

We apply these same phase scatterometry algorithms to measured data from 8 targets on the TaN mask, collected over the course of 4 different dates. In our analysis here we used only a single illumination angle of 6^o (although more illumination angles were collected in the experiment), 41 wavelengths from $13 - 14$nm, and ± 5 diffraction orders.

The recovered phase vs pitch in Figure 4.18 (a) shows fairly similar trends for all 3 algorithms, all fairly consistent with the nominal value based on RCWA + reflectometry. There is very little difference between the coherent and partially coherent linearized RCWA (which is expected because we are using a relatively coherent synchrotron source). All 3 algorithms track qualitatively with the nominal RCWA prediction, with the linearized RCWA generally slightly closer than NLSQ + DblSc; it is expected that linearized RCWA would be closer to the nominal RCWA, because they are both based on the same scattering model, and as shown in Figure 4.17 this leads to greater accuracy. There are two points ($p = 280, 440$nm) where the phase exhibits an anomalous dip for the linearized RCWA

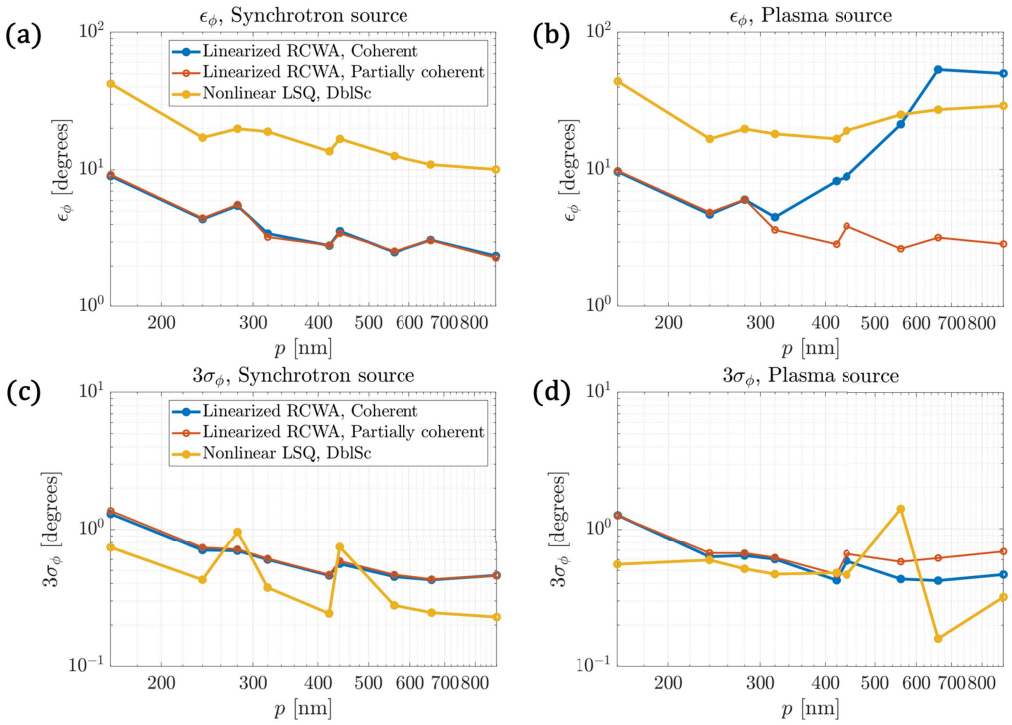

Figure 4.17: Scatterometry accuracy and precision for 3 algorithms and 2 light sources. (a) Using a synchrotron source, the simulations suggest very similar accuracy for linearized RCWA whether or not coherence is considered in the training set, with linearized RCWA giving about a factor of 4 better accuracy than NLSQ + DblSc. (b) Using a plasma source the results are similar, except that for larger targets the accuracy gets much worse for linearized RCWA if coherence is not considered in training, because the diffraction orders of larger features are merging together in this less coherent source. But in terms of precision (c)-(d) all 3 algorithms give comparable performance, with NLSQ + DblSc often the most precise (but not significantly). This implies that DblSc is less accurate, but has similar or slightly better precision than linearized RCWA.

algorithms; the same dip is observed in NLSQ + DblSc, at 280, but not at 440. In addition to the trend of phase vs pitch, a key metric is the precision of each method, in Figure 4.18 (b). The phase precision is computed as 3σ of phase values for the same target collected on the same day (we separate the days in this way to avoid biasing the precision estimate with actual physical changes on the mask). Similar precision is achieved for all 3 algorithms, with the NLSQ + DblSc method being slightly more precise than linearized RCWA, but not substantially. Both the findings of a larger (almost constant) offset from RCWA for DblSc, as well as slightly better precision, are consistent with the simulations summarized in Figure 4.17, which suggests that NLSQ + DblSc is substantially less accurate but slightly more

precise than linearized RCWA.

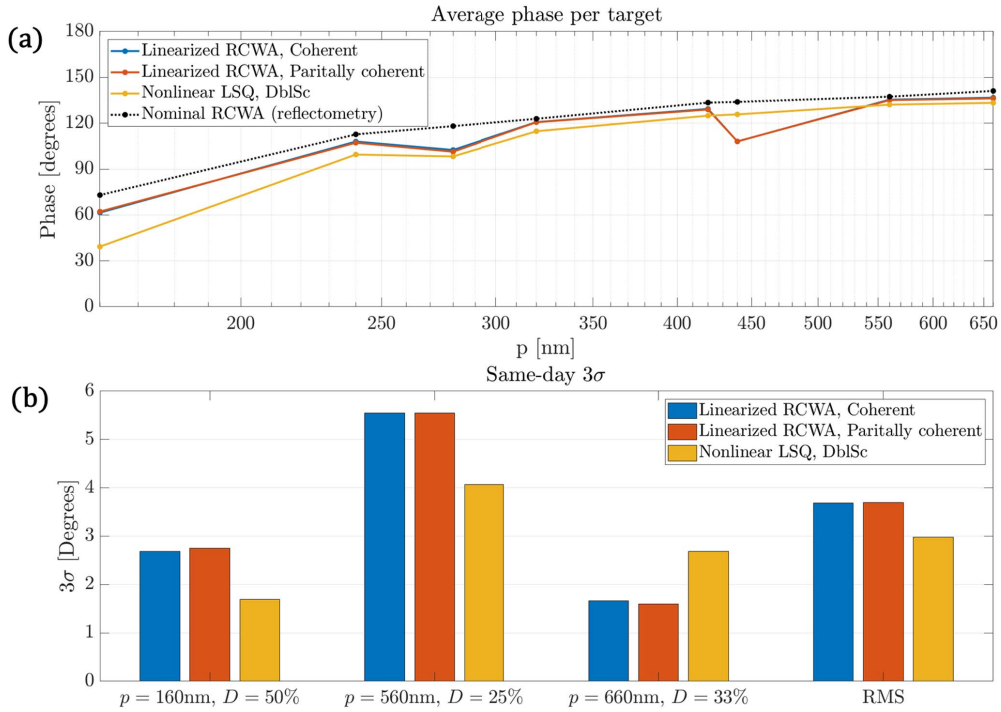

Figure 4.18: Scatterometry experiment. (a) Phase vs pitch for 3 scatterometry algorithms, plus the nominal value based on RCWA + reflectometry. All 3 algorithms track qualitatively with the nominal RCWA prediction, with the linearized RCWA generally slightly closer than NLSQ + DblSc. A similar anomaly is observed in all 3 algorithms at $p = 280$nm. The linearized RCWA algorithms both have a similar anomaly at $p = 440$nm, but this is not seen in NLSQ + DblSc. (b) Phase precision, computed as 3σ of phase values for the same target collected on the same day. Similar precision is achieved for all 3 algorithms, with the NLSQ + DblSc method being slightly more precise than linearized RCWA, but not substantially.

Aside from the absolute phase shift, a crucial task for a phase scatterometry would be to monitor phase changes over time (for instance due to contamination growing on the absorber or multilayer). We have 3 targets with multiple measurements; in Figure 4.19, we plot the recovered phase for each, after subtracting off the initial phase value. In Figure 4.19(a) we show the smallest feature, $p = 160$nm $D = 50\%$, which was measured 5 times over 3 sessions. Generally linearized RCWA predicts much larger phase deviations than NLSQ + DblSc, but the trends are qualitatively similar, just scaled. On the other hand Figure 4.19 (b) and (c) show two larger features, $p = 560$nm $D = 25\%$ and $p = 660$nm $D = 33\%$, which were each measured 10 times over 4 sessions. They both exhibit similar behavior, where the phase deviations are closer in magnitude than Figure 4.19(a), but still somewhat larger for

linearized RCWA. However, now we see that the trends are almost anti-correlated, meaning the deviations in NLSQ + DblSc are roughly the negative of linearized RCWA.

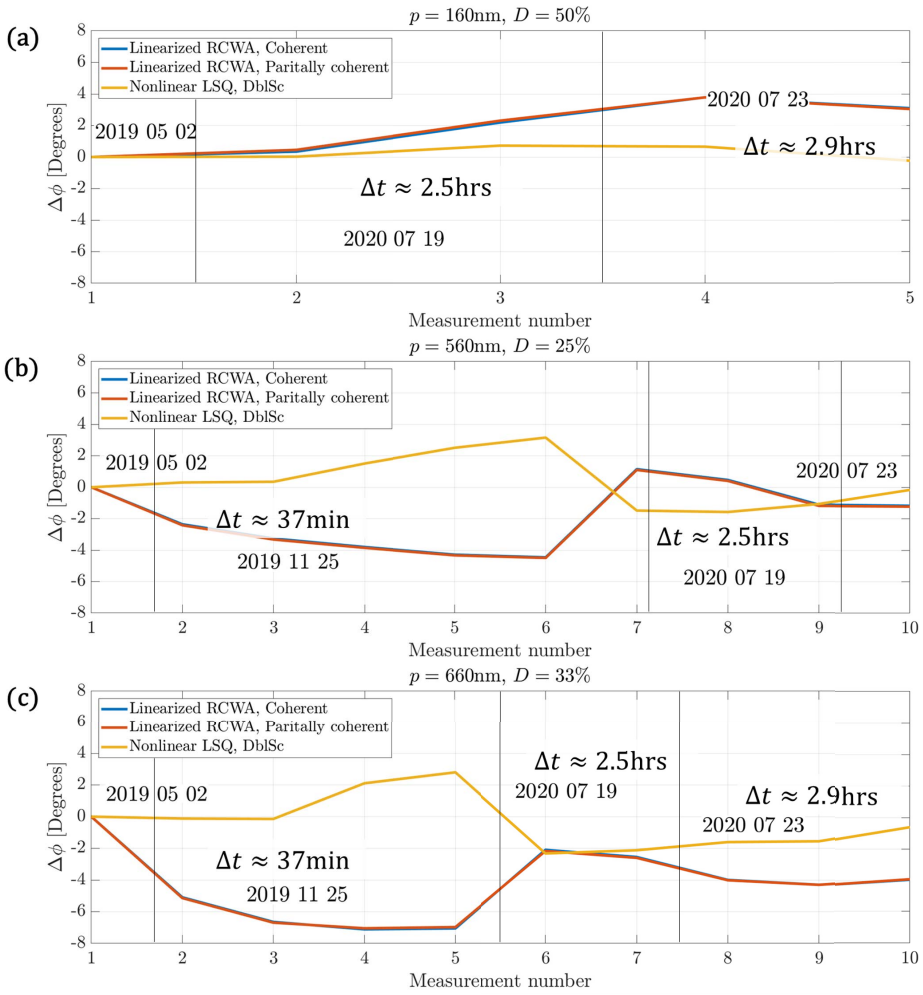

Figure 4.19: Scatterometry experiment, phase deviation from initial value for 3 features that were measured multiple times. (a) $p = 160$nm $D = 50\%$, measured 5 times over 3 sessions. Generally much larger phase deviations are predicted by linearized RCWA than by NLSQ + DblSc, but the trends are similar. (b) $p = 560$nm $D = 25\%$, measured 10 times over 4 sessions. Slightly larger phase deviations are predicted by linearized RCWA. The trend for NLSQ + DblSc shows almost the negative of the trend for linearized RCWA. (c) $p = 660$nm $D = 33\%$, measured 10 times over 4 sessions. Very similar to (b), where linearized RCWA has slightly larger phase deviations, and NLSQ + DblSc often has almost the negative trend of linearized RCWA.

4.3 Phase Imaging

In this section we explore computational phase imaging for EUV masks. Reflectometry can measure the relative phase of the Fresnel coefficients of absorber and multilayer; however, as demonstrated in Figure 2.9, 3D effects alter the phase as a function of pitch. Scatterometry can in principle be sensitive to feature-dependent phase effects, but hurdles with computation and modeling can make the results inconclusive or difficult to interpret. The key advantage of phase imaging is that it does not require any prior assumptions about the object. Whereas both reflectometry and scatterometry require detailed physical models of many parameters to be at least approximately known in advance, phase imaging can measure the scattered field (including amplitude and phase) even without prior knowledge of the physical structure being imaged.

In computational phase imaging one measures a set of images under multiple imaging conditions and uses these images to recover the amplitude and phase of the sample. In Section 4.3.1 we discuss the PhaseLift convex phase retrieval algorithm, which has the attractive mathematical property of convexity, meaning that any initial guess should lead to a global optimum. This is quite different from the more commonly used gradient based nonlinear least-squares algorithms which are nonconvex and therefore can fall into local minima. Although PhaseLift is often computationally intractable, our problem of measuring a discrete set of diffraction orders of a 1D grating is small enough to allow us to employ this algorithm. In Section 4.3.2 we present a novel hardware approach for phase imaging using a set of Zernike Phase Contrast (ZPC) zone plates, and an extension to hyperspectral imaging with hyperspectral ZPC (hZPC) which allows us to independently probe the complex response in image space and illumination space. In ZPC we image the same feature under the same illumination with multiple different zone plates. Each zone plate is encoded with a unique phase shift in a predefined region aligned with the illumination. This allows the 0 order forward scattered light to be phase shifted while the rest of the diffracted orders are imaged normally. We show how this ZPC method is more accurate than through-focus imaging because the quadratic defocus function is not sufficiently sensitive to the phase of low spatial frequencies (unless a very large focus range is collected). Because ZPC uses a discontinuous transfer function, the same level of phase diversity can be obtained for both low and high spatial frequencies, which greatly improves the accuracy. In Section 4.3.3 we describe the 6 custom fabricated ZPC zone plates used in ZPC and hZPC experiments. In Section 4.3.4 we show simulations quantifying calibration requirements for ZPC phase imaging. Finally in Section 4.3.5 we present hZPC reconstructions for horizontal and vertical lines on the 3 experimentally measured photomasks (a standard TaN absorber plus two aPSMs), where we see promising agreement with reflectometry in the trend of phase vs wavelength, particularly for vertical lines and spaces.

4.3.1 Convex phase retrieval

In this work, we adopt the PhaseLift optimization framework for phase retrieval as opposed to more commonly used nonlinear least-squares approaches[57]. PhaseLift is a convex formulation of phase retrieval[6], meaning that the optimization is guaranteed to converge to a global optimum regardless of the initial guess[5]. Figure 4.20 shows a convex and a non-convex function in 1D. The simplest description of a (twice differentiable) convex function is that its curvature is never negative; a strongly convex function has curvature that is always positive while a weakly convex function can have 0 curvature. For any convex function (even weakly convex), all local minima are global minima–however a weakly convex function can have multiple global minima (multiple points all with the same function value). A strongly convex function on the other hand can have only one local minimum, which is the global minimum. In either case any initialization can be guaranteed to converge to a global minimum, and in the case of a strongly convex function to the unique global minimum[5]. No such guarantee can be made for a nonconvex problem, where depending on the initialization it cannot be guaranteed whether an iterative minimization algorithm will converge to the correct local minimum.

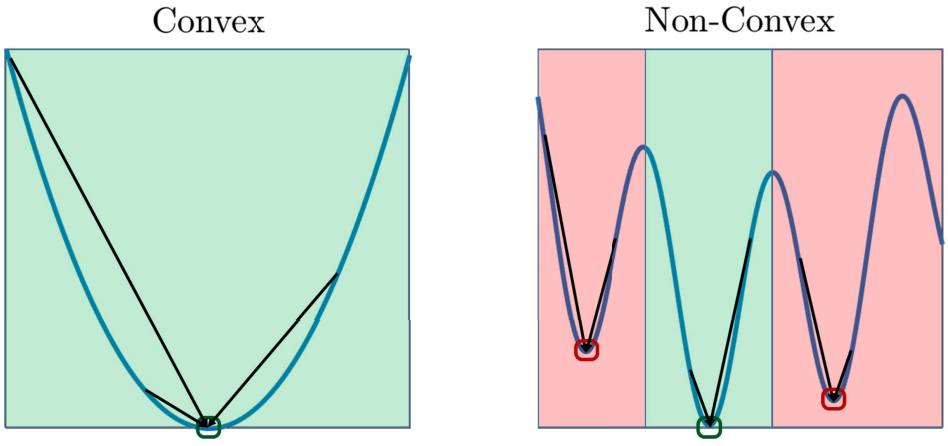

Figure 4.20: (a) Example of a convex function, which has only global minima, implying that any initialization will converge to a global minimum. This example of a quadratic function is strongly convex (meaning curvature is always strictly positive for a twice-differentiable function), in which case a stronger statement can be made that there is a unique global minimum, which can be reached by any initialization. (b) Example of a non-convex function, which has multiple local minima which are not all global minima. Some initializations will converge to the global minimum, but others will get stuck in one of the spurious local minima.

Unfortunately, the attractive mathematical property of convexity comes with a heavy computational cost for the phase retrieval problem. Indeed, the unknown variable must be

redefined from a vector $\mathrm{E} \in \mathbb{C}^n$ of diffraction orders to a Hermitian positive semi-definite (PSD) matrix $\mathbf{X} \in \mathbb{C}_+^{n\times n}$ representing the autocorrelation of diffracted waves, $\mathbf{X} = \mathrm{E}\mathrm{E}^*$, where $\mathbb{C}_+^{n\times n} \subset \mathbb{C}^{n\times n}$ is the set of $n \times n$ Hermitian PSD matrices. This lifting of the problem to a higher dimension allows one to solve a linear least-squares problem in the unknown matrix $\mathbf{X}$, as opposed to a nonlinear least-squares problem in the unknown vector E[6].
In the nonlinear least-squares approach we define our problem as[57]:

$$\min_{\mathrm{E}\in\mathbb{C}^n} \sum_{i=1}^{N} \left\| \sqrt{\mathrm{I}_i} - |\mathbf{F}^* \mathrm{diag}(\mathrm{h}_i)\mathrm{E}| \right\|_2^2 + \rho\left(E\right)$$

where E is the Fourier-transform of our unknown field, I_i is the measured intensity for image i; $\mathbf{F}$ is the matrix representation of the Fourier-transform; $\mathrm{diag}(\cdot)$ is a function that maps a vector to a diagonal matrix and vice-versa; h_i is the known pupil function for image i; and $\rho(\cdot)$ is a regularizer function to enforce prior knowledge about E. Although imaging is linear in electric field, it is nonlinear in amplitude due to the absolute-value. Therefore, reconstructing the field from intensity images requires solving nonlinear least-squares, and is generally non-convex. The primary advantage of lifting our problem to a higher dimension is that now a linear operator maps the unknown variable to the measured intensity: $\mathcal{L} : \mathbb{C}^{n\times n} \to \mathbb{R}^{m\times N}$, where m is the number of pixels per intensity image, and N is the number of intensity images. Mathematically, the mapping is:

$$\left(\mathrm{I}_i\right)_j = \mathrm{e}_j^* \mathbf{F}^* \mathrm{diag}\left[\mathrm{h}_i\right] \mathbf{X} \mathrm{diag}\left[\mathrm{h}_i^*\right] \mathbf{F} \mathrm{e}_j = \mathrm{a}_{i,j}^* \mathrm{vec}\left[\mathbf{X}\right]$$

where $\left(\mathrm{I}_i\right)_j$ is the measured intensity in image i at pixel j; e_j is the j^{th} standard basis vector; and $\mathrm{a}_{i,j}^* \in C^{1\times n^2}$ is the matrix representation of the functional that maps the (vectorized) matrix $\mathbf{X}$ to measurement (i, j). Note that $\mathrm{a}_{i,j}^*$ is one row of the matrix representation of $\mathcal{L}$. In this work, we explicitly construct the matrix representation of $\mathcal{L}$ by evaluating the linear operator on a basis for the set of $n \times n$ Hermitian matrices. This is often a computationally impractical approach for 2D images, but is not overwhelming in our case of a 1D image. Our optimization problem can now be written as:

$$\min_{\mathbf{X}\in\mathbb{C}_+^{n\times n}} \left\|\mathrm{vec}\left[\mathbf{I}\right] - \mathbf{L}\mathrm{vec}\left[\mathbf{X}\right]\right\|_2^2 + \rho\left(E\right)$$

One additional complication in PhaseLift is that, while the problem is convex, it is weakly convex because the matrix $\mathbf{L}$ is not full-rank. This means that, rather than a unique global minimum, there will be an entire subspace of global minima. However, we have a very strong piece of information about $\mathbf{X}$: namely that the true $\mathbf{X} = \mathrm{E}\mathrm{E}^*$ is a rank-1 matrix. It has been shown that by imposing a nuclear-norm penalty with sufficient phase diversity in the measurements, exact recovery of any low-rank matrix is guaranteed with overwhelming probability, even for an under determined linear system[6]. The nuclear-norm prior is a penalty of the form $\|\mathbf{X}\|_* = \|\mathrm{diag}[\mathbf{S}]\|_1 = \sum_i \left(\mathbf{S}\right)_{ii}$, where $\mathbf{U}\mathbf{S}\mathbf{V}^* = \mathbf{X}$ is the singular-value decomposition of $\mathbf{X}$. Note that in our special case of $\mathbf{X} \in \mathbb{C}_+^{n\times n}$ we may write

$\|\mathbf{X}\|_* = \text{trace}[\mathbf{X}]$ because the trace is the sum of the eigenvalues, which are equivalent to the singular values for a Hermitian PSD matrix[53]. Although non-differentiable, the nuclear-norm penalty is convex, and can be enforced with a proximal update consisting of soft-thresholding the singular values: $\mathbf{X} \mapsto \mathbf{U}(\text{softThresh}[\mathbf{S}, \alpha])\mathbf{V}^*$ [6], where $\text{softThresh}(x, \alpha) = (x/|x|)\max[0, |x| - \alpha]$ is the proximal operator for $\|\cdot\|_1$ [5].

4.3.2 Zernike phase contrast (ZPC)

One practical way to get phase diversity in the measurements is to vary the focus. This produces a roughly parabolic phase versus spatial frequency. Unfortunately, this means that very low spatial frequencies require increasingly large defocus range to get a fixed amount of phase diversity—the required range scales with the pitch squared, which quickly can become impractical. An alternative approach we explore is instead to use a discontinuous transfer function, where the phase suddenly changes around the 0 order, so that the 0 order receives a specified phase shift, and the rest of the orders are imaged normally. This allows all spatial frequencies to receive the same amount of phase diversity relative to the 0 order, unlike through-focus imaging.

The PhaseLift algorithm is convex, which in a formal mathematical sense guarantees that the algorithm will not get stuck in a local minimum. But unfortunately convexity by itself does not imply that the algorithm will converge to a unique true solution because the linear operator is generally rank-deficient. Mathematically, the problem is generally weakly convex, meaning there could be an infinite number of solutions that each exactly satisfy the measurements. While in theory the nuclear-norm prior can disambiguate many of these solutions, this does not hold true if there is not sufficient phase diversity in the measurements. Again, a key example of a hardware method that struggles to get sufficient phase diversity is through-focus imaging. Because of the quadratic relationship between phase and spatial frequencies, very low spatial frequencies require a very large focus range to sensitively detect a phase shift from the 0 order. This can make through-focus phase retrieval ill-suited for imaging samples that have significant power in low spatial frequencies, which is often the case for general samples.

This brings us to another hardware approach for phase imaging, which can achieve isotropic phase diversity for all spatial frequencies, Zernike Phase-Contrast (ZPC) imaging[58]. A Zernike Phase-Contrast (ZPC) image uses a phase-shifting region in the aperture aligned with the illumination, which produces a 0-order phase-shift relative to the rest of the image. In our case, we encode 6 different phase-shifts into a set of off-axis zone-plates that operate at $\lambda = 13.5$nm, with 0.55/4 NA and 6^o chief-ray angle. In total, our ZPC imaging data set consists of 90 ZPC images: 6 0-order phase-shifts in steps of 60^o, each with 15 focus steps of 300nm. The images are used to recover the amplitude and phase of each scattered wave in the far-field for our feature of interest. The advantage of this model-free approach is that it can equally well capture any M3D effects that may be present, without the need for an accurate prior model.

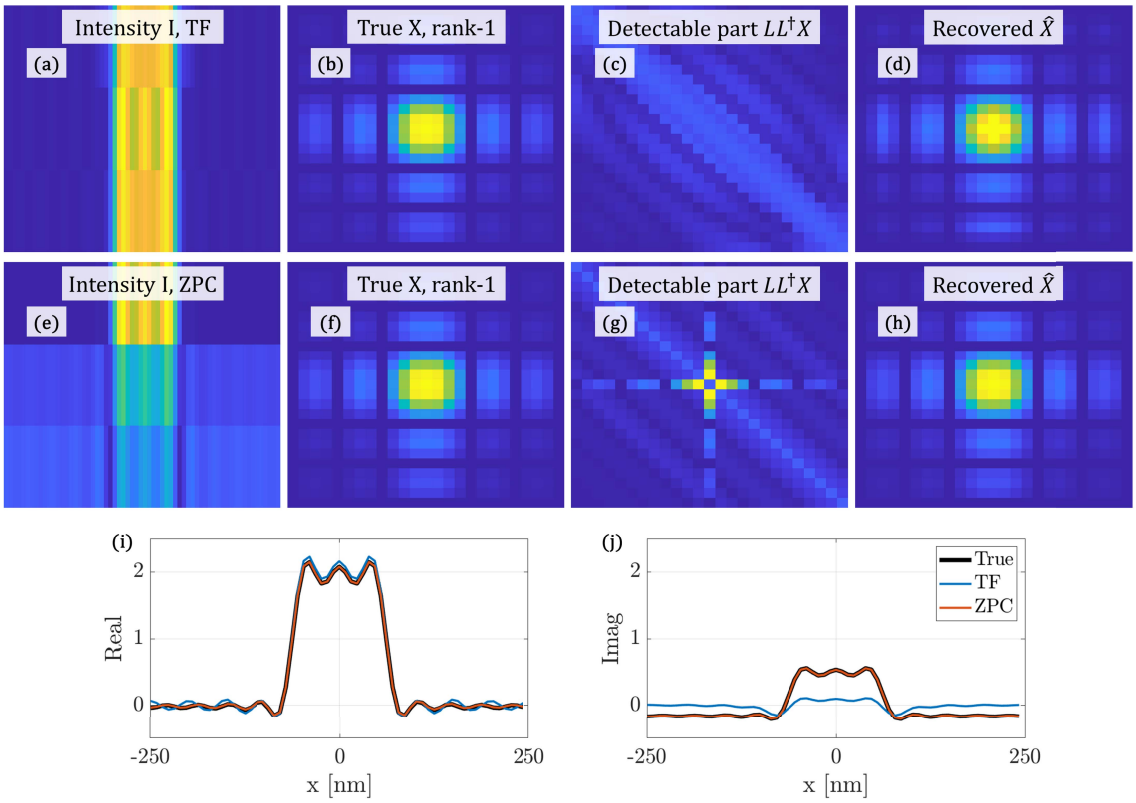

Figure 4.21: The through-focus approach suffers from errors recovering the low spatial frequencies, whereas ZPC can faithfully recover the entire complex field. (a) Intensity images for through-focus. (b) Unknown true autocorrelation $X = EE^*$. (c) Projecting onto the range of the imaging operator L. (d) Recovered $\hat{X}$ from PhaseLift for TF. (e) Intensity images for ZPC. (f) True X, same as (b). (g) Projection onto the range of the imaging linear operator L. (h) Recovered $\hat{X}$ from PhaseLift for ZPC. (i)-(j) Real and imaginary parts of reconstructed field for TF and ZPC. TF contains low-frequency errors in the imaginary part indicative of low-frequency phase errors.

4.3.3 Measurements

The key to our technique for measuring the amplitude and phase response under a given illumination condition is to measure multiple images taken under the same illumination conditions, but with different imaging transfer functions. To this end we have a set of 6 ZPC zone plates, which each has a 6^o chief-ray and 0.55/4 NA, but in the center of each zone plate there is a phase-shifted region with a unique phase shift. The phase shifts range from 0^o to 300^o in steps of 60^o. As explored in Section 4.3.4, in simulation an accurate reconstruction requires 2-4 phase shifts depending on the level of calibration, so we choose to measure 6

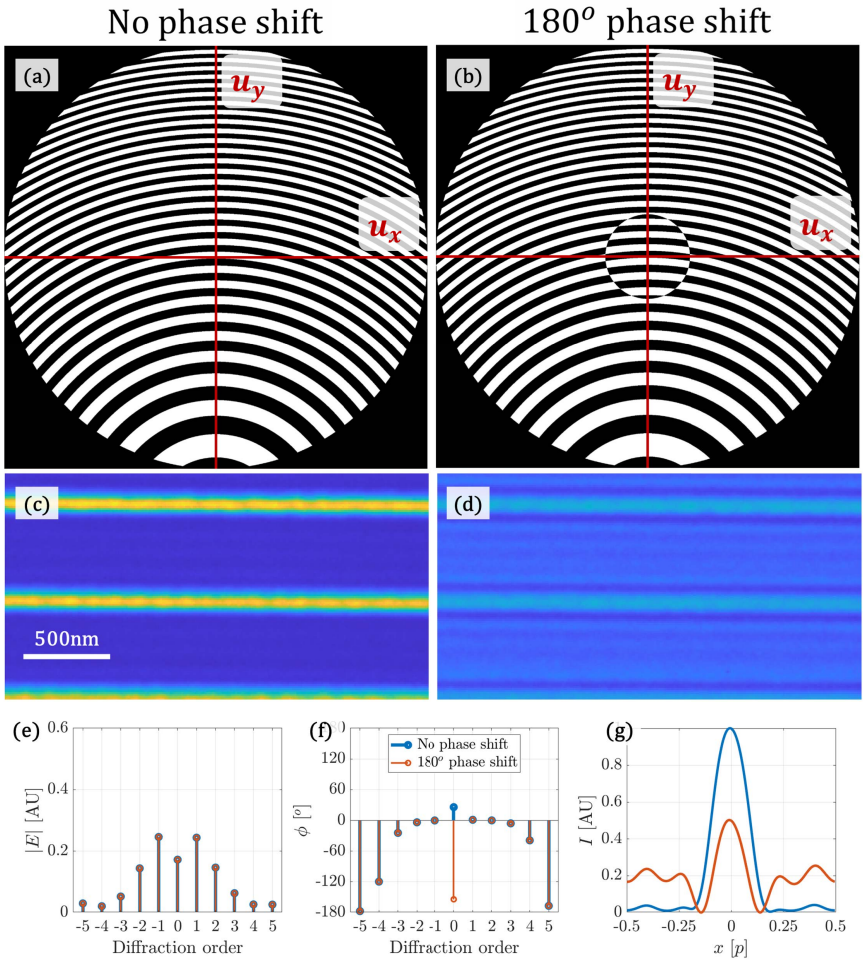

Figure 4.22: Schematic description of ZPC, with experimental data of no phase shift vs 180 degree phase shift on the 0 order. (a) Zone plate with no phase shift on 0 order. (b) Zone plate with 180^o phase shift on 0 order. (c) Experimental image, no phase shift. (d) Experimental image, 180^o phase shift. (e) Amplitude of far-field diffraction orders (same for both). (f) Phase of far-field diffraction orders are the same for both, except for the 0 order, which is shifted by 180^o. (g) Far-field intensity images, showing how a 180^o phase shift on the 0 order greatly reduces the intensity of the bright parts of the image and increases intensity of dark parts; this happens because the 180^o phase shift turns constructive interference between the 0 order and scattered light into destructive interference, and vice-versa.

phase shifts in an effort to accurately reconstruct the complex field even in the presence of realistic calibration errors such as alignment and focus. We illuminate with a plane-wave aligned with the chief-ray, such that the forward-scattered light (0 order diffraction) passes

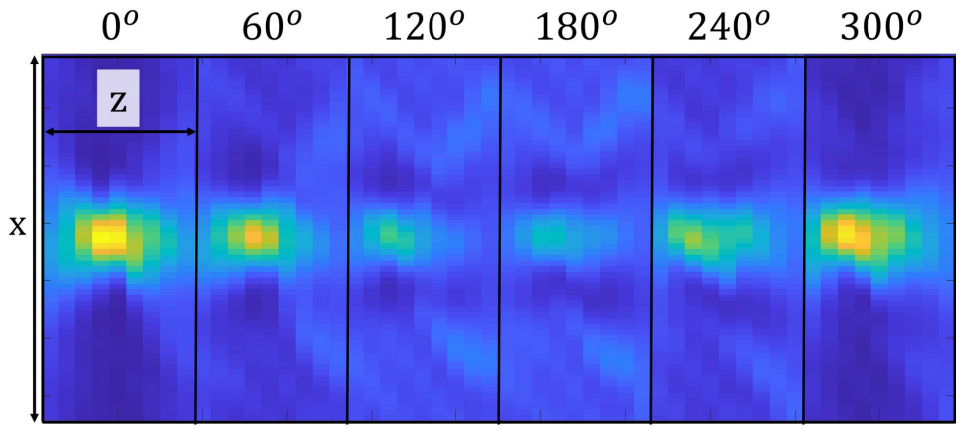

Figure 4.23: ZPC dataset for single illumination condition.

through the phase shift. On the other hand, light scattered away from the 0 order will pass through the outer part of the zone plate and be imaged with no phase shift. Figure 4.22 shows two examples of ZPC imaging with a 0^o and 180^o phase shift, clearly demonstrating how the 0 order phase shift greatly impacts the image contrast by creating either constructive or destructive interference of the scattered light with the background. In addition to the different phase shifts, we also acquire a focal stack for each zone plate, both to account for focus alignment and to increase the number of different transfer functions with which we image the mask. A full dataset with 6 phase shifts is shown in Figure 4.23.

Furthermore, by varying the wavelength of illumination, we can extend ZPC to hyperspectral ZPC (hZPC), which enables independently probing the phase response of a mask in image-space and illumination-space. Experimentally, the hZPC measurement is identical to the original ZPC measurement, except that in addition to scanning the 0 order phase shift and the focus, we also scan the illumination wavelength. An example hZPC dataset with 6 phase shifts and 7 wavelengths is shown in Figure 4.24. Note that we must offset the focus for each wavelength scan, because the focal length of a zone plate varies as a function of wavelength; remaining focus discrepancies for each wavelength and zone plate can then be removed in the alignment procedure. For both ZPC and hZPC, the feature being considered is a 560nm (mask pitch) 3:1 line-space in the horizontal (shadowing) orientation.

4.3.4 Calibration requirements

In this section we attempt to determine the measurement and calibration requirements for ZPC such as the number of phase shifts and focus steps, the photon dose per pixel, and random errors in lateral alignment per image, and in axial alignment per image and per zone plate. We define a nominal data collection scheme with 6 phase shifts and 11 focus steps evenly spaced by 500nm. We define 3 levels of calibration: “Ideal” with no errors in lateral

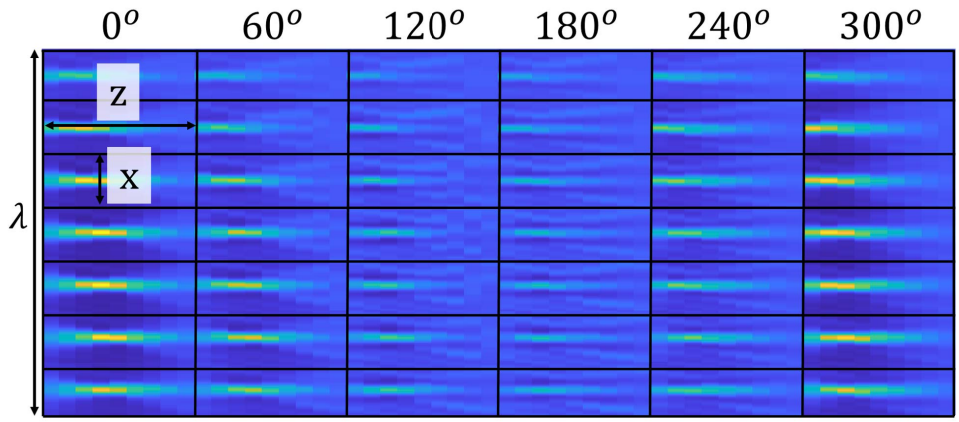

Figure 4.24: hZPC dataset: ZPC datasets for each of 7 different wavelengths of illumination.

or axial alignment and 1e4 photons per pixel; "Realistic" with 10nm lateral alignment error, 100nm focus error per image, 250nm focus error per zone plate, and 1e2 photons per pixel; and "Improved" with 2nm alignment error, 20nm focus error per image, 50nm focus error per zone plate, and 1e2 photons per pixel.

For each case (ideal, realistic, or improved calibration) we loop through the measurement and calibration parameters and experiment with varying one at a time, and run 10 realizations of the random process for each parameter value. At each realization we simulate ZPC images based on the measurements, calibration errors, and photon level, and then run a PhaseLift reconstruction to obtain the fitted electric field. We then compute 3 metrics for accuracy and precision: the absolute accuracy of the complex field ($\|\hat{E} - E\|/\|E\|$), the absolute phase accuracy ($|\hat{\phi} - \phi_{true}|$), and the phase precision ($3\sigma_\phi$). For accuracy metrics we take the RMS of the 10 realizations, while for precision we take the standard deviation.

Figure 4.25 shows this analysis for 3 parameters: The number of phase shifts N_ϕ, which are equally spaced by $2\pi/N_\phi$; the number of z-planes, N_z, on each side of 0, making the total number of focus steps $2N_z + 1$; and the average number of photons per pixel, $N_{photons}$. There is a large improvement in (a) moving from $N_\phi = 1$ to 2 phase shifts (zone plates) for all calibration levels, and a further improvement moving to 3 phase shifts for realistic calibration (but not ideal or improved). There are not substantial improvements for 4 or more phase shifts at any calibration level. This suggests that in a practical system, the number of phase shifts could be reduced from 6 down to just 2-3 without much loss of accuracy or precision. For the number of focus steps, with realistic calibration, there is constant improvement in (d) up to $N_z = 10$; but with ideal or improved calibration there is little improvement after 2. On the other hand in (e) and (f) all 3 calibration levels show improvements up to 10, which corresponds to 21 total z steps. 21 z steps may be too many for a practical system, but depending on the accuracy and precision required as few as 5 might be suitable (corresponding to about $3\sigma_\phi = 1^o$ for improved calibration). For the photon

dose, as low as $N_{photon} = 10$ photons per pixel may be sufficient. The absolute accuracy in field will not be much improved beyond 1-10 photons per pixel, as shown in (g); the phase accuracy and precision could be further improved with ideal calibration, but with improved calibration the further improvements to phase accuracy and precision would be marginal, and 10 photons per pixel is already sufficient to achieve $\leq 1^o$ accuracy and precision. All this suggests that the number of measurements as well as the photon dose can be substantially reduced without dramatically impacting the accuracy or precision. For example, one could employ 3 phase shifts, 5 focus steps, and 10 photons per pixel with improved calibration and still remain close to $3\sigma_\phi = 1^o$. A higher number of z-steps would probably provide the greatest improvement to phase accuracy, while a higher photon dose would likely provide the greatest improvement to phase precision.

Figure 4.26 shows the same analysis for 3 calibration parameters: the random focus error per zone plate $\sigma_{z,ZP}$; the random focus error per image σ_z; and the random alignment error per image σ_x. For the focus error per zone plate we see in (c) that errors will need to be kept under 62.5nm for ideal or improved calibration, or 31.25nm for realistic calibration, in order to keep $3\sigma_\phi \leq 1^o$. The same requirement of 62.5nm for $3\sigma_\phi \leq 1^o$ also holds for focus error per image in (f) in the cases of ideal or improved calibration; with realistic calibration there is no level low enough to achieve this, and there is minimal improvement below 250nm. For lateral alignment the requirement for $3\sigma_\phi \leq 1^o$ is roughly 4-8nm for ideal alignment or 2-4nm for improved alignment. There is not substantial improvement for realistic alignment below about 16nm. All this suggests that the calibration levels must be substantially improved from the current levels of approximately 250nm focus error per zone plate, 100nm focus error per image, and 10nm lateral alignment error per image. The focus error per zone plate should be brought down roughly to 62.5nm; a similar target of about 62.5nm should also be required for the focus error per image; and the lateral alignment should be reduced to below 4nm. With these levels at 10-100 photons per pixel, phase precision around 1^o should be attainable.

4.3.5 Experimental results, 3 masks

Reconstructed hyperspectral amplitude and phase for 3 masks are shown in Figures 4.27-4.29 in both the horizontal and vertical orientations. All 6 cases are for a line-space grating with $p = 560$nm, $D = 25\%$. The vertical orientation is laterally symmetric because the diffraction is perpendicular to the plane of incidence, but the horizontal orientation is asymmetric because diffraction in the plane of incidence breaks the symmetry.

Figure 4.30 shows the raw data for the 3 masks and 2 orientations measured. For each mask and orientation, there is a 1D line scan of a single grating period, a through-focus stack, a wavelength scan, and 6 zone plates with evenly-spaced 60^o phase shifts.

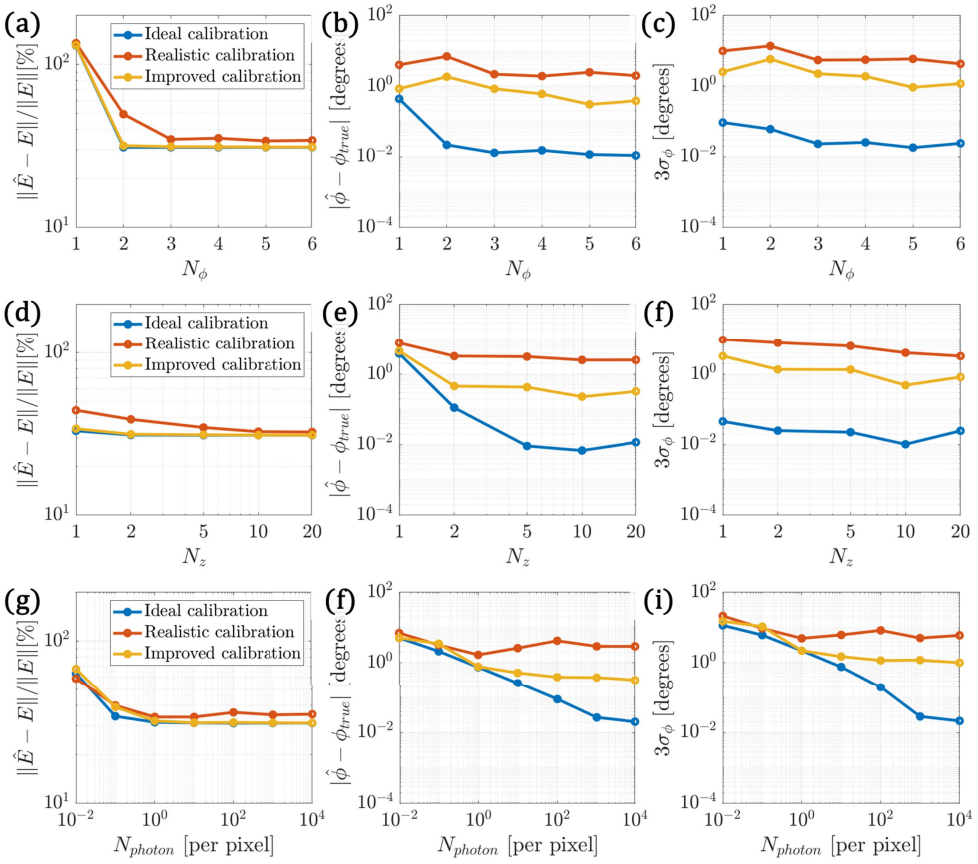

Figure 4.25: Sensitivity to measurement settings at 3 levels of calibration: ideal, realistic, and improved. Three error metrics are displayed on the y-axes: absolute accuracy of the complex field ($\|\hat{E} - E\|/\|E\|$), absolute phase accuracy ($|\hat{\phi} - \phi_{true}|$), and phase precision ($3\sigma_\phi$). (a)-(c) x-axis shows number of phase shifts N_ϕ (equally spaced by $2\pi/N_\phi$). With realistic calibration there is noticeable improvement in (a) up to 3 phase shifts; with ideal or improved calibration there is little improvement after 2 phase shifts. (d)-(f) x-axis shows number of z-planes N_z on each side of 0 (total number of focus steps is $2N_z + 1$). With realistic calibration, there is constant improvement in (d) up to $N_z = 10$, but with ideal or improved calibration there is little improvement after 2. On the other hand in (e) and (f) all 3 calibration levels show improvements up to 10. (g)-(i) x-axis shows number of photons $N_{photons}$ per pixel. In terms of absolute accuracy, (g) shows a leveling off above about 1 photon per pixel. On the other hand, the absolute accuracy in (h) and precision in (i) show continued improvement up to 10^4 photons per pixel.

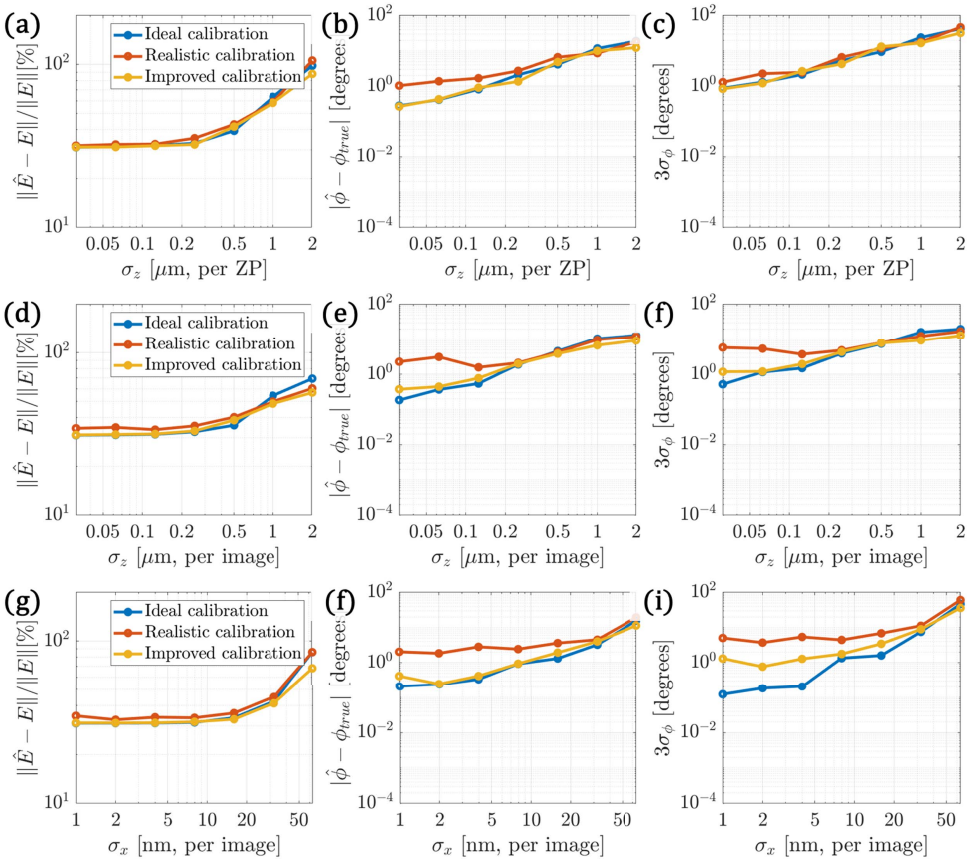

Figure 4.26: Sensitivity to calibration parameters at 3 levels of calibration: ideal, realistic, and improved. Three error metrics are displayed on the y-axes: absolute accuracy of the complex field ($\|\hat{E} - E\|/\|E\|$), absolute phase accuracy ($|\hat{\phi} - \phi_{true}|$), and phase precision ($3\sigma_\phi$). (a)-(c) x-axis shows the random z error in each zone plate, $\sigma_{z,ZP}$. The absolute accuracy is nearly constant below 250nm, but the phase precision at this level of calibration error is nearly 10^o; obtaining precision around 1^o would require the z error per zone plate be below 62.5nm. (d)-(f) x-axis shows the random z error in each image, σ_z. The results are very similar to (a)-(c), except the phase accuracy (e) and precision (f) for realistic calibration here stops decreasing below about 250nm; with improved or ideal calibration there is improvement down to at least 62.5nm. (g)-(i) x-axis shows the random x error in each image, σ_x. The complex field accuracy (g) stops improving below about 16nm. With realistic calibration there is little improvement in the phase accuracy (f) and precision (i) below 8nm. With ideal or improved calibration there is improvement down to at least 2nm.

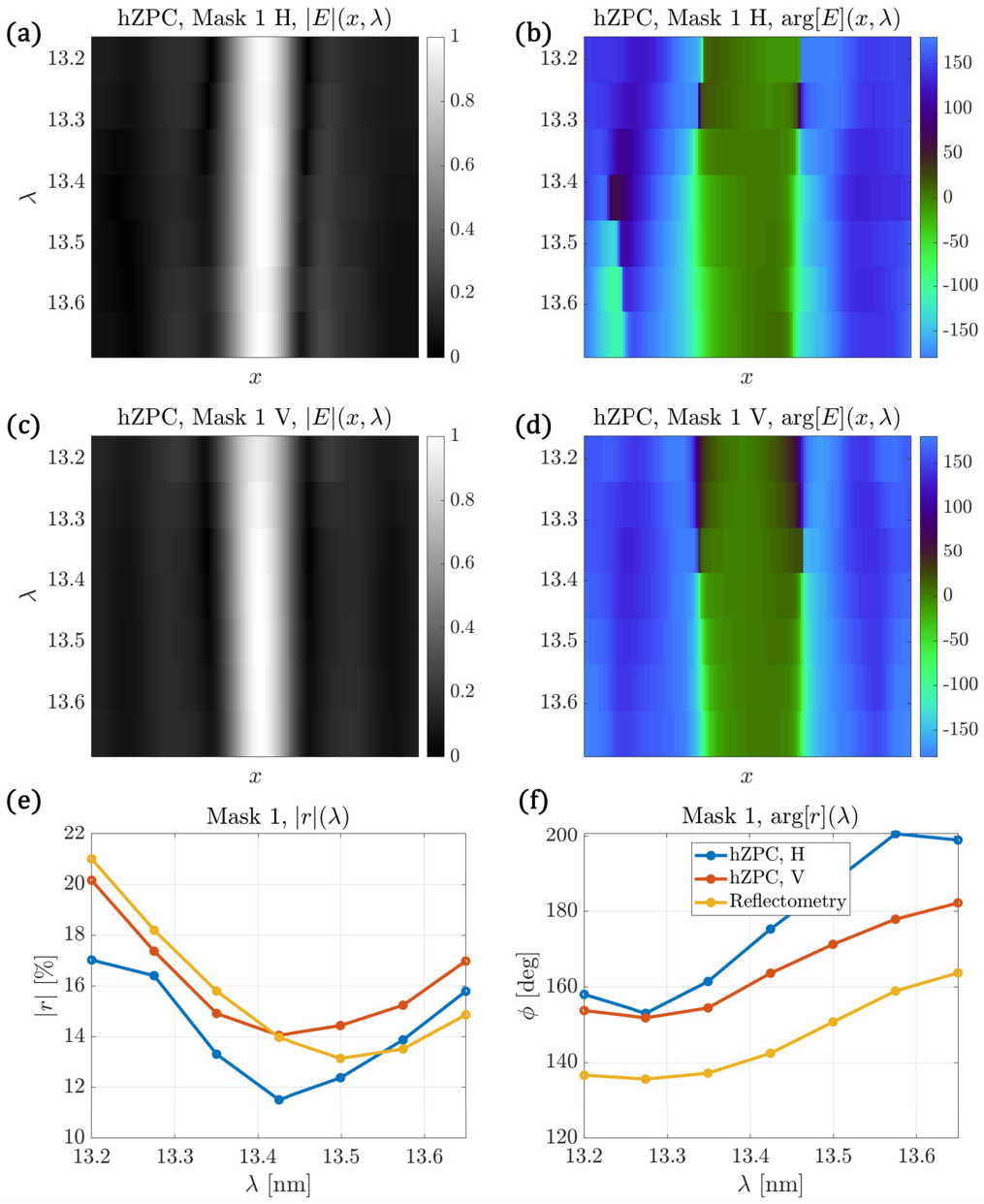

Figure 4.27: Reconstructed hyperspectral amplitude (a), (c) and phase (b), (d) for TaN absorber, in horizontal (a)-(b) and vertical (c)-(d) orientations. Relative amplitude (e) and phase (f) of average absorber field divided by average multilayer field.

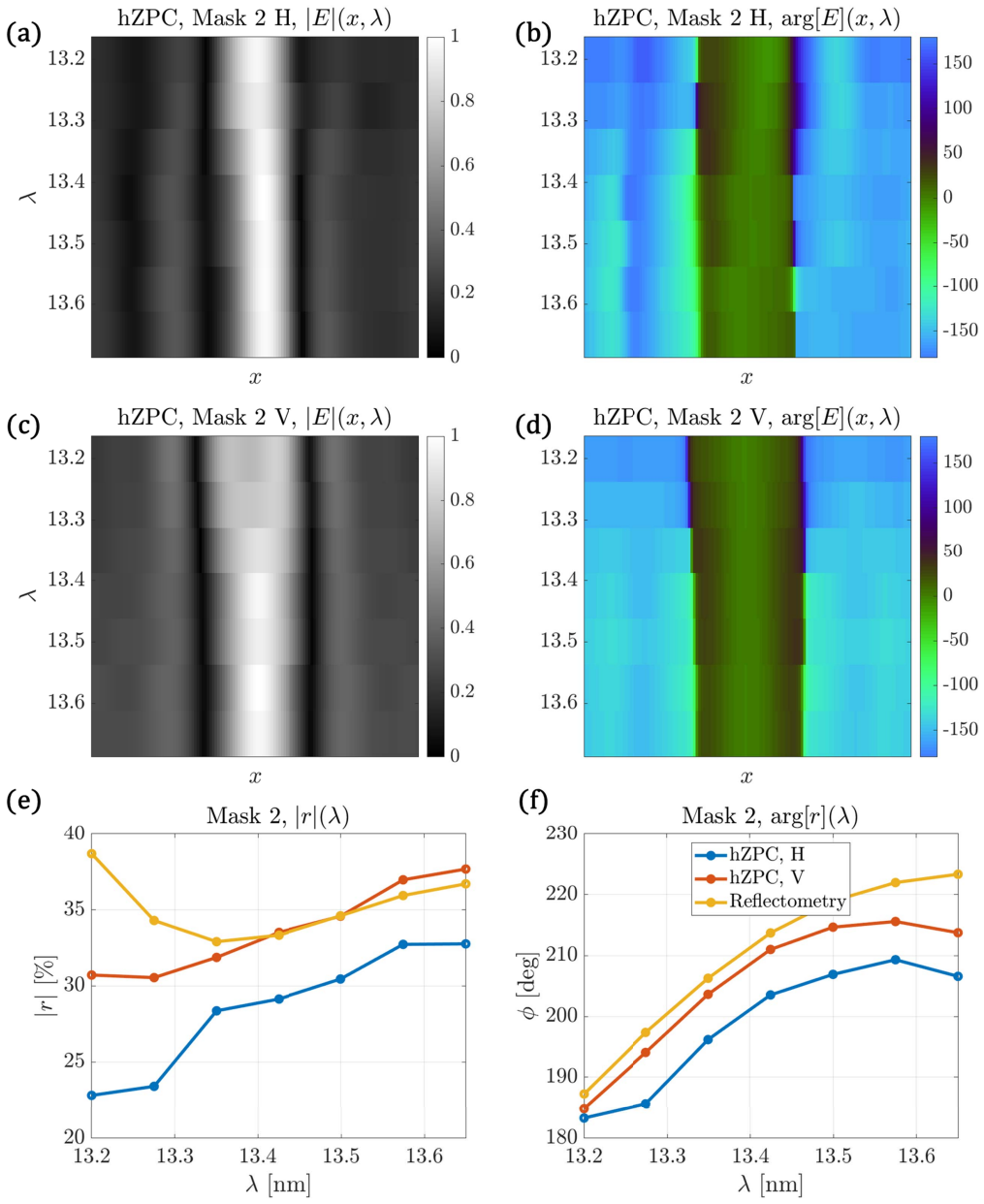

Figure 4.28: Reconstructed hyperspectral amplitude (a), (c) and phase (b), (d) for 2-layer aPSM absorber, in horizontal (a)-(b) and vertical (c)-(d) orientations. Relative amplitude (e) and phase (f) of average absorber field divided by average multilayer field.

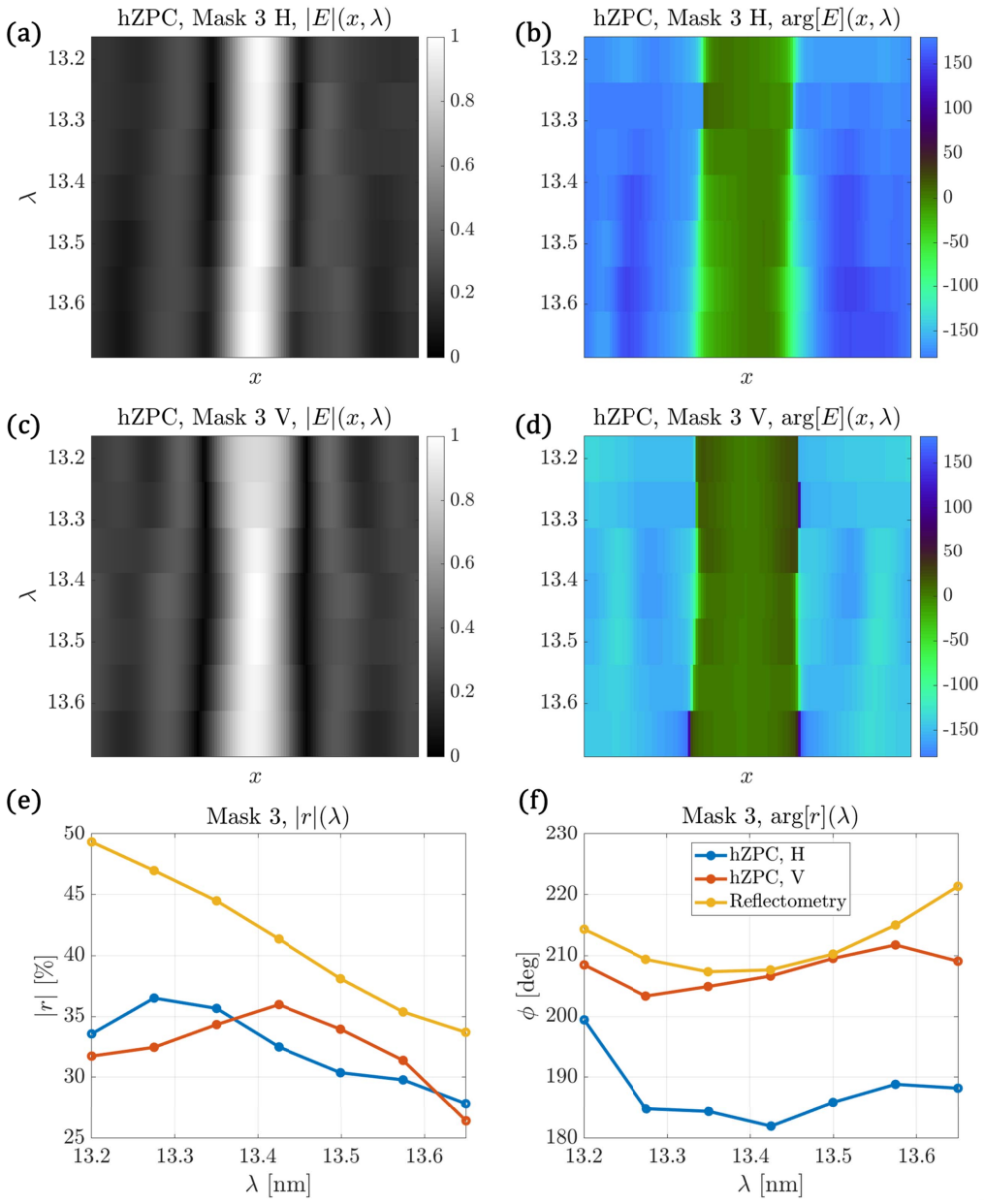

Figure 4.29: Reconstructed hyperspectral amplitude (a), (c) and phase (b), (d) for 3-layer aPSM absorber, in horizontal (a)-(b) and vertical (c)-(d) orientations. Relative amplitude (e) and phase (f) of average absorber field divided by average multilayer field.

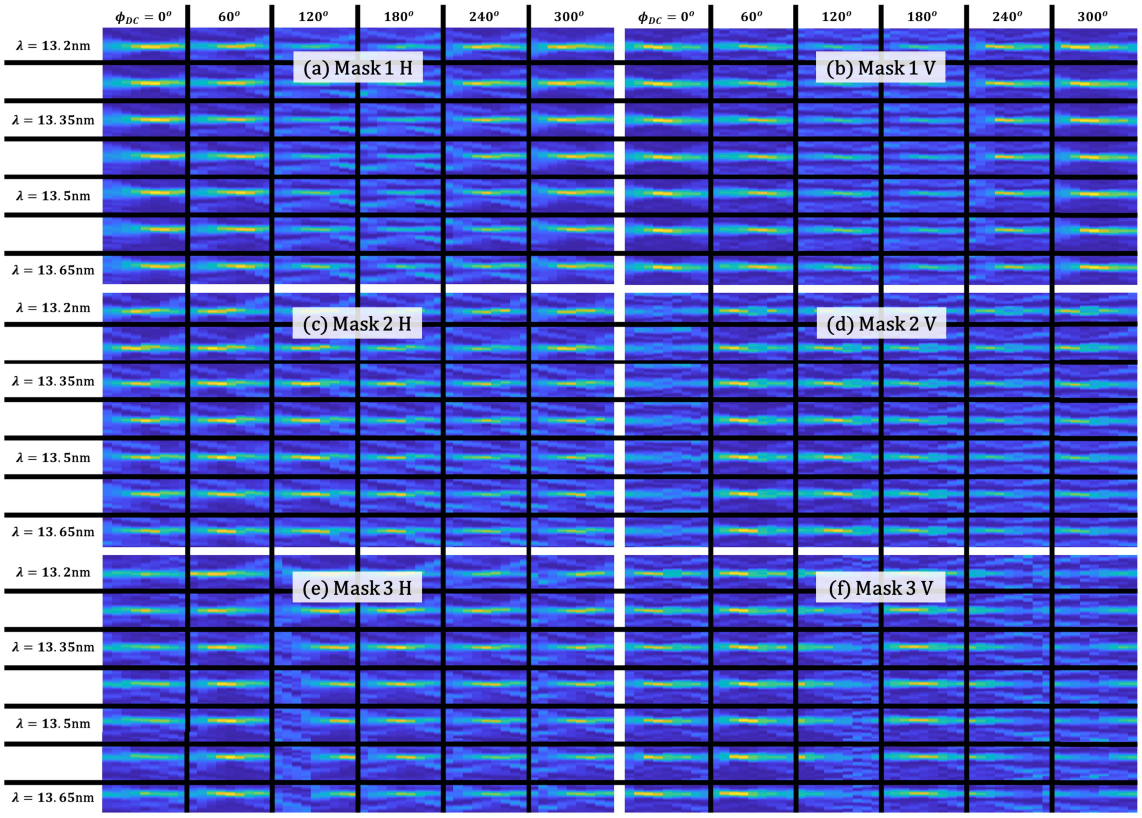

Figure 4.30: Raw data for 3 masks in horizontal and vertical orientations. Each case consists of a 4 dimensional scan: a 1D line scan of a single grating period, a through-focus stack, a wavelength scan, and 6 zone plates with evenly spaced 60^o phase shifts

Bibliography

[1] Konstantinos Adam and Andrew R Neureuther. "Domain Decomposition Methods for the Rapid Electromagnetic Simulation of Photomask Scattering". In: *Journal of Micro/Nanolithography, MEMS, and MOEMS* 1.3 (2002), pp. 253–269.

[2] RK Ahrenkiel. "Modified Kramers-Kronig Analysis of Optical Spectra". In: *JOSA* 61.12 (1971), pp. 1651–1655.

[3] AL Aquila et al. "Developments in Realistic Design for Aperiodic Mo/Si Multilayer Mirrors". In: *Optics express* 14.21 (2006), pp. 10073–10078.

[4] David Attwood and Anne Sakdinawat. *X-Rays and Extreme Ultraviolet Radiation: Principles and Applications*. Cambridge university press, 2017.

[5] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[6] Emmanuel J Candes, Thomas Strohmer, and Vladislav Voroninski. "Phaselift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming". In: *Communications on Pure and Applied Mathematics* 66.8 (2013), pp. 1241–1274.

[7] Yen-Ching Chang. "N-Dimension Golden Section Search: Its Variants and Limitations". In: *2009 2nd International Conference on Biomedical Engineering and Informatics*. IEEE. 2009, pp. 1–6.

[8] Chris Heinz Clifford. "Simulation and Compensation Methods for EUV Lithography Masks with Buried Defects". PhD thesis. UC Berkeley, 2010.

[9] Nick Cobb. "Sum of coherent systems decomposition by SVD". In: *Berkeley CA* 94720 (1995), pp. 1–7.

[10] Andreas Erdmann et al. "Topography Effects and Wave Aberrations in Advanced PSM-Technology". In: *Proc. SPIE*. Vol. 4346. 2001, pp. 345–355.

[11] Andreas Erdmann et al. "Attenuated PSM for EUV: Can they Mitigate 3D Mask Effects?" In: *Extreme Ultraviolet (EUV) Lithography IX*. Vol. 10583. International Society for Optics and Photonics. 2018, p. 1058312.

[12] Alain Gibaud and Guillaume Vignaud. "Specular Reflectivity from Smooth and Rough Surfaces". In: *X-ray and Neutron Reflectivity*. Springer, 2009, pp. 85–131.

[13] Bob Gleason and Wen-Hao Cheng. "Optical Properties of Alternating Phase-Shifting Masks". In: *26th Annual BACUS Symposium on Photomask Technology*. International Society for Optics and Photonics. 2006, 63491B–63491B.

[14] Joseph W Goodman. *Introduction to Fourier Optics*. Roberts and Company Publishers, 2005.

[15] Eric M Gullikson, Stanley Mrowka, and Benjamin B Kaufmann. "Recent Developments in EUV Reflectometry at the Advanced Light Source". In: *Emerging Lithographic Technologies V*. Vol. 4343. International Society for Optics and Photonics. 2001, pp. 363–373.

[16] Sang-In Han et al. "Design and Method of Fabricating Phase-Shift Masks for Extreme-Ultraviolet Lithography by Partial Etching into the EUV Multilayer Mirror". In: *Microlithography 2003*. International Society for Optics and Photonics. 2003, pp. 314–330.

[17] E. Hecht. *Optics*. Pearson education. Addison-Wesley, 2002. ISBN: 9780321188786.

[18] Burton L Henke, Eric M Gullikson, and John C Davis. "X-Ray Interactions: Photoabsorption, Scattering, Transmission, and Reflection at E= 50-30,000eV, Z= 1-92". In: *Atomic data and nuclear data tables* 54.2 (1993), pp. 181–342.

[19] Harold Horace Hopkins. "On the diffraction theory of optical images". In: *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 217.1130 (1953), pp. 408–432.

[20] Koen van Ingen Schenau et al. "Imaging Performance of the EUV High NA Anamorphic System". In: *31st European Mask and Lithography Conference*. International Society for Optics and Photonics. 2015, 96610S–96610S.

[21] Rik Jonckheere et al. "Study of EUV Reticle Storage Effects through Exposure on EBL2 and NXE". In: *Extreme Ultraviolet Lithography 2020*. Vol. 11517. International Society for Optics and Photonics. 2020, 115170Z.

[22] Atsuhiko Kato. *Chronology of Lithography Milestones*. 2007. URL: `http://www.lithoguru.com/scientist/litho_history/Kato_Litho_History.pdf`.

[23] Charalambos C Katsidis and Dimitrios I Siapkas. "General Transfer-Matrix Method for Optical Multilayer Systems with Coherent, Partially Coherent, and Incoherent Interference". In: *Applied optics* 41.19 (2002), pp. 3978–3987.

[24] Jack Kiefer. "Sequential Minimax Search for a Maximum". In: *Proceedings of the American mathematical society* 4.3 (1953), pp. 502–506.

[25] Thomas Kuhlmann et al. "Design and Fabrication of Broadband EUV Multilayer Mirrors". In: *Proc. SPIE*. Vol. 4688. 2002, pp. 509–515.

[26] M-Claire van Lare, Frank J Timmermans, and Jo Finders. "Alternative Reticles for Low-k1 EUV Imaging". In: *International Conference on Extreme Ultraviolet Lithography 2019*. Vol. 11147. International Society for Optics and Photonics. 2019, p. 111470D.

[27] Harry J Levinson. *Principles of Lithography*. Vol. 146. SPIE press, 2005.

[28] Burn J Lin. "Optical lithography with and without NGL for single-digit nanometer nodes". In: *Optical Microlithography XXVIII*. Vol. 9426. International Society for Optics and Photonics. 2015, p. 942602.

[29] Chris Mack. *Fundamental Principles of Optical Lithography: The Science of Microfabrication*. John Wiley & Sons, 2008.

[30] Greg McIntyre et al. "Enhancing Resolution with Pupil Filtering for Projection Printing Systems with Fixed or Restricted illumination angular distribution". In: *Proc. SPIE*. Vol. 8679. 2013, 86792N.

[31] Paul B Mirkarimi, Sasa Bajt, and Mark A Wall. "Mo/Si and Mo/Be Multilayer Thin Films on Zerodur Substrates for Extreme-Ultraviolet Lithography". In: *Applied optics* 39.10 (2000), pp. 1617–1625.

[32] Gordon E Moore et al. *Cramming More Components onto Integrated Circuits*. 1965.

[33] Jan Mulkens et al. "Overlay and Edge Placement Control Strategies for the 7nm Node using EUV and ArF Lithography". In: *Extreme Ultraviolet (EUV) Lithography VI*. Vol. 9422. International Society for Optics and Photonics. 2015, 94221Q.

[34] Patrick Naulleau et al. "Ultrahigh Efficiency EUV Contact-Hole Printing with Chromeless Phase Shift Mask". In: *Photomask Japan 2016*. International Society for Optics and Photonics. 2016, 99840P–99840P.

[35] Patrick P Naulleau et al. "Fabrication of High-efficiency Multilayer-Coated Gratings for the EUV Regime Using E-Beam Patterned substrates". In: *Optics communications* 229.1 (2004), pp. 109–116.

[36] Jens Timo Neumann et al. "Imaging Performance of EUV Lithography Optics Configuration for Sub 9nm Resolution". In: *Proc. SPIE*. Vol. 9422. 2015, 94221H.

[37] Jens Timo Neumann et al. "Interactions of 3D Mask Effects and NA in EUV Lithography". In: *Proc. SPIE*. Vol. 8522. 2012, p. 852211.

[38] Vicky Philipsen et al. "Reducing EUV Mask 3D Effects by Alternative Metal Absorbers". In: *Extreme Ultraviolet (EUV) Lithography VIII*. Vol. 10143. International Society for Optics and Photonics. 2017, p. 1014310.

[39] Alberto Pirati et al. "EUV Lithography Performance for Manufacturing: Status and Outlook". In: *Proc. SPIE*. Vol. 9776. 2016, 97760A.

[40] William H Press et al. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge university press, 2007.

[41] Stuart Sherwin, Andrew Neureuther, and Patrick Naulleau. "Actinic EUV Scatterometry for Parametric Mask Quantification". In: *Extreme Ultraviolet (EUV) Lithography IX*. Vol. 10583. International Society for Optics and Photonics. 2018, 105831G.

[42] Stuart Sherwin, Andrew Neureuther, and Patrick Naulleau. "Modeling High-Efficiency Extreme Ultraviolet Etched Multilayer Phase-Shift Masks". In: *Journal of Micro/Nanolithography, MEMS, and MOEMS* 16.4 (2017), p. 041012.

[43] Stuart Sherwin et al. "Advanced Multilayer Mirror Design to Mitigate EUV Shadowing". In: *Extreme Ultraviolet (EUV) Lithography X*. Vol. 10957. International Society for Optics and Photonics. 2019, p. 1095715.

[44] Stuart Sherwin et al. "Measuring the Phase of EUV Photomasks". In: *International Conference on Extreme Ultraviolet Lithography 2019*. Vol. 11147. International Society for Optics and Photonics. 2019, 111471F.

[45] Stuart Sherwin et al. "Picometer Sensitivity Metrology for EUV Absorber Phase". In: *Journal of Micro/Nanopatterning, Materials, and Metrology* 20.3 (2021), p. 031011.

[46] Stuart Sherwin et al. "Quantitative Phase Retrieval for EUV Photomasks". In: *Extreme Ultraviolet (EUV) Lithography XI*. Vol. 11323. International Society for Optics and Photonics. 2020, p. 1132315.

[47] SherJang Singh et al. "A Study on EUV Reticle Surface Molecular Contamination under Different Storage Conditions in a HVM Foundry Fab". In: *Extreme Ultraviolet (EUV) Lithography VIII*. Vol. 10143. International Society for Optics and Photonics. 2017, 101431T.

[48] Bruce W Smith. *Optics for Photolithography*. Marcel Dekker, 1998.

[49] FJ Timmermans et al. "Alternative Absorber Materials for Mitigation of Mask 3D Effects in High NA EUV Lithography". In: *34th European Mask and Lithography Conference*. Vol. 10775. International Society for Optics and Photonics. 2018, 107750U.

[50] Frank J Timmermans et al. "EUV Phase Shift Mask Requirements for Imaging at Low-k1". In: *Photomask Technology 2020*. Vol. 11518. International Society for Optics and Photonics. 2020, p. 1151807.

[51] Kenny K Toh et al. "Chromeless Phase-Shifted Masks: a New Approach to Phase-Shifting Masks". In: *10th Annual Symposium on Microlithography*. International Society for Optics and Photonics. 1991, pp. 27–53.

[52] *Transistor Count*. Nov. 2021. URL: `https://en.wikipedia.org/wiki/Transistor_count#Microprocessors`.

[53] Lloyd N Trefethen and David Bau III. *Numerical Linear Algebra*. Vol. 50. Siam, 1997.

[54] Paul Tseng. "Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization". In: *Journal of optimization theory and applications* 109.3 (2001), pp. 475–494.

[55] Obert Wood et al. "Improved Ru/Si Multilayer Reflective Coatings for Advanced Extreme Ultraviolet Lithography photomasks". In: *SPIE Advanced Lithography*. International Society for Optics and Photonics. 2016, pp. 977619–977619.

[56] Chien-Ching Wu et al. "Lifetime Test on EUV Photomask with EBL2". In: *Photomask Japan 2019: XXVI Symposium on Photomask and Next-Generation Lithography Mask Technology*. Vol. 11178. International Society for Optics and Photonics. 2019, 111780E.

[57] Li-Hao Yeh et al. "Experimental Robustness of Fourier Ptychography Phase Retrieval Algorithms". In: *Optics Express* 23.26 (2015), pp. 33214–33240.

[58] F Zernike. "Phase contrast". In: *Z. Tech. Physik.* 16 (1935), p. 454.