Towards diffraction-limited short-wavelength imaging systems

Gautam Gunjala

Electrical Engineering and Computer Sciences
University of California, Berkeley

Technical Report No. UCB/EECS-2022-117
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2022/EECS-2022-117.html

May 13, 2022

Towards diffraction-limited short-wavelength imaging systems

by

Gautam K. Gunjala

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Laura Waller, Chair
Professor Andrew Neureuther
Professor Austin Roorda

Spring 2022

Towards diffraction-limited short-wavelength imaging systems

Abstract

Towards diffraction-limited short-wavelength imaging systems

by

Gautam K. Gunjala

Doctor of Philosophy in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Laura Waller, Chair

Modern applications of optics, especially those which require shorter wavelengths of light, place ever-increasing demands on the performance of optical tools and systems. Working with extreme ultraviolet, soft x-ray and hard x-ray light poses complex limitations and challenges to diagnosing and maintaining diffraction-limited performance by measuring and controlling optical aberrations. By utilizing computational methods such as optimization and machine learning, we show that some of these limitations can be circumvented without sacrificing accuracy or precision.

In this work, we discuss a method for aberration measurement that is based on an analysis of speckle images acquired in situ. By using a stationary random object, our method eliminates the need for precise manufacturing and alignment of a test target. Moreover, the method provides a full, dense characterization of the optical system under test using relatively few images. The method has been successfully applied to an EUV microscope system, and is shown to be accurate to within $\lambda/180$. We also discuss a method for aberration compensation via the characterization and control of an adaptive optical element for x-ray optical systems. Adaptive x-ray optics are a relatively new technology, and our work aims to enable their use within the specifications of synchrotron beamline systems. To this end, we demonstrate the ability to experimentally predict and control the behavior of the glancing-incidence deformable mirror surface to within 2 nm rms, allowing the application of sub-wavelength corrections to an incident wavefront.

To all of my teachers

Contents

List of Figures

Acknowledgments

First and foremost, I would like to thank my adviser Prof. Laura Waller for 9 years of fantastic guidance and mentorship. I remember starting to work with Laura in my second year of undergrad, and the combination of our weekly meetings, her wonderful group environment, and her enthusiasm in presenting the beauty of optics in class and elsewhere truly made the time I spent in her group one of the defining experiences of my undergraduate years at UC Berkeley. As a graduate student, I appreciated Laura's emphasis on effective communication and presentation and her ability to keep the broader impacts of our research in perspective. I remember frequently leaving my meetings with her with a renewed sense of purpose and motivation. I also appreciated Laura's trust and support as exploring the applications of my research led me to shorter wavelengths and additional collaborators. Laura has been an amazing role model for me, and I am grateful for the values she has instilled in me as a researcher.

I would like to thank my collaborators at Lawrence Berkeley National Laboratory, Dr. Antoine Wojdyla and Dr. Kenneth A. Goldberg. A fortuitous opportunity to test the methods we had been developing on their extreme ultraviolet microscope setup led to years of collaboration on several fascinating research projects. I am incredibly grateful that I had the chance to learn from their experiences with EUV and x-ray optics, to discuss papers and research ideas, and to enjoy an occasional espresso and the view of the bay from the lab. Our meetings were always fun and something I looked forward to.

I would like to thank my collaborators at Argonne National Laboratory, Dr. Zhi Qiao, Dr. Xianbo Shi, and Dr. Lahsen Assoufid. Through a collaboration between the two national laboratories on their respective synchrotron upgrade projects, I was able to work with them on x-ray adaptive optics. In addition to their insights during our meetings, their efforts in setting up a remotely controllable experimental system was crucial to successfully performing this research during the COVID-19 pandemic.

Next I would like to thank all of the members of the Computational Imaging Lab I got to know and work with over the years, starting with my undergraduate mentor Aamod Shanker. "Think about diffusers" were his words that began this entire journey. The other group members at the time, especially Li-Hao Yeh, Michael Chen, Herbert Liu, Nick Antipa, Zack Phillips and Lei Tian were also all very friendly and supportive, and fun to hang out with at the graduate social hours. In graduate school, I enjoyed working with Stuart Sherwin on the measurement of aberrations; our conversations and collaborative efforts on the visible-light experimental setup really helped iron out the details of that project. I also enjoyed (frequent) conversations with the rest of my cohort: David Ren, Linda Liu, Kyrollos Yanny and Kristina Monakhova, and others in the lab or office: Grace Kuo, Ruiming Cao and Emrah Bostan.

I would like to thank my friends from graduate school: Alan Dong, Phong Nguyen, Stan Smith, Alex Reinking, Alain Anton, Kieran Peleaux, Chandan Singh, Michael Dennis and Andrea Bajcsy. Our antics were always a ton of fun, and you all definitely helped me maintain my sanity through grad school. I also want to thank my friends from elsewhere:

Nitin Sadras, Ananth Subramaniam, Mason Uemura, Harold Chun, Stacey Jou, Kaustubh Deshpande, Alex Kim, Katy Sun, Judy Chen, Amanda Yip, Angela Hsiao, Michael Dong, Jason Qin, Andy Shi, Jess Liu and Palak Singal.

Finally I would like to thank my family for their support over the years. Thanks to my parents, Uttam and Radhika, for inspiring me to always do my best, and thanks to my uncle, Prasad, for sharing his often amusing experiences with academia.

Chapter 1

Introduction

Many applications of modern science and engineering have come to rely on short wavelength optics. For example, in the space of microscopy and metrology, shorter wavelengths enable the visualization of smaller sample features. Imaging techniques based on x-ray illumination are commonly employed in the life and materials sciences. In spectroscopy, many insights can be drawn from the unique interactions between soft x-rays and matter. Projection imaging using extreme ultraviolet light has been a relatively new innovation in photolithography, helping maintain a decades-long trend of patterning semiconductors with higher density and improving the performance of electronics.

While there are many motivations for using shorter wavelengths of light, working with them in practice often poses significant challenges. As we push towards wavelengths of just a few nanometers, our tolerance for errors in the alignment, fabrication and metrology of optical systems is typically sub-nanometer. Errors on the same order of magnitude as the wavelength are often prohibitive to the operation of the optical system. Additionally, the absorption and transmission properties for short wavelengths of light and optical surfaces can limit the design and placement of optical elements in a system. For example, extreme ultraviolet light is significantly absorbed by most matter, which limits the number of elements that can be used while maintaining sufficient photon throughput. On the other hand, most matter is transmissive to hard x-ray radiation, and so most focusing elements (e.g. Kirkpatrick-Baez mirrors) operate at grazing incidence. In some cases, the high energies associated with e.g. x-ray sources can also cause heat deformation in optical elements, resulting in dynamic errors. Compared to visible light optics, where many of these problems either do not exist or are much more easily solved, the design and maintenance of short wavelength systems free of errors, or aberrations, remains an active area of research.

1.1 The diffraction limit and aberrations

Many of the challenges that arise when working with short wavelengths can be attributed to the desire for "diffraction-limited" optical systems. The diffraction limit is a theoretical

limit on the resolution of an optical system, defined:

$$r = k \cdot \frac{\lambda}{\mathrm{NA}} \tag{1.1}$$

where k is a context-specific constant factor, λ is the illumination wavelength, and NA is the numerical aperture (angular acceptance) of the system. The quantity r represents the size of the smallest resolvable detail, defined using one of several metrics. The resolution of an optical system determines, for example, the image of a point source (point spread function) and similarly the smallest possible focal spot for an incident plane wave. A practical system (which generally has a fixed source wavelength and NA) whose performance closely aligns with this theoretical ideal is called "diffraction-limited" [32, 23].

It is often the case however that practical systems deviate from this theoretical ideal, and in such cases, the minimum resolvable detail or minimal spot size are larger than expected. Errors in the fabrication and alignment of optical elements often result in this loss of performance, and the resulting deviation from an ideal wavefront (aberrations) can be precisely specified. It then follows that access to this information and the ability to act on it are essential to the efficient operation of practical optical systems. In this thesis, we will explore techniques for the measurement and correction of aberrations, with a particular emphasis on extreme ultraviolet ($\lambda \in [10, 120]$ nm) and soft x-ray ($\lambda \in [0.1, 10]$ nm) systems.

1.2 Aberration measurement

Chapters 2 and 3 of this thesis will discuss a technique for the measurement of aberrations in imaging systems and its application to an extreme ultraviolet (EUV) microscope. While there are numerous existing methods for aberration measurement, they are generally difficult, costly or impractical to apply to specialized modalities such as EUV imaging. Interferometric techniques [29, 61, 8, 62, 77, 48, 47] or adaptive optics [78, 5, 37, 63], both widely used in the visible spectrum, require the insertion of complicated and expensive hardware in the optical system. These techniques are impractical in the EUV regime since the same range of optical elements does not exist for this wavelength, and moreover, EUV systems typically exist in a vacuum, making modification of the beam path impractical.

One measurement approach used in EUV is to image precisely-calibrated test objects [28, 16] such as gratings [53, 52], contact arrays [54, 83], or custom features [43], which do not require modifying the system hardware. However, known test objects containing features with sizes near the resolution limit of the imaging system are difficult or expensive to fabricate with high fidelity. Fabrication errors and the 3D structure of test objects can also introduce model mismatch and complicate analysis [65, 60]. Furthermore, many of these test objects are not ideal for aberration and wavefront recovery. For example, periodic targets only probe discrete points in frequency space, and custom features are generally designed to retrieve a single component of the wavefront error. In both of these cases, a combination of multiple feature sizes, orientations and focal positions must be measured in order to retrieve

a relatively small amount of information [53]. In other words, these techniques are generally not data-efficient and provide a limited, albeit highly accurate, characterization of an optical system.

Another class of aberration measurement techniques operates by jointly estimating both the sample of interest and the system aberrations using optimization [57, 11]. An inverse problem is generally formulated using the imaging forward model and regularized to enforce priors on the sample, aberrations or both. This method has the advantages of not requiring any additional hardware or test targets and also capturing a complete characterization of the aberrations and the sample image simultaneously. However, the use of strong regularizers and priors often leads to a biased reconstruction of the aberration function that is highly dependent on the sample being imaged.

In this work, we develop a method for characterizing the aberrations of a full-field imaging system that does not require hardware modifications or the fabrication of test objects. To implement, we acquire speckle images of a suitable diffuser object at multiple angles of plane-wave illumination. The object being imaged does not need to be precisely fabricated; it only needs to have a pseudo-random surface, weak phase and sufficient power-spectral density extending to the imaging system's resolution limit. A blank EUV photomask conveniently meets these requirements, due to intrinsic surface roughness [55, 81].

Our method differs from a number of existing speckle-based wavefront sensing techniques [10, 6] in that we rely on Fourier space properties of speckle imaged near-focus rather than real-space properties of propagated speckle. Similar methods have been demonstrated in electron microscopy, where amorphous carbon is imaged (i.e the Zemlin tableau method) [84, 80, 74, 15] as part of a calibration procedure; however, these are generally not based on inversion of a precise forward model and are used only to determine low-order aberrations.

Our method has several advantages over other approaches that entail imaging test objects. First, no special fabrication is required, as suitable objects can be found opportunistically. Second, unlike periodic objects, uncorrelated surface roughness provides isotropic sampling of frequency space. Third, no registration or alignment of the test object is required, as the statistics of the roughness should not change across the object. Finally, our method is data-efficient; in our implementation, we use 10 speckle images to recover all field-varying aberrations of up to order 5. We apply the method to the SHARP High-NA Actinic Reticle Review Project (SHARP) [21], an EUV microscope that operates near 13.5 nm wavelength, but we emphasize that it is suitable for any full-field imaging system that has coherent, steerable illumination.

1.3 Aberration correction

Chapters 4 and 5 of this thesis will discuss adaptive optics for rapid, open-loop aberration correction in x-ray systems. In particular, we will emphasize applications to synchrotron and free-electron laser (FEL) facilities as aberration control is crucial to their operation. The next generation of light sources, including FELs and diffraction-limited storage rings

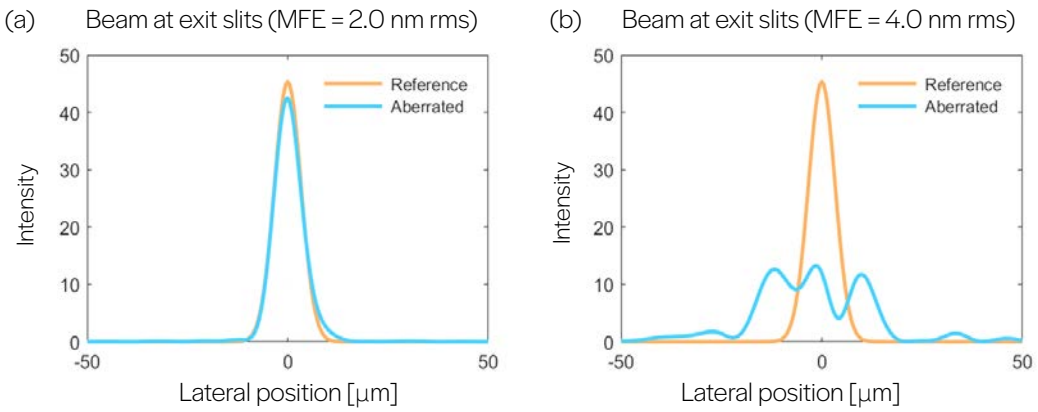

Figure 1.1: **X-ray beam profile at focus after aberrated mirror.** (a) A mirror figure error (MFE) of 2.0 nm rms causes a slight reduction in peak intensity with respect to reference (no aberrations); however the beam can still be considered in focus. (b) An MFE of 4.0 nm rms results in very low peak intensity, and the beam is not focused.

(DLSR) at synchrotrons, are poised to produce x-ray beams of unprecedented brightness and coherent flux. Development of these upgraded light sources is already underway around the world, including the Advanced Light Source Upgrade (ALS-U) and Advanced Photon Source Upgrade (APS-U) projects in US national laboratories. The improved source characteristics aim to enable fast experiments where wavefront phase information can be used to study matter in exquisite detail.

Given these advances in beam quality, it becomes increasingly important to ensure that this quality is not lost through aberrations in the beamline optics downstream. The scale of tolerances for surface height errors in these short-wavelength imaging optics presents a significant challenge, which is demonstrated in Fig. 1.1. Simulating a focusing element with the intention of bringing incident illumination with a wavelength of 1.25 nm to a diffraction-limited spot, we can see that while 2 nm rms of simulated figure error can be generally tolerated, merely 4 nm rms of error render the element unable to focus light. In addition to nominal surface figure errors, aberration correction is further complicated by dynamic conditions such as heatload deformation owing to the energy of the incident illumination.

Under these circumstances, adaptive optics techniques have emerged as an attractive solution for the design of beamlines with wavefront shaping capabilities [66, 59, 35, 41, 13, 69]. Over the last decade, significant advances [51, 67] have led to the commercial availability of piezo-bimorph mirrors [1, 36] and their successful deployment on several beamlines [49, 76]. These mirrors have symmetrically placed bimorph elements attached to silicon mirror substrates, which allow these systems to maintain thermal stability while providing one-dimensional shape actuation. Investigations of the mirrors' linear response demonstrate that their shape can be controlled to a nanometer level in a predictable way [79, 1] as

required for diffraction-limited performance. Studies of the mirrors' dynamic response show that appreciable shape changes on the scale of 1 second are possible [2].

In this work, we propose a two-part framework for the open-loop operation of an x-ray deformable mirror, involving: (1) approximating the nonlinear system dynamics using a feed-forward neural network, and (2) control to a desired surface shape using nonlinear quadratic cost regulation over a finite time horizon. We will present an experimental demonstration of our approach performed using an ex-situ visible-light Fizeau interferometer to record the behavior of an adaptive mirror driven through various shape transitions. We note that our methodology is broadly applicable to x-ray or other optical systems utilizing an adaptive element, independent of the optical configurations and wavelength ranges.

Chapter 2

Imaging systems, aberrations and speckle

In this chapter, we will present the theoretical framework surrounding our method for aberration measurement in optical imaging systems based on the acquisition of speckle images under varying angles of illumination. Specifically, we will construct a mathematical model for diffusers – weak-phase, stationary scattering media – and use it to develop a general image formation model. We will show that under specific conditions, the system aberrations are observable in intensity measurements, setting the stage for development of a reconstruction algorithm.

2.1 Modeling imaging systems

We will begin with a Fourier optics model for imaging systems [23]. This model is built upon the idea that signals (images) can be represented as a linear combination of components with unique, single frequencies (plane waves). In this model, the action of an optical system is analogous to low-pass filtering, where the frequencies permitted by the system are related to its numerical aperture (NA). This relationship is illustrated in Fig. 2.1

Mathematically, the imaging system is expressed as a transfer function, $P(\cdot)$, acting on time-averaged electric fields, $E(\cdot)$, and the measured intensity, $I(\cdot)$, is obtained by computing the squared magnitude of the output electric field. In general, we have:

$$I(\boldsymbol{x}) = |E_i(\boldsymbol{x})|^2 = |P(\boldsymbol{x}) * E_o(\boldsymbol{x})|^2, \tag{2.1}$$

where $\boldsymbol{x} \in \mathbb{R}^2$ is the spatial coordinate, $*$ denotes convolution and subscripts i and o indicate the image and object planes, respectively. Taking the Fourier transform of Eq. (2.1) results in an autocorrelation, which can be rewritten as a convolution. Expanding this, we obtain:

$$\widehat{I}(\boldsymbol{u}) = \widehat{E_i}(\boldsymbol{u}) * \widehat{E_i^*}(-\boldsymbol{u}) = \widehat{P}(\boldsymbol{u})\widehat{E_o}(\boldsymbol{u}) * \widehat{P^*}(-\boldsymbol{u})\widehat{E_o^*}(-\boldsymbol{u}), \tag{2.2}$$

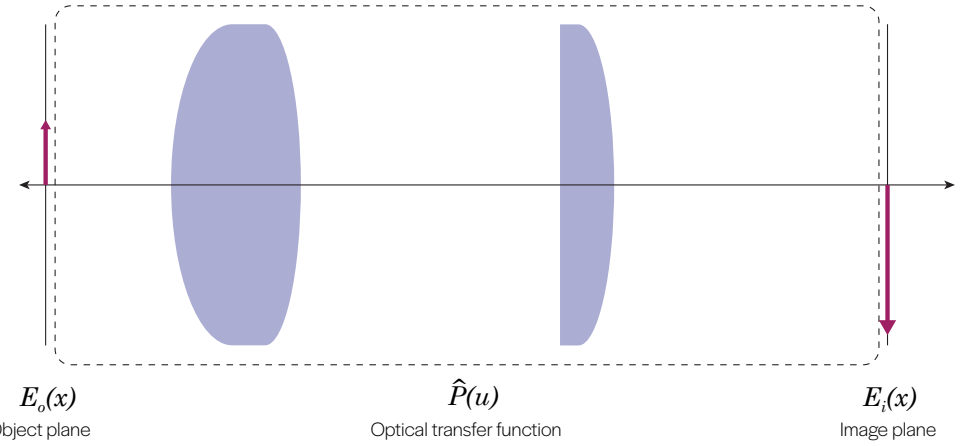

Figure 2.1: **A general imaging system.** The optical transfer function fully describes the relationship between the electric fields at the object and image planes. A sensor placed at the image plane records intensity.

where $\boldsymbol{u} \in \mathbb{R}^2$ is the NA-normalized spatial frequency coordinate and hatted quantities are related to their counterparts by Fourier transform [24].

The function $\widehat{P}(\boldsymbol{u})$ is known as the optical transfer function, or the pupil function. This complex-valued function generally consists of an indicator function that selects frequencies permitted by the system's aperture, an amplitude function that describes attenuation of certain frequency components and a phase function which encodes system aberrations. We will assume that the imaging system we wish to describe has a circular aperture, which is the case for a majority of systems in practice. Mathematically, we have:

$$\widehat{P}(\boldsymbol{u}) = A(\boldsymbol{u}) \cdot \exp\left[i \cdot W(\boldsymbol{u})\right] \cdot \mathrm{Circ}(\boldsymbol{u}), \tag{2.3}$$

where $A(\cdot)$ is the amplitude function, i is the imaginary unit and the $\mathrm{Circ}(\cdot)$ function is the characteristic function of the unit disk. The real-valued function $W(\cdot)$ represents the aberrations present in the imaging system, and will be referred to as the wavefront error function (WEF).

In our analysis, we will assume that the amplitude term of the pupil function, $A(\cdot)$, is negligible. This is often true for visible light systems, for which typical optical elements (*e.g.* lenses, mirrors) have very little absorption. In the case of the EUV system to which our technique was applied, this was also approximately true due to its very low NA. We will

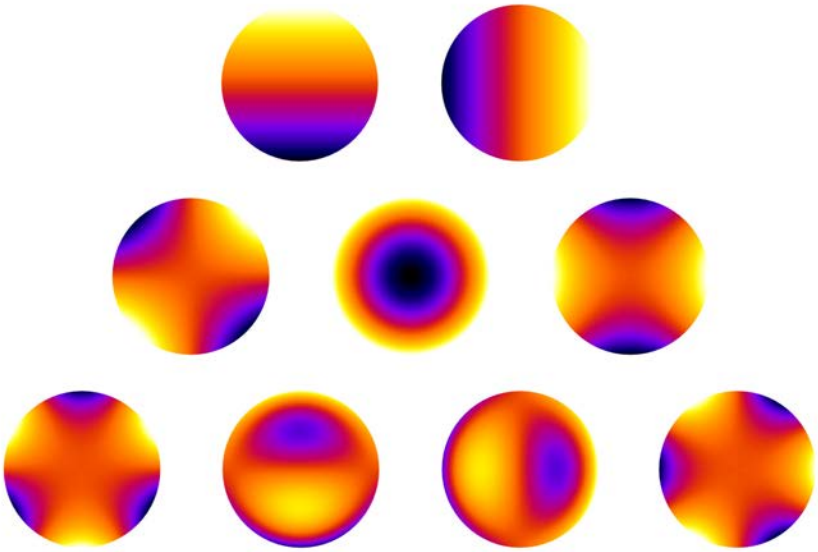

Figure 2.2: **Zernike polynomials.** Plots depict the first-, second- and third-order polynomials in the Zernike basis, which consist of tilt, defocus, astigmatism, coma and trefoil.

then turn our focus to the WEF, which our method aims to recover.

2.2 Aberrations in imaging systems

The wavefront error function (WEF) is a quantitative description of the aberrations of an imaging system. This function is often expressed using the Zernike polynomial basis (see Fig. 2.2), which is an orthonormal basis over the unit disk. This representation is convenient, as many of the basis elements (*e.g.* defocus, coma) correspond to commonly occurring alignment errors in imaging systems. For example, a system whose sample plane is axially displaced will have significant defocus, which can be described by a parabolic phase error. Moreover, a parametric representation of the WEF greatly reduces the dimensionality of an algorithm which seeks to recover it. Rather than solving for the graph of the WEF over an arbitrarily fine mesh, we can simply search for the best set of a few coefficients which generate the graph.

Considering Eq. 2.3, we see that if aberrations are absent or negligible (*i.e.* $W(\boldsymbol{u}) = 0$), then the complex exponential term results in unity and we are left with an ideal low-pass filter based on the system's aperture. Under these conditions, the image of a point emitter would be an Airy function, and the system would be considered diffraction-limited. It follows that a departure from these conditions ($W(\boldsymbol{u}) \neq 0$) implies imperfect imaging and can result

in a system which is no longer diffraction-limited.

The severity of aberrations in an imaging system can be described by computing the root-mean-square (RMS) magnitude of the WEF. That is, if c is a vector of Zernike polynomial coefficients such that $W(\boldsymbol{u}) = \sum_i c_i \cdot Z_i(\boldsymbol{u})$, then:

$$m = \pi \cdot \|c\|_2 = \pi \cdot \sqrt{\sum_i c_i^2} \tag{2.4}$$

is the magnitude of the system aberrations and has units of radians RMS. A practical point of reference for interpreting the magnitude of aberrations in a system is known as the *Maréchal criterion* [46]. This states that if the aberrations are limited to within $\lambda/14$ nanometers RMS, then the system is virtually diffraction limited. This suggests that for an aberration measurement method to be useful in practice, it must achieve adequate sensitivity for measurements at this scale. We will also use the Maréchal criterion to comment on the results of our application to EUV microscopy.

Note 1: Zernike polynomials represent functions over two spatial dimensions and are thus ambiguously ordered. In this work, we will use the OSA/ANSI ordering.

Note 2: The standard normalization of Zernike polynomials is such that the norm of each polynomial is π, resulting in the scalar term in Eq. 2.4 [25].

2.3 Modeling a diffuser

The crux of our aberration measurement method is the use of a weak diffuser as the sample, whose properties enable simplifications to the imaging forward model. To start, we assume that the randomness of the scattering medium is stationary, meaning that its scattering properties do not vary spatially [68, 26]. Assuming plane wave illumination, we construct a general model $E(\cdot)$ for the electric field immediately beyond a diffuser as follows:

$$E(\boldsymbol{x}) = \exp\left\{i \cdot \left(\exp\left\{-a|\boldsymbol{x}|^2 * \eta(\boldsymbol{x})\right\}\right)\right\} \tag{2.5}$$

where evaluations of $\eta(\boldsymbol{x}) \sim \mathcal{N}(0, \sigma^2)$ for unique spatial positions are independent and identically distributed. The parameters a and σ are related to the width (lateral) and height, respectively, of an average feature on the rough diffuser surface. The phase component of Eq. 2.5 can be interpreted as "smoothed noise" in that the noise term $\eta(\cdot)$ is smoothed by a normalized Gaussian function parametrized by a. Large values of a result in a rapidly decaying Gaussian and hence small features in the phase.

Additionally, we assume that the diffuser phase is weak, which has two major consequences [27, 75, 64, 40]. First, weak phase implies very small spatial variation in optical path distance, and so, we can view a weak phase diffuser as a small perturbation to a transmission flat or flat mirror in transmission- and reflection-mode imaging, respectively. As a

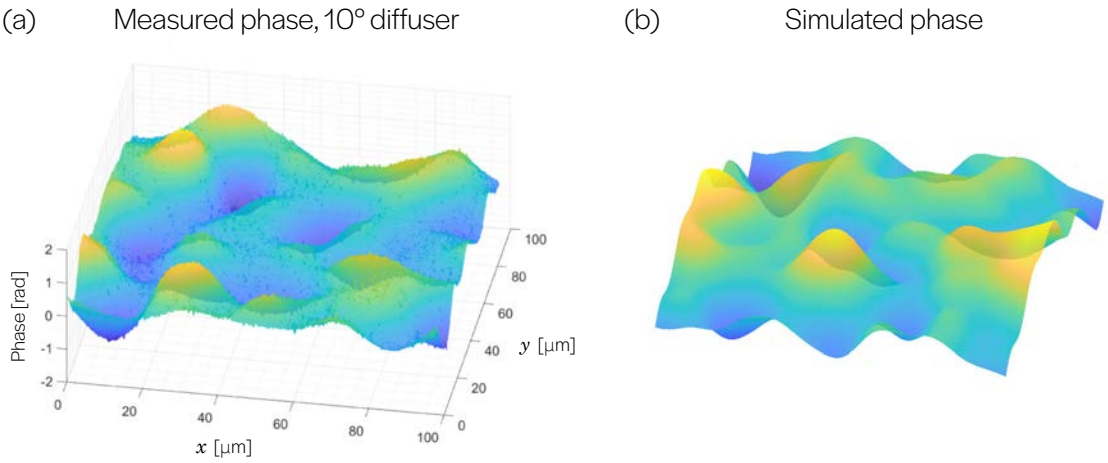

Figure 2.3: **Measured phase of holographic diffuser compared to simulated phase.** (a) Phase of a 10° holographic diffuser (Edmund Optics #54-493) measured using Fourier ptychographic microscopy. (b) Phase simulated using Eq. 2.5 (arbitrary units).

result, we have omitted an amplitude (attenuation) term from Eq. 2.5. Second, weak phase allows for accurate approximation via first-order Taylor expansion. Mathematically, we have:

$$E(\boldsymbol{x}) \approx 1 + i \cdot \left(\exp\left\{-a|\boldsymbol{x}|^2\right\} * \eta(\boldsymbol{x})\right). \tag{2.6}$$

This linearization of the sample phase is crucial to simplifying the imaging forward model. To validate our model for the (real space) phase of a diffuser, Eq. 2.5, we obtain a quantitative phase measurement of a 10° holographic diffuser using Fourier ptychographic microscopy [85, 57, 12]. We compare this measured phase to the surface generated by the model in Fig. 2.3. Note that the units in the simulated phase are arbitrary, as this can be scaled to match experimental conditions via the parameters a and σ.

To aid in our derivation of the imaging forward model, we will now consider the Fourier transform of Eq. 2.6. We have:

$$\widehat{E}(\boldsymbol{u}) = \delta(\boldsymbol{u}) + i \cdot \exp\left\{-b|\boldsymbol{u}|^2\right\} \cdot \widehat{\eta}(\boldsymbol{u}) \tag{2.7}$$

where the convolution in the imaginary term of Eq. 2.6 is now a multiplication. Note that the scalar relationships between Eq. 2.6 and Eq. 2.7 have been omitted. It is worth discussing here the nature of the noise term, $\widehat{\eta}(\boldsymbol{u})$, which is the Fourier transform of $\eta(\boldsymbol{x})$ in Eq. 2.5.

We first decompose $\eta(\boldsymbol{x})$ into the sum of its even ($f(x) = f(-x)$) and odd ($f(x) = -f(-x)$) components. Assuming that $\eta(\boldsymbol{x}) \sim \mathcal{N}(0, \sigma^2)$, we can show that the even and odd components are uncorrelated.

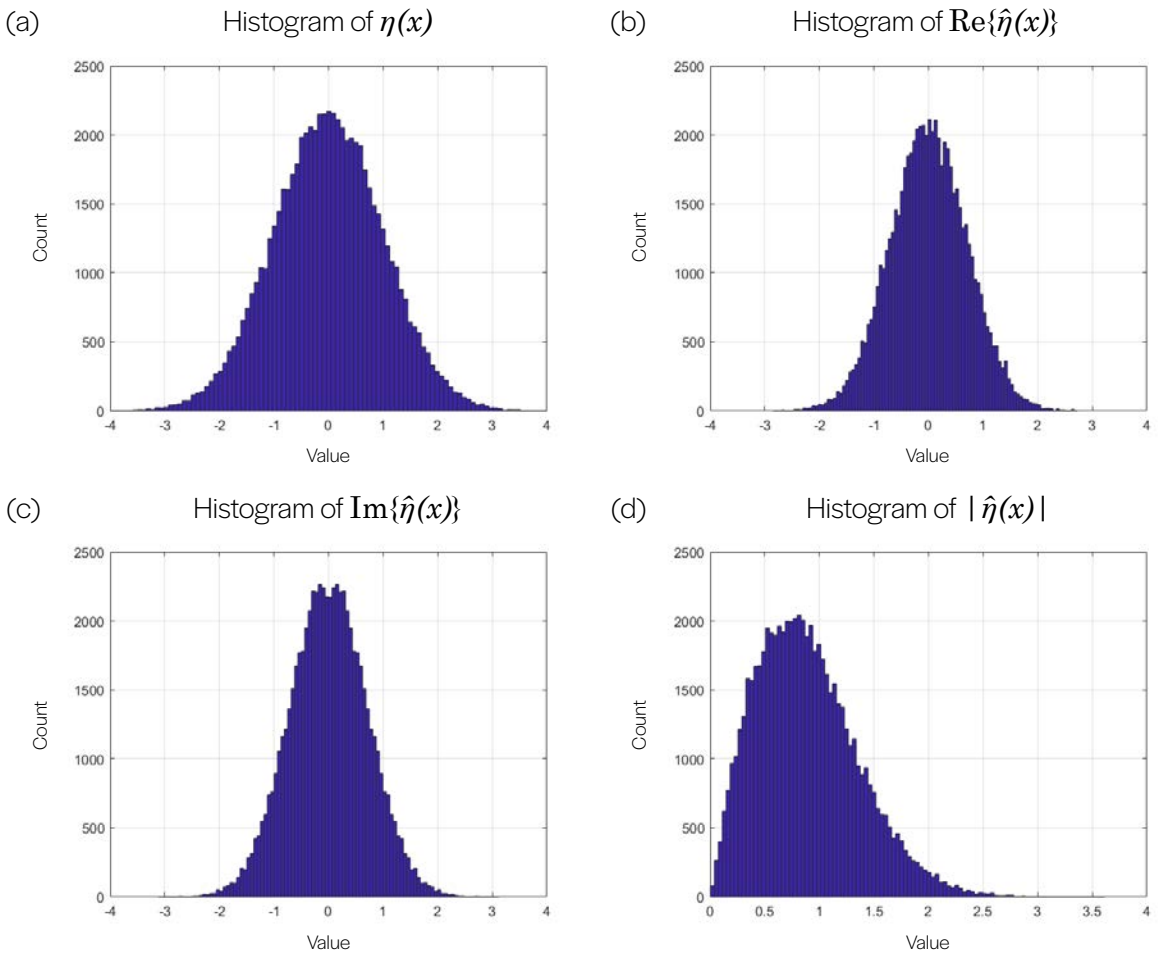

Figure 2.4: **Distributions of Fourier transform of Gaussian white noise.** (a) We begin with a representative sampling of Gaussian white noise, $\eta(\boldsymbol{x})$ in Eq. 2.5. (b-c) The distributions of the real and imaginary parts of the Fourier transform of Gaussian white noise, $\widehat{\eta}(\boldsymbol{u})$, both follow a Gaussian distribution. (d) The magnitude $|\widehat{\eta}(\boldsymbol{u})|$ follows a Rayleigh distribution.

$$\begin{aligned}
&\operatorname{Cov}\left[\frac{\eta(\boldsymbol{x})+\eta(-\boldsymbol{x})}{2}, \frac{\eta(\boldsymbol{x})-\eta(-\boldsymbol{x})}{2}\right] \\
&= \mathbb{E}\left[\left(\frac{\eta(\boldsymbol{x})+\eta(-\boldsymbol{x})}{2} - \mathbb{E}\left[\frac{\eta(\boldsymbol{x})+\eta(-\boldsymbol{x})}{2}\right]\right)\left(\frac{\eta(\boldsymbol{x})-\eta(-\boldsymbol{x})}{2} - \mathbb{E}\left[\frac{\eta(\boldsymbol{x})-\eta(-\boldsymbol{x})}{2}\right]\right)\right] \\
&= \frac{1}{4}\cdot\mathbb{E}\left[\left(\eta(\boldsymbol{x})+\eta(-\boldsymbol{x})-\mathbb{E}\left[\eta(\boldsymbol{x})\right]-\mathbb{E}\left[\eta(-\boldsymbol{x})\right]\right)\left(\eta(\boldsymbol{x})-\eta(-\boldsymbol{x})-\mathbb{E}\left[\eta(\boldsymbol{x})\right]+\mathbb{E}\left[\eta(-\boldsymbol{x})\right]\right)\right]
\end{aligned}$$

We know that $\mathbb{E}\left[\eta(\boldsymbol{x})\right] = \mathbb{E}\left[\eta(-\boldsymbol{x})\right] = 0$ and can cancel terms accordingly.

$$\begin{aligned}
\operatorname{Cov}\left[\frac{\eta(\boldsymbol{x})+\eta(-\boldsymbol{x})}{2}, \frac{\eta(\boldsymbol{x})-\eta(-\boldsymbol{x})}{2}\right] &= \frac{1}{4}\cdot\mathbb{E}\left[\left(\eta(\boldsymbol{x})+\eta(-\boldsymbol{x})\right)\left(\eta(\boldsymbol{x})-\eta(-\boldsymbol{x})\right)\right] \\
&= \frac{1}{4}\cdot\mathbb{E}\left[\eta(\boldsymbol{x})^2-\eta(-\boldsymbol{x})^2\right] \\
&= \frac{1}{4}\cdot\mathbb{E}\left[\eta(\boldsymbol{x})^2\right]-\mathbb{E}\left[\eta(-\boldsymbol{x})^2\right] \\
&= \frac{1}{4}\cdot\left[\sigma^2-\sigma^2\right] = 0. \quad \blacksquare
\end{aligned}$$

Note that $\eta(\boldsymbol{x})$ and $\eta(-\boldsymbol{x})$ are independent for all nonzero $\boldsymbol{x}$, since they represent i.i.d. white noise. This implies that the odd and even components both follow a Gaussian distribution almost everywhere. We also note that the Fourier transform of a real, even function is purely real, and the Fourier transform of a real, odd function is purely imaginary. Due to the linearity of the Fourier transform, it follows that the transforms of the even and odd components maintain their Gaussian distributions (almost everywhere) in frequency space. Thus, we have $\operatorname{Re}\{\widehat{\eta}(\boldsymbol{u})\} \sim \mathcal{N}(0,\rho^2)$ and $\operatorname{Im}\{\widehat{\eta}(\boldsymbol{u})\} \sim \mathcal{N}(0,\rho^2)$. This implies that the frequency space white noise, $\widehat{\eta}(\boldsymbol{u})$, has a magnitude which follows the Rayleigh distribution:

$$|\widehat{\eta}(\boldsymbol{u})| \sim \text{Rayleigh}(\rho). \tag{2.8}$$

Visual depictions of simulated white noise components, their Fourier transforms and their distributions are shown in Fig. 2.4. Using these models and assumptions about the diffuser, we may proceed with the derivation of our imaging forward model.

2.4 Derivation of the imaging forward model

In this section, we will derive the image formation model and show that information pertaining to system aberrations is present in images of weak diffusers. We will begin with Eq. 2.2, reproduced below:

$$\widehat{I}(\boldsymbol{u}) = \widehat{P}(\boldsymbol{u})\widehat{E_o}(\boldsymbol{u}) * \widehat{P^*}(-\boldsymbol{u})\widehat{E_o^*}(-\boldsymbol{u}). \tag{2.9}$$

From the previous sections, we may substitute specific expressions for $\widehat{P}(\cdot)$ and $\widehat{E_o}(\cdot)$. Prior to this substitution, however, we will generalize for the case of off-axis plane-wave illumination. In real space, this illumination may be viewed as a multiplication of the object function with a phase ramp, i.e.:

$$E_o(\boldsymbol{x}) = \exp\left[i\ 2\pi\left(\boldsymbol{u}_0^{\mathsf{T}}\boldsymbol{x}\right)\right] \cdot E(\boldsymbol{x}). \tag{2.10}$$

Its Fourier transform is then:

$$\widehat{E_o}(\boldsymbol{u}) = E(\boldsymbol{u} - \boldsymbol{u}_0). \tag{2.11}$$

Combining Eq. 2.7 and Eq. 2.11, we have:

$$\widehat{E_o}(\boldsymbol{u}) = \delta(\boldsymbol{u} - \boldsymbol{u}_0) + i \cdot \exp\left\{-b|\boldsymbol{u} - \boldsymbol{u}_0|^2\right\} \cdot \widehat{\eta}(\boldsymbol{u} - \boldsymbol{u}_0). \tag{2.12}$$

We now substitute Eq. 2.3 and Eq. 2.12 into Eq. 2.2, recalling that the amplitude term of the general pupil function is negligible within our assumptions. This results in:

$$\begin{aligned}
\widehat{I}(\boldsymbol{u}) = &\left[\delta(\boldsymbol{u} - \boldsymbol{u}_0)\widehat{P}(\boldsymbol{u}) * \delta(-\boldsymbol{u} - \boldsymbol{u}_0)\widehat{P^*}(-\boldsymbol{u})\right] \\
&+ i\left[\delta(-\boldsymbol{u} - \boldsymbol{u}_0)\widehat{P^*}(-\boldsymbol{u}) * \widehat{\varphi}(\boldsymbol{u} - \boldsymbol{u}_0)\widehat{P}(\boldsymbol{u})\right] \\
&- i\left[\delta(\boldsymbol{u} - \boldsymbol{u}_0)\widehat{P}(\boldsymbol{u}) * \widehat{\varphi^*}(-\boldsymbol{u} - \boldsymbol{u}_0)\widehat{P^*}(-\boldsymbol{u})\right] \\
&+ \left[\widehat{\varphi}(\boldsymbol{u} - \boldsymbol{u}_0)\widehat{P}(-\boldsymbol{u}) * \widehat{\varphi^*}(-\boldsymbol{u} - \boldsymbol{u}_0)\widehat{P^*}(-\boldsymbol{u})\right],
\end{aligned} \tag{2.13}$$

where

$$\widehat{\varphi}(\boldsymbol{u}) = \exp\left\{-b|\boldsymbol{u}|^2\right\} \cdot \widehat{\eta}(\boldsymbol{u}) \tag{2.14}$$

is the phase of the diffuser prior to linearization. Since the weak phase of the object is noise-like and hence, uncorrelated, we can assume that the interactions between the phase of the object and itself, captured by the fourth term in Eq. (2.13), is a delta-like function scaled by some scalar γ. As such, simplification of the above results in:

$$\begin{aligned}
\widehat{I}(\boldsymbol{u}) = &\,\delta(\boldsymbol{u})\left[\widehat{P}(\boldsymbol{u}_0)\widehat{P^*}(\boldsymbol{u}_0) + \gamma\right] \\
&+ i\widehat{P^*}(\boldsymbol{u}_0)\left[\widehat{\varphi}(\boldsymbol{u})\widehat{P}(\boldsymbol{u} + \boldsymbol{u}_0)\right] \\
&- i\widehat{P}(\boldsymbol{u}_0)\left[\widehat{\varphi^*}(-\boldsymbol{u})\widehat{P^*}(-\boldsymbol{u} + \boldsymbol{u}_0)\right].
\end{aligned} \tag{2.15}$$

Note that we have also made a change of variable in the arguments of $\widehat{P}(\cdot)$ and $\widehat{\varphi}(\cdot)$. In doing so, we interpret the effect of off-axis illumination as the interaction of a fixed object

function with a shifted pupil, rather than a shifted object function with a fixed pupil. Since $\varphi(\boldsymbol{x})$ is a real-valued function, we know that its Fourier transform, $\widehat{\varphi}(\boldsymbol{u})$, satisfies conjugate symmetry, i.e. $\widehat{\varphi}(\boldsymbol{u}) = \widehat{\varphi^*}(-\boldsymbol{u})$. Using this fact, we can factor Eq. 2.15 to obtain:

$$\widehat{I}(\boldsymbol{u}) = \delta(\boldsymbol{u}) \left[|\widehat{P}(\boldsymbol{u}_0)|^2 + \gamma \right] + i\widehat{\varphi}(\boldsymbol{u}) \left[\widehat{P^*}(\boldsymbol{u}_0)\widehat{P}(\boldsymbol{u}+\boldsymbol{u}_0) - \widehat{P}(\boldsymbol{u}_0)\widehat{P^*}(-\boldsymbol{u}+\boldsymbol{u}_0) \right]. \quad (2.16)$$

For pupil recovery, we will primarily be concerned with the latter term in Eq. 2.16. We define:

$$\widehat{I_\varnothing}(\boldsymbol{u}) = i\widehat{\varphi}(\boldsymbol{u}) \left[\widehat{P^*}(\boldsymbol{u}_0)\widehat{P}(\boldsymbol{u}+\boldsymbol{u}_0) - \widehat{P}(\boldsymbol{u}_0)\widehat{P^*}(-\boldsymbol{u}+\boldsymbol{u}_0) \right], \quad (2.17)$$

where $\widehat{I_\varnothing}(\cdot)$ represents the DC-suppressed counterpart of $\widehat{I}(\cdot)$.

To further simplify the above equation, we introduce the symmetric (even) and anti-symmetric (odd) decomposition operators, denoted by $\mathcal{S}\{\cdot\}$ and $\mathcal{A}\{\cdot\}$, respectively:

$$\mathcal{S}\{f(\boldsymbol{u})\} = \frac{f(\boldsymbol{u}) + f(-\boldsymbol{u})}{2} \quad , \quad \mathcal{A}\{f(\boldsymbol{u})\} = \frac{f(\boldsymbol{u}) - f(-\boldsymbol{u})}{2}. \quad (2.18)$$

Recall that any function can be decomposed as $f(\boldsymbol{u}) = \mathcal{S}\{f(\boldsymbol{u})\} + \mathcal{A}\{f(\boldsymbol{u})\}$.

Using this notation and substituting Eq. 2.3 into Eq. 2.17, we obtain:

$$\widehat{I_\varnothing}(\boldsymbol{u}) = 2i \cdot \exp\left[i\mathcal{A}\{W(\boldsymbol{u}+\boldsymbol{u}_0)\}\right] \cdot \widehat{\varphi}(\boldsymbol{u}) \cdot \sin\left(\mathcal{S}\{W(\boldsymbol{u}+\boldsymbol{u}_0)\}\right). \quad (2.19)$$

We remark that this simplification requires both difference operands in Eq. 2.17 to be nonzero, such that interference can occur and produce contrast in intensity. Since each instance of the pupil $\widehat{P}(\cdot)$ is band-limited by the Circ$(\cdot)$ function, the region where the interference condition is met is given by the set:

$$\mathscr{U} = \{\boldsymbol{u} : \text{Circ}(\boldsymbol{u}+\boldsymbol{u}_0) = 1\} \cap \{\boldsymbol{u} : \text{Circ}(\boldsymbol{u}-\boldsymbol{u}_0) = 1\}. \quad (2.20)$$

As a result, Eq. 2.19 is valid only within the region where the shifted pupils overlap. We also note that the even part of the WEF exists inside a real-valued multiplicative factor, while the odd part exists inside a complex exponential. The complex exponential structure makes the recovery of odd aberration functions challenging due to its interactions with the complex-valued noise term of $\widehat{\varphi}(\cdot)$ in Eq. 2.14. As a result, we exclude this term by considering the magnitude of Eq. 2.19. Although we lose information about the odd WEF components, we will demonstrate in the following section that this information remains accessible via off-axis illumination.

Taking the magnitude of both sides of Eq. 2.19, we arrive at our real-valued forward model:

$$|\widehat{I_\varnothing}(\boldsymbol{u})| = 2 \cdot |\widehat{\varphi}(\boldsymbol{u})| \cdot |\sin\left(\mathcal{S}\{W(\boldsymbol{u}+\boldsymbol{u}_0)\}\right)|, \ \boldsymbol{u} \in \mathscr{U}. \quad (2.21)$$

2.5 Properties of the imaging forward model

In this section, we will examine the structure of the imaging forward model, Eq. 2.21, and demonstrate its behavior through examples. We will first establish the relationship between the WEF, $W(\cdot)$, and the spectrum of measured intensity. For simplicity, we will consider the case of on-axis illumination, i.e. $\boldsymbol{u}_0 = 0$, for a system whose aberrations are purely defocus with varying magnitude. The WEF in this case is:

$$W(\boldsymbol{u}) = z_4 \cdot 2\sqrt{3}|\boldsymbol{u}|^2, \tag{2.22}$$

where z_4 is a Zernike coefficient with units π rad rms.

The simulated results of imaging a diffuser through such a system are presented in Fig. 2.5. The left column shows the WEF as we change the Zernike coefficient from 0.5π to 3.5π rad rms. The middle column shows the Fourier transform of an intensity measurement, simulated *without* using our derived forward model (i.e. without explicitly making the weak phase approximation). As the coefficient z_4 increases, we see more rings in the Fourier spectrum. This is due the sinusoid in the forward model interacting with the defocus function, a paraboloid, resulting in a radial chirp pattern. This sinusoidal component is simulated and shown in the right column, where a visual correspondence is evident. We note that the randomness of the diffuser results in a noisy realization of the predicted ring pattern. Recall from Eq. 2.14 that this noise follows a Rayleigh distribution.

Next we will examine the effects of off-axis illumination. We will consider a system whose aberrations are a fixed amount of pure horizontal coma. The WEF in this case is:

$$W(\boldsymbol{u}) = z_8 \cdot 3\sqrt{8}|\boldsymbol{u}|^2 \cdot u_x. \tag{2.23}$$

The simulated results of imaging a diffuser through such a system are presented in Fig. 2.6. The left column shows a schematic of incident light on the surface of the diffuser (not to scale). The middle column again shows the Fourier transform of an intensity measurement, simulated *without* using our derived forward model. The right column shows our forward model, evaluated in the region $\mathscr{U}$ where it is valid (see Eq. 2.20). We note here that off-axis illumination has the effect of shifting the two copies of the pupil function present in the Fourier transform of intensity in opposite directions by a distance that scales with the illumination angle. The result is a self-interference pattern whose fringes are determined by the underlying polynomial. Information about the odd components of the *original* WEF can be inferred from the even components of the *shifted* polynomial, which are visible in these measurements.

It is important to consider the magnitude of illumination angles used. Comparing the cases of $\theta_x = -3.0°$ and $\theta_x = -1.5°$, we can see that the former results in a larger shift and hence smaller area of overlap. In the limiting case, an illumination angle at precisely the numerical aperture of the imaging system would result in only a single point of overlap. This single value cannot represent an interference pattern. Practically, one must consider the sampling of the sensor and the NA of the imaging system when choosing illumination

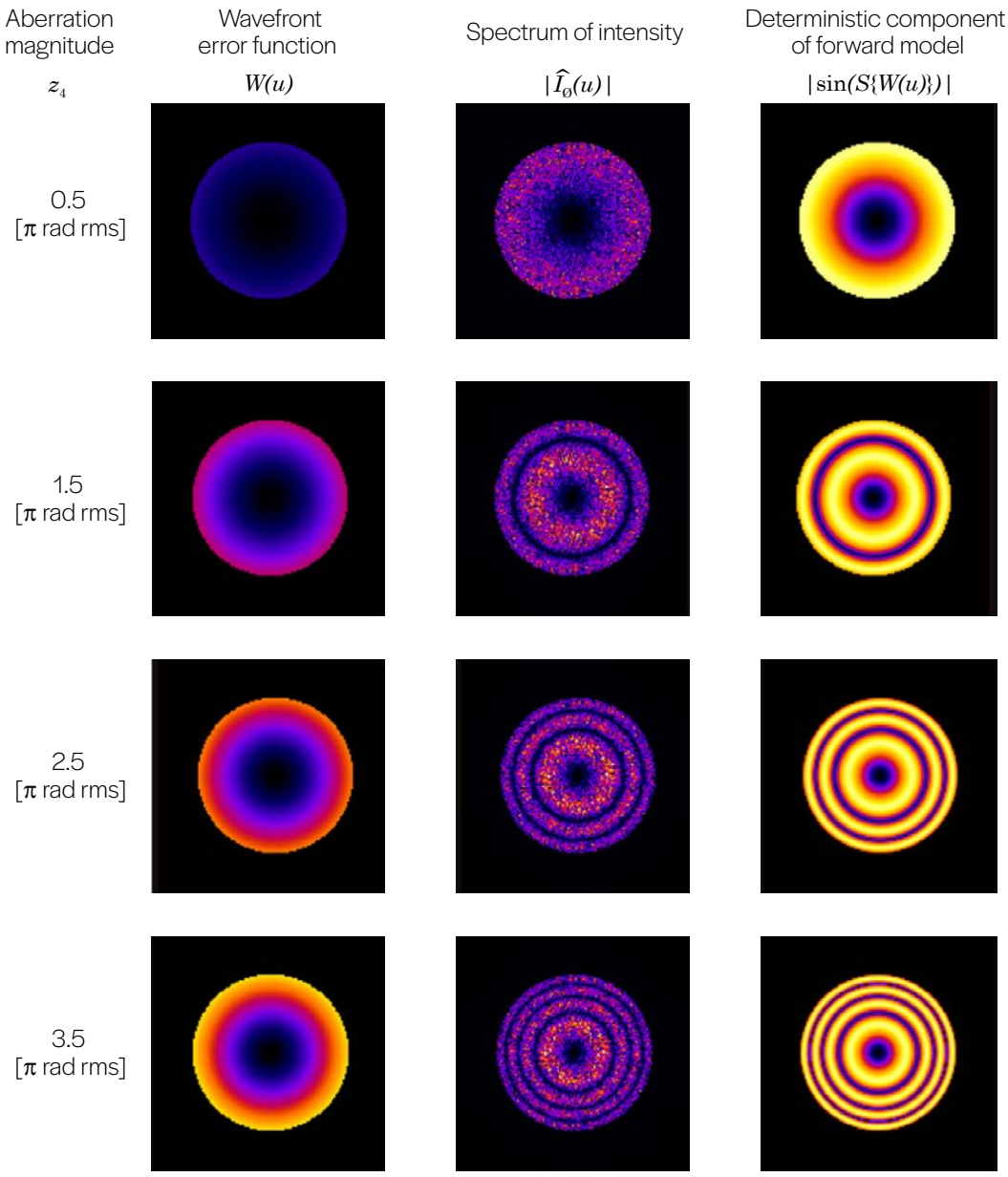

Figure 2.5: **Imaging a diffuser through a system with varying defocus and on-axis illumination.** *(left column)* The wavefront error function is given by varying magnitudes of defocus. *(center column)* Fourier spectrum of intensity measurements simulated independently of the derived forward model (Eq. 2.21). *(right column)* Deterministic sinusoidal component of the derived forward model.

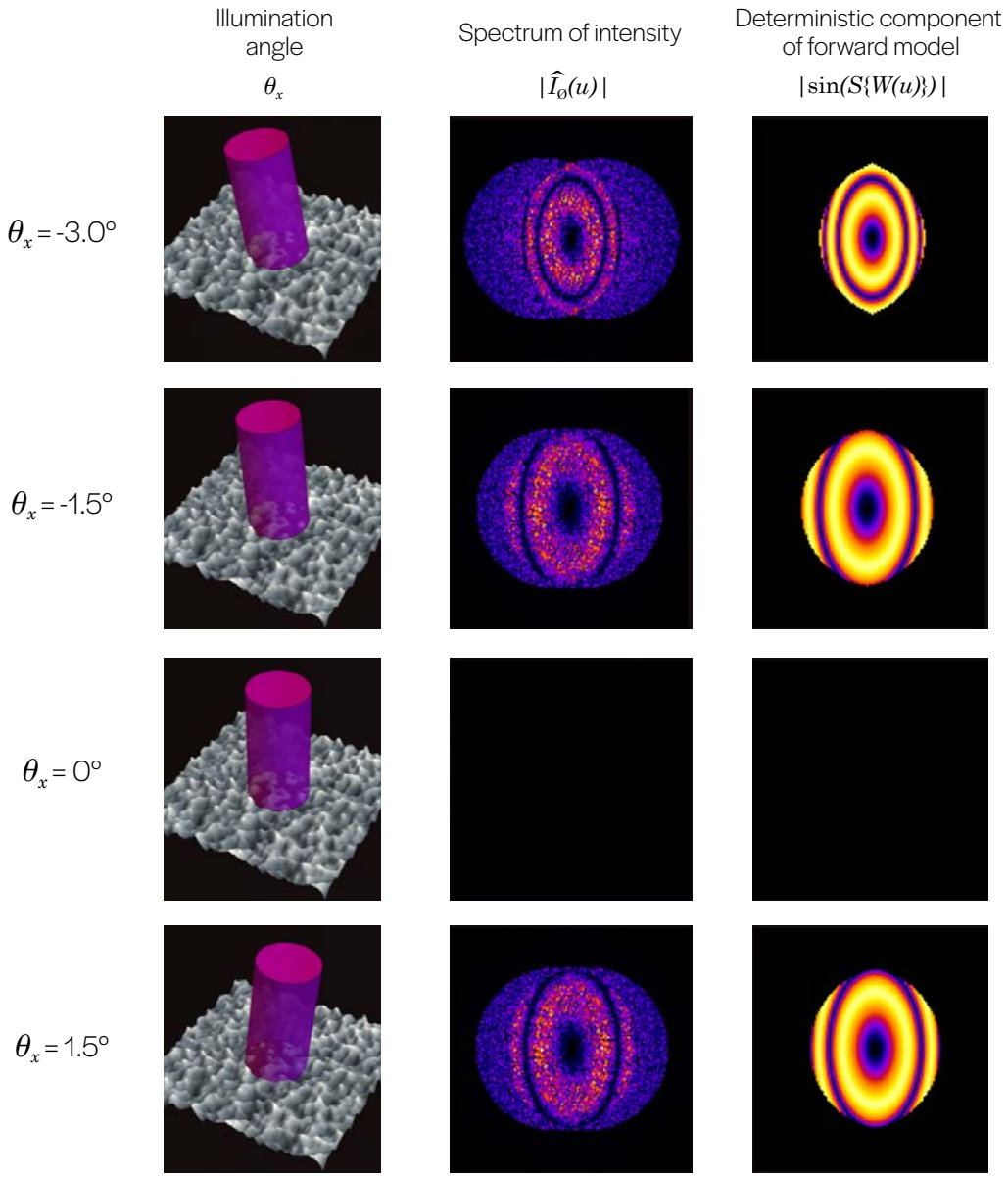

Figure 2.6: **Imaging a diffuser through a system with coma and varying illumination angle.** *(left column)* The wavefront error function is given by varying magnitudes of defocus. *(center column)* Fourier spectrum of intensity measurements simulated independently of the derived forward model (Eq. 2.21). *(right column)* Deterministic sinusoidal component of the derived forward model.

angles, such that there are enough values from which to infer the system aberrations. In general, the illumination angles used for our method are small, and can be viewed as slight perturbations with respect to on-axis illumination.

We note that no information exists in the spectrum of the on-axis measurement ($\theta_x = 0$). This is due to our assumption that the system contained pure coma, which is represented by an odd polynomial. Thus, the even part of the WEF is the zero function, and no interference occurs. We also note that the measurements for $\theta_x = -1.5^\circ$ and $\theta_x = 1.5^\circ$ have identical spectra. Because the two shifted copies of the pupil move in opposite directions in Fourier space, there is an inherent symmetry with respect to illumination angles. This is useful to consider as illumination angles that are equal but opposite do not carry new information. Practically, this case may be avoided for speed or included for redundancy.

We now consider the contribution of the diffuser phase, $\widehat{\varphi}(\cdot)$ to the forward model.The diffuser essentially behaves as an all-pass filter in expectation, but for any given realization, we can think of the diffuser as corrupting the self-interference fringe patterns of our forward model with Rayleigh-distributed multiplicative noise. The results of imaging multiple realizations of simulated diffusers (with the same statistical parameters) are shown in Fig. 2.7. Each row represents a new realization of a diffuser surface. We note that the Fourier transforms of the simulated intensity measurements all display an identical ring structure characteristic of defocus, but the function is not identically sampled in each case. We also note that, due to the nature of *multiplicative* noise, the regions of the image where the signal is low (i.e. zeros) are particularly well-defined.

As a final remark, we observe that the forward model cannot distinguish between wavefront error functions of opposite sign. Because we are only sensitive to the even part of the real-valued WEF, there is an inherent sign ambiguity such that $W(u)$ and $-W(u)$, probed under the same conditions, will generate identical Fourier spectra. This is unlikely to be prohibitive in practice, as in practical situations, it is fairly evident which of the two possible solutions is correct. In addition, external or prior knowledge can be used to rule out the solution with incorrect sign.

2.6 Summary

In this section, we laid the groundwork for modeling imaging systems, aberrations and weak-phase diffusers in the context of Fourier optics. We then derived a simplified model for the case of imaging a weak-phase diffuser, demonstrating that the wavefront error function (i.e. aberrations) is directly observable in the computed Fourier spectra of intensity measurements. We examined the various components of our derived model and their impact on the measured images. In the next chapter, we will use our derived models to formulate an inverse problem for the recovery of imaging system aberrations, and we will demonstrate its effectiveness in simulations and experimental conditions.

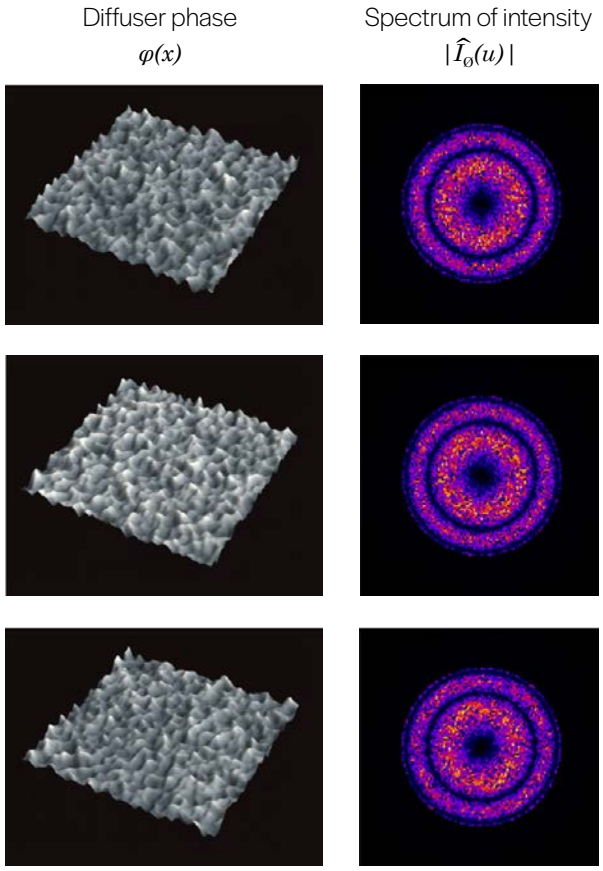

Figure 2.7: **Imaging multiple diffusers through a system with defocus and on-axis illumination.** *(left column)* The wavefront error function is given by varying magnitudes of defocus. *(right column)* Deterministic sinusoidal component of the derived forward model.

2.7 List of symbols

Symbol	Description
Coordinates	
$\boldsymbol{x} = (x, y)$	Real domain coordinates
$\boldsymbol{u} = (u_x, u_y)$	Frequency (Fourier) domain coordinates
Imaging system	
$I(\boldsymbol{x})$	Intensity
$E(\boldsymbol{u})$	Electric field
$E_o(\boldsymbol{u})$	Electric field at object (sample) plane
$E_i(\boldsymbol{u})$	Electric field at image (sensor) plane
$\widehat{P}(\boldsymbol{u})$	Optical transfer function (pupil function)
$W(\boldsymbol{u})$	Wavefront error function
$\mathrm{Circ}(\cdot)$	Characteristic function of unit disk
$\widehat{I_{\varnothing}}(\boldsymbol{u})$	DC-suppressed Fourier transform of intensity
Diffuser	
$\varphi(\boldsymbol{x})$	Phase of diffuser
$\eta(\boldsymbol{x})$	Random component of diffuser phase model
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution
Operators	
$\mathrm{Cov}\left[\cdot, \cdot\right]$	Covariance
$\mathbb{E}\left[\cdot\right]$	Expectation
$\mathrm{Re}\{\cdot\}$	Real part
$\mathrm{Im}\{\cdot\}$	Imaginary part
$\mathcal{S}\{\cdot\}$	Symmetric (even) part
$\mathcal{A}\{\cdot\}$	Anti-symmetric (odd) part

Chapter 3

Aberration recovery using weak diffusers

In this chapter, we will use the theoretical models derived previously to formulate an inverse problem for the recovery of imaging system aberrations, or the wavefront error function (WEF). We will then develop an optimization based approach for solving the inverse problem. We will show that our algorithm is able to accurately recover aberrations in both simulations and proof-of-concept visible light experiments, and we will show the results of the application of our method to an EUV imaging system. Finally, we will conclude the chapter with a discussion of error performance and feasibility for extreme ultraviolet (EUV) systems and their modern applications.

3.1 Inverse problem formulation

In the previous chapter, we derived a simplified imaging forward model by assuming a weak diffuser at the sample plane; this model is reproduced here:

$$|\widehat{I}_{\varnothing}(\boldsymbol{u})| = 2 \cdot |\widehat{\varphi}(\boldsymbol{u})| \cdot |\sin\left(\mathcal{S}\left\{W(\boldsymbol{u}+\boldsymbol{u}_0)\right\}\right)|\,,\ \boldsymbol{u} \in \mathscr{U}\,, \tag{3.1}$$

where we recall that the set $\mathscr{U}$ is given by the overlap region of two circles that are shifted based on the angle of incidence of plane waves on the diffuser. Using this model, we can formulate an inverse problem for the recovery of system aberrations. Generally, we will define a scalar loss function that penalizes mismatch between measurements and the forward model according to an appropriate metric. Such a loss function will have a small value when the parameters of the forward model accurately approximate the measurements. Hence, we will solve the inverse problem by minimizing the loss function with respect to the parameters of interest in the forward model, typically performed using gradient-based optimization. This section will outline the formulation of this problem.

Parametrization of solution space

The first step we must perform is the discretization of Eq. 3.1. While inverse problems in imaging often treat images as vectors of their pixel values, our specific application allows for a parametrization of the quantity of interest, $W(\boldsymbol{u})$, greatly reducing the dimension of the solution space and the burden of computation. In addition, many of the functions in the forward model which represent polynomial operations can be written elegantly as linear operators.

In most practical situations, we can expect that the wavefront errors arising from the manufacturing and alignment of optics are smooth functions, and so the WEF is typically expressed in a Zernike series:

$$W(\boldsymbol{u}) = \sum_{i=3}^{N} c_i \cdot Z_i(\boldsymbol{u}) = \mathbf{Z}\mathbf{c}, \tag{3.2}$$

where $\mathbf{Z}$ is a matrix whose columns are the unrolled evaluations of Zernike polynomials in OSA/ANSI order and $\mathbf{c}$ is a vector of corresponding coefficients. We note that the summation in Eq. 3.2 begins at 3 instead of 0, due to piston (degree 0) and tilt (degree 1) terms not being recoverable using our model. Consider Eq. 2.17, reproduced below:

$$\widehat{I_{\varnothing}}(\boldsymbol{u}) = i\widehat{\varphi}(\boldsymbol{u}) \left[\widehat{P^*}(\boldsymbol{u}_0)\widehat{P}(\boldsymbol{u}+\boldsymbol{u}_0) - \widehat{P}(\boldsymbol{u}_0)\widehat{P^*}(-\boldsymbol{u}+\boldsymbol{u}_0) \right]. \tag{3.3}$$

For constant $W(\boldsymbol{u})$, and hence constant $P(\boldsymbol{u})$, the bracketed term in this equation will cancel, and there will be no contrast in the intensity spectrum. For tilt terms, we note that these have degree 1 and are thus odd functions. A shift applied to an odd function can only produce even terms of lower degree than the original function. As a result, any new information brought about by off-axis illumination can only exist as a constant term, which we previously showed cannot produce intensity contrast. As such, the minimum degree of a recoverable WEF is 2, corresponding to defocus and astigmatism. For the majority of our simulations and experiments, we recover the set of Zernike coefficients corresponding to polynomials of degree 5 and lower. These are the polynomials Z_3 to Z_{21}, and hence, $\mathbf{c}$ is a vector of length 18.

While the Zernike basis is most convenient for describing 2D wavefront error functions in optical systems, it is not necessarily the simplest basis for performing operations on polynomials. As such, we will define the change-of-basis matrix, $\mathbf{E}$, which converts a vector from the Zernike basis to the standard basis defined over a finite polynomial space. The elements of this basis are the functions $\{1, x, y, x^2, xy, y^2, ...\}$. Note that there is again ambiguity in the ordering of standard basis polynomials in 2 dimensions. In this work, we will order them in increasing degree, with higher powers of x taking priority. If $\mathbf{c}$ is a vector a Zernike coefficients and $\mathbf{d}$ is a vector of standard basis coefficients, then the following hold:

$$\mathbf{d} = \mathbf{E}\mathbf{c} \quad , \quad \mathbf{c} = \mathbf{E}^{-1}\mathbf{d}. \tag{3.4}$$

Translation of polynomials as a linear operator

Now that we can represent polynomials in the standard basis, and easily convert to and from the Zernike basis, we can formulate the polynomial shift operator as a linear operation on standard basis coefficients. That is, we want a matrix representation $\mathbf{H}_k$ for the function:

$$H_{\boldsymbol{k}}\{f\} : f(\boldsymbol{x}) \mapsto f(\boldsymbol{x} + \boldsymbol{k}) \tag{3.5}$$

We note that this mapping is only linear as long as $\boldsymbol{k}$ is fixed; this can safely be assumed in our context as we attribute a single illumination angle to an acquired image or image segment.

For clarity, we will derive the construction of $\mathbf{H}_k$ for the case of polynomials in 1D, noting that the extension to 2D is simply a matter of carefully ordering polynomials in two variables and iterating over separate indices for x and y terms. Transformation of monomial terms (standard basis elements) results in expressions of the form $(x+k)^n$, which can be expanded in general using the binomial theorem. The i-th row of the matrix determines the coefficient of x^{i-1} in the output polynomial, and the j-th column acts on the coefficient of x^{j-1} in the input polynomial. Hence we have an upper triangular matrix whose entries are given by:

$$[\mathbf{H}_k]_{i,j} = \binom{j-1}{i-1} k^{j-i}, \quad j \geq i. \tag{3.6}$$

As an example, we will construct $\mathbf{H}_2$ for the space of degree 3 polynomials. Evaluation of Eq. 3.6 yields the matrix:

$$\begin{bmatrix} 1 & 1\cdot k & 1\cdot k^2 & 1\cdot k^3 \\ 0 & 1 & 2\cdot k & 3\cdot k^2 \\ 0 & 0 & 1 & 3\cdot k \\ 0 & 0 & 0 & 1 \end{bmatrix}_{k=2} = \begin{bmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 4 & 12 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now if we consider a polynomial, e.g. $f(x) = 2x^3 + 3x^2 - x + 1$, we have:

$$f(x+2) = 2(x+2)^3 + 3(x+2)^2 - (x+2) + 1 = 2x^3 + 15x^2 + 35x + 27$$

Using our matrix, we compute the same set of coefficients:

$$\begin{bmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 4 & 12 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 27 \\ 35 \\ 15 \\ 2 \end{bmatrix}. \blacksquare$$

The following code is used to compute the coefficients of a translated 2D polynomial, given the initial coefficients.

```
function [ M ] = shiftOperator( maxDeg, a, b, x0, y0 )
% M = shiftOperator( maxDeg, a, b, x0, y0 )
%
% Creates a matrix operator, M, which maps coefficients of a polynomial of
% degree maxDeg to coefficients of the transformed polynomial
% g(x,y) = f(ax + x0, by + y0) with respect to the standard basis

% Compute size of matrix
N   = 1/2 * (maxDeg + 1) * (maxDeg + 2);
M   = zeros(N);
I   = eye(N);

% Populate columns by applying transformation to standard basis elements
for i = 1 : N
    M(:,i)  = getCoefs2DShiftedPoly( I(:,i), maxDeg, x0, y0, a, b );
end
end
```

```
function [ s ] = getCoefs2DShiftedPoly( c, maxDeg, dx, dy, sx, sy )
% dx, dy: shifts in x, y
% sx, sy: scale in x, y
s = zeros(size(c));
N = maxDeg;

for i = 0 : N
    for j = 0:(N-i)
        idx = (i+j)*(i+j+1)/2 + j + 1;
        coef = c( idx );
        for n = 0 : i
            for m = 0 : j
                idx0 = (n+m)*(n+m+1)/2 + m + 1;
                s(idx0) = s(idx0) + ...
                    coef*nchoosek(i,n)*sx^n*dx^(i-n) * ...
                    nchoosek(j,m)*sy^m*dy^(j-m);
            end
        end
    end
end
end
```

Additional linear operators

Once the shifted polynomial coefficients have been computed, there are two remaining linear operations in the forward model, Eq. 3.1, that act on them. The first will take the even part of the function represented by the coefficients, and the second will evaluate the polynomial on the normalized image frequency space grid, which is defined by the physical dimensions of the detector and the NA of the imaging system.

Taking the even part of a polynomial is straightforward since the elements of the standard basis are all either even or odd functions. Hence, we can define the matrix $\mathbf{S}$ as a diagonal matrix with 1's corresponding to the even basis elements and 0's corresponding to the odd basis elements. As an example, we will consider the space of 2D polynomials of degree at most 2. The standard basis elements for this space are $\{1, x, y, x^2, xy, y^2\}$ in this order. From the parity of these functions, it follows that:

$$\mathbf{S} = \operatorname{diag}\left(\begin{bmatrix}1 & 0 & 0 & 1 & 1 & 1\end{bmatrix}\right).$$

Note that in a practical implementation, the rows and columns zeroed by $\mathbf{S}$ can be omitted from computation for speed.

Evaluation of polynomials onto the image grid is also a straightforward procedure. We construct the matrix $\mathbf{\Psi}$, known as the evaluation map, whose columns are the vectorized evaluations of the standard basis elements. We note that these columns must be ordered to correspond with the coefficient vector (in our case, increasing degree with priority given to x). We also note that vectorizing the image grid may be done in either row-major or column-major fashion, but consistency must be maintained throughout the application of the technique.

Handling uncertainty in diffuser phase

We will now turn our attention to the contribution of the sample being imaged – in our case, a diffuser – to the Fourier spectrum of the recorded image. In Eq. 3.1, this is given by the term $|\widehat{\varphi}(\boldsymbol{u})|$. We recall from the previous chapter (Eq. 2.14) that the Fourier space model for the diffuser consists of a deterministic Gaussian window related to the average lateral size of surface features and an instance of i.i.d. Rayleigh-distributed multiplicative noise. We will write this as:

$$|\widehat{\varphi}(\boldsymbol{u})| = \widehat{\varphi_d}(\boldsymbol{u}) \cdot \widehat{\eta}(\boldsymbol{u}), \quad \widehat{\eta}(\boldsymbol{u}) \sim \text{Rayleigh}(\sigma), \tag{3.7}$$

where $\widehat{\varphi_d}(\boldsymbol{u})$ represents the deterministic Gaussian window and $\widehat{\eta}$ the noise. The multiplicative nature of the noise in our problem makes recovery of the aberration parameters challenging. A rigorous formulation of an inverse problem given this forward model would typically involve the derivation of a likelihood function containing logarithms [4]. However, because the deterministic component of our signal is the magnitude of a sinusoid (which has zeros), doing so would result in an unstable reconstruction algorithm. Instead, we approximate the noise using an additive model, treating the true Rayleigh distribution as a Gaussian distribution of equivalent mean and variance:

$$\widehat{\eta}(\boldsymbol{u}) \approx \mathbb{E}\left[\widehat{\eta}(\boldsymbol{u})\right] + \xi(\boldsymbol{u}) \ , \quad \xi(\boldsymbol{u}) \sim \mathcal{N}\left(0, \frac{4-\pi}{2}\sigma^2\right). \tag{3.8}$$

Using an unbiased estimator for the Rayleigh parameter, σ [71]:

$$\widehat{\sigma} = \left(\sum_{j=1}^{K} \frac{\eta(\boldsymbol{u})^2}{2K} \right)^{1/2} \frac{4^K K!(K-1)!\sqrt{K}}{(2K)!\sqrt{\pi}} \approx \left(\sum_{j=1}^{K} \frac{\eta(\boldsymbol{u})^2}{2K} \right)^{1/2} \frac{e^1\sqrt{K}}{\sqrt{K-1}} \left(\frac{K-1}{K} \right)^K, \quad (3.9)$$

the expected value of the Rayleigh noise can be computed as:

$$\mathbb{E}\left[\widehat{\eta}\right] = \widehat{\sigma} \left(\frac{\pi}{2} \right)^{1/2}. \quad (3.10)$$

Note that in computing the unbiased estimate of the Rayleigh distribution parameter, an approximation is made using Stirling's formula [17] to numerically evaluate factorials.

Discrete forward model

With all of our major operators defined and the noise reformulated, we can proceed with the discretization of our forward model for application to image data. Upon this discretization, we will drop the spatial coordinate $\boldsymbol{u}$. We have:

$$|\widehat{\boldsymbol{I}_{\varnothing}}| = 2 \cdot \mathbf{1}_{\mathscr{U}_{\boldsymbol{k}}} \cdot \widehat{\boldsymbol{\varphi}_{\boldsymbol{d}}} \cdot \left(\mathbb{E}\left[\widehat{\eta}\right] + \xi \right) \cdot \left| \sin\left(\mathbf{\Psi S H}_{\boldsymbol{k}} \mathbf{E c} \right) \right|, \quad (3.11)$$

where $\cdot$ indicates element-wise multiplication. Upon expansion of the parenthetical term, we note that we are left with heteroscedastic additive noise [70], meaning that its variance changes across space. This variance is signal-dependent, proportional to $|\sin\left(\mathbf{\Psi S H}_{\boldsymbol{k}} \mathbf{E c}\right)|^2$. However, because our model function is sinusoidal with a range of $[0, 1]$, we can further approximate this term as simple additive Gaussian noise. Practically, the reconstruction errors introduced by doing so are negligible, especially for the case of small aberrations, where the lack of zeros in the signal (caused by wrapping of strong wavefront phase) results in more uniformity of noise statistics. As a result, we have the discrete forward model:

$$|\widehat{\boldsymbol{I}_{\varnothing}}| \approx \mathbf{1}_{\mathscr{U}_{\boldsymbol{k}}} \cdot \left(2 \cdot \widehat{\boldsymbol{\varphi}_{\boldsymbol{d}}} \cdot \mathbb{E}\left[\widehat{\eta}\right] \cdot \left| \sin\left(\mathbf{\Psi S H}_{\boldsymbol{k}} \mathbf{E c} \right) \right| + \xi \right). \quad (3.12)$$

Objective (Loss) function

To solve for the system aberrations – the vector $\mathbf{c}$ in Eq. 3.12 – we will write down an objective (or loss) function based on our forward model that aims to penalize model mismatch. Minimization of this objective function will yield the aberration coefficients which minimize the error between the measured data and model predictions. In our case, we will use a mean squared error (MSE) objective function. We denote the Fourier transform of the measurement acquired under the k-th illumination angle as $\widehat{\boldsymbol{m}}_k$. Then the loss function is:

$$\mathcal{L}(\mathbf{c}) = \sum_{k=1}^{N} \mathbf{1}_{\mathscr{U}_{\boldsymbol{k}}} \cdot \| \widehat{\boldsymbol{m}}_k - 2 \cdot \widehat{\boldsymbol{\varphi}_{\boldsymbol{d}}} \cdot \mathbb{E}\left[\widehat{\eta}\right] \cdot \left| \sin\left(\mathbf{\Psi S H}_{\boldsymbol{k}} \mathbf{E c} \right) \right| \|_2^2 \quad (3.13)$$

and the inverse problem is written:

$$\mathbf{c}^\star = \arg\min_{\mathbf{c}} \mathcal{L}(\mathbf{c}), \tag{3.14}$$

assuming that a total of N images are acquired.

3.2 Solving the inverse problem

Having formulated the inverse problem, Eq. 3.14, we will now discuss methods for solving for the optimal aberration coefficients $\mathbf{c}^\star$. Generally, this involves the use of iterative, gradient-based optimization techniques to minimize the loss function, Eq. 3.13. However, there are two major caveats to this approach that we must keep in mind.

First, gradient-based optimization is only guaranteed to converge to an optimal solution when the objective function is convex. A convex function, f, has the property that any local minimum of f is also the global minimum of f. The oscillatory nature of the sine function in our forward model produces local extrema, and thus results in a non-convex loss function. Second, in order to use gradient-based methods, the gradient of the loss function must be well-defined for all values in the solution space for the variable $\mathbf{c}$. However, due to the absolute value function in our forward model, the point $\mathbf{c} = 0$ does not have a well-defined gradient.

The latter can be handled by generalizing our notion of a gradient to a set of "sub-differentials" [34]. Where a differentiable function will have a well-defined gradient related

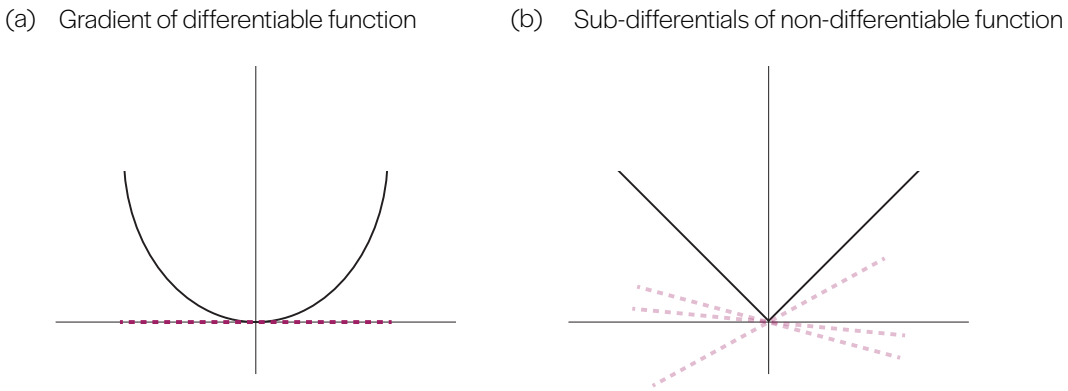

Figure 3.1: **Derivatives, tangent lines and sub-differentials.** (a) A differentiable 1D function has a unique tangent line associated with its derivative at zero. (b) A non-differentiable function has a set of sub-differentials associated with all lines that do not intersect the graph except at zero.

to a unique tangent hyperplane to its graph, a function that is not strictly differentiable at a point will have a set of hyperplanes that *only* intersect the graph of the function at that point. All hyperplanes in this set are associated with a sub-differential. Example sub-differentials of the absolute value function in a single variable are shown in Fig. 3.1b. In the case of our forward model, we simply need to choose a value within the sub-differential set, such that a gradient-based algorithm may proceed if an element of $\mathbf{c}$ is zero. Typically, we choose zero as our sub-differential. Practically speaking, it is improbable that the algorithm will ever encounter this situation, as long as we do not initialize with any zero values in $\mathbf{c}$.

The former caveat is more challenging in that it cannot be completely addressed. We must accept that no guarantee of global optimality of recovered solutions exists when applying convex optimization techniques to non-convex objective functions. Despite a lack of provable optimality, we can still increase the robustness of our algorithm to local minima by using accelerated gradient techniques, in which an update is not only on the gradient but also previous steps taken. This can be thought of as an update based on "momentum". Additionally, we can use multiple random initializations spread throughout the solution space, such that some may avoid paths which converge to local minima. One would then select the solution which converged to the minimum overall cost. In our case, we use 100 random initializations to solve for the set of coefficients, $\mathbf{c}$.

Derivation of loss function gradient

Here we derive the gradient of our loss function, Eq. 3.13, which is the crux of our gradient descent algorithm. We will start by making some modifications to the loss function. First, we can simplify the expression by combining matrices that represent known quantities into a single matrix:

$$\mathbf{A}_{\boldsymbol{k}} = \boldsymbol{\Psi}\mathbf{S}\mathbf{H}_{\boldsymbol{k}}\mathbf{E}. \tag{3.15}$$

Additionally, rather than including $\widehat{\boldsymbol{\varphi}_{\boldsymbol{d}}}$ as part of the forward model, we can divide the measurement by this (deterministic) quantity in a process known as "noise whitening" [31]. This is a safe operation since sufficiently small surface roughness features ensure a wide full-width at half-maximum (FWHM) of $\widehat{\boldsymbol{\varphi}_{\boldsymbol{d}}}$, so that we are generally dividing by values close to 1. We will define a whitened measurement as:

$$\widehat{\boldsymbol{m}}_k^{'} = \frac{\widehat{\boldsymbol{m}}_k}{\widehat{\boldsymbol{\varphi}_{\boldsymbol{d}}}}, \tag{3.16}$$

where division is performed element-wise. Finally, for simplicity of differentiation, we will rewrite $|\sin(x)|$ as $\sqrt{\sin^2(x)}$. The loss function is then rewritten:

$$\mathcal{L}(\mathbf{c}) = \sum_{k=1}^{N} \mathbf{1}_{\mathscr{U}_{\boldsymbol{k}}} \cdot \|\widehat{\boldsymbol{m}}_k{}' - 2 \cdot \mathbb{E}\left[\widehat{\eta}\right] \cdot \sqrt{\sin^2\left(\mathbf{A}_{\boldsymbol{k}}\mathbf{c}\right)}\|_2^2. \tag{3.17}$$

To simplify notation, we will define:

$$\boldsymbol{e}_k = \widehat{\boldsymbol{m}}_k{}' - 2 \cdot \mathbb{E}\left[\widehat{\eta}\right] \cdot \sqrt{\sin^2\left(\mathbf{A_k c}\right)}. \tag{3.18}$$

Then we have [58]:

$$\begin{aligned}
\nabla_{\mathbf{c}} \mathcal{L} &= \sum_{k=1}^{N} -2 \cdot 2 \cdot \mathbb{E}\left[\widehat{\eta}\right] \cdot \boldsymbol{e}_k^{\intercal} \, \frac{d}{d\mathbf{c}} \left[\sqrt{\sin^2\left(\mathbf{A_k c}\right)}\right] \\
&= -4 \cdot \mathbb{E}\left[\widehat{\eta}\right] \cdot \sum_{k=1}^{N} \boldsymbol{e}_k^{\intercal} \, \operatorname{diag}\left[\frac{\sin\left(\mathbf{A_k c}\right) \cdot \cos\left(\mathbf{A_k c}\right)}{\sqrt{\sin^2\left(\mathbf{A_k c}\right)}}\right] \mathbf{A_k} \\
&= -4 \cdot \mathbb{E}\left[\widehat{\eta}\right] \cdot \sum_{k=1}^{N} \boldsymbol{e}_k^{\intercal} \, \operatorname{diag}\left[\frac{\sin\left(\mathbf{A_k c}\right) \cdot \cos\left(\mathbf{A_k c}\right)}{|\sin\left(\mathbf{A_k c}\right)|}\right] \mathbf{A_k} \\
&= -4 \cdot \mathbb{E}\left[\widehat{\eta}\right] \cdot \sum_{k=1}^{N} \boldsymbol{e}_k^{\intercal} \, \operatorname{diag}\left[\operatorname{sgn}\left(\sin\left(\mathbf{A_k c}\right)\right) \cdot \cos\left(\mathbf{A_k c}\right)\right] \mathbf{A_k},
\end{aligned} \tag{3.19}$$

where $\operatorname{sgn}(\cdot)$ is the sign function, returning ± 1.

Calibration of known quantities

In order to fully formulate our inverse problem and solve for the aberration coefficients in practice, certain quantities in the model must be defined through a separate calibration process. Fortunately, the required information can be recovered from the same or similar image data to that of the aberration estimation process, and so calibration adds minimal overhead to our method. The specific quantities we must specify are the plane wave illumination angles used in each measurement (used in defining $\mathbf{H_k}$), and the parameters defining the surface roughness characteristics, $\widehat{\boldsymbol{\varphi}_{\boldsymbol{d}}}$ and $\mathbb{E}\left[\widehat{\eta}\right]$. In this section we will outline these calibration procedures.

First, we discuss the calibration of illumination angles for each measurement. Practically, a desired illumination angle is either manually adjusted or input to a system controller; however, any deviation between what is physically realized and this desired input results in aberration reconstruction errors. As a result, we generally infer illumination angles from the acquired data, using our desired values as an approximate guess. As shown in Fig. 2.6, the Fourier transform of our weak-phase speckle images under off-axis illumination is highly structured, demonstrating two circles of equal radius which overlap. We also know their centers depart from the origin with equal distance and opposite directions. This well-defined structure informs image processing algorithms that can identify the placement of these circles and thus, recover the illumination angle.

For large images, for which the circle boundaries are clearly defined with sufficient sampling, a method for recovering illumination angles is given in [14]. At a high level, this method operates by evaluating a set of candidate circle center points against the image

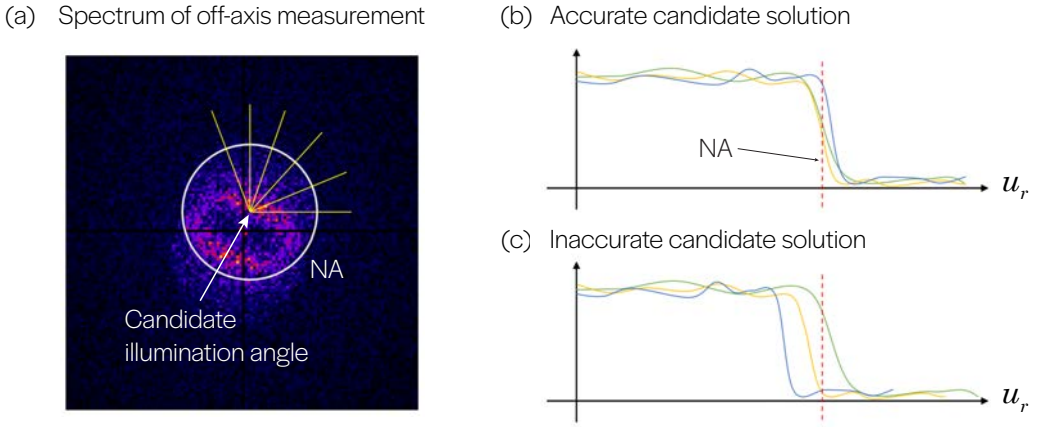

Figure 3.2: **Illumination angle estimation, method #1.** (a) A candidate illumination angle defines a point in 2D Fourier space at which our image formation model predicts a circle with radius defined by the system NA. (b) If the candidate is accurate, interpolation of the Fourier spectrum along radial lines will have a sharp and consistent change coincident with the NA-defined radius. (b) If the candidate is inaccurate, this change will be inconsistent.

data. The image spectrum is interpolated along radial lines originating at the candidate center. For an erroneous candidate center, these radial evaluations will reach the circle's radius (determined by the imaging system NA) at varying positions, while for the true center, all radial evaluations should experience a sharp decrease precisely at the radius. This method is visualized in Fig. 3.2.

While the previous method is reliable when there are plenty of pixels which constitute the circle boundaries, interpolation along radial lines is more prone to errors when using images with a pupil of much lower resolution. Use of an incorrect angle of illumination results in an incorrect crop of the interference region, and thus an incorrect interpretation of the interference pattern, as shown in Fig. 3.3. Hence, we often use a different procedure for finding the circle centers which does not require interpolation of the spectrum.

We begin with the observation that the general spectrum of an off-axis speckle measurement ($\boldsymbol{u}_0 \neq 0$) will have exactly two axes of symmetry, as shown in Fig. 3.4a. The projection of the domain of interest along one axis will always be equal to the diameter of one circle, which can be derived from the system NA; we will refer to this axis as the minor axis. The major axis then defines the line along which the two circle centers lie, and the additional information of its total length fully specifies the angle of incident illumination. The problem is thus reduced to finding axes of symmetry and projection lengths. For the former, we will take advantage of the relationship between axes of symmetry and moments of inertia for a spatial distribution of mass.

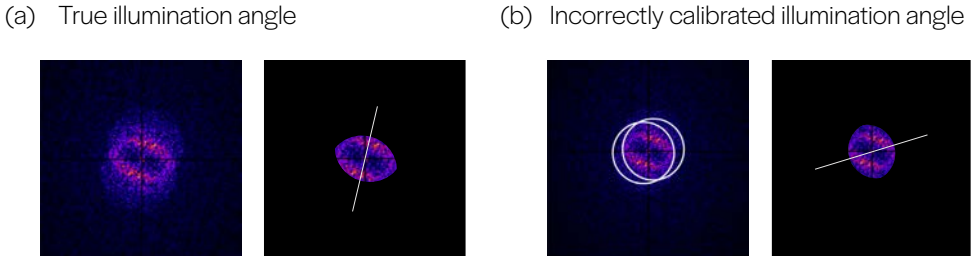

Figure 3.3: **Incorrectly calibrated illumination angle.** (a) The true angle of incident illumination correctly identifies and crops the domain of the interference pattern. (b) Using an incorrect angle results in a different fringe pattern, and eventually, a different recovered aberration polynomial.

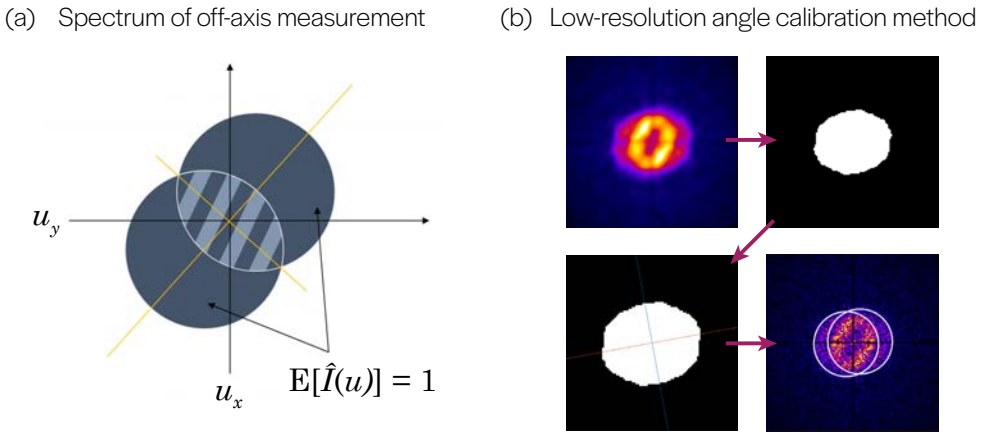

Figure 3.4: **Illumination angle estimation, method #2 (for low-resolution pupil images).** (a) The general structure of the spectrum of an intensity image of a weak diffuser will have two axes of symmetry, shown in yellow. (b) After smoothing and thresholding the intensity spectrum, we identify axes of symmetry and major/minor axis lengths by treating the binary image as a thin mass with moments of inertia. This informs the placement of circle centers, and hence, illumination angles.

We first perform Gaussian smoothing of the Fourier spectrum of the intensity measurement due to its noisy nature, owing to the surface roughness being imaged. This ensures that the domain of interest upon thresholding will be contiguous. We then threshold the smooth image to acquire a binary representation of the domain of interest; this should roughly re-

semble two circles. Various image processing techniques may be applied at this stage to refine the binarization process. For example, a combination of erosion and dilation may be used to exclude outliers near the edges of the domain, and a flood-fill algorithm may be used to fill in any holes resulting from applying the threshold to the interference region. We note that threshold selection can be difficult to automate, but images acquired under similar circumstances using the same imaging system will generally have a consistent threshold. The threshold must be able to distinguish the regions of the image where signal is present (i.e. within the NA-defined circles) and background noise which generally exists outside this domain.

Once the image is binarized, it is treated as a spatial distribution of mass, and its moments of inertia are found by computing the eigenvectors of the following matrix:

$$I = \begin{bmatrix} \int u_y^2 \, dS & -\int u_x u_y \, dS \\ -\int u_x u_y \, dS & \int u_x^2 \, dS \end{bmatrix}. \tag{3.20}$$

The eigenvectors and their corresponding eigenvalues can be used to discern the major and minor axes of symmetry. If we define θ as the angle between the major axis and the u_x-axis, and we define the length of the major axis (interval occupied by the projection of the image onto the major axis) as $2 + 2r$, then the coordinate (r, θ) specifies the incident illumination angle exactly. We note that there is sign ambiguity in both of the angle calibration methods presented. However, as this angle is only used to crop the data to the interference region, this does not affect our aberration recovery technique.

We will now discuss the calibration of the diffuser surface roughness, in order to specify $\widehat{\boldsymbol{\varphi}_{\boldsymbol{d}}}$ and $\mathbb{E}[\widehat{\eta}]$. Unlike the broader class of aberration measurement techniques which rely on known test targets, our technique only requires a *statistical* description of the target which is much easier to calibrate. We acquire one speckle image with the surface roughness significantly out of focus to ensure that the defocus aberration term will be the dominant component of the wavefront error function. Using the Fourier spectrum of this single image, $\widehat{\boldsymbol{m}}_{\text{cal}}$, we solve the following simplified inverse problem:

$$\{\alpha^\star, \beta^\star, z^\star\} = \arg\min_{\alpha,\beta,z} \|\widehat{\boldsymbol{m}}_{\text{cal}} - \alpha \cdot \exp\left\{-i\beta\|\boldsymbol{u}\|^2\right\} \cdot |\sin\left(z\|\boldsymbol{u}\|^2\right)|\|^2, \tag{3.21}$$

where the scalar parameters α, β and z fit both the chirp pattern caused by significant defocus and the Gaussian decay due to surface roughness feature size. Due to the small number of relevant parameters in this non-convex optimization problem, we find that using a derivative-free optimization technique, such as the Nelder-Mead simplex method [82], is faster and more robust than gradient methods. Once the optimal set of parameters is found, we have:

$$\widehat{\boldsymbol{\varphi}_{\boldsymbol{d}}} = \alpha \cdot \exp\left\{-i\beta\|\boldsymbol{u}\|^2\right\}, \tag{3.22}$$

A comparison of an evaluation of the simplified forward model to a simulated Fourier spectrum is shown in Fig. 3.5a. We then use this fitted model to isolate the residual noise values

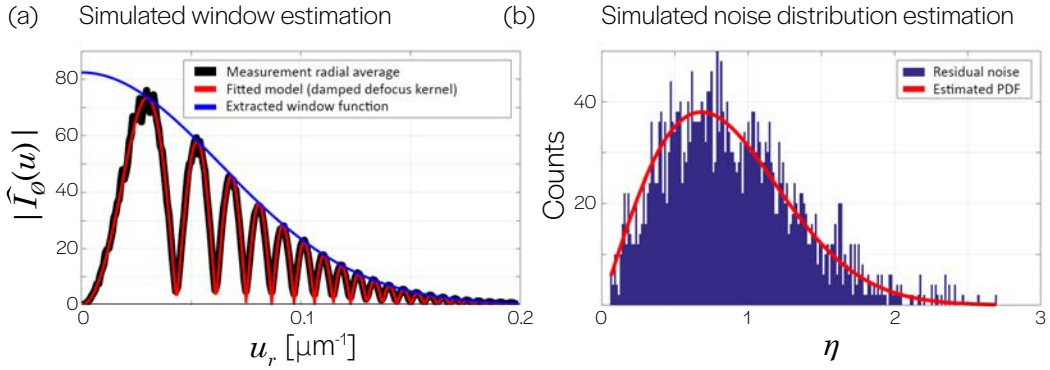

Figure 3.5: **Diffuser calibration process.** (a) The deterministic window function, $\widehat{\varphi_d}$, is fit alongside a defocus model for a significantly defocused, on-axis illuminated calibration image. (b) Division by the window function (where stable) results in a set of i.i.d. samples of Rayleigh noise, from which the distribution parameter is estimated. Note that fitting is performed in 2D, and radial averaging is only used for visualization.

and thus characterize the distribution of the Rayleigh white noise. We divide the spectrum of the measurement by the deterministic terms where division is numerically stable to obtain independent samples of white noise drawn from the Rayleigh distribution:

$$\frac{\widehat{\boldsymbol{m}}_{\text{cal}}}{2\,|\widehat{\boldsymbol{\varphi}_d}|\;|\sin\left(z\|\boldsymbol{u}\|^2\right)|} = \boldsymbol{\eta} \sim \text{Rayleigh}(\rho). \tag{3.23}$$

Generally, we perform this division wherever evaluation of the sinusoidal component of the model does not result in values close to zero. Once we obtain this set of i.i.d. Rayleigh noise, the distribution parameter is estimated using Eq. 3.9 and Eq. 3.10. An example of this distribution characterization is shown in Fig. 3.5b. We note that the parameter α and the expected value of the noise are both multiplicative components in the full imaging forward model, and thus they do not have precise physical meaning themselves. The product of these parameters accounts for all phenomena that scale the sinusoidal fringe pattern, including surface roughness characteristics, image exposure time and scaling conventions in numerical Fourier transforms.

With the constant factors of our forward model calibrated and the loss function gradient derived, we can proceed to solve our optimization problem via gradient descent.

3.3 Recovery of simulated aberrations

In this section, we will apply the techniques outlined in the previous section to the recovery of aberrations in simulated imaging systems. We will simulate (without using the weak phase approximation) a set of speckle images acquired using an aberrated imaging system, from which we will attempt to recover the wavefront error function. This set of consists of a number of images acquired with the diffuser surface in focus and distinct illumination angles and one additional image with the diffuser surface significantly out of focus and on-axis illumination for calibration, as described in the previous section. In this section, we consider reconstruction under ideal circumstances and do not account for noise due to the sensor. We will assume that the imaging system aberrations can be described using a 5-th order Zernike polynomial; since piston and tilt do not influence our forward model they are irrecoverable, and we are hence solving for a vector, $\mathbf{c}$, of 18 coefficients.

We define our real and Fourier coordinate systems and normalize Fourier space using the known NA of the system; these coordinate systems are used to define the matrices of Eq. 3.13. We then apply our diffuser calibration procedure to the defocused image and our angle estimation procedure to the remaining images. With the model and inverse problem fully specified, we are ready to solve for the wavefront error function coefficients. Due to the non-convexity of the problem, we choose 100 points within a bounded norm at which to initialize gradient descent. Furthermore, we use backtracking line-search [82] to improve speed of convergence, and perhaps, robustness to local minima. We run 200 iterations of gradient descent with backtracking line-search for each of the 100 initialization points. Of our 100 converged solutions, we report the one associated with the minimum cost as defined by the loss function.

An example of aberration recovery performed on simulated data is shown in Fig. 3.6. Four speckle images such as Fig. 3.6a are generated by an aberrated imaging system with wavefront error function (WEF) given by Fig. 3.6b. The peak-valley (PV) wavefront deviation of the simulated aberrations is approximately 1.6 rad rms. The DC-suppressed Fourier spectra of the simulated intensity measurements are shown in Fig. 3.6c, and their whitened and cropped counterparts are shown in Fig. 3.6d. Note that the whitening process results in larger values near the edges of the pupil, where the Gaussian decay due to surface feature size most significantly impacts the spectrum. Also note that for off-axis measurements, the non-interference regions are cropped out as they do not contain information about the system aberrations. The result of evaluating the forward model using the true WEF is shown in Fig. 3.6e. Note that these images closely resemble the measurements albeit without the Rayleigh noise caused by the rough surface. The close agreement (3.11% relative error) between the recovered coefficients and the true coefficients is shown via evaluation of the forward model in Fig. 3.6f and by plotting the coefficient values in Fig. 3.6g.

Note that we did not introduce sensor noise into our simulated measurements, which implies that we do not achieve a perfect reconstruction even in ideal circumstances. This is again due to the noisy nature of the spectra due to the diffuser surface. The *expectation* of our measurements is a smooth interference pattern which can uniquely identify the WEF;

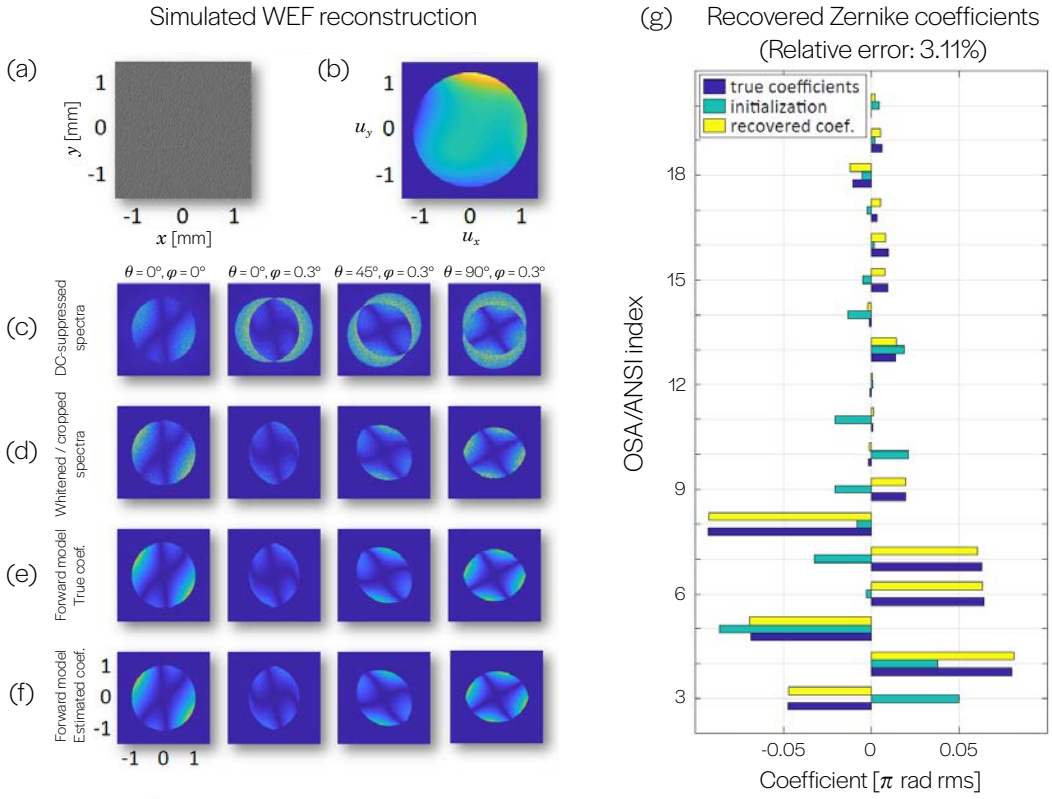

Figure 3.6: **Wavefront error function recovery from simulated data.** (a) Example intensity measurement, simulated without using simplified weak-phase forward model. (b) Wavefront error function composed of 18 random Zernike coefficients (OSA/ANSI indices 3-20). (c)) Fourier spectra of 4 intensity images with illumination angles described by azimuthal angle θ and deflection angle (from optical axis) ϕ. (d) Processed spectra, $\boldsymbol{m}_k^{'}$, as defined in Eq. 3.16; the spectrum has been whitened by removing a Gaussian decay and cropped to the interference region. (e) Evaluation of the imaging forward model using the true Zernike coefficients. (f) Evaluation of the forward model using the recovered Zernike coefficients. (g) Recovered coefficients demonstrate accurate recovery (3.11% relative error) of 5th order polynomial using four input images for a particular initialization.

however, practically, the Rayleigh noise randomly corrupts the information contained in the spectra, leading to reconstruction errors. These errors can be mitigated by averaging images, which we will describe in more detail at the end of this chapter.

3.4 Experimental aberration recovery: visible light

To demonstrate the accuracy of our method experimentally, we constructed a visible-light imaging system and conducted a series of proof-of-concept experiments. A diagram of the experimental setup is shown in Fig. 3.7. The key features of the system are coherent, steerable laser illumination, an index-matched holographic diffuser in the object plane, and a deformable mirror (DM) in the Fourier plane to generate ground-truth wavefront error functions. An off-the-shelf diffuser generally does not satisfy our conditions for a weak-phase diffuser, and so we index match its rough surface using oil under a glass coverslip (see Fig. 3.8). It is important for the deviation in refractive index between the diffuser and the oil to be small, so that the weak object approximation is valid. In our experiment, we have $n_{\text{diffuser}} = 1.5805$ and $n_{\text{oil}} = 1.5890$. Additionally, we assume aberrations inherent to the $4f$ system are negligible with respect to those introduced by the DM.

We applied a series of aberration functions to the DM and acquired speckle images from 3 different illumination angles, which we used to recover the system aberrations via our

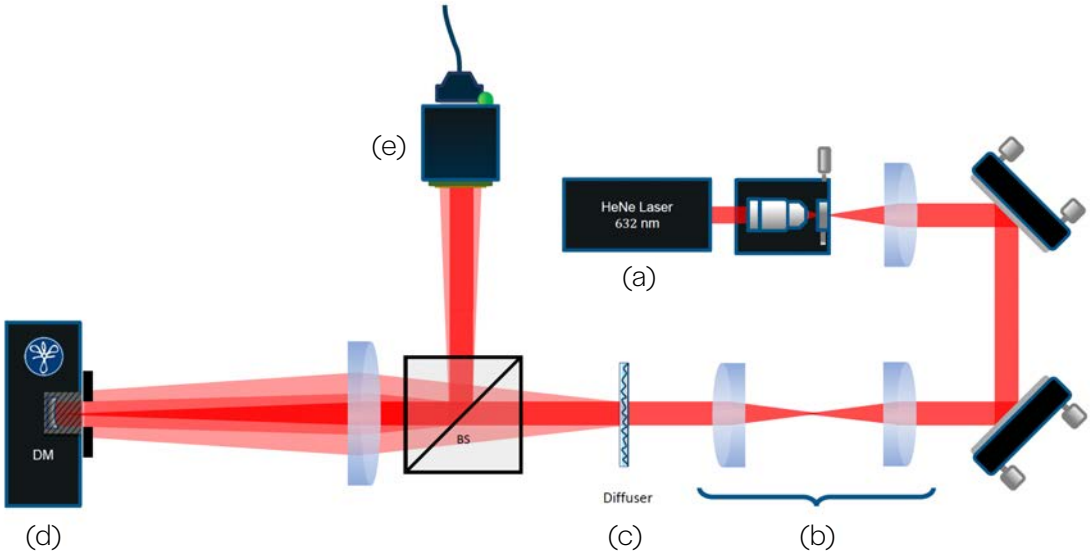

Figure 3.7: **Setup for experimental aberration recovery in a visible-light system.** (a) Melles-Griot HeNe laser (632 nm). (b) $4f$ system with mirror tilt used to control illumination angle at the object (diffuser) plane. (c) 10° holographic diffuser (Edmund Optics, #54-493) index-matched with oil. (d) Deformable mirror (Iris AO PTT111, 7 mm pupil diameter, gold-coated) in Fourier plane to introduce controlled aberrations in a unit-magnification imaging system. (e) ThorLabs DCC1240C CMOS camera (1280×1024 pixels, 5.3 μm pixels).

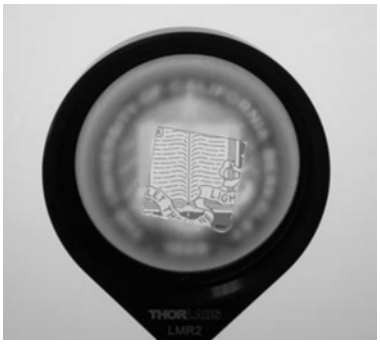

Figure 3.8: **Index-matched holographic diffuser.** Holographic diffuser ($n_{\text{diffuser}} = 1.5805$) with index matching oil ($n_{\text{oil}} = 1.5890$) underneath a square glass coverslip.

reconstruction algorithm; this result was compared to our original inputs to the DM. Each aberration function consisted of a single Zernike coefficient (OSA/ANSI indices 3-9), and our reconstruction algorithm assumed the solution was a general 2D polynomial of order 3. For this experimental data, we modified the image spectrum processing procedure to remove a small neighborhood of low frequencies rather than simply suppressing the DC value, since correlated surface roughness produces a peak of finite width.

The results of these experiments are shown in Fig. 3.9. In the left column, we have recovered Zernike coefficients plotted against the recovered coefficients; ideally, these plots should be close to the line $y = x$. To the right we show examples of processed speckle image spectra that are used in our reconstruction algorithm, along with evaluations of the forward model using the converged coefficients. Note the visual correspondence between the measured information and the predicted fringe patterns. Similar performance is achieved for Z3 and Z5; however, results for Z6 are somewhat degraded and those for Z7 are significantly worse. Differences in performance among the aberrations appear to be related to the fringe contrast in experimental measurements. We see that Z3 and Z5 have high contrast fringe patterns, Z6 has lower contrast, and Z7 has substantially less contrast, particularly at the edge of the pupil. In all cases, Fig. 3.9 shows agreement between the measured and best-fit intensity spectra in areas with high fringe contrast, demonstrating the accuracy of our model and the capabilities of our solver.

Despite successfully demonstrating our modeled phenomenon and recovering aberrations, there are a few hardware limitations of our experiment that are worth noting. These limitations are due to the design of the deformable mirror, which consists of hexagonal reflective panels with independent piston, x-tilt and y-tilt actuators. An immediate consequence is that our desired aberration functions are experimentally realized as the closest piecewise-linear function within the reachable set of the actuators. This, along with some amount

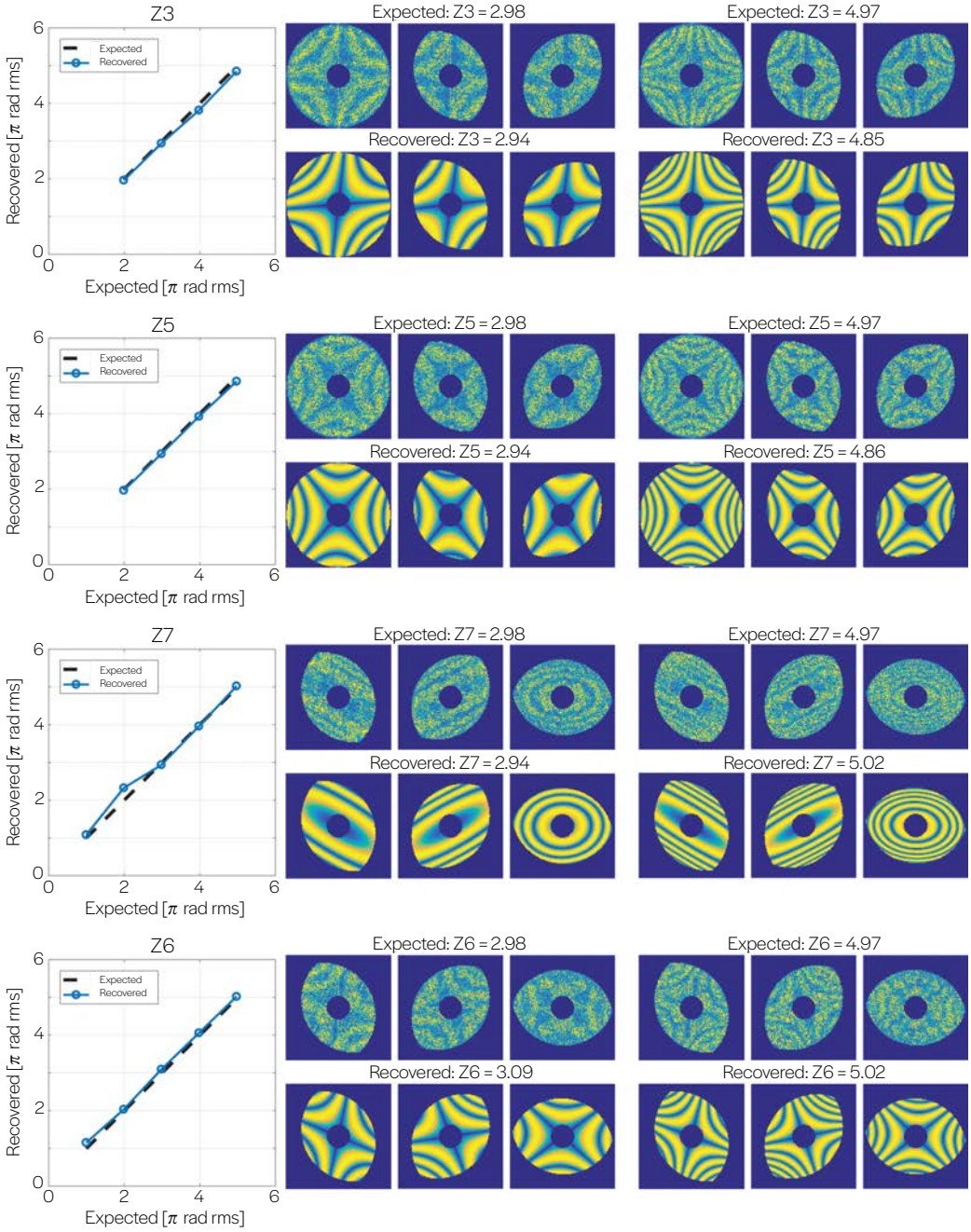

Figure 3.9: **Visible-light aberration recovery results.** Results of recovering aberrations introduced by the deformable mirror, including oblique astigmatism (Z3), vertical astigmatism (Z5), vertical coma (Z7) and vertical trefoil (Z6).

of spacing between adjacent reflective panels, contributes somewhat to reconstruction errors as the true aberration function likely contains significant higher-order terms outside of our 3rd-order approximation. Additionally, the hexagonal packing of mirror panels and the spacing between panels impose restrictions on the achievable illumination angles. Since the deformable mirror exists at the pupil plane of the imaging system, we do not want light to focus between panels. For this reason, a second camera (not shown in Fig. 3.7) was installed to view the pupil plane (mirror surface).

3.5 Experimental aberration recovery: EUV

Having accurately recovered imaging system aberrations in our visible light proof-of-concept experiments, our technique was then implemented in an extreme ultraviolet (EUV) imaging system. This was made possible by the existence of a convenient, readily-available sample which satisfies the requirements for a weak-phase diffuser at the scale of 13.5 nm EUV light – an EUV photomask blank. These are highly polished surfaces which have atomic-scale roughness [18], with height variations on the order of 0.2 nm and a mean lateral bump size of 50 nm. In this section, we will describe the EUV system characterized using our method, outline the characterization process and present our main result.

SHARP EUV microscope

The imaging system that we characterized is known as SHARP (SHARP High-NA Actinic Reticle review Project) and is a synchrotron-based, full-field EUV imaging system at Lawrence Berkeley National Laboratory, designed to emulate aerial image formation in industrial photolithography scanners. It also serves as a research platform for answering modern questions about materials, defects and illumination conditions in semiconductor patterning.

The imaging objective characterized in our experiments is an off-axis zone plate with a focal length of 500 µm, that is manufactured at the Center for X-Ray Optics. The zone plate achieves an NA of 0.082, and its 6° off-axis geometry prevents the specular beam from reaching the sensor (see Fig. 3.10). The image of the sample is formed on a back-thinned CCD camera (PIXIS:2048, Princeton Instruments) located 450 mm downstream, providing an effective 900× magnification. The illumination angle-scanning mirror is a 1 mm x 1 mm MEMS device (Mirrorcle Technologies) coated with an Mo/Si reflective multilayer tuned for the 55° nominal angle of operation. A elliptical condenser mirror is placed such that the angle-scanning mirror is conjugate to the object plane. Angle scanning during image acquisition is used to improve the uniformity of the illumination and reduce coherent artifacts. The microscope operates at a wavelength of 13.5 nm (91.7 eV) with a bandwidth of 1:1450, under ultra-high vacuum conditions.

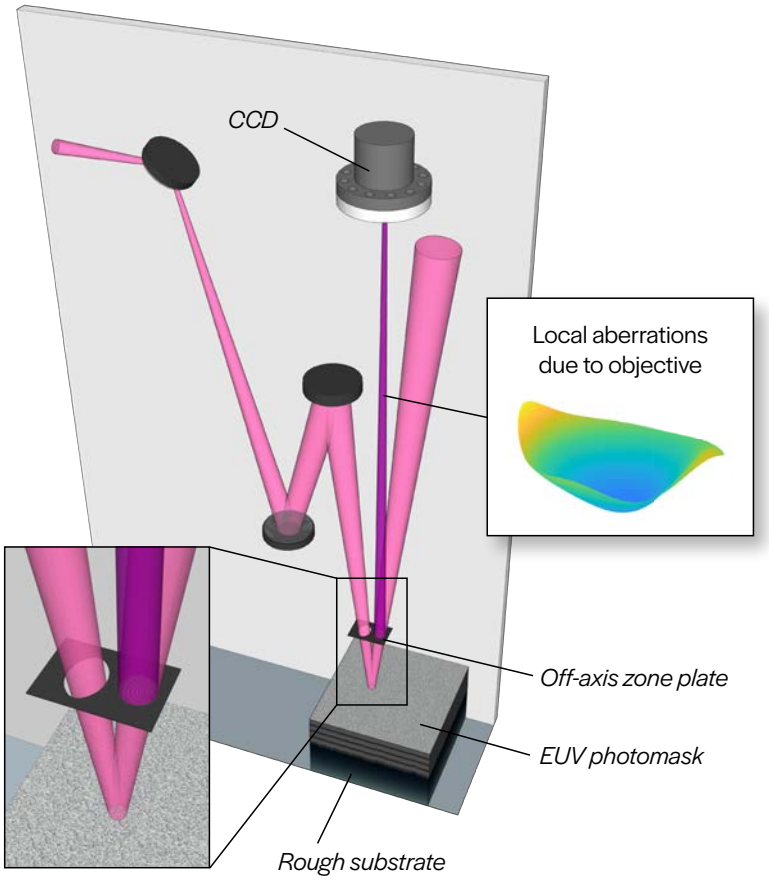

Figure 3.10: **SHARP EUV microscope imaging configuration.** A mirror conjugated with the object plane (which contains a blank EUV photomask) allows control over illumination angle. The objective lens (an off-axis Fresnel zone plate) images the beam scattered by the mask blank onto the sensor. The system suffers from field-dependent aberrations, primarily due to Petzval curvature.

Objective aberration characterization

Our general procedure for aberration recovery closely follows that of the visible-light experiment. We acquire a total of 10 images: one defocused calibration image and 9 in-focus images with varying illumination angle. These images have 2048×2048 pixels with an effective pixel size of 15 nm, accommodating a field of view of approximately 30.72×30.72 µm. These images are shown in Fig. 3.11a. The dark region at the bottom of each image is caused by occluding optical elements; we are primarily interested in analyzing the lighter

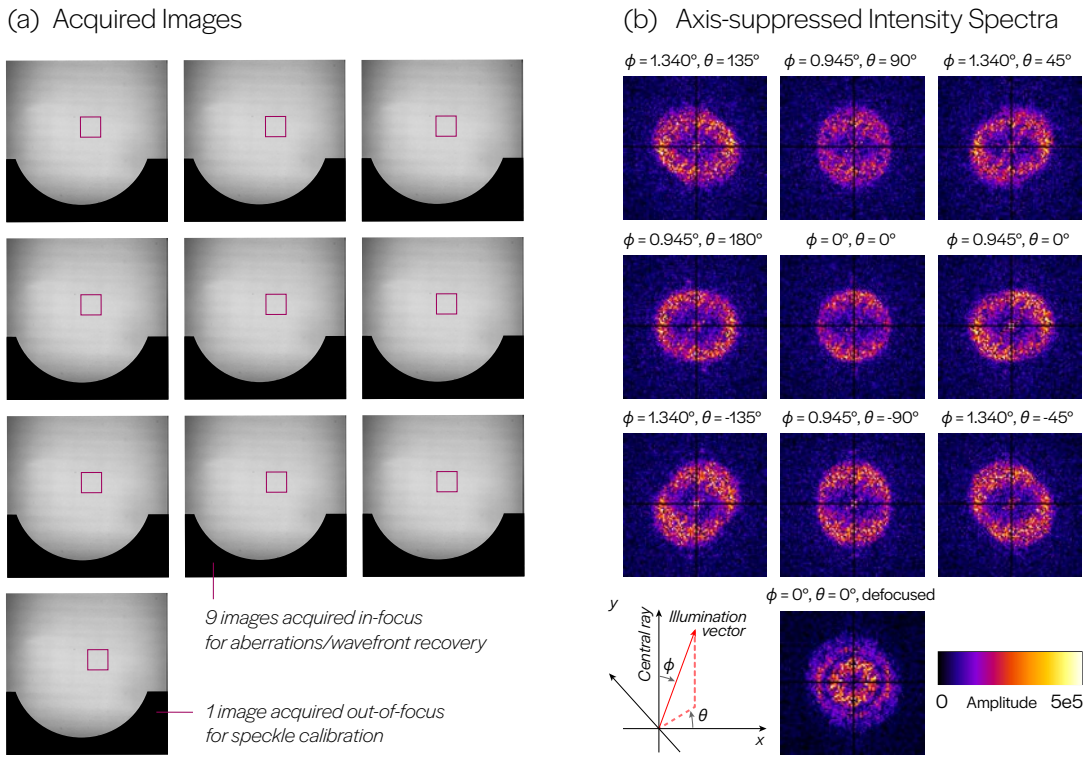

Figure 3.11: **Measurements and computed Fourier spectra.** (a) Ten intensity images (9 with varying illumination angle, 1 with defocus) are acquired by SHARP with a blank EUV photomask as the object. (b) Spatial spectra computed at the sub-region of the full field indicated by the magenta square in (a). Illumination angles are given with respect to the 6° central ray angle, as shown by the schematic in the lower left. Note that these angles should be treated as inputs in the acquisition process, and do not account for angle variations across the FOV due to wavefront curvature.

gray region of the images which exhibit speckle patterns.

For this EUV experiment, we modify the scope of our inverse problems and reconstructions on account of significant field variation in aberrations. Due to the single-lens imaging geometry, we expect aberrations which are minimized near the center of the field of view and increasing in magnitude radially outward. In addition, due to the off-axis imaging geometry, we expect a focal gradient along one axis. For this reason, we segment the full 2048×2048 pixel field of view into 256×256 pixel sub-regions with 50% overlap in the horizontal and vertical directions. Sub-regions in which the photomask was occluded or did not provide sufficient contrast (near the image boundaries) were excluded.

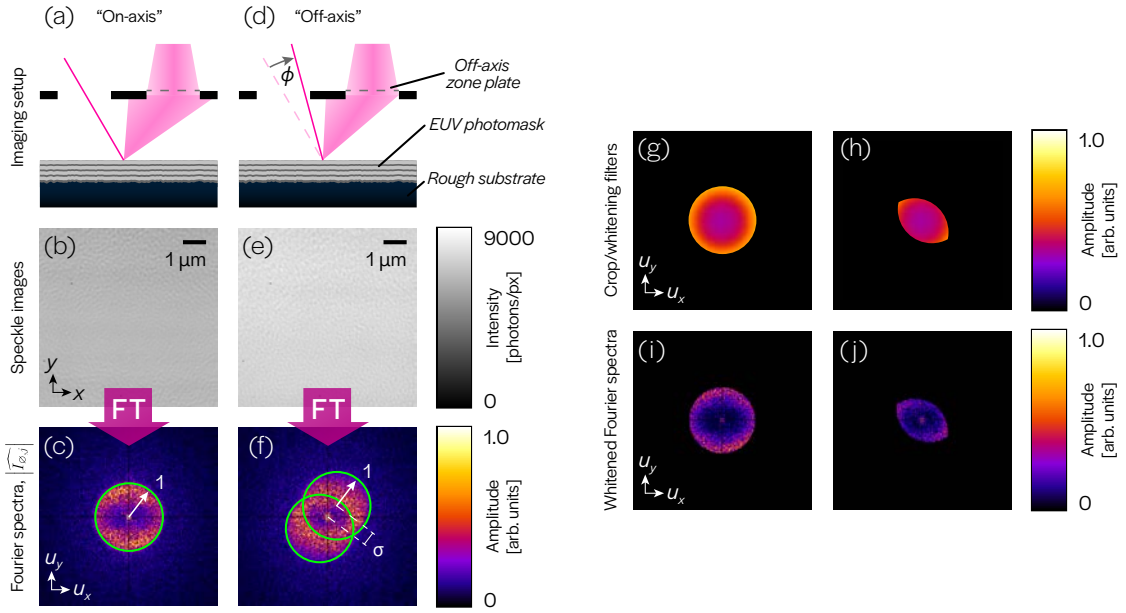

Figure 3.12: **Speckle images and processing for acquisitions under on- and off-axis coherent illumination.** (a-b) Illumination with a 6° central ray angle produces an effective on-axis speckle image of the photomask. (c) Fourier spectrum magnitude of on-axis measurement. (d-e) Illumination with deviation of ϕ from the 6° central ray angle produces an off-axis image of the photomask. (f) Fourier spectrum magnitude demonstrating a shift of $\sigma = \sin\phi/\mathrm{NA}_{\mathrm{obj}}$. (g-h) Filters that crop the spectral magnitudes to their interference regions and divide out the Gaussian window $|\widehat{\varphi}_d|$ to whiten the residual noise. (i-j) Whitened Fourier spectra computed by applying filters (g) and (h) to spectra (c) and (f), respectively.

Corresponding sub-regions of the acquired images are marked by magenta squares in Fig. 3.11a, and their Fourier transforms are shown in Fig. 3.11b. This set of image spectra represents the input to one instance of our aberration recovery inverse problem. This problem is formulated and solved for each sub-region of the full microscope field of view. Note that due to the characteristics of our image spectra, we do not need to delete as large a neighborhood around the DC as we did in our visible light measurements; however, we delete values along the frequency space axes due to noise.

We determine the statistical properties of the photomask surface roughness using the defocused calibration image and we estimate the illumination angle for each sub-region spectrum using the procedure detailed in the previous section. This information is used to produce whitened [30] and appropriately cropped spectra, as shown in Fig. 3.12 for the case of on- and off-axis illumination. Note that the final processed spectra (Fig. 3.12i,j) do not show oscillatory fringe patterns, which suggests that the aberrations in SHARP are not significant

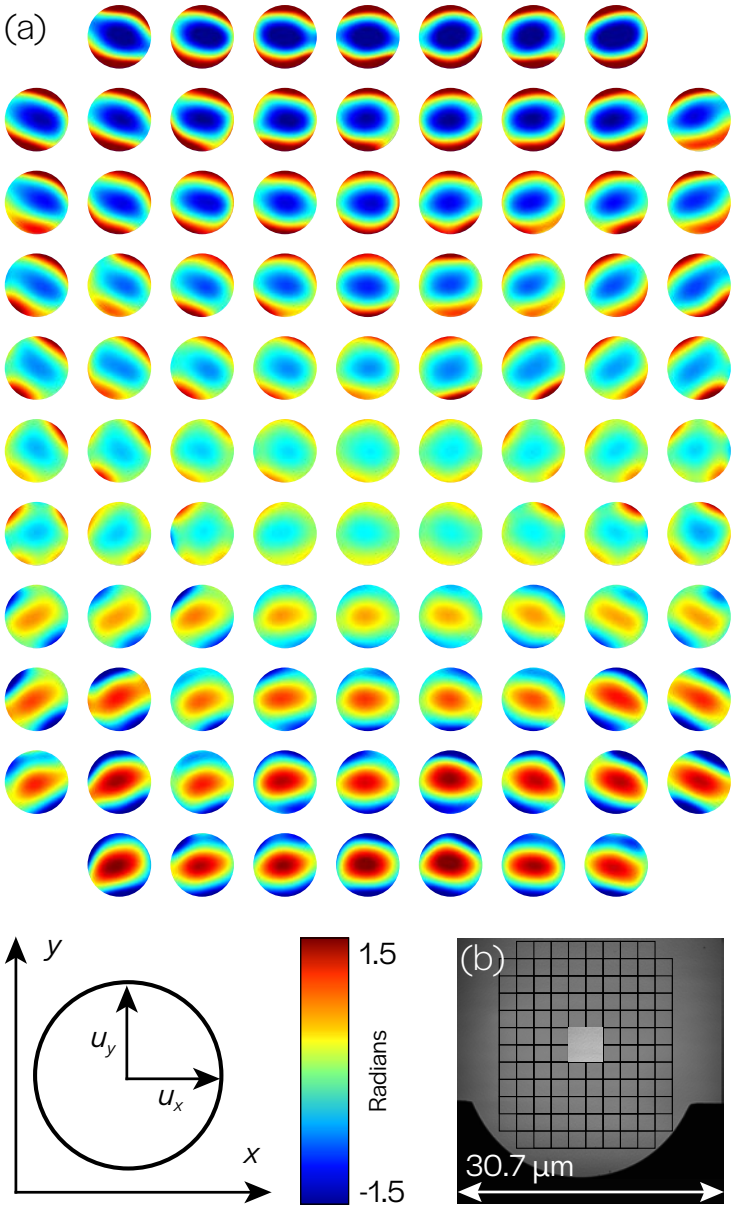

Figure 3.13: **Field-dependent aberrations.** (a) Wavefront error functions (WEFs) plotted across the field-of-view (FOV), demonstrating minimal aberrations in the center and an increase in magnitude at edges. Each WEF is a function of (u_x, u_y) within the unit disk and corresponds to an (x, y) position in the FOV, as shown in the schematic in the lower left. (b) Square sub-regions of the full FOV show the (x, y) positions represented by the WEFs, with one sub-region highlighted to demonstrate size.

enough to result in phase wrapping. Under these circumstances, the statistical calibration step is of greater importance as it determines the scaling of the fitted polynomial.

Once the calibration steps are complete, we proceed with the computational reconstruction algorithm. For each image sub-region, we generate 125 random vectors with Gaussian-distributed elements as initialization points. From each, we compute 200 iterations of gradient descent using backtracking line search and select the resulting coefficient vector with minimum cost, given by Eq. 3.17.

The recovered wavefront error functions for all sub-regions of the field of view are shown in Fig. 3.13. They show that aberrations reach a minimum at the center of the FOV and increase progressively outward, as expected [20]. Averaged over the central 5-µm region, the total wavefront error was 0.0476 ± 0.0055 waves rms (after the removal of residual defocus), corresponding to $\lambda/21$ rms with a measurement accuracy within $\lambda/182$ (see following section for error analysis). This indicates that the central region comfortably satisfies the Maréchal criterion ($\lambda/14$), and this result agrees with the nominal performance for a single-lens design, for which the region where the aberrations are contained below $\lambda/20$ is approximately 5×5 µm^2 [56, 22]. Additionally, defocus dominates along the vertical direction due to the off-axis geometry, as expected. We emphasize that these results were achieved without the use of any additional optical hardware or any precise knowledge, fabrication or alignment of the sample.

3.6 Analysis of error and algorithmic performance

In this section, we will describe additional simulation studies that were done to estimate the error bounds for recovered aberrations, and also to assess convergence behaviors across multiple initializations.

To estimate the reconstruction error in our aberration recovery algorithm, we simulated a set of measurements based on the sub-region size we consider (256×256) and the parameters of the SHARP imaging system (NA, wavelength, magnification and illumination angles). To characterize the effects of shot noise, we simulated measurements of a fixed aberration polynomial with various levels of photon counts per pixel and attempted to recover the WEF. For each level, we initialized the algorithm with 50 randomly chosen points, selected the converged result with minimum cost and recorded its error. The aberration magnitude used roughly corresponds to the value measured in the sweet spot of SHARP (roughly 0.158 waves-rms, or $\approx \lambda/6$ rms, including defocus). The results of these trials are shown in Fig. 3.14a, in which the vertical black line corresponds to the imaging conditions of SHARP—roughly 6,300 photons/pixel. At this level, the absolute reconstruction error ($\varepsilon_a = \frac{1}{2}||\mathbf{c} - \mathbf{c}^*||_2$) of the reported coefficient vector was 0.0069λ waves rms ($\lambda/145$ rms), corresponding to a relative error ($\varepsilon_r = ||\mathbf{c} - \mathbf{c}^*||_2/||\mathbf{c}||_2$) of 4.3%.

To characterize the performance of our algorithm at various magnitudes of system aberrations, we generated 25 datasets at each of 11 levels of rms wavefront error. For each dataset, we initialize our algorithm with 50 random vectors with approximately the same

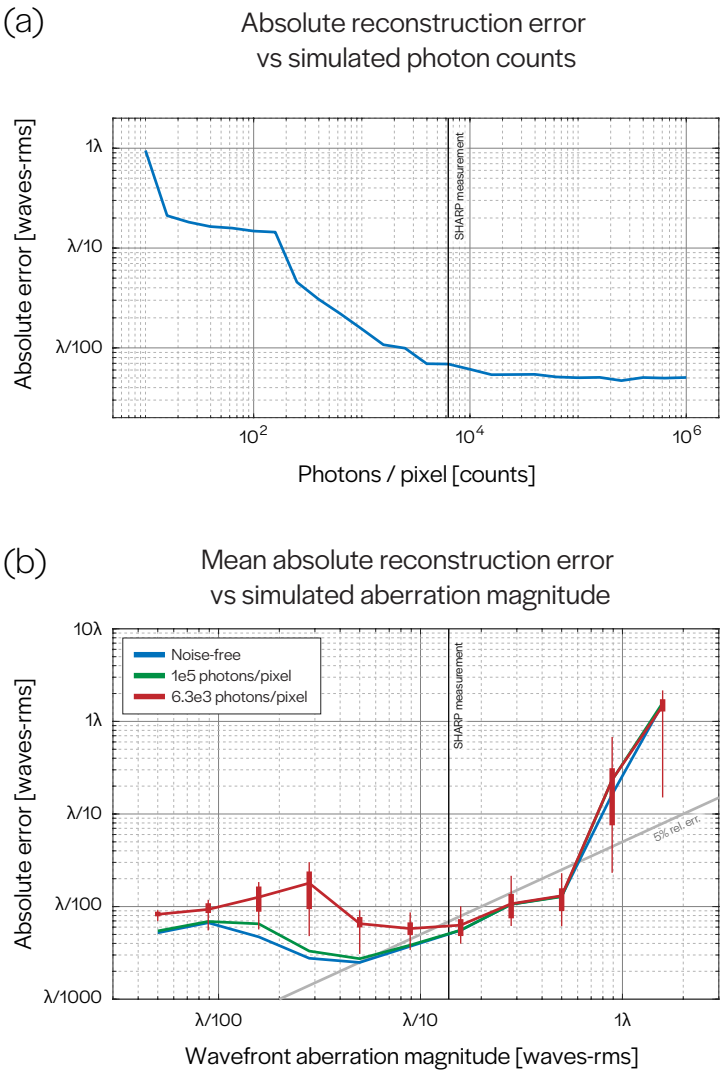

Figure 3.14: **Reconstruction error analysis.** (a) Relative reconstruction error for a single simulated WEF (magnitude 0.158 waves rms) and images corrupted by varying levels of shot noise. The vertical black line represents the imaging conditions of SHARP, roughly 6,300 photons/pixel. (b) Mean absolute reconstruction error for 25 independent WEFs at each of several magnitudes of rms wavefront error. The analysis was performed without adding noise to images (blue), simulating experimental conditions of 6,300 photons/pixel (red) and simulating 10^5 photons/pixel – corresponding roughly to averaging 16 images at each illumination angle. The minima, maxima and interquartile ranges of absolute errors for simulations at 6,300 photons/pixel are also shown.

magnitude as the true coefficient vector, and we report the converged solution with minimum cost (see Eq. 3.17) as the recovered Zernike coefficient vector. We then note the absolute reconstruction errors for each of the 25 reported solutions. We performed an identical analysis under three different levels of simulated shot noise: noise-free, 10^5 photons/pixel and experimental conditions (6,300 photons/pixel). The mean absolute reconstruction errors are shown in Fig. 3.14b, along with the minima, maxima and interquartile ranges for simulations under experimental illumination conditions. The vertical black line corresponds to the experimentally obtained aberration magnitude of SHARP (including defocus). At the nearest sampled aberration magnitude to this level (roughly 0.158 waves rms, approximately $\lambda/6$ rms), the simulated aberration polynomials were reconstructed with a mean absolute error of 0.0063 waves-rms ($\lambda/159$ rms), corresponding to a mean relative error of 4.0%. As a result, we claim that the true aberrations in the sweet spot of SHARP lie within 4.0% of our reconstruction. In the sweet spot of SHARP, we recover a local aberration WEF magnitude of 0.138 waves rms (approximately $\lambda/7$ rms), which is mostly due to a defocus coefficient of 0.130 waves rms. Computing a 4.0% relative error, we have an uncertainty of 0.0055 waves rms ($\lambda/182$ rms), which we reported in the previous section.

In addition, we further examine the simulated reconstruction of aberrations at a specific magnitude (0.2π rad rms) and exposure level similar to SHARP in Fig. 3.15. We run 100 initializations for each of 25 unique datasets and record the converged solution of minimum cost for each. The mean overall reconstruction error across those 25 solutions is shown in

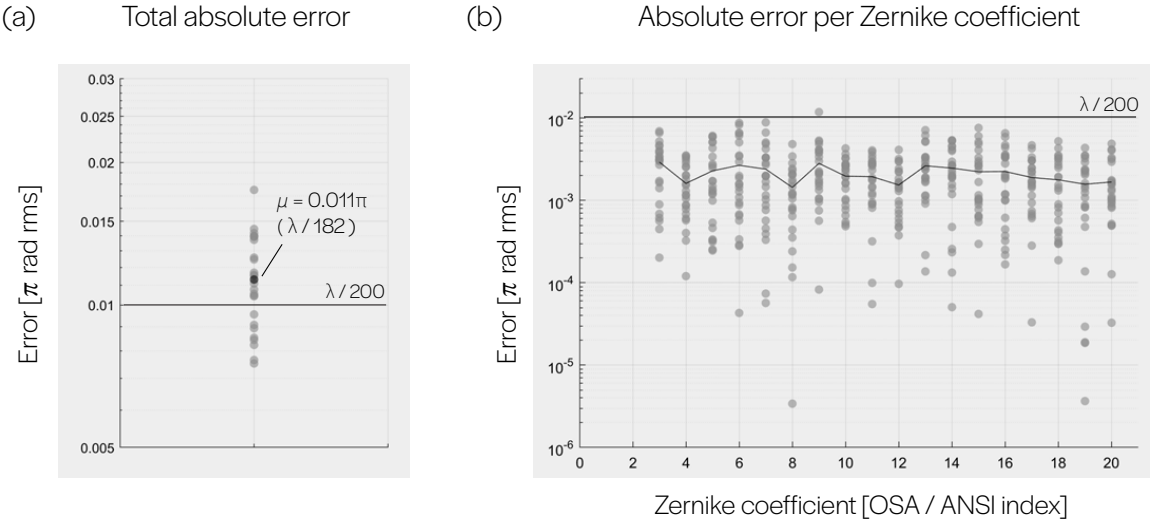

Figure 3.15: **Total reconstruction error and error per coefficient.** (a) Total reconstruction error for 25 datasets of simulated speckle images for imaging systems with aberrations of magnitude 0.2π rad rms (from 100 initializations). (b) Reconstruction error per coefficient for the same set of simulated reconstructions.

Fig. 3.15a. In Fig. 3.15b, we examine the reconstruction error for each individual Zernike coefficient of the 18 that we typically solve for. We find that this error is relatively uniform, i.e. no particular coefficients have a significantly larger spread than others.

In this context, it is also worth noting the two main sources of noise in acquired images: measurement noise resulting from low photon counts, and noisy sampling of polynomial patterns due to surface roughness. The latter is intrinsic to our use of speckle to probe the phase of the pupil plane. While the effects of photon noise on reconstruction error are shown in Fig. 3.14a, here we consider the effects of both noise sources on the spectra themselves. In Fig. 3.16a, we see that low photon counts reduce spectral fringe contrast significantly, which inevitably leads to poor fitting of polynomials and large reconstruction errors. We also see that as the number of photons per pixel tends to infinity, the zeros of the fringe patterns become more and more clearly defined, but peaks in the spectrum remain noisy. This is due to the multiplicative Rayleigh noise produced by the surface roughness. In Fig. 3.16b, we see that as multiple realizations of speckle images *for the same imaging system and illumination angle* are averaged, the Rayleigh noise begins to average out. In the limiting case as the number of realizations averaged tends to infinity, the resulting image would be a scaled (biased) but noise-free image of the fringe pattern, resembling the deterministic part of the forward model.

With these effects in mind, we can revisit Fig. 3.14b and analyze the end behavior. On the right side, we see a sharp increase in reconstruction error for aberrations of increasing magnitude. This is due to inadequate sampling of fringe patterns with significant phase wrapping for the imaging system NA and pixel size in the SHARP system. These imaging system properties specify the number of pixels available within the NA to represent the self-interference patterns containing aberration information. In the case of very strong aberrations, this information is simply lost and our method cannot recover the wavefront error function accurately. However, on the left side, we see a slight increase in reconstruction error for very small aberrations. This is likely due to the lack of true fringes for aberration phase that does not wrap, resulting in a signal that is below the floor of the previously described sources of noise. In this case, averaging multiple acquisitions to artificially raise the photon count, or translating the photomask surface to acquire and average images with unique speckle can potentially improve reconstruction error at the cost of total speed of image acquisition.

3.7 Summary

In this section, we discussed in detail our method of aberration measurement for optical imaging systems based on imaging a weak diffuser, including the discretization of the previously described imaging forward model, simulations across a range of aberration magnitudes and experimental application to both visible light and EUV imaging systems. Finally, we performed an analysis of simulated reconstructions to estimate the accuracy of the proposed technique.

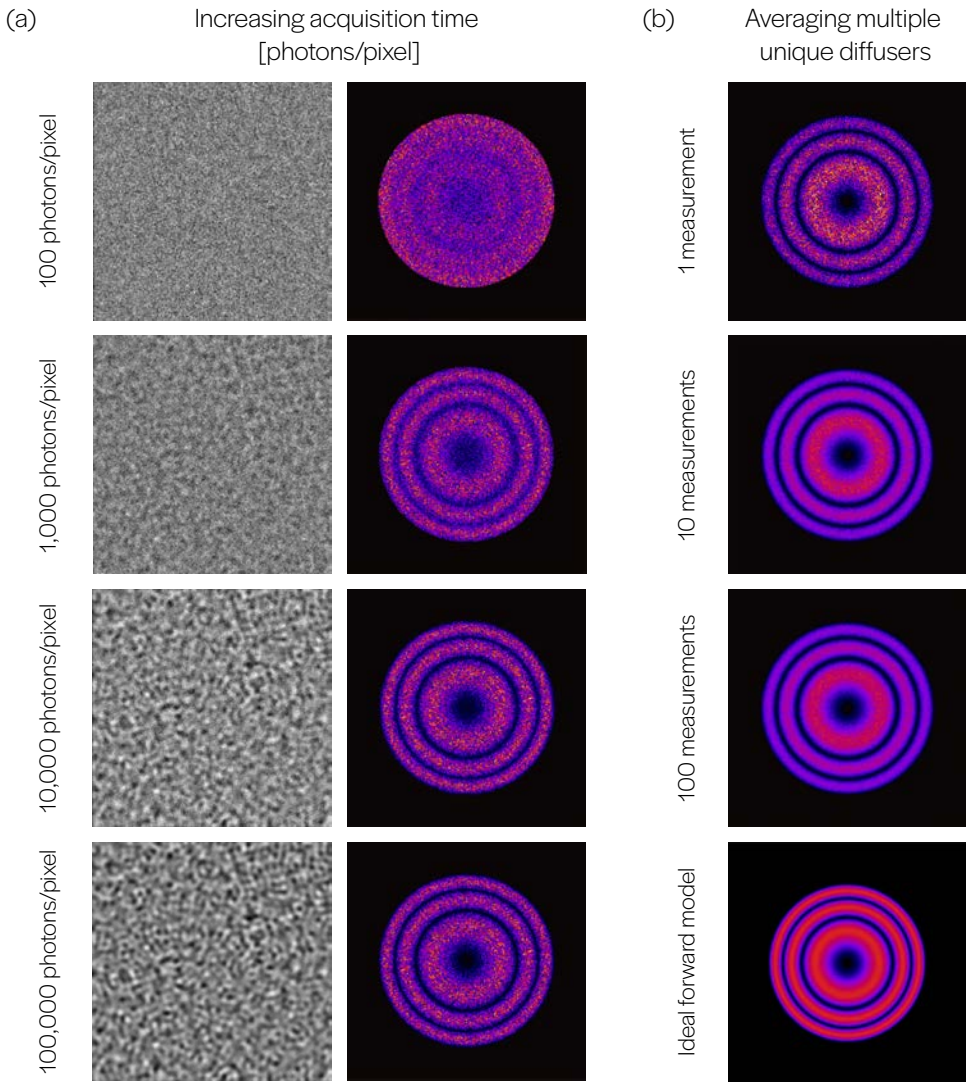

Figure 3.16: **Sources of noise in speckle image spectra.** (a) Effect of increasing simulated exposure time on speckle image spectra. (b) Effect of averaging multiple measurements acquired using unique realizations of diffuser surfaces, with the ideal forward model (no Rayleigh noise) shown as the limit of averaging.

3.8 List of symbols

Symbol	Description
Matrix / vector quantities	
$\mathbf{c}$	Zernike coordinate vector
$\mathbf{E}$	change of basis matrix operator, Zernike to standard
$\mathbf{H}_k$	matrix operator defining translation of a polynomial as an action on its coefficients; related to the k-th illumination angle
$\mathbf{S}$	matrix operator that selects the even part of a polynomial via its coefficients
$\mathbf{\Psi}$	evaluation mapping from polynomial coefficients to graph of function on a specified grid
$\widehat{\boldsymbol{\eta}}$	Rayleigh noise due to diffuser surface, vectorized
$\widehat{\boldsymbol{\varphi}_d}$	deterministic part of diffuser model in Fourier space (Gaussian)
$\mathbf{1}_{\mathscr{U}_k}$	indicator function of interference region for the k-th illumination angle, vectorized
$\widehat{\boldsymbol{I}_{\varnothing}}$	DC- (or axis-) suppressed Fourier transform of speckle intensity measurement, vectorized
$\mathbf{A}_k$	composite matrix operator, $\mathbf{A}_k = \mathbf{\Psi S H}_k \mathbf{E}$
$\widehat{\boldsymbol{m}_k}$	DC- (or axis-) suppressed Fourier transform of the speckle intensity measurement acquired under the k-th illumination angle, i.e. $\widehat{\boldsymbol{I}_{\varnothing,k}}$
$\widehat{\boldsymbol{m}_k}'$	whitened Fourier spectrum of k-th speckle intensity image, i.e. $\widehat{\boldsymbol{m}_k}' = \frac{\widehat{\boldsymbol{m}_k}}{\widehat{\boldsymbol{\varphi}_d}}$
$\widehat{\boldsymbol{m}}_{\text{cal}}$	Fourier spectrum of defocused speckle image used for diffuser calibration
Operators	
$\mathcal{L}\{\cdot\}$	Loss function, or objective function

Chapter 4

Aberration correction in short-wavelength beamline systems

In this chapter, we will outline our methods for the data-driven characterization and control of a prototype deformable mirror, including both the modeling of the system dynamics and formulation of the optimal control problem.

4.1 Modeling the dynamics of a deformable mirror

The primary challenge in the implementation of adaptive optics for beamline systems lies in driving the deformable mirror to a desired surface profile both quickly *and* accurately. While demonstrations of a repeatable linear response have been shown, this effect requires a very long duration of actuation, i.e. it describes steady-state behavior. For example, the predicted surface profile using a linear model may only be reached after waiting for 15 minutes or longer. On the other hand, while the surface profile does have an appreciable transient response, this is much more difficult to model and includes nonlinear and hysteretic effects. Our approach aims to model these nonlinear dynamics and use this model to obtain the optimal control for a desired shape change.

A secondary challenge when working with these short-wavelength beamline systems is the relative lack of non-invasive wavefront sensing techniques. As a result, it is often inconvenient or impossible to simultaneously use a light source for imaging or other experimental purposes while continuously monitoring its beam profile. In these circumstances, an even greater importance is placed on having accurate models for the dynamics of the adaptive optics, since model mismatch cannot be easily compensated using feedback control. In other words, it is imperative for the dynamics model to demonstrate predictive performance with minimal error.

In this section, we will first describe the linear approach used in a variety of existing demonstrations of x-ray adaptive optics and characterizing their long-term behavior. Following that, we will describe our data-driven approach for modeling the short-term dynamics

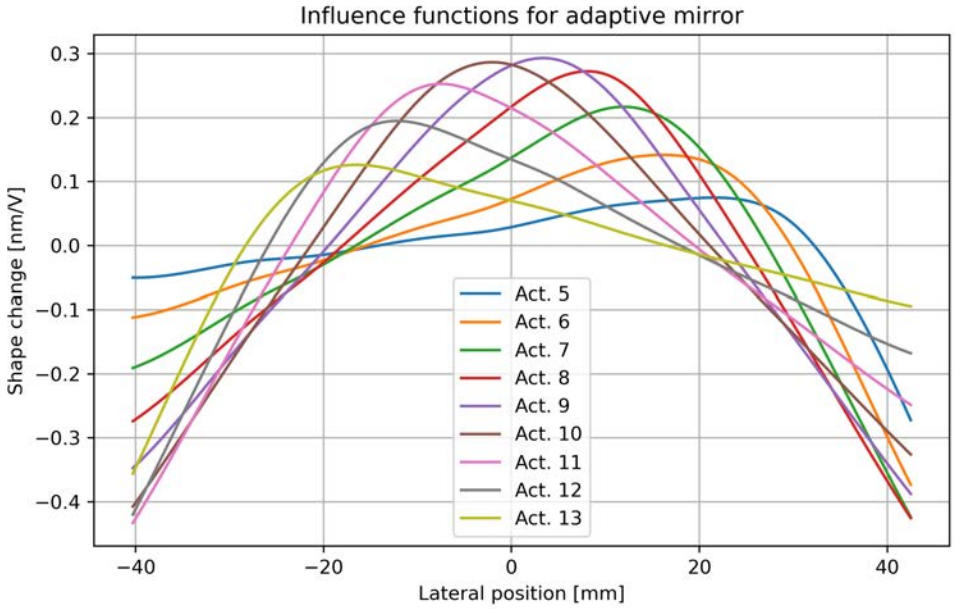

Figure 4.1: **Influence functions measured for JTEC adaptive mirror.** By applying voltage changes to each of the 9 central actuators in isolation, the resulting change of the mirror surface per volt is estimated. Note that piston and tilt terms are excluded from all measurements and computation.

of a similar deformable mirror system.

A linear approach

A large body of existing work in adaptive x-ray optics is based on the approximation of deformable reflectors as linear systems for very small changes in inputs and a very long duration of settling time. One can think of such a linear characterization as describing the steady state response of a mirror to an input or change in inputs. A basis of "influence functions" (also called "actuator response functions" or "characteristic functions") [33, 19, 50] is constructed by supplying voltage to individual actuators in isolation and measuring the resulting surface once the mirror has settled. The basis can then be used for prediction, making the assumption that the response to a linear combination of input voltages is a linear combination of these influence functions.

A basis of influence functions measured for the JTEC adaptive optic is shown in Fig. 4.1. Each curve represents the predicted shape change across the visible area of the mirror surface per volt of change applied to the corresponding actuator. Due to measurement sensitivity constraints, and in the interest of generalizing to a larger neighborhood of applied voltage, the

curves are generated by applying voltages in the set $\{-100, -50, 0, 50, 100\}$V and computing the average per-volt shape change between consecutive measurements. For example, for a single actuator k, we computed

$$\mathbf{b}_k = \frac{1}{50} \cdot (S_{-100,k} - S_{-50,k}), \tag{4.1}$$

using surface measurements S to estimate the shape change per volt, and we averaged 4 such measurements to arrive at the influence function for that actuator. The mirror was allowed to settle for one minute after an input was applied. Of the 18 actuators which influence local surface curvature, only the central 9 produced measurable changes to the mirror surface beyond piston and tilt within the imaged field of view.

To use the measured influence functions for shape prediction, we will construct the matrix $\mathbf{B}$ which consists of all measured influence functions as column vectors. For an initial mirror shape S_{init} and change in voltage Δv, we have:

$$S_{\text{final}} = S_{\text{init}} + \mathbf{B}\Delta v. \tag{4.2}$$

We applied this predictive model to a series of experimental shape changes, noting the root mean square error (RMSE) between the measured and predicted shapes. A total of 497 shape changes were observed over 3 separate experiments in which the applied input voltage was held for 2, 12 and 30 seconds, respectively. The results of these experiments are shown in Fig. 4.2.

As expected, linear prediction generally becomes more accurate as the duration of time increases, evidenced by the reduction in spread of the distributions in Fig. 4.2. Of the three experiments, the one corresponding to 30-second input intervals was the closest in reproducing the steady-state responses that the linear model is designed to predict. At the other extreme, the experiment with an input interval of 2 seconds shows that the linear model is a very poor approximation of the system's highly nonlinear transient response.

Despite the model's performance in representing some steady state behavior, the linear method is fundamentally limited in its ability to make accurate predictions over short time scales and across a large range of motion. This is demonstrated in a few selected examples of shape changes observed after 30 seconds and associated linear predictions in Fig. 4.3. In the top example, we see that significant prediction error arises from cases in which the curvature of the mirror is greatly altered. In the following two examples, we see that actuation corresponding to surface height changes in excess of 100 nm are similarly difficult to predict with the linear model. Note that as the influence functions were determined by the application of voltages in range $[-100, 100]$V, no actuator was driven outside of this range in our experimental studies.

The underlying assumptions in linear modeling are that input voltage is the only factor which determines the shape of the mirror and that the response is consistent across varying initial positions (i.e. that the change in shape from -100V to -50V is identical to the change in shape from 50V to 100V). Given observations of hysteresis in these adaptive mirror systems [3], it is clear that linear models are incomplete and, like any linearizations of nonlinear

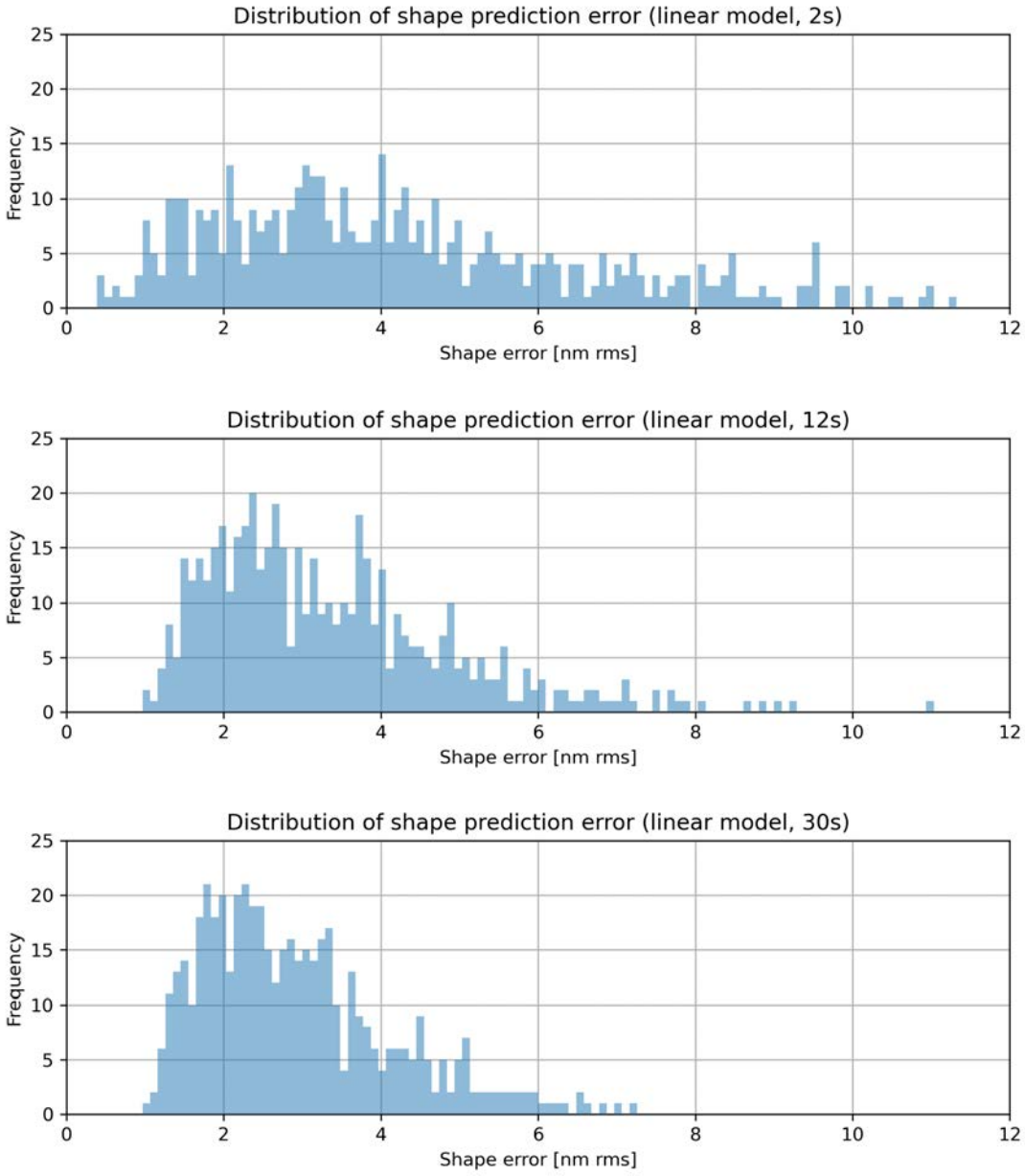

Figure 4.2: **Linear prediction errors using influence functions.** Distribution of prediction errors when using linear influence function model. Errors are defined with respect to measurements taken 2, 12, and 30 seconds after inputs are applied.

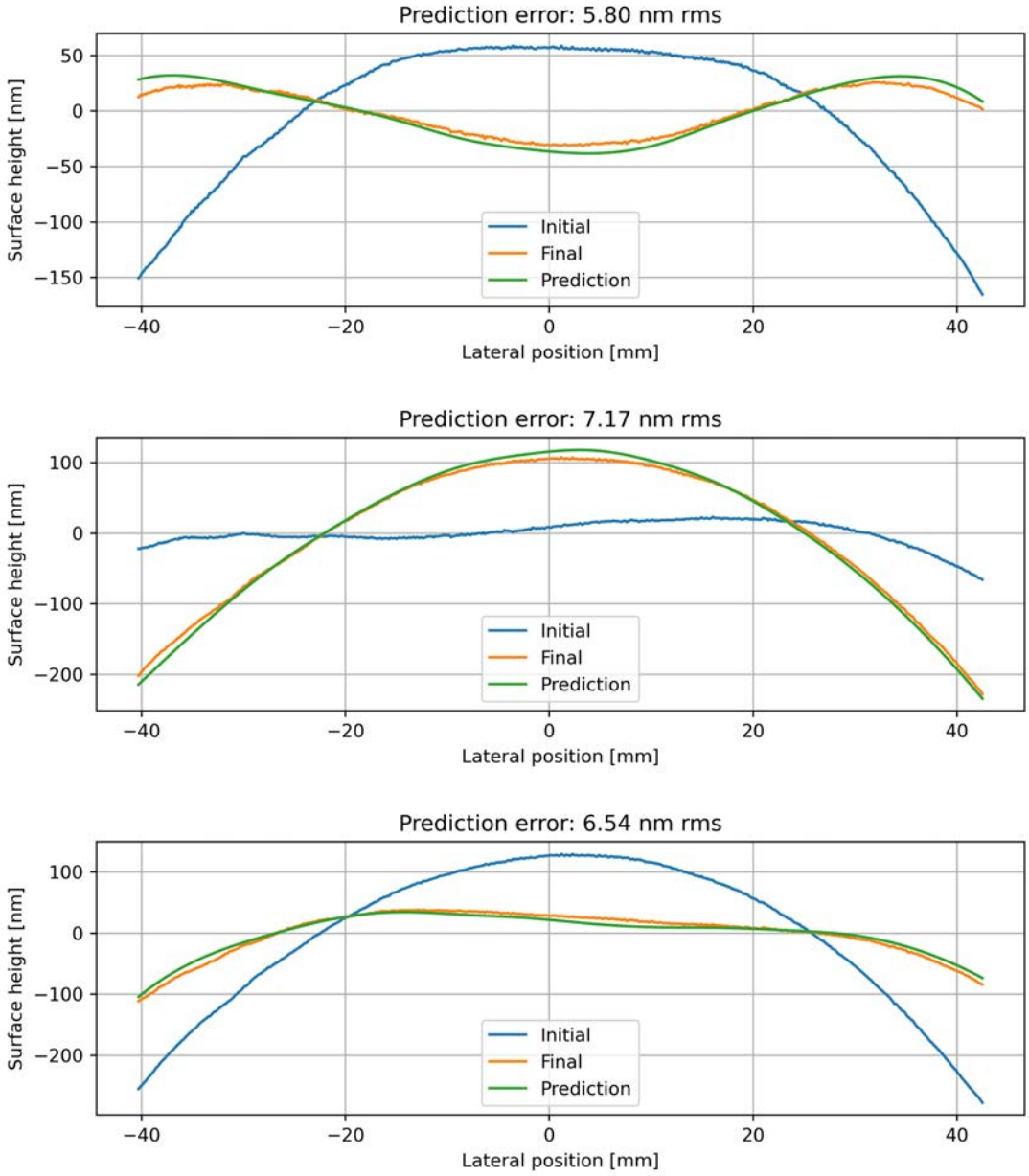

Figure 4.3: **Selected examples of linear prediction performance.** Observed initial mirror surface shapes and final shapes after 30s of actuation, along with linear predictions. Prediction error is generally large when the mirror is driven through large changes in surface height or curvature / convexity.

systems, only valid within a small neighborhood of the calibration data by which they are defined.

In practice, a linear model such as this may still be used effectively in control systems where state feedback is readily available. The linear model is very fast in both forward and inverse computations, and state feedback can to some extent make up for inaccuracies in system modeling. However, this method is poorly suited for the conditions and challenges of x-ray adaptive optics, which justifies the use of a more complex but complete model.

Approximation using a neural network

Given the shortcomings of linear modeling, it is clear that we would like to have a model for the mirror system dynamics with the ability to describe nonlinear effects – specifically hysteresis and drift. A number of approaches [72, 38] have been demonstrated in which the goal was to obtain a physically precise model for the behavior of similar hysteretic devices. These approaches often begin from first principles and use models based on, for example, finite element methods to fully describe the deformation process. While such models can be highly accurate and are physically transparent, they require complicated characterization procedures for practical implementation which are not always feasible, and they are often too unwieldy for applications such as inverse modeling and control. Furthermore, despite attempts to be as inclusive as possible to physical effects in the system, some level of model mismatch or perturbation must still be calibrated and accounted for.

As an alternative, we propose the use of machine learning to model the mirror surface deformation and address some of the challenges in general modeling of nonlinear systems. Machine learning techniques have been widely applied in recent years to a vast set of problems in classification and regression. Similar complex systems with strict demands on performance, such as synchrotron storage rings, are now using techniques derived from machine learning to improve their stability [42]. Given their success, we aim to apply similar data-driven techniques to the operation of adaptive x-ray optics. We hope that this method will circumvent the limited scope of linear modeling, and at the same time enable a lightweight forward model for fast and efficient control.

As a first step, we will define the specific problem of modeling the system dynamics and discuss the resulting model parameters to consider. For simplicity, we will use a discrete-time model with the assumption that time steps will be small. Hence the model will be designed to predict the mirror surface shape after a fixed duration of time, and predictions over longer periods of time are obtained via iterative application of the forward model. We will also assume that the effects of hysteresis and drift can be sufficiently modeled using a series of prior inputs applied and prior surface shapes. It has been shown [36] that the mirror's response to inputs is repeatable, and hence it can be reasonably assumed that no additional factors significantly influence the surface shape. We will assume that environmental factors such as temperature are constant.

Based on the assumptions above, we are looking for a function $f(\cdot)$ such that:

$$S_{t+\Delta t} = f(S_t, S_{t-\Delta t}, S_{t-2\Delta t}, ..., v_t, v_{t-\Delta t}, v_{t-2\Delta t}, ...), \tag{4.3}$$

where S_t is a representation of the mirror surface at time t and v_t is a representation of inputs applied at time t. Note that at time t, the input v_t is instantaneously applied and has not yet perturbed the mirror surface. Representation of inputs is straightforward, as each actuator is associated with a scalar voltage input; hence, each input v is a vector of length 9. Representation of the mirror surface profile on the other hand requires some processing. Measurements of the mirror surface are obtained as images via Fizeau interferometery and are compressed into a spline curve representation of the 1D surface along the dimension of interest. This prevents the function $f(\cdot)$ from scaling with the number of pixels in measured images and allows for faster training and evaluation of the model. It was determined from data that surface representations using greater than 14 parameters offered negligible improvements in performance, and so each surface S is a vector of length 14.

With our inputs and outputs defined, we then consider the structure of the function $f(\cdot)$. We will use a feedforward neural network [7] (also known as a multi-layer perceptron, or MLP) as our discrete-time forward model, with five fully-connected layers and exponential linear unit (ELU) activation functions. The relevant parameters for this model are the dimensions of each layer, the number of layers and the number of surface and voltage vectors used as input. For simplicity, we maintain a constant layer width for all intermediate layers in the network, Through an exhaustive search over this parameter space, it was found that improvements in performance were limited when layer width exceeded 150 and greater than 5 layers were used. Additionally, it was found that the set of inputs:

$$\{S_t, S_{t-\Delta t}, S_{t-2\Delta t}, v_t, v_{t-\Delta t}, v_{t-2\Delta t}, v_{t-3\Delta t}\}$$

produced significant improvements over the set of inputs:

$$\{S_t, S_{t-\Delta t}, v_t, v_{t-\Delta t}, v_{t-2\Delta t}\},$$

but using additional inputs resulted in negligible improvements in accuracy and adversely affected the speed of inverse computations for control.

The final discrete-time system dynamics model and neural network architecture are illustrated in Fig. 4.4. A history of 3 inputs and measurements, in addition to the input at the current time, are concatenated as a vector of length 78; this is treated as the input to the neural network. Each layer of the network computes a vector of length 150 and applies an ELU nonlinearity. (These nonlinearities introduced between layers allow for the approximation of nonlinear functions.) An ELU was chosen since it is an invertible function, which will be important when using this learned model for control. The final layer outputs a vector of length 14, which corresponds to the representation of the final surface profile. Note that a skip connection was introduced, which appends an initial input to the vector at an intermediate layer. This greatly improved the model's ability to predict when the mirror was at rest (i.e. predicting that when $S_t = S_{t-\Delta t} = S_{t-2\Delta t}$ and $v_t = v_{t-\Delta t} = v_{t-2\Delta t} = v_{t-3\Delta t}$, then $S_{t+\Delta t} = S_t$) or approaching a steady-state surface shape.

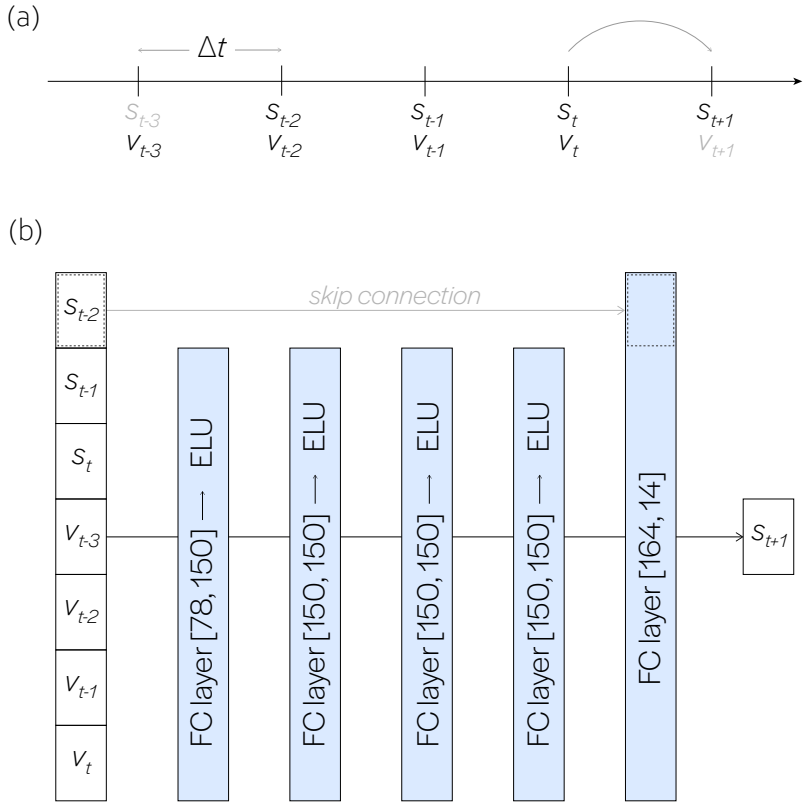

Figure 4.4: **Discretization and neural network architecture.** (a) Our discrete-time model aims to predict the shape of the mirror at a time Δt in the future (s_{t+1}) using a finite history of mirror shapes and voltages input to actuators. Note that s_{t-3} and v_{t+1} are not used in the prediction of s_{t+1}. (b) Our learned system dynamics model consists of 5 fully connected (FC) layers ([input dimension, output dimension]) followed by exponential linear unit (ELU) activation functions. Additionally, a skip connection was introduced which greatly improved its ability to predict when the mirror was at or close to rest.

4.2 Control of a deformable mirror

In this section, we will describe our method for the control of the deformable mirror, which is formulated as an inverse problem using our predictive model. We will assume that the forward model $f(\cdot)$, defined in the previous section, has been specified using acquired data and is able to make accurate predictions of the surface response to inputs. We can then formulate and solve an optimization problem to obtain a sequence of inputs, to be applied at consecutive time steps in accordance with our discrete-time model, which will drive the

mirror to a desired surface shape.

Linear quadratic regulator

Our method draws inspiration from the linear quadratic regulator (LQR) [9, 73], which is an optimization-based technique that can generate the optimal sequence of inputs that drives a linear system to a desired state. Specifically, we will be looking at the discrete-time finite-horizon variant of the LQR algorithm, which is applied to systems of the form:

$$\begin{aligned} \mathbf{x}[k+1] &= \mathbf{A}\mathbf{x}[k] + \mathbf{B}\mathbf{u}[k] \\ \mathbf{y}[k] &= \mathbf{C}\mathbf{x}[k] + \mathbf{D}\mathbf{u}[k] \end{aligned} \tag{4.4}$$

where $\mathbf{x}[\cdot]$ is the state variable, $\mathbf{u}[\cdot]$ is the input to the system, $\mathbf{y}[\cdot]$ is the measurable quantity, and $\{\mathbf{A}, \mathbf{B}\}$ and $\{\mathbf{C}, \mathbf{D}\}$ describe the state dynamics and output readout relationships, respectively. In our application, this problem is somewhat simplified in that we will assume access to the state variable, $\mathbf{x}[\cdot]$. In practice, this could mean direct imaging of the mirror surface using e.g. interferometry, or a combination of wavefront sensing and back-propagation to an effective mirror plane. Hence, we can ignore the second part of Eq. 4.4.

Since we have access to the state variable, we can directly formulate the quadratic cost function in terms of the quantities in the state dynamics equation. This cost function will be minimized with respect to the input sequence $\{\mathbf{u}[k]\}$, and the optimization problem (for the case of driving the state variable to the zero vector) is written:

$$\{\mathbf{u}[k]\}_{k=0}^{N-1} = \min_{\{\mathbf{u}\}} \mathbf{x}[N]^\intercal \mathbf{Q}\mathbf{x}[N] + \sum_{j=0}^{N-1} \left(\mathbf{x}[j]^\intercal \mathbf{Q}\mathbf{x}[j] + \mathbf{u}[j]^\intercal \mathbf{R}\mathbf{u}[j] + \mathbf{x}[j]^\intercal \mathbf{N}\mathbf{u}[j]\right) \tag{4.5}$$

where N represents the number of time steps (or horizon) to solve for, and the matrices $\{\mathbf{Q}, \mathbf{R}, \mathbf{N}\}$ are specified to adjust the meaning of optimality in the context of the problem. For example, $\mathbf{Q}$ determines the penalty applied to state error (distance from zero) at every time step, and $\mathbf{R}$ determines the penalty applied to inputs for applications in which there is a cost associated with actuation. In our problem, we will assume there is no penalty to applying input voltage and will thus ignore $\mathbf{R}$ and $\mathbf{N}$ and drop their quadratic forms in Eq. 4.5. The variables $\mathbf{x}[j]$ are fully constrained by the system dynamics and initial condition.

Simplifying the full formulation of LQR for our problem context, we have:

$$\{\mathbf{u}[k]\}_{k=0}^{N-1} = \min_{\{\mathbf{u}\}} \mathbf{x}[N]^\intercal \mathbf{Q}\mathbf{x}[N] + \sum_{j=0}^{N-1} \mathbf{x}[j]^\intercal \mathbf{Q}\mathbf{x}[j]. \tag{4.6}$$

This formulation can be viewed as a minimization of the cost-to-go, meaning that penalty or cost is associated with both the final state error (i.e. how much $\mathbf{x}[N]$ deviates from zero) and

remaining progress at intermediate steps. Intuitively, a minimum cost would be achieved if an input could drive the system to zero with only one step and all future inputs would completely eliminate further motion of the system; however, this is generally not feasible under the constraints of system dynamics and so a more gradual convergence is expected.

The LQR problem can be solved analytically by formulating and solving the Riccatti equation [9]. It is important to note that this method is guaranteed to output the optimal set of inputs $\{\mathbf{u}[k]\}$ for the specific formulation of the objective function. This guarantee does not exist in the context of nonlinear systems, but the technique is still practical and often yields favorable results.

Application to nonlinear systems often involves a technique known as the Iterative LQR [45], in which a nonlinear system is linearized about its initial condition, and the LQR problem is then formulated and analytically solved using the linearized system dynamics. Due to the possibility that the solved system trajectory extends beyond the neighborhood of the initial condition (where the linearized dynamics are sufficiently accurate), this procedure is applied iteratively. The state variable may be probed frequently, leading to a new linearization of the system dynamics and computation of an updated trajectory.

Non-convex optimization for mirror control

While the iterative LQR technique has been demonstrated in practice for the control of nonlinear systems, we note that it requires frequent sampling of the state variable to update both the linearized system dynamics and the trajectory toward a desired state. In our implementation, we chose instead to apply convex optimization techniques to the full, nonconvex optimization problem that follows from the nonlinear system dynamics.

Specifically, we formulate the control problem as:

$$\begin{gathered}
\{\mathbf{u}[k]\}_{k=0}^{N-1} = \min_{\{\mathbf{u}\}} \sum_{j=1}^{N} w_j \|\mathbf{x}[j]\|^2 \ \text{ subject to:} \\
\mathbf{x}[k+1] = f(\mathbf{x}[k], \mathbf{x}[k-1], \mathbf{x}[k-2], \mathbf{u}[k], \mathbf{u}[k-1], \mathbf{u}[k-2], \mathbf{u}[k-3]) \\
\mathbf{x}[k] = \mathbf{x}[0] \ \ \forall k < 0 \\
\mathbf{u}[k] = \mathbf{u}[-1] \ \ \forall k < 0.
\end{gathered} \tag{4.7}$$

In the notation above, the state variable $\mathbf{x}[k]$ is the vector representation of the mirror surface S_t, and e.g. $\mathbf{x}[k+1]$ corresponds to $S_{t+\Delta t}$. Similarly, the input variable $\mathbf{u}[k]$ is the vector of voltages applied to mirror actuators v_t, and e.g. $\mathbf{u}[k+1]$ corresponds to $v_{t+\Delta t}$. The constraints listed in Eq. 4.7 consist of the system dynamics, with the relationship specified by the neural network architecture, and an assumption that the mirror begins at rest (steady state) when the control is applied. Note that $\mathbf{x}[0]$ has been excluded from the computation of cost since this is an initial condition. Also note that we assume the matrix $\mathbf{Q}$ in Eq. 4.6 is a diagonal matrix, and have simplified the objective function to contain scalar weights.

In our experiments, these weights are all set to a single constant value. Solution of this optimization problem will be discussed in the following chapter.

4.3 Summary

In this section, we motivated the problem of aberration correction in short wavelength systems such as synchrotron and free-electron laser facilities, along with its challenges and constraints. We discussed early attempts at prediction and control using linear models, and the incompatibility of these methods with large shape changes and open-loop (feedback-free) operation. We then outlined our general approach, involving: (1) approximating the nonlinear system dynamics using a feedforward neural network, and (2) control to a desired surface shape using nonlinear quadratic cost regulation over a finite time horizon. This sets the stage for the following chapter, in which we discuss the experimental implementation of our method and analyze the results.

4.4 List of Symbols

Symbol	Description
Matrix / vector quantities	
$\mathbf{b}_k$	influence function associated with the k-th actuator of the mirror
$\mathbf{B}$	basis of influence functions for approximation of mirror as a linear system
$\mathbf{x}[k]$	state variable (instantaneous mirror shape) at time step k in discrete-time model
$\mathbf{u}[k]$	input variable (actuator voltages) applied at time step k in discrete-time model
Functions	
$f(\mathbf{x}[\cdot], ..., \mathbf{u}[\cdot], ...)$	nonlinear system dynamics describing mirror surface response to inputs (neural network)

Chapter 5

Characterization and control of an X-ray deformable mirror

In this chapter, we will present our experimental implementation of our proposed characterization and control algorithms for an adaptive x-ray optics system. We will provide a description of the experimental setup, discuss the specific procedure that we followed, and present and analyze the results of our experiments. We will estimate the errors that can be expected in both the prediction and control problems, and we will show the algorithms' performance in a specific use case of interest in beamline facilities.

5.1 Adaptive mirror and experimental setup

The mirror used in our experiments was a prototype piezo-bimorph mirror [36] fabricated by JTEC Corporation. The mirror substrate (enclosed in the mirror box as shown in Fig. 5.1) has an overall dimension of 160 mm (length) $\times$ 50 mm (width) $\times$ 10 mm (thickness). The platinum-coated optical surface is 150 mm (length) $\times$ 8 mm (width), centered on the top side of the mirror. In its intended use case, short-wavelength radiation is incident at a grazing angle along the length of the optical surface.

Deformations of the mirror surface are enabled by piezoelectric actuators that are bonded to the substrate and locally perturb the curvature (second derivative) of the optical surface. Two piezo strips (one strip is visible in Fig. 5.1), each comprised of 18 separate electrodes, sandwich the optical surface. The two electrodes at a corresponding position of the two strips comprise one electrode channel (referred to as actuator). There are a total of 18 such actuator channels (CH1 to CH18). There are two additional piezo strips glued to the bottom surface of the mirror, which act as a single electrode channel (CH20) that globally perturbs the curvature optical surface. The grounded channel (CH19) is on the backside of all piezo strips.

We define the resting state of the mirror as the steady-state surface shape when all 20 channels are at the same voltage. In this work, we kept CH19 and CH20 at the fixed voltage

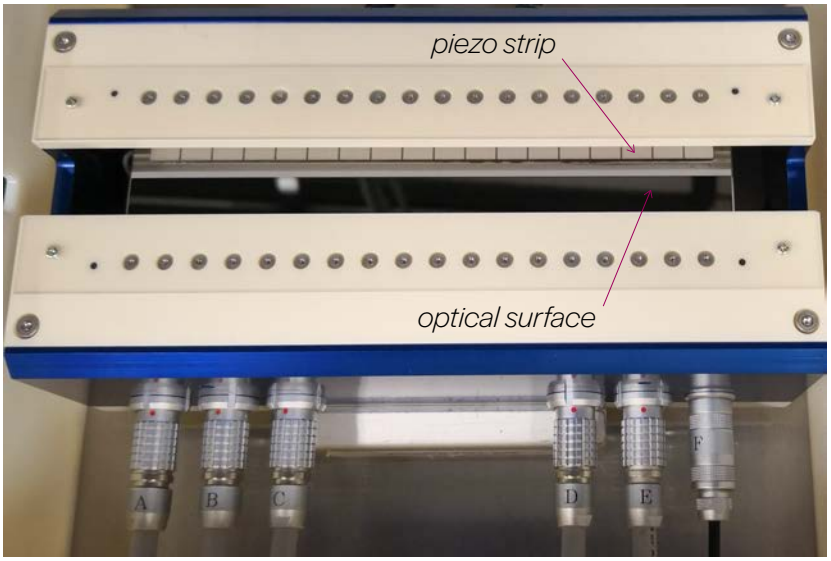

Figure 5.1: **Piezo-bimorph deformable x-ray mirror.** Adaptive x-ray mirror (JTEC Corporation) that was studied and characterized in our experiments. A segmented piezo strip enables spatially-localized deformations of the optical surface.

of 500 V and only varied the top-surface channels in a relative range of [-100 V, 100 V] around their nominal value of 500 V. We note that in this resting position, the mirror is not flat.

Measurements of the mirror surface were performed using a visible-light metrology setup designed in-house at the Advanced Photon Source at Argonne National Laboratory. The setup includes a FizCam 2000 Fizeau interferometer (4D Technology) with 100 mm aperture mounted vertically above the sample. The mirror is placed on a manual tip-tilt stage underneath the interferometer with the optical surface facing up. The mirror, stage and transmission flat of the interferometer are enclosed by Lexan polycarbonate to reduce the effects of air turbulence and temperature fluctuation on the measurements. An image of the setup is shown in Fig. 5.2a.

An example image acquired by the interferometer and surface reconstruction are shown in Fig. 5.2b. The interferometer simultaneously collects 4 phase-shifted fringe patterns like the one shown via polarization-based multiplexing onto the sensor pixels. This allows for rapid acquisition of surface profiles without moving parts, and is crucial to our ability to characterize the mirror system dynamics over short time scale. Due to vibrations in the system setup, and some loss of information in recovering surface height information from interferograms, we remove tilt and piston from all recovered surface profiles. In practice, tilt and piston are likely to be measured and compensated by a separate mechanism, and so we

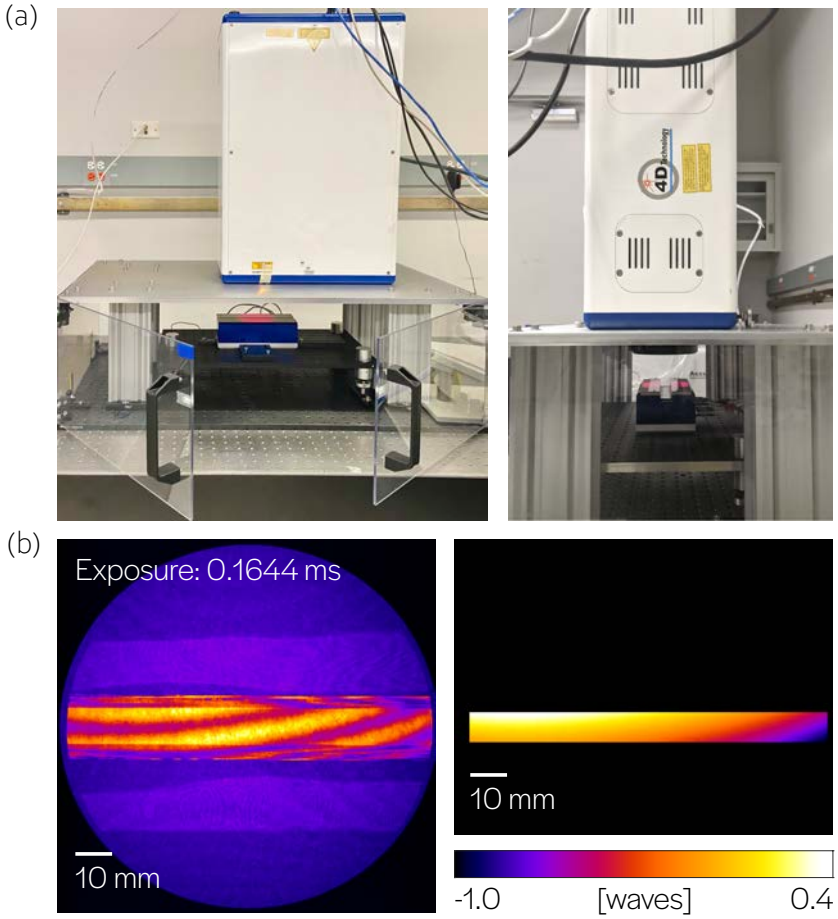

Figure 5.2: **Experimental setup.** (a) Mirror surface profiling with a Fizeau interferometer ($\lambda = 658$ nm) mounted vertically above the bimorph mirror. (b) An example interferogram. Four such measurements are acquired simultaneously and used to recover a surface profile, which is cropped to the active area of the mirror.

will restrict the scope of wavefront perturbation via adaptive optics to higher-order effects. We also note that the aperture of the Fizeau interferometer (100 mm) is smaller than the length of the mirror, resulting in actuators outside the field of view contributing only tilt to the surface profile. Since tilt is excluded from computation, we restrict actuation to the 9 central channels (CH5 to CH13) for all of our experiments.

5.2 Data acquisition for model training

The data used to train our neural network model was acquired by applying sequences of random voltages to the central 9 mirror actuators and recording the resulting mirror shapes after a fixed, 2 s, time interval. An acquisition consists of 10 images captured with an exposure time of 0.126 ms and averaged together. Acquisitions are triggered such that this exposure time does not contribute delays to the 2 s interval. The voltages are uniform random values in the range $[-100, 100]$V and are rounded to the nearest tenth of a volt. A time interval of 2 s between the application of voltage inputs and surface measurement is maintained throughout the acquisition of training data. Note that this corresponds to $\Delta t = 2$ s in our discrete-time model (see Fig. 4.4a). We remark that the choice of time step was limited by the speed at which inputs are sent to the prototype mirror, and not the speed of acquisition of surface measurements.

We acquire 8 separate sequences of surface measurements, starting with an initial measurement followed by 500 voltage changes. We acquire an additional 4 sequences of 501 measurements in which voltage changes are held for 5 time steps (i.e. the same voltage input is applied for 5 consecutive measurements). We believe this data is vital to informing the model of system convergence behavior. Since each training example requires 4 sequential measurements (one state to be predicted from three inputs), we can extract a maximum of 497 training examples from each of the 501-measurement sequences. In total, the 12 sequences give 5,964 training examples.

After the set of surface profiles is processed (tilt and piston removed, masked to optical surface of mirror), the 2D images are averaged along the narrow, sagittal direction to produce 1D curves along the dimension of actuation. These curves are then compressed into the 14-parameter spline representation described in the previous chapter.

We used the PyTorch library as the machine learning framework. To train our dynamics model, we perform 15-fold cross validation on the training dataset with 3000 iterations of Adam per fold. To test the trained model's performance, we acquire another independent sequence of measurements and divide into sub-sequences, as described above, resulting in 497 test examples. Additionally, we repeated the acquisition of the test set with identical voltage changes and longer time intervals (12s, 30s) between measurements. This was done to test the model's ability to predict convergence behaviors.

5.3 Predictive performance of data-driven model

Some examples of shape prediction are demonstrated in Fig. 5.3a, and the corresponding prediction errors for neural network (proposed method) and linear prediction are labeled. Linear prediction is performed using the basis of influence functions shown in Fig. 4.1. The aggregate performance over the entire test dataset is shown in Fig. 5.3b. Overall, the mean prediction error for our method was 1.26 nm root mean square (RMS), compared to 4.20 nm RMS for linear prediction. Our neural network model demonstrated lower prediction error

than the linear model in 460/497 test examples. The cases for which linear prediction demonstrated lower prediction error generally involved very small changes in input voltage.

We then attempt surface profile prediction with longer time intervals, and again compare with linear modeling. In Fig. 5.4, the prediction errors for both models are plotted against the magnitude of observed shape change for a variety of time scales. Testing the model's predicted system response with a 2 s time interval gives the highest measured accuracy because the test data is acquired through a procedure identical to that of the training data. At this short time scale, errors from the linear model are essentially random.

For longer time scales, we apply our learned model iteratively, using predicted, intermediate surface shapes as input for subsequent steps toward the goal shape. For 12 s intervals (6 prediction steps), we see that the neural network prediction performs worse on a somewhat sparse set of examples, and linear prediction begins to exhibit a linear correlation between the shape change observed (requested) and error. For 30 s intervals (15 prediction steps), neural network predictive performance degrades over a larger set of examples, and linear prediction maintains its linear correlation, albeit with lesser slope.

Unsurprisingly, the tests show that the predictive performance of our neural network model is best at the time interval used to acquire the training data; its performance degrades when used at longer time intervals. This is likely due to the much more limited representation of repeated inputs in our training data set; the data extracted from measurements with repeated inputs (no voltage input change) comprises only one-third of the total training data set. Also, the maximum number of consecutive measurements with repeating voltage inputs is only five, corresponding to a maximum of 10 s mirror relaxation without varying voltages. While the network effectively predicts shape changes when all voltage inputs are updated, more data is likely required to inform the network of convergence behaviors at different states when the inputs are held constant.

The results also demonstrate the relationship between the magnitude of surface shape change and linear prediction error. While Fig. 4.2 demonstrated a reduction in spread of errors with increasing settling time after an input, Fig. 5.4 clearly shows that there is limited room for improvement. Even after waiting for long time intervals, the linear model is still unable to accurately predict large shape changes. This shows that linear-model prediction is primarily effective for small shape changes (< 20 nm RMS) over long time scales.

5.4 Directed shape control

Once the parameters of the nonlinear system dynamics model are learned, the model is used to determine a sequence of voltage inputs that will drive the mirror from a measured initial state to a desired final state in a given, finite number of steps. In our experiments, we examine the performance of the control algorithm with a 10-step horizon. Our algorithm directly minimizes the non-convex objective function (Eq. 4.7) using Adam [39], a variant of stochastic gradient descent. While this formulation does not theoretically guarantee optimality of the

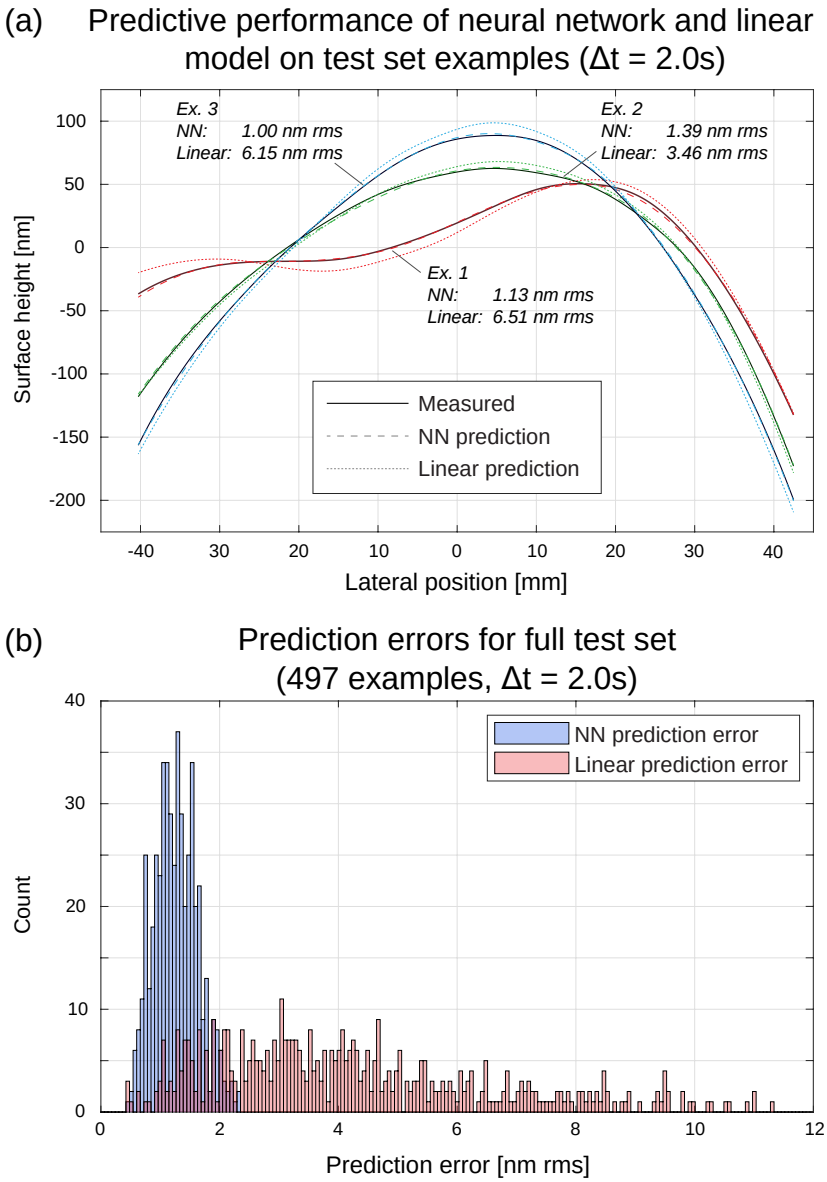

Figure 5.3: **Performance of predictive model on single time-step test data.** (a) Predicted curves from the neural network and linear models for 3 examples in the test dataset. (b) Predictive performance of neural network and linear models across full test dataset.

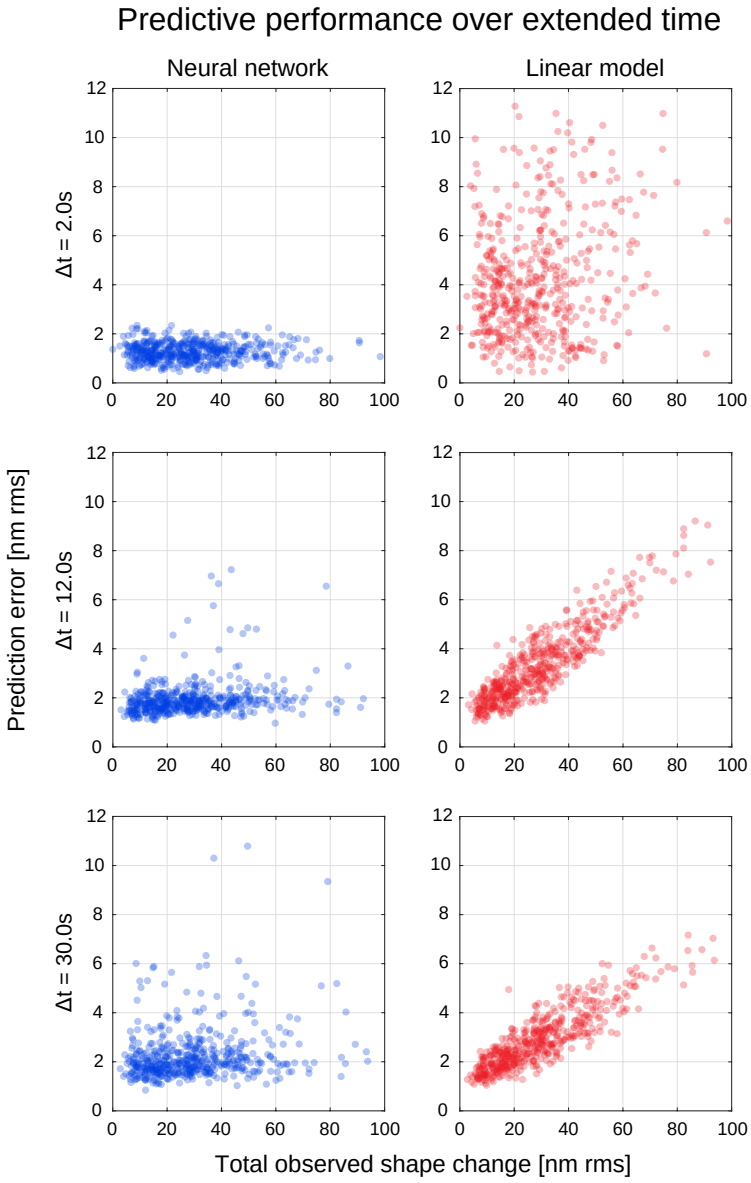

Figure 5.4: **Performance of predictive model on multiple time-step test data.** Prediction errors of neural network and linear models plotted against the observed RMS change in surface shape for 3 different time scales. For time scales larger than 2 s, the neural network model is iteratively applied.

converged solution, we show that the observed performance meets our specifications in a majority of experiments.

In practice, mirror shape control will be used to compensate phase errors in the wavefront of a focused beam. Therefore, it is of central importance to be able to direct the mirror to achieve and hold arbitrary shapes within its capabilities. As a demonstration, we test the ability of our control algorithm to direct the mirror to a series of 50 random, prescribed shapes. For each target shape, an initial surface measurement is acquired and used to generate a 10-step sequence of voltage inputs, to achieve the shape and stabilize the mirror. We allow 150 s to elapse between experiments so that the initial conditions (see Eq. 4.7) are approximately true.

The results of these experiments are shown in Fig. 5.4. A selected sequence of transitions to 3 prescribed shapes is shown in Fig. 5.4a, where the measured surface profiles (colored, dashed) closely approximate the desired shapes (black, solid). Figure 5.4b shows the voltages applied to each of the 9 actuators over the 10 steps that were generated by the optimization algorithm.

We observe that these voltage sequences sometimes demonstrate oscillatory behavior, suggesting that the model is accommodating dynamic effects such as overshoot and creep. In Fig. 5.6a, we see that the algorithm drives the mirror close to the goal after the first step, with the remaining 9 steps being used to maintain the position. Some overshoot may still occur, as the error with respect to the prescribed shape is often slightly larger after 10 steps than after the first step. This may be caused by a combination of the non-convexity of the optimization problem and the lack of guaranteed optimality, and by any residual errors in the predictive capability of our learned system dynamics model. The former can be somewhat addressed by changing the parameters of the Adam algorithm (learning rate, iterations) or the weights w_k in Eq. 4.7, or by adding regularization to the objective function, e.g. penalizing RMS differences between time-adjacent voltage inputs or predictions ('velocity'). Figure 5.6b shows the aggregate performance of our control algorithm across the 50 test cases. The mean RMS errors between the measured and prescribed shapes are 1.70 nm after the first step and 1.91 nm after 10 steps.

Use case: varying focal distance

Among the directed shape-control tests, we drove the mirror to a set of cylindrical shapes with prescribed radii of curvature from 2 km to 6 km. This emulates the case of an adaptive mirror used to vary the focal distance, as in Ref. [76]. In our applications, we consider these to be relatively large moves, with central surface height changes from 146.3 nm in the 6 km case, to 429.0 nm in the 2 km case. Test results are shown in Fig. 5.7. We observe that the mean RMS errors between the measured and prescribed shapes are 1.44 nm after 1 step and 1.51 nm after 10 steps.

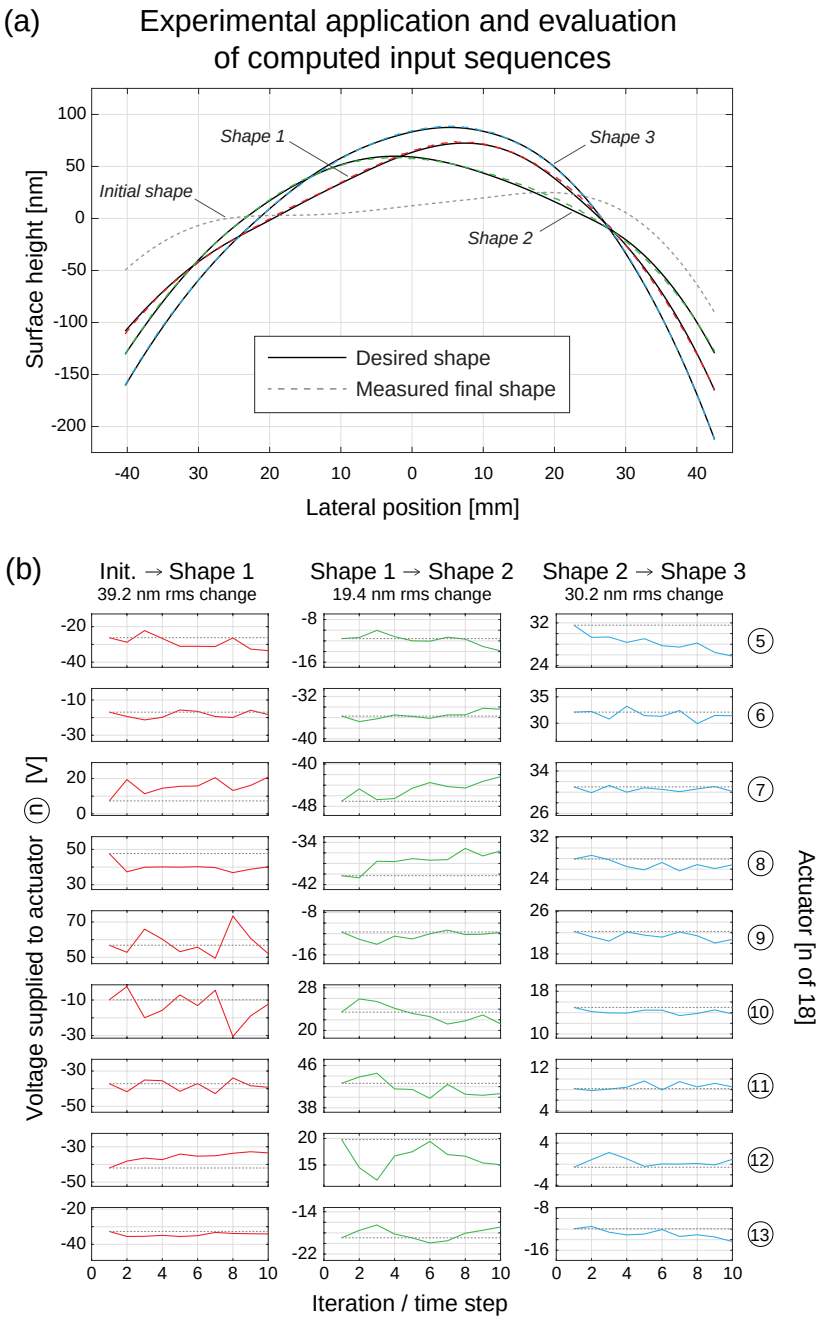

Figure 5.5: **Controlling mirror to desired surface shapes.** (a) Measurements of the mirror surface taken over a sequence of 3 prescribed shapes. (b) Voltages applied to actuators over 10 solved time steps for the 3 shape transitions.

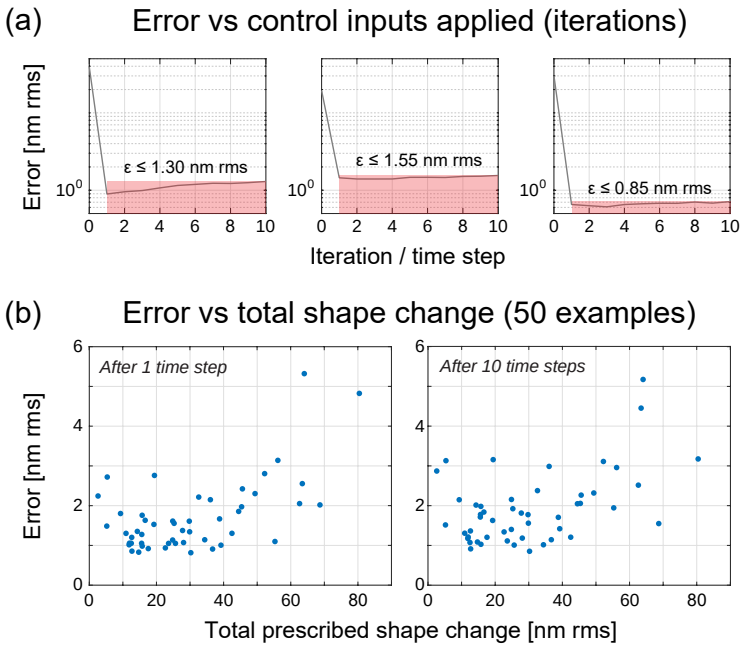

Figure 5.6: **Error performance for mirror control.** (a) Total shape error in nm RMS over 10 time steps. (b) Aggregate performance of control algorithm for the full test set of 50 prescribed random shapes.

5.5 Summary

In this section, we discussed our experimental implementation of the characterization and control of an adaptive x-ray mirror. We provided a description of the mirror used in our experiments and the imaging system used to perform rapid interferometric surface metrology. We then outlined our data acquisition procedure for the training of our data-driven neural network model for the system dynamics, and presented an investigation of the trained model's accuracy for both single and multiple time-step predictions. We then outlined our procedure for control over a 10-step horizon and presented the results of both random and simulated use-case (tunable focal length) experiments.

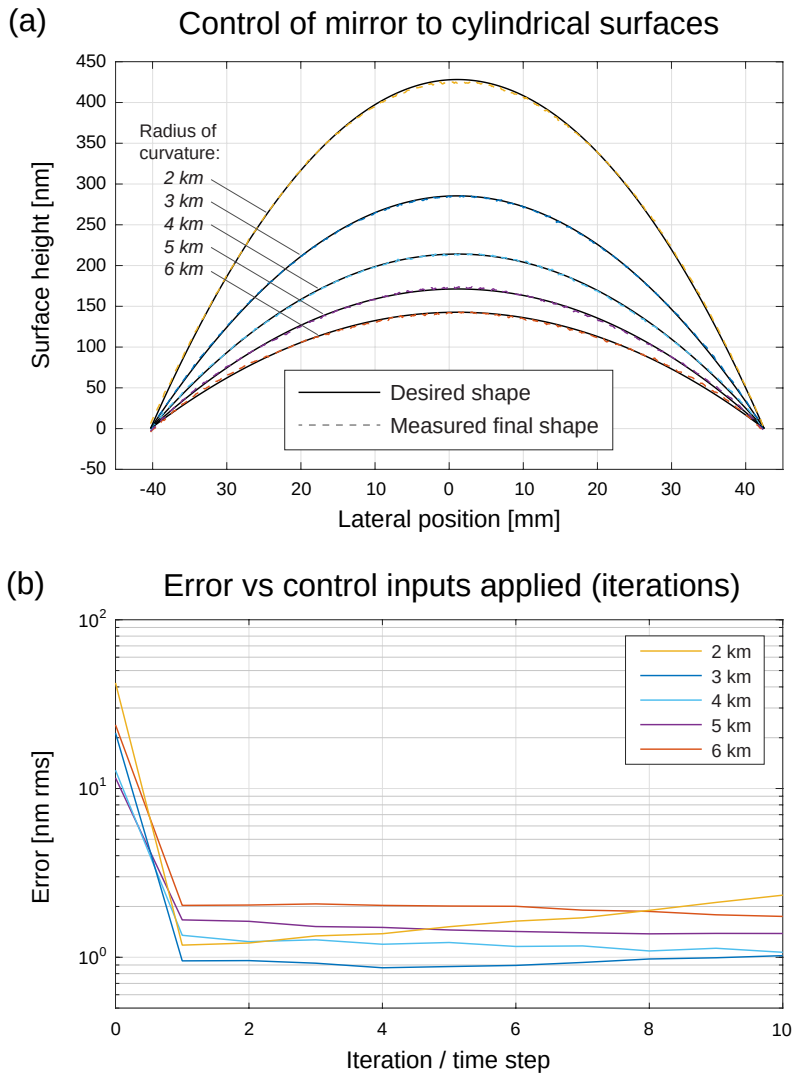

Figure 5.7: **Controlling mirror to cylindrical surface shapes.** (a) Measurements of the mirror surface taken after 10-step control algorithm is used to achieve cylindrical surface shapes. Piston was added to desired and measures surface profiles for visualization only. (b) Total shape error in nm RMS over 10 time steps.

Chapter 6

Conclusions and future work

In this chapter, we will present concluding remarks for each of the two main projects: aberration measurement in EUV systems, and aberration compensation using adaptive x-ray optics. We will also point out some avenues for further exploration.

6.1 Aberration measurement

We have demonstrated the recovery of imaging system aberrations using an in-situ speckle imaging technique. By relying on a stationary random surface, the technique does not suffer from manufacturing tolerances and uncertainties associated with known test targets. The sample is also an unbiased probe of the system aberrations, which is not the case in joint (pupil and sample) optimization techniques. We verified the accuracy of our reconstructions of system in aberrations both in simulations and in visible-light experiments.

We are able to reconstruct the field-dependent aberrations of a full-field EUV microscope using the atomic-scale roughness of photomask blanks and no additional hardware. Our results demonstrate that SHARP achieves diffraction-limited performance, with wavefront errors below $\lambda/21$ averaged over the center 5 µm × 5 µm region of the total captured field-of-view. We also demonstrated a measurement accuracy better than 4.0% ($\lambda/181$). This analysis was performed using only images acquired under standard operation of the microscope, and is useful when invasive techniques are difficult or impossible to implement, as is often the case for systems in ultra-high vacuum.

This work demonstrates that our technique is suitable for evaluating the performance of the next generation of industrial-grade microscopes that will be used in semiconductor manufacturing. In addition, as x-ray light source facilities progress towards diffraction-limited storage rings and free electron lasers, with high brightness, this versatile, in-situ technique will prove increasingly valuable in the characterization of coherent sources and beamline optical systems.

Future work

A major question that merits further exploration is the extent of information contained in images of speckle. We showed that the Zernike polynomials that comprise field-varying aberrations could be recovered from speckle images. Further analysis of those coefficients could then be performed to estimate more tangible properties of the imaging system, such as field curvature (field varying defocus).

Additionally, our low-order speckle calibration method may be performed at various positions across the field of view to determine field variation of surface roughness characteristics. For example, if the sample is known to be stationary but speckle analysis reveals field variation in estimated speckle size, this indicates that the imaging system has measurable distortion.

While the primary conclusions of our work resulted in an efficient and robust method for aberration measurement, we gained additional insights into the "illumination wavefront" via alternate image processing techniques. The basis for obtaining this information is shown in Fig. 6.1.

Moving left to right through the windowed sub-regions of the speckle image (see Fig. 6.1a) and observing their corresponding Fourier transforms, we see a counter-clockwise rotation in the orientation of the two shifted pupil circles. This may indicate that the local angle of illumination is changing across the imaging field of view. This suggests the possibility of a wavefront sensing technique based on speckle data. Such a technique would infer local angles of incident illumination from the Fourier transforms of speckle, and the wavefront would be reconstructed in a manner similar to that of a Shack-Hartmann sensor. However, a speckle image could be freely processed with nonuniform or adaptive sampling to retrieve the wavefront, rather than the uniformly spaced grid of measurements enforced by the Shack-Hartmann hardware.

However, a major complication to the proposal above is the fact that the direct relationship between speckle spectra across the field and local illumination angle is only true in a telecentric imaging system. In the case of SHARP - a single lens imaging system - this relationship is further complicated by non-telecentricity. Considering these effects, we expect to see quadratic variation of these "illumination angles" derived from pupil circle positions across the field. This can be fitted and removed from measurements, and the residuals may be interpreted as the true illumination incident on the sample. The results of applying this procedure are shown in Fig. 6.2 Upon removal of a strong quadratic component, the recovered incident wavefront is roughly flat. In addition, the fitted quadratic component may reveal information about the transverse alignment of the lens.

We emphasize that the analysis of illumination angles for wavefront information is entirely separate from the determination of aberrations from the interference patterns in the pupil circle overlap. While a non-planar wavefront can influence the positions of pupil circles in speckle image spectra across the field, this does not change the underlying polynomial that determines the interference pattern within their overlap.

Further exploration into speckle-based metrology could yield interesting and practical

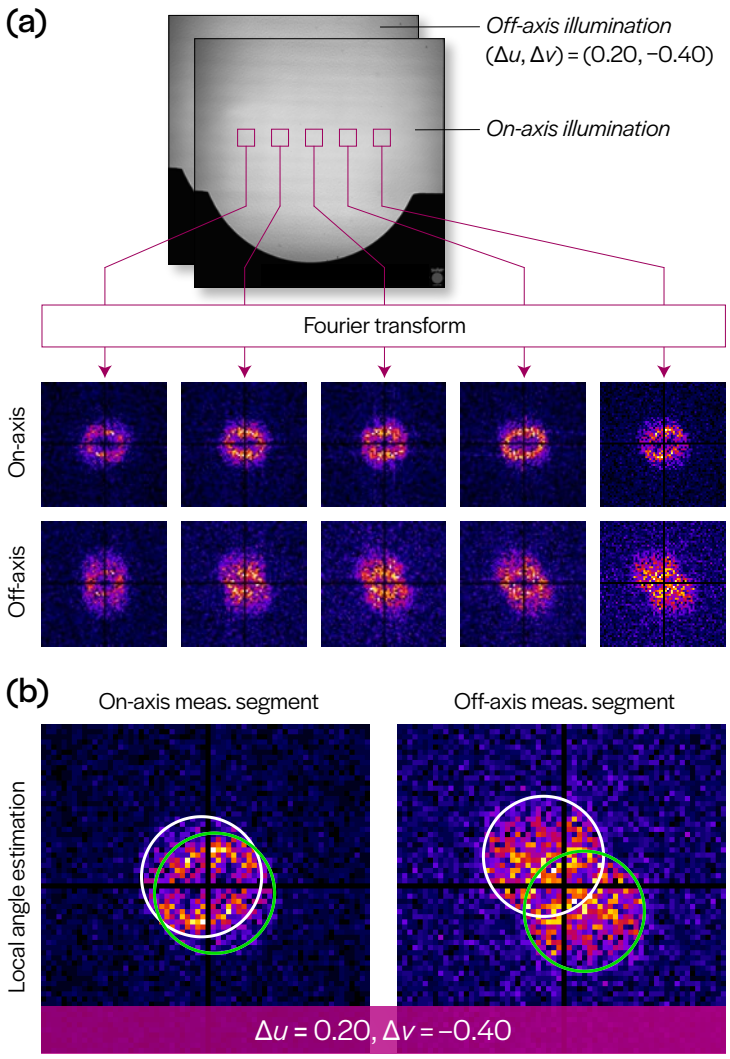

Figure 6.1: **Field variation and disambiguation of local illumination angle.** (a) Demonstration of the variation in computed Fourier spectrum magnitudes across the field in a single image, shown for both on- and off-axis illuminated images. (b) For a pair of images of the same sub-region, a known change in illumination angle $(\Delta u, \Delta v)$ disambiguates which circles correspond to the correct illumination angles (green) since only one pair of circles will have the correct relative shift.

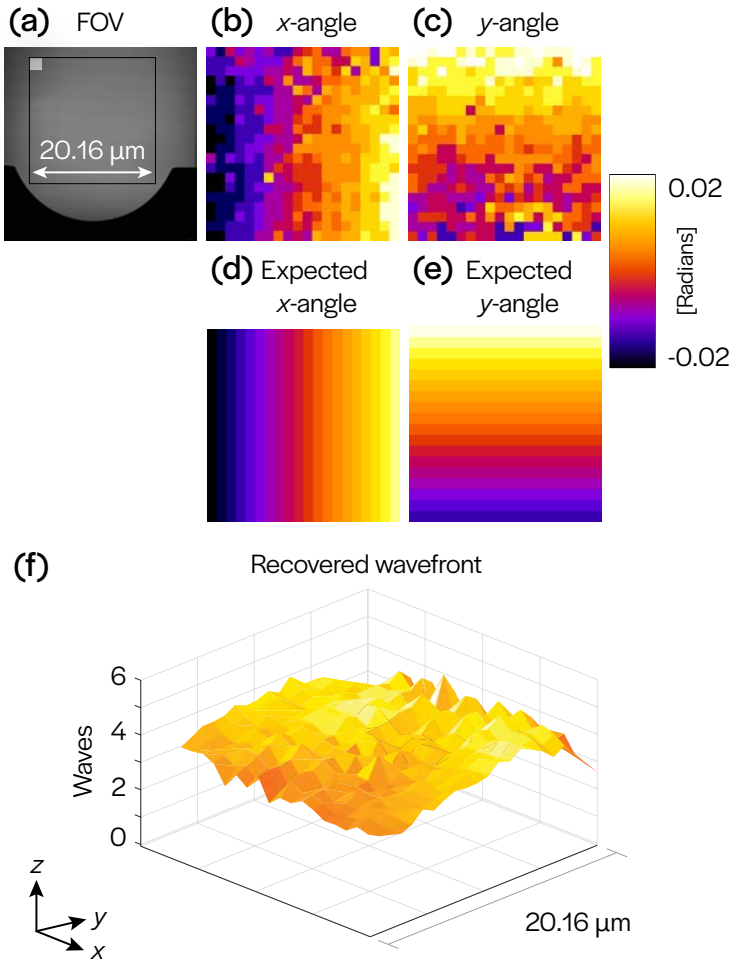

Figure 6.2: **Reconstruction of the illumination wavefront.** (a) Area of the FOV under consideration, with a single sub-region highlighted to demonstrate size. (b-c) Local angles of the wavefront surface normal in x and y, respectively. Each sample corresponds to a sub-region of the FOV. (d-e) Expected local illumination angles of the wavefront surface normal, based on the geometry of a single lens imaging system. (f) Illumination wavefront reconstructed via numerical integration of the difference between expected and recovered local illumination angles.

techniques for short-wavelength systems, especially where the only viable alternative to computational techniques is the use of highly specialized or invasive optical components.

An additional avenue for future work is the application of imaging system information derived from our speckle technique to the correction of aberrated images acquired by the microscope. This is particularly useful for computational imaging techniques, such as Fourier ptychography, where the image of the sample is reconstructed from both measurements and prior information. The results of calibrating system aberrations and using this information for image recovery has been simulated [44], showing improvements in the magnitude of tolerated aberrations in image reconstruction compared to joint recovery techniques [57]. Using calibrated field-varying aberrations as a prior in image reconstruction could enable diffraction-limited imaging performance over a much larger field of view than is currently usable in SHARP, thus greatly increasing throughput.

6.2 Aberration correction

We have shown that the combination of a data-driven model for piezo-bimorph adaptive mirror shape dynamics and an optimization-based control strategy were able to reduce residual mirror figure errors below 2 nm RMS, outperforming linear models and achieving the shape-control accuracy required to achieve diffraction-limited performance in the x-ray regime.

Our method effectively accounts for creep and hysteresis, nonlinear properties that currently limit the performance of such devices in open-loop operation. Accurate predictive modeling to achieve stable, arbitrary surface shapes is essential for effective deployment on high-coherent-flux x-ray beamlines where continuous feedback may be difficult to implement.

This calibration method is simple to implement and easily automated, requiring only a sequence of random shape commands and surface profile measurements. The data can be collected ex-situ, as presented in this paper, or even in-situ with a wavefront sensor, where the phase of the beam can be mapped back into the mirror shape if required. The method is also robust, providing accurate predictions and control across the full range of operation of the mirror. Other types of adaptive mirrors, such as resistive-element mirrors [13] can also be characterized with this technique.

The number of shape measurements required to build the training dataset is larger than what is required to acquire the characteristic functions in the linear model, but the training data can gathered during routine operation, over time. There is some flexiblity around the structure of the neural network itself (hyperparameters such as number of inputs, layers), but the performance level we found is very close to the noise level of our sensor, and needed no further refinement despite being rather economical.

Future work

Several aspects of our data-driven control technique can be further investigated for potential improvements in performance, both accuracy and speed. The structure of the learned

model (currently a fully-connected neural network) may be refined and its relevant hyperparameters may be adjusted. Additional data may be collected and used to train the model. In particular, this data may be acquired to describe the mirror over a greater range of operation than $\pm$100V, or it may be acquired to track convergence behavior over longer periods. Since the current input model uses 3 consecutive state (shape) measurements at fixed time intervals, one strategy to account for more diverse input conditions could be to add some delay between the application of an input and the series of measurements. Without applying new inputs and by varying the measurement delay, this could be used to capture convergence behavior in the neighborhood of the steady state surface profile, rather than the initial surface. Inclusion of this additional data would likely improve predictive performance of the model.

It would additionally be useful to study the feasibility of this data-driven control technique for a larger variety of experimental conditions to ensure that mirror states can be achieved without feedback. For example, an experiment may require that a series of measurements be taken with varying mirror surface curvature. It would be useful to determine the performance and limitations of planning multi-step trajectories.

Finally, we may also consider alternate methods of mirror surface metrology that may be simpler to integrate into practical systems. For example, at-wavelength Talbot interferometry can be used to reconstruct the mirror shape without the aperture limitations of our current visible-light Fizeau interferometry setup. However, this suffers from a spatial resolution that varies with the curvature of the surface, and backpropagation to an effective mirror plane introduces additional sources of error. It would be useful to determine whether this data may be used similarly to train a predictive model, and how the control performance compares. Along these lines, we may also consider "lossy" observation models, where we do not have direct access to the mirror surface shape. Effectively controlling the the system to an improved beam profile with limited state space information is another interesting research direction and highly useful for beamline operation.

Bibliography

[1] Simon G. Alcock et al. "Dynamic adaptive X-ray optics. Part I. Time-resolved optical metrology investigation of the bending behaviour of piezoelectric bimorph deformable X-ray mirrors". In: *Journal of Synchrotron Radiation* 26.1 (2019), pp. 36–44. DOI: 10.1107/S1600577518015953.

[2] Simon G. Alcock et al. "Dynamic adaptive X-ray optics. Part II. High-speed piezoelectric bimorph deformable Kirkpatrick–Baez mirrors for rapid variation of the 2D size and shape of X-ray beams". In: *Journal of Synchrotron Radiation* 26.1 (2019), pp. 45–51. ISSN: 16005775. DOI: 10.1107/S1600577518015965.

[3] Simon G. Alcock et al. "High-speed adaptive optics using bimorph deformable x-ray mirrors". In: *Review of Scientific Instruments* 90.2 (2019), p. 021712. DOI: 10.1063/1.5060737. eprint: https://doi.org/10.1063/1.5060737. URL: https://doi.org/10.1063/1.5060737.

[4] Gilles Aubert and Jean-Francois Aujol. "A variational approach to removing multiplicative noise". In: *SIAM Journal on Applied Mathematics* 68.4 (2008), pp. 925–946.

[5] Horace W Babcock. "Adaptive Optics Revisited". In: *Science* 249.4966 (1990), pp. 253–257. DOI: 10.1126/science.249.4966.253. URL: https://doi.org/10.1126/science.249.4966.253.

[6] Sébastien Bérujon et al. "Two-dimensional x-ray beam phase sensing". In: *Physical Review Letters* 108.15 (2012), pp. 1–5. ISSN: 00319007. DOI: 10.1103/PhysRevLett.108.158102. URL: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.108.158102.

[7] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN: 0387310738.

[8] J. H. Bruning et al. "Digital Wavefront Measuring Interferometer for Testing Optical Surfaces and Lenses". In: *Appl. Opt.* 13.11 (Nov. 1974), pp. 2693–2703. DOI: 10.1364/AO.13.002693. URL: http://ao.osa.org/abstract.cfm?URI=ao-13-11-2693.

[9] Frank M. Callier and Charles A. Desoer. *Linear Systems Theory*. Springer-Verlag New York, 1991. ISBN: 9781461269618.

[10] R. Cerbino et al. "X-ray-scattering information obtained from near-field speckle". In: *Nature Physics* 4.3 (2008), pp. 238–243. ISSN: 1745-2473. DOI: 10.1038/nphys837. URL: http://www.nature.com/doifinder/10.1038/nphys837.

[11] Jaebum Chung et al. "Computational aberration compensation by coded-aperture-based correction of aberration obtained from optical Fourier coding and blur estimation". In: *Optica* 6.5 (May 2019), pp. 647–661. DOI: 10.1364/OPTICA.6.000647. URL: http://opg.optica.org/optica/abstract.cfm?URI=optica-6-5-647.

[12] Jaebum Chung et al. "Wide field-of-view fluorescence image deconvolution with aberration-estimation from Fourier ptychography". In: *Biomed. Opt. Express* 7.2 (Feb. 2016), pp. 352–368. DOI: 10.1364/BOE.7.000352. URL: http://opg.optica.org/boe/abstract.cfm?URI=boe-7-2-352.

[13] Daniele Cocco et al. "Adaptive shape control of wavefront-preserving X-ray mirrors with active cooling and heating". In: *Optics Express* 28.13 (2020), p. 19242. ISSN: 1094-4087. DOI: 10.1364/oe.394310.

[14] Regina Eckert, Zachary F. Phillips, and Laura Waller. "Efficient illumination angle self-calibration in Fourier ptychography". In: *Appl. Opt.* 57.19 (July 2018), pp. 5434–5442. DOI: 10.1364/AO.57.005434. URL: http://ao.osa.org/abstract.cfm?URI=ao-57-19-5434.

[15] Rolf Erni. *Aberration-Corrected Imaging in Transmission Electron Microscopy: An Introduction*. World Scientific, 2010.

[16] James R Fienup et al. "Hubble Space Telescope characterized by using phase-retrieval algorithms". In: *Applied Optics* 32.10 (1993), pp. 1747–1767. DOI: 10.1364/AO.32.001747. URL: https://doi.org/10.1364/AO.32.001747.

[17] Walter Gander, Martin J. Gander, and Felix Kwok. *Scientific Computing*. Springer, 2014.

[18] Simi A. George et al. "Extreme ultraviolet mask substrate surface roughness effects on lithographic patterning". In: *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* 28.6 (2010), C6E23. ISSN: 10711023. DOI: 10.1116/1.3502436. URL: http://scitation.aip.org/content/avs/journal/jvstb/28/6/10.1116/1.3502436%20http://link.aip.org/link/JVTBD9/v28/i6/pC6E23/s1%7B%5C&%7DAgg=doi.

[19] Kenneth A. Goldberg and Valeriy V. Yashchuk. "Optimized mirror shape tuning using beam weightings based on distance, angle of incidence, reflectivity, and power". In: *Review of Scientific Instruments* 87.5 (2016), p. 051805. DOI: 10.1063/1.4950747. eprint: https://doi.org/10.1063/1.4950747. URL: https://doi.org/10.1063/1.4950747.

[20] Kenneth A. Goldberg et al. "Benchmarking EUV mask inspection beyond 0.25 NA". In: 7122 (2008). Ed. by Hiroichi Kawahira and Larry S. Zurbrick, 71222E. DOI: 10.1117/12.801529. URL: https://doi.org/10.1117/12.801529.

[21] Kenneth A. Goldberg et al. "The SEMATECH high-NA actinic reticle review project (SHARP) EUV mask-imaging microscope". In: *Photomask Technology 2013*. Ed. by Thomas B. Faure and Paul W. Ackmann. Vol. 8880. International Society for Optics and Photonics. SPIE, 2013, pp. 95–103. DOI: 10.1117/12.2026496. URL: https://doi.org/10.1117/12.2026496.

[22] Kenneth Alan Goldberg et al. "An EUV Fresnel zoneplate mask-imaging microscope for lithography generations reaching 8 nm". In: *SPIE*. Ed. by Michael L Rieger. Vol. 7969. Mar. 2011, pp. 796910–796910–12. DOI: 10.1117/12.881651. URL: http://spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.881651%20http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1350027.

[23] Joseph Goodman. *Introduction to Fourier Optics*. McGraw-hill, 2008.

[24] Joseph W Goodman. *Statistical Optics*. John Wiley & Sons, 2015.

[25] Eric P. Goodwin and James C. Wyant. *Field Guide to Interferometric Optical Testing*. SPIE Publications, 2006.

[26] Gautam Gunjala et al. "Optical transfer function characterization using a weak diffuser". In: *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXIII*. SPIE. 2016, p. 971315.

[27] D K Hamilton, C J R Sheppard, and T Wilson. "Improved imaging of phase gradients in scanning optical microscopy". In: *Journal of Microscopy* 135.3 (1984), pp. 275–286. DOI: 10.1111/j.1365-2818.1984.tb02533.x. URL: https://doi.org/10.1111/j.1365-2818.1984.tb02533.x.

[28] Bridget M Hanser et al. "Phase-retrieved pupil functions in wide-field fluorescence microscopy". In: *Journal of Microscopy* 216.1 (2004), pp. 32–48. DOI: 10.1111/j.0022-2720.2004.01393.x. URL: https://doi.org/10.1111/j.0022-2720.2004.01393.x.

[29] P Hariharan, BF Oreb, and Tomoaki Eiju. "Digital phase-shifting interferometry: a simple error-compensating phase calculation algorithm". In: *Applied Optics* 26.13 (1987), pp. 2504–2506. DOI: 10.1364/AO.26.002504. URL: https://doi.org/10.1364/AO.26.002504.

[30] M.H. Hayes. *Statistical Digital Signal Processing and Modeling*. Wiley India Pvt. Limited, 2009. ISBN: 9788126516100. URL: https://books.google.com/books?id=z0GqhOe9GNQC.

[31] Monson H. Hayes. *Statistical Digital Signal Processing and Modeling*. Wiley, 1996.

[32] E. Hecht. *Optics*. 4th. Addison-Wesley, 1998.

[33] Olivier Hignette, Andreas K. Freund, and Elia Chinchio. "Incoherent x-ray mirror surface metrology". In: *Materials, Manufacturing, and Measurement for Synchrotron Radiation Mirrors*. Ed. by Peter Z. Takacs and Thomas W. Tonnessen. Vol. 3152. International Society for Optics and Photonics. SPIE, 1997, pp. 188–199. DOI: 10.1117/12.295559. URL: https://doi.org/10.1117/12.295559.

[34] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of Convex Analysis*. Springer-Verlag, 2001.

[35] Lei Huang, Junpeng Xue, and Mourad Idir. "Controlling X-ray deformable mirrors during inspection". In: *Journal of Synchrotron Radiation* 23.6 (2016), pp. 1348–1356. ISSN: 16005775. DOI: 10.1107/S1600577516014600.

[36] Yoshio Ichii et al. "Development of a glue-free bimorph mirror for use in vacuum chambers". In: *Review of Scientific Instruments* 90.2 (2019), pp. 1–5. DOI: 10.1063/1.5066105.

[37] Na Ji, Daniel E Milkie, and Eric Betzig. "Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues". In: *Nature Methods* 7.2 (2010), pp. 141–147. DOI: 10.1038/nmeth.1411. URL: https://doi.org/10.1038/nmeth.1411.

[38] Hui Jiang et al. "A piezoelectric deformable X-ray mirror for phase compensation based on global optimization". In: *Journal of Synchrotron Radiation* 26.3 (2019), pp. 729–736. DOI: 10.1107/S1600577519003047.

[39] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization." In: *CoRR* abs/1412.6980 (2014). URL: https://doi.org/10.48550/arXiv.1412.6980.

[40] Earl J Kirkland. *Advanced Computing in Electron Microscopy*. Springer Science & Business Media, 2010.

[41] Ombeline de La Rochefoucauld et al. "Developments of EUV/x-ray wavefront sensors and adaptive optics at Imagine Optic". In: *Adaptive X-Ray Optics V*. Ed. by Daniele Spiga and Hidekazu Mimura. Vol. 10761. International Society for Optics and Photonics. SPIE, 2018, 107610E. DOI: 10.1117/12.2320927. URL: https://doi.org/10.1117/12.2320927.

[42] S. C. Leemann et al. "Demonstration of Machine Learning-Based Model-Independent Stabilization of Source Properties in Synchrotron Light Sources". In: *Physical Review Letters* 123.19 (2019), p. 194801. DOI: 10.1103/PhysRevLett.123.194801. URL: https://doi.org/10.1103/PhysRevLett.123.194801.

[43] Zachary Levinson et al. "Measurement of EUV lithography pupil amplitude and phase variation via image-based methodology". In: *Journal of Micro/Nanolithography, MEMS and MOEMS* 15.2 (2016). DOI: 10.1117/1.JMM.15.2.023508.

[44] Eric Li et al. *Exceeding the limits of algorithmic self-calibration in super-resolution imaging*. 2021. DOI: 10.48550/ARXIV.2109.07188. URL: https://arxiv.org/abs/2109.07188.

[45] Weiwei Li and Emanuel Todorov. "Iterative Linear Quadratic Regulator Design for Nonlinear Biological Movement Systems". In: *ICINCO*. 2004.

[46] D. D. Lowenthal. "Maréchal Intensity Criteria Modified for Gaussian Beams". In: *Appl. Opt.* 13.9 (Sept. 1974), pp. 2126–2133. DOI: 10.1364/AO.13.002126. URL: http://opg.optica.org/ao/abstract.cfm?URI=ao-13-9-2126.

[47] Daniel Malacara. *Optical Shop Testing (Wiley Series in Pure and Applied Optics)*. USA: Wiley-Interscience, 2007. ISBN: 0471484040.

[48] Daniel Malacara-Hernandez. "Review of interferogram analysis methods". In: *Optical Testing and Metrology III: Recent Advances in Industrial Optical Inspection*. Ed. by Chander Prakash Grover. Vol. 1332. International Society for Optics and Photonics. SPIE, 1991, pp. 678–689. DOI: 10.1117/12.51118. URL: https://doi.org/10.1117/12.51118.

[49] Satoshi Matsuyama et al. "Nearly diffraction-limited X-ray focusing with variable-numerical- aperture focusing optical system based on four deformable mirrors". In: *Scientific Reports* 6 (2016), p. 24801. DOI: 10.1038/srep24801.

[50] Daniel J. Merthe et al. "Methodology for optimal in situ alignment and setting of bendable optics for diffraction-limited focusing of soft x-rays". In: *Advances in Metrology for X-Ray and EUV Optics IV*. Ed. by Lahsen Assoufid, Peter Z. Takacs, and Anand Krishna Asundi. Vol. 8501. International Society for Optics and Photonics. SPIE, 2012, pp. 70–85. DOI: 10.1117/12.930023. URL: https://doi.org/10.1117/12.930023.

[51] Hidekazu Mimura et al. "Breaking the 10nm barrier in hard-X-ray focusing". In: *Nature Physics* 6.2 (2010), pp. 122–125. DOI: 10.1038/nphys1457. URL: http://dx.doi.org/10.1038/nphys1457.

[52] Ryan Miyakawa and Patrick Naulleau. "Preparing for the Next Generation of EUV Lithography at the Center for X-ray Optics". In: *Synchrotron Radiation News* 32.4 (2019), pp. 15–21. DOI: 10.1080/08940886.2019.1634432.

[53] Ryan Miyakawa et al. "AIS wavefront sensor: a robust optical test of exposure tools using localized wavefront curvature". In: *Extreme Ultraviolet (EUV) Lithography V*. Ed. by Obert R. Wood II and Eric M. Panning. Vol. 9048. International Society for Optics and Photonics. SPIE, 2014, 90483A. DOI: 10.1117/12.2048389. URL: https://doi.org/10.1117/12.2048389.

[54] Iacopo Mochi et al. "Improving the performance of the actinic inspection tool with an optimized alignment procedure". In: *Alternative Lithographic Technologies*. Ed. by Frank M. Schellenberg and Bruno M. La Fontaine. Vol. 7271. International Society for Optics and Photonics. SPIE, 2009, p. 727123. DOI: 10.1117/12.814261. URL: https://doi.org/10.1117/12.814261.

[55] Patrick Naulleau et al. "Extreme ultraviolet mask roughness: requirements, characterization, and modeling ". In: *Photomask and Next-Generation Lithography Mask Technology XXI*. Ed. by Kokoro Kato. Vol. 9256. International Society for Optics and Photonics. SPIE, 2014, 92560J. DOI: 10.1117/12.2070303. URL: https://doi.org/10.1117/12.2070303.

[56] Patrick P Naulleau, Iacopo Mochi, and Kenneth A. Goldberg. "Optical modeling of Fresnel zoneplate microscopes". In: *Applied optics* 50.20 (July 2011), pp. 3678–84. ISSN: 1539-4522. DOI: 10.1364/AO.50.003678. URL: http://www.ncbi.nlm.nih.gov/pubmed/21743581.

[57] Xiaoze Ou, Guoan Zheng, and Changhuei Yang. "Embedded pupil function recovery for Fourier ptychographic microscopy". In: *Optics Express* 22.5 (2014), pp. 4960–4972.

[58] K. B. Petersen and M. S. Pedersen. *The Matrix Cookbook*. Version 20121115. Nov. 2012. URL: http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html.

[59] Lisa A. Poyneer et al. "X-ray metrology and performance of a 45-cm long x-ray deformable mirror". In: *Review of Scientific Instruments* 87.5 (2016). ISSN: 10897623. DOI: 10.1063/1.4950739. URL: http://dx.doi.org/10.1063/1.4950739.

[60] Sudharshanan Raghunathan et al. "Experimental measurements of telecentricity errors in high-numerical-aperture extreme ultraviolet mask images". In: *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena* 32 (2014), 06F801. ISSN: 2166-2746. DOI: 10.1116/1.4901876. URL: http://scitation.aip.org/content/avs/journal/jvstb/32/6/10.1116/1.4901876.

[61] Matthew P Rimmer and James C Wyant. "Evaluation of large aberrations using a lateral-shear interferometer having variable shear". In: *Applied Optics* 14.1 (1975), pp. 142–150. DOI: 10.1364/AO.14.000142. URL: https://doi.org/10.1364/AO.14.000142.

[62] Vasco Ronchi. "Forty Years of History of a Grating Interferometer". In: *Appl. Opt.* 3.4 (Apr. 1964), pp. 437–451. DOI: 10.1364/AO.3.000437. URL: http://ao.osa.org/abstract.cfm?URI=ao-3-4-437.

[63] Austin Roorda et al. "Adaptive optics scanning laser ophthalmoscopy". In: *Optics Express* 10.9 (2002), pp. 405–412. DOI: https://doi.org/10.1364/OE.10.000405. URL: https://doi.org/10.1364/OE.10.000405.

[64] H Rose. "Nonstandard imaging methods in electron microscopy". In: *Ultramicroscopy* 2 (1976), pp. 251–267. DOI: 10.1016/S0304-3991(76)91538-2. URL: https://doi.org/10.1016/S0304-3991(76)91538-2.

[65] Johannes Ruoff. "Impact of mask topography and multilayer stack on high NA imaging of EUV masks". In: *SPIE Photomask Technology*. Vol. 7823. 2010, 78231N. ISBN: 9780819483379. DOI: 10.1117/12.864120. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=724666%20http://spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.864120.

[66] Manuel Sanchez Del Rio et al. "Compensation of heat load deformations using adaptive optics for the ALS upgrade: A wave optics study". In: *Journal of Synchrotron Radiation* 27 (2020), pp. 1141–1152. ISSN: 16005775. DOI: 10.1107/S1600577520009522.

[67] Kawal Sawhney, Simon G Alcock, and Riccardo Signorato. "A novel adaptive bimorph focusing mirror and wavefront corrector with sub-nanometre dynamical figure control". In: *Proc. SPIE*. Vol. 7803. 2010, p. 780303. DOI: 10.1117/12.861593. URL: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.861593.

[68] Aamod Shanker et al. "Off-axis aberration estimation in an EUV microscope using natural speckle". In: *Imaging and applied optics Congress*. 1. 2016, pp. 2–5. ISBN: 9781943580156. DOI: 10.1364/ISA.2016.ITh1F.2.

[69] Xianbo Shi et al. "Prototype design and experimental tests of a zoom mirror system for the APS upgrade". In: *Proc. SPIE*. Vol. 11491. 2020, p. 1149110. DOI: 10.1117/12.2569129.

[70] Robert H Shumway and David S Stoffer. *Time Series Analysis and its Applications*. Springer, 2011.

[71] MM Siddiqui. "Statistical inference for Rayleigh distributions". In: *Journal of Research of the National Bureau of Standards, Sec. D* 68.9 (1964), p. 1007.

[72] H. Song et al. "Extracting hysteresis from nonlinear measurement of wavefront-sensorless adaptive optics system". In: *Optics Letters* 34.1 (2009), p. 61. DOI: 10.1364/OL.34.000061. URL: https://www.osapublishing.org/abstract.cfm?URI=ol-34-1-61.

[73] Eduardo D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Springer New York, 1998. ISBN: 9781461268253.

[74] John CH Spence. *Experimental High-Resolution Electron Microscopy*. Oxford University, 1988.

[75] Norbert Streibl. "Three-Dimensional Imaging by a Microscope". In: *JOSA A* 2.2 (1985), pp. 121–127. DOI: 10.1364/JOSAA.2.000121. URL: https://doi.org/10.1364/JOSAA.2.000121.

[76] John P. Sutter et al. "1 m long multilayer-coated deformable piezoelectric bimorph mirror for adjustable focusing of high-energy X-rays". In: *Optics Express* 27.11 (2019), p. 16121. DOI: 10.1364/OE.27.016121. URL: https://www.osapublishing.org/abstract.cfm?URI=oe-27-11-16121.

[77] Mitsuo Takeda, Hideki Ina, and Seiji Kobayashi. "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry". In: *J. Opt. Soc. Am.* 72.1 (Jan. 1982), pp. 156–160. DOI: 10.1364/JOSA.72.000156. URL: http://www.osapublishing.org/abstract.cfm?URI=josa-72-1-156.

[78] Robert K Tyson. *Principles of Adaptive Optics*. CRC, 2015.

[79] Maurizio Vannoni et al. "Characterization of a piezo bendable X-ray mirror". In: *Journal of Synchrotron Radiation* 23.1 (Jan. 2016), pp. 169–175. ISSN: 1600-5775. DOI: 10.1107/S1600577515019803. URL: http://scripts.iucr.org/cgi-bin/paper?S1600577515019803.

[80] E Voelkl. "Using diffractograms to evaluate optical systems with coherent illumination". In: *Optics Letters* 28.23 (2003), pp. 2318–2320. DOI: 10.1364/OL.28.002318. URL: https://doi.org/10.1364/OL.28.002318.

[81] Yow-Gwo Wang. "Key Challenges in EUV Mask Technology: Actinic Mask Inspection and Mask 3D Effects". PhD thesis. UC Berkeley, 2017.

[82] Stephen Wright and Jorge Nocedal. *Numerical Optimization*. Springer Science, 1999.

[83] Kenji Yamazoe, Iacopo Mochi, and Kenneth Alan Goldberg. "Gradient descent algorithm applied to wavefront retrieval from through-focus images by an extreme ultraviolet microscope with partially coherent source". In: *Journal of the Optical Society of America A* 31.12 (Nov. 2014), B34. ISSN: 1084-7529. DOI: 10.1364/JOSAA.31.000B34. URL: http://www.opticsinfobase.org/abstract.cfm?URI=josaa-31-12-B34.

[84] F Zemlin et al. "Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms". In: *Ultramicroscopy* 3 (1978), pp. 49–60.

[85] Guoan Zheng, Roarke Horstmeyer, and Changhuei Yang. "Wide-field, high-resolution Fourier ptychographic microscopy". In: *Nature Photonics* 7.9 (2013), p. 739.