AFRL-RY-WP-TR-2021-0132

ITERATIVE PHASE ESTIMATION ALGORITHMS IN INTERFEROMETRIC SYSTEMS (Preprint)

Wesley Farriss
LADAR Technology Branch
Multispectral Sensing & Detection Division

MAY 2021
Final Report

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
May 2021	Thesis	21 May 2021 –21 May 2021

4. TITLE AND SUBTITLE
ITERATIVE PHASE ESTIMATION ALGORITHMS IN INTERFEROMETRIC SYSTEMS (Preprint)

5a. CONTRACT NUMBER
N/A

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
N/A

6. AUTHOR(S)
Wesley E. Farriss

5d. PROJECT NUMBER
N/A

5e. TASK NUMBER
N/A

5f. WORK UNIT NUMBER
N/A

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Sensors Directorate (AFRL/RYMM)
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command
United States Air Force

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Sensors Directorate
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command
United States Air Force

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/RYMM

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RY-WP-TR-2021-0132

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
PAO case number AFRL-2021-1575, Clearance Date 21 May 2021. Submitted in partial fulfillment of the requirements for the degree Doctor of Philosophy to the Institute of Optics Arts, Sciences and Engineering, Edmund A. Hajim School of Engineering and Applied Sciences, University of Rochester, Rochester, NY. The U.S. Government is joint author of this work and has the right to use, modify, reproduce, release, perform, display, or disclose the work. Report contains color.

14. ABSTRACT
Coherent optical interferometry has a long history of enabling extremely precise measurements at length scales of less than the wavelength of light used in the interferometer. It is the ability of these systems to measure both the relative phase and amplitude information of the optical field by interfering two beams of light which makes them so useful. Interferometric techniques have also been adopted for use in both imaging/ sensing technologies. For imaging systems in ideal conditions, the ability to measure both phase and amplitude information in one transverse plane allows for the calculation of that field's phase and amplitude distribution in any other transverse plane. However, the presence of atmospheric turbulence unpredictably alters the index of refraction in the propagation medium thereby adversely affecting the fidelity with which the phase and amplitude in a transverse plane can be calculated. …… continued on back of page.

15. SUBJECT TERMS
3D, algorithim development, phase retrieval, phase screens, turbulence mitigation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Wesley E. Farriss
Unclassified	Unclassified	Unclassified	SAR	219	19b. TELEPHONE NUMBER *(Include Area Code)* N/A

14. ABSTRACT, continued

To address this problem, we demonstrate iterative sharpness maximization (ISM) correction of anisoplantic turbulence effects in simulated range-compressed holography (RCH) fields and their corresponding range images. Our turbulence correction estimated four phase screens distributed along the path of optical propagation using nonlinear optimizations aided by the method of sieves technique. We conducted a study of range images created from single speckle realization 3D RCH fields subjected to twenty different turbulence profiles at five different strengths of turbulence, D/r0 = 7, 14, 21, 28, and 36. Range images showed significant improvement for all strengths of turbulence. To assist in correction, we introduced a novel constraint limiting the spread of energy in the corrected pupil. Corrected range images were qualitatively indistinguishable from unaberrated range images in all but the most severe turbulence case, D/r0 = 36. Additionally, our algorithm was tested for fields affected by shot noise. Mean target photons per speckle ranged from 10−2 to 102 in these simulations. For an effective D/r0 = 36, range images corrected from fields with 102 mean photons per speckle had very similar RMSE when compared to corrected noiseless range images. On aver-age, corrected range images created from fields with 1 mean target photons per speckle differed by less than 5% RMSE from noiseless corrected range images.

We went on to construct a RCH system in a laboratory setting using a linear frequency modulated CW laser and a high frame rate camera which allowed us to create 3D images of laboratory targets. Data was collected both with and without the effects of turbulence. In the former, multiple Lexitek turbulence screens were used to aberrate the image fields of our lab target at two different effective strengths of anisoplanatic turbulence, D/r0 = 7 and D/r0 = 16, respectively. Both of these sets of real aberrated image fields showed profound improvement in quality after correction with our phase ISM turbulence mitigation algorithm.

Novel interferometric systems are also being developed which enable modal analysis of an optical field. This generalized optical interferometry (GOI) treats coherent optical fields as a linear superposition of transverse modes and recovers the amplitudes of modal weighting coefficients. In order to maximize the utility of these systems, we used phase retrieval by nonlinear optimization to recover the phase of these modal weighting coefficients.

Algorithms were developed both for array detector and for bucket detector systems. Information diversity increased the robustness of both algorithms by better constraining the solutions. In our array detection phase retrieval, the algorithm was able to recover nearly all coefficient phases for simulated fields consisting of up to 21 superpositioned Hermite Gaussian modes from simulated data and proved to be resilient to shot noise.

Similarly, the algorithm we developed using data from a simulated bucket detector was able to consistently recover better than 95% of coefficient phases for simulated random fields consisting of up to 21 superpositioned Hermite Gaussian modes using between three and seven measurements per unknown phase coefficient. With shot noise, the algorithm achieved performance on par with noiseless simulations with 106 mean signal photons per measurement. The role played by number of measurements per unknown (mpu), photons per unknown per measurement (ppu), and order of superposition in the bucket detection algorithm's performance was also explored.-

Iterative Phase Estimation Algorithms in Interferometric Systems

by

Wesley E Farriss

Submitted in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

James R. Fienup

The Institute of Optics

Arts, Sciences and Engineering

Edmund A. Hajim School of Engineering and Applied Sciences

University of Rochester Rochester, New York

2021

Dedicated to my father, C.W. Farriss II, and a man who was like a father to me, Gerald "Butch" Spangler Sr. The first gave me my love of science and math. The second nurtured and enabled my love of the algorithm. Thanks, Dad. Rest in peace, Butch.

Table of Contents

Biographical Sketch

Wesley Farriss was born in Cuyahoga Falls, OH, USA on November 22, 1982. He attended the University of Akron, in Ohio, and graduated with a Bachelor of Science degree in Physics in 2009. After graduation, he commissioned as an active duty officer in the US Army. Wesley served four years on active duty in both CBRN and intelligence roles and completed a twelve month combat tour in Afghanistan's Khost province as an air assault mission planner, supporting US special operations forces while serving with the 10th Mountain Division's Combat Aviation Brigade. Wesley left active duty service to pursue his love of physics and mathematics and joined the Army reserves in Fall 2013. He began doctoral studies in Optics that same Fall at the University of Rochester's world-renowned Institute of Optics. He was awarded an MS in Optics in December 2018. He conducted research in support of the Office of Naval Research and the Air Force Research Laboratories in the fields wavefront retrieval/sensing, computational imaging, and turbulence mitigation, under the guidance of Dr. James R. Fienup.

Peer-reviewed Publications

1. W. E. Farriss, J. R. Fienup, T. Malhotra, and A. N. Vamivakas, Phase retrieval in generalized optical interferometry systems, Opt. Express **26**, 2191–2202 (2018)

2. T. Malhotra, W. E. Farriss, J. Hassett, A. F. Abouraddy, J. R. Fienup, and A. N. Vamivakas, Interferometric spatial mode analyzer with a bucket detector, Opt. Express **26**, 8719–8728 (2018)

Conference Publications

1. T. Malhotra, W. E. Farriss, J. R. Fienup, A. F. Abouraddy, and A. N. Vamivakas, Optical beam modal analysis via generalized two-path interferometry, in Frontiers in Optics, (Optical Society of America, 2015), pp. FW4G–4

2. T. Malhotra, W. E. Farriss, J. R. Fienup, A. F. Abouraddy, and A. N. Vamivakas, Optical beam spatial modal analysis using a two-path generalized Michelson interferometer, in Frontiers in Optics, (Optical Society of America, 2016), pp. FF5G–2

3. W. E. Farriss, J. R. Fienup, T. Malhotra, and A. N. Vamivakas, Robustness of joint-plane phase retrieval in generalized two-path interferometry, in Coherent Laser Radar Conference, (2016)

4. W. E. Farriss, J. R. Fienup, T. Malhotra, and A. N. Vamivakas, Multiplane phase retrieval in generalized two-path interferometry, in Frontiers in Optics, (Optical Society of America, 2016), pp. FF1G–2

5. W. E. Farriss, T. Malhotra, A. N. Vamivakas, and J. R. Fienup, Phase retrieval in generalized two-path interferometry, in Computational Optics and Sensing, (Optical Society of America, 2016), pp. CT4C–3

6. W. E. Farriss, J. R. Fienup, T. Malhotra, and A. N. Vamivakas, Single-pixel phase retrieval in generalized interferometry, in Computational Optical Sensing and Imaging, (Optical Society of America, 2017), pp. CTu1B–2

7. W. E. Farriss and J. R. Fienup, Application of algorithmic differentiation to turbulence mitigation in digital holography, (2018)

8. W. E. Farriss, J. R. Fienup, J. W. Stafford, and N. J. Miller, Sharpness-based correction methods in holographic aperture ladar (HAL), Proc. SPIE **10772**, 107720K (2018)

9. W. E. Farriss and J. R. Fienup, Multiplane turbulence correction in range-compressed holography, in Computational Optical Sensing and Imaging, (Optical Society of America, 2020), pp. JTu4A–3

Acknowledgments

A doctoral education is a team effort. My graduate education was probably the truest example of this fact. I began my PhD in the Institute of Optics at the University of Rochester while on terminal leave from active duty service in the United States Army. Prior to beginning the program, I knew this adjustment would be challenging. I quickly discovered that the scale of this challenge was far larger than anticipated. The amount of care, attention, and patience extended to me by the Institute of Optics faculty, staff, and students has been truly remarkable.

Professor James R. Fienup, my advisor, has been the cornerstone of this undertaking. There are many tangible benefits to being Dr. Fienup's graduate student, his technical expertise and detail-oriented instruction, a plethora of career and research opportunities, frequent access to fattening foods, and professional career guidance based on a lifetime of hard work and achievement. Indeed, these were the reasons why I chose him as my advisor during my first year in the PhD program.

Far more important than all of those tangible benefits has been the profound decency and patience he has shown me. Serving as anyone's graduate advisor is a difficult task at times, I suspect. Serving as my graduate advisor required a larger expenditure of both time and effort. Dr. Fienup, thank you for making me a member of your team and, perhaps more importantly, thank you for continuing to embrace me as a member of your team during the more challenging times. Without your grace and compassion, this document would not exist and I would be much poorer in both knowledge and spirit.

I would be remiss if I did not specifically acknowledge professors Nicholas George, Carlos Stroud, Miguel Alonso, and Nick Vamivakas by name. I am forever grateful to each of you for seeing potential in me when it may not have been otherwise apparent.

To my knucklehead year-group buddies, Robby Pettit, Rodrigo Gutierrez, Aku Antikainen, Charlie Granger, Dan Nikolov, José Carlos Pérez. Thanks for everything, gents. I miss you all a great deal. You're a fine bunch of bastards and I'd do optics with you on a cocktail napkin anytime. The next round is on me, should you actually read this.

Next, I'd like to single out my best friend, research group-mate, office-mate, work-husband, and fellow dog-lover, Scott Paine. I've been extraordinarily lucky to find a friend as selfless and tolerant of my blatherskite as you. Anywhere, anytime, just say the word and I'll be there to help you dispose of the bodies.

To past and present members of Dr. Fienup's research group, Alden Jurling, Dustin Moore, Alex Iachetta, Zach DeSantis, Aaron Michalko, Scott Will, Joseph Tang, and Matt Banet, your technical expertise and solidarity in attending years of group meetings has been essential to my success. I look forward to seeing what each of you does with your extraordinary skill set.

Additionally, I would like my colleagues and collaborators at the Air Force Research Labs Sensors Directorate Lidar Branch (RYMM), Nick Miller, Jason Stafford, Dave Rabb, and Matt Gebhardt. You've given me space, time, assistance, and opportunity to excel. Thank you.

To my undergraduate professors at the University of Akron, Robert Mallik, Alper Buldum, Ben Hu, Ang Chen, and Neil Raber. I took a round-about way in getting here, but I've finally made use of the foundational skills you taught me.

Most importantly, to my puppies Kora(RIP), Summer, and Wilson. I really only did this PhD so I could afford to buy you furniture to nap on, cars for you to ride in, and windows for you to stare out of.

Finally, to anyone I may have neglected to thank, please do not misinterpret my having forgotten to thank you with a lack of gratitude. There are a truly astounding number of people I owe thanks to for their aid during the last eight years.

Abstract

Coherent optical interferometry has a long history of enabling extremely precise measurements at length scales of less than the wavelength of light used in the interferometer. It is the ability of these systems to measure both the relative phase and amplitude information of the optical field by interfering two beams of light which makes them so useful.

Interferometric techniques have also been adopted for use in both imaging/sensing technologies. For imaging systems in ideal conditions, the ability to measure both phase and amplitude information in one transverse plane allows for the calculation of that field's phase and amplitude distribution in any other transverse plane. However, the presence of atmospheric turbulence unpredictably alters the index of refraction in the propagation medium thereby adversely affecting the fidelity with which the phase and amplitude in a transverse plane can be calculated.

To address this problem, we demonstrate iterative sharpness maximization (ISM) correction of anisoplantic turbulence effects in simulated range-compressed

holography (RCH) fields and their corresponding range images. Our turbulence correction estimated four phase screens distributed along the path of optical propagation using nonlinear optimizations aided by the method of sieves technique. We conducted a study of range images created from single speckle realization 3D RCH fields subjected to twenty different turbulence profiles at five different strengths of turbulence, $D/r_0 = 7,\ 14,\ 21,\ 28$, and 36. Range images showed significant improvement for all strengths of turbulence. To assist in correction, we introduced a novel constraint limiting the spread of energy in the corrected pupil. Corrected range images were qualitatively indistinguishable from unaberrated range images in all but the most severe turbulence case, $D/r_0 = 36$. Additionally, our algorithm was tested for fields affected by shot noise. Mean target photons per speckle ranged from 10^{-2} to 10^{2} in these simulations. For an effective $D/r_0 = 36$, range images corrected from fields with 10^2 mean photons per speckle had very similar RMSE when compared to corrected noiseless range images. On average, corrected range images created from fields with 1 mean target photons per speckle differed by less than 5% RMSE from noiseless corrected range images.

We went on to construct a RCH system in a laboratory setting using a linear frequency modulated CW laser and a high frame rate camera which allowed us to create 3D images of laboratory targets. Data was collected both with and without the effects of turbulence. In the former, multiple Lexitek turbulence screens were used to aberrate the image fields of our lab target

at two different effective strengths of anisoplanatic turbulence, $D/r_0 = 7$ and $D/r_0 = 16$, respectively. Both of these sets of real aberrated image fields showed profound improvement in quality after correction with our phase ISM turbulence mitigation algorithm.

Novel interferometric systems are also being developed which enable modal analysis of an optical field. This generalized optical interferometry (GOI) treats coherent optical fields as a linear superposition of transverse modes and recovers the amplitudes of modal weighting coefficients. In order to maximize the utility of these systems, we used phase retrieval by nonlinear optimization to recover the phase of these modal weighting coefficients.

Algorithms were developed both for array detector and for bucket detector systems. Information diversity increased the robustness of both algorithms by better constraining the solutions. In our array detection phase retrieval, the algorithm was able to recover nearly all coefficient phases for simulated fields consisting of up to 21 superpositioned Hermite Gaussian modes from simulated data and proved to be resilient to shot noise.

Similarly, the algorithm we developed using data from a simulated bucket detector was able to consistently recover better than 95% of coefficient phases for simulated random fields consisting of up to 21 superpositioned Hermite Gaussian modes using between three and seven measurements per unknown phase coefficient. With shot noise, the algorithm achieved performance on par with noiseless simulations with $\mathbf{10^6}$ mean signal photons per measure-

ment. The role played by number of measurements per unknown (mpu), photons per unknown per measurement (ppu), and order of superposition in the bucket detection algorithm's performance was also explored.

Contributors and Funding Sources

This work was supported by a dissertation committee consisting of Professor James Fienup (advisor), Professor A. Nick Vamivakas, and Professor Scott Carney of The Institute of Optics, Professor Mujdat Cetin of the Department of Electrical and Computer Engineering department, and Dr. Jason Stafford of the US Air Force Research Labs (AFRL).

Work in Chapters 2, 3, and 4 would not have been possible without the support AFRL's Sensors Directorate lidar branch (RYMM). In particular, I need to acknowledge that the experimental work in Chapter 4 was made possible only through the use of AFRL's facilities and equipment located at Wright-Patterson Air Force Base in Dayton, OH.

Nicholas Miller (AFRL) made significant contributions to the experimental work in Chapter 4. Nick was the lead researcher on the effort to construct our range-compressed holography lab apparatus detailed in this chapter as well as offering indispensable assistance in the writing of this chapter, as well as providing later simulated data sets for work in Chapter 3. Jason Stafford's

contribution of earlier data sets for work in Chapter 3 were also significant and are greatly appreciated.

Tanya Malhotra and Prof. A. Nick Vamivakas were both active collaborators on the efforts in Chapter 5.

Work in Chapters 2, 3, and 4 was funded by the Air Force Research labs Sensors Directorate lidar branch (RYMM).

Work in Chapter 5 was funded by U. S. Navy Office of Naval Research grant (N00014-14-1-0260).

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Interferometry

Interferometry is an important measurement technique which has been in regular use for well over a century [1, 2]. It is integral to many sub-fields of astronomy [3], physics [4], and engineering [5]. In the most general sense, interferometry is used to analyze physical phenomena which can be described by wave mechanics, e.g. sound waves, light waves, particles, etc. In classical wave optics, it is conducted by measuring a superposition of light waves (electromagnetic fields) incident on a photodetector.

These light waves are composed of coupled electric and magnetic fields moving (propagating) through some medium or free space. Since the electric and magnetic fields are inextricably linked by Maxwell's equations, only the state of either the electric field or the magnetic field needs to be known in

order to mathematically describe how light travels [6]. The convention in the field of optics is to describe the electric field, and to omit mention of the magnetic field component of light. In this work we use the terms, light, light waves, electric fields, and fields interchangeably.

Electric fields are generally complex-valued. As such, the field's values can be expressed in the form $a \exp(ib)$, where a is a non-negative real number, b is a real number, and i is the imaginary number. This is known as phasor notation of a complex number. A monochromatic scalar electric field (electric fields where we disregard the orientation of the electric field, i.e. polarization) can be expressed as [7],

$$U(\mathbf{r}_\perp, z, t) = A(\mathbf{r}_\perp, z) \exp\left\{i[\phi(\mathbf{r}_\perp, z) - 2\pi\nu t]\right\}, \tag{1.1}$$

where $A(\mathbf{r}_\perp, z)$ is the non-negative real-valued amplitude, $\phi(\mathbf{r}_\perp, z)$ is the spatially varying phase, ν is the constant frequency, t is time, the vector $\mathbf{r}_\perp = (x, y)$ describes a point in the transverse plane orthogonal to the optical axis, and z is the location of the transverse plane along the optical axis. All fields in this work are assumed to be expressed in phasor notation. Additionally, the spatial phase will more succinctly be referred to as simply "phase" and the non-negative real-valued amplitude as "amplitude." The energy of the field is directly proportional to the modulus-squared of the field,

$$I(\mathbf{r}_\perp, z, t) = \left|U(\mathbf{r}_\perp, z, t)\right|^2. \tag{1.2}$$

We refer to this quantity as the field's "intensity".

The mathematical operation which describes the process of interfering fields is addition; fields add. The mathematical expression for the intensity of two interfered monochromatic fields, $u_1(\mathbf{r}_\perp)$ and $u_2(\mathbf{r}_\perp)$, in a transverse plane is

$$\begin{aligned} I_{12}(\mathbf{r}_\perp) &= |U_1(\mathbf{r}_\perp) + U_2(\mathbf{r}_\perp)|^2 \\ &= A_1^2(\mathbf{r}_\perp) + A_2^2(\mathbf{r}_\perp) \\ &\quad + 2A_1(\mathbf{r}_\perp)A_2(\mathbf{r}_\perp)\cos\left[\phi_1(\mathbf{r}_\perp) - \phi_2(\mathbf{r}_\perp)\right]. \end{aligned} \tag{1.3}$$

The sum of the first two terms is known as the bias and the third term is the interference or modulation term. If $A_1(\mathbf{r}_\perp) = A_2(\mathbf{r}_\perp) = A(\mathbf{r}_\perp)$, this periodic expression oscillates between zero and $4A^2(\mathbf{r}_\perp)$. The intensity distributions from interfered coherent electric fields typically form what are called fringe patterns. Two examples of different types of interferometric fringe pattern intensities are shown in Fig. 1.1. The light and dark regions in the fringe patterns of Fig. 1.1 are indicative of whether the waves in that region add constructively (the peak of one wave aligning with the peak of the other) or the waves add destructively (the peak of one wave aligning with the trough of the other), respectively. Fig. 1.1 (a) is an example of a case where the number of fringes can be used to determine the surface height deviations in either transmissive or reflective surfaces. This works best for surfaces which are generally smooth on the order of a wavelength of the illumination. We

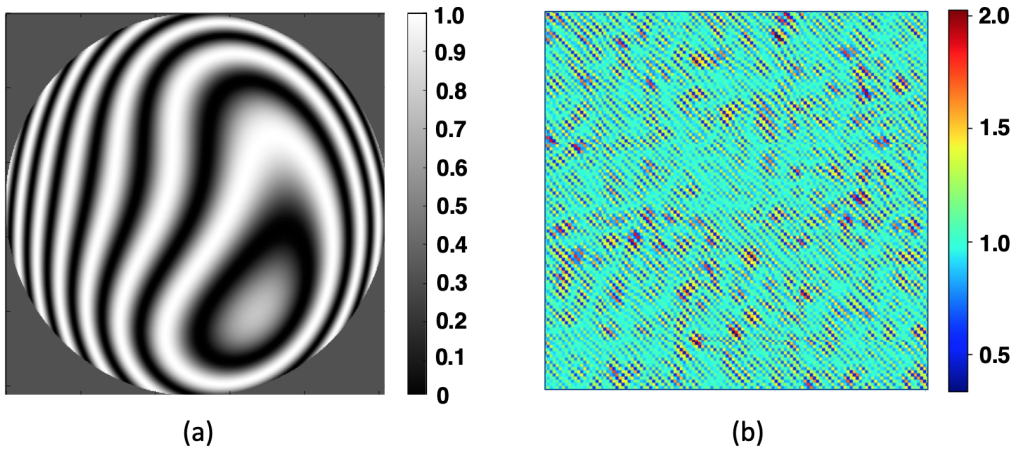

Figure 1.1: (a) Simulated smooth surface homodyne interferogram. (b) Simulated rough surface heterodyne interferogram.

see in this fringe pattern that the minimum intensity value is effectively zero, implying that the amplitudes of the two waves must be approximately equal. Fig. 1.1 (b) shows a fringe pattern of light which has been scattered from a target whose surface is rough on the order of a wavelength. The differing amplitudes of the two fields are apparent, as the minimum normalized intensity value is nonzero. When we say a field is coherent, it is sufficient for our purposes to think of this as meaning the light is monochromatic. This implies that the distance between the wave peaks in our electric fields remain constant in time regardless of the relative time delay between waves and regardless of the point in space where interference is occurring, and also that the shape of the waveform, remains constant over time and space [8]. Perfectly coherent beams generate the best possible fringe contrast, meaning that fringe maxima and minima are determined exclusively by the amplitude

of the interfered beams, as was assumed in Eq. (1.3). We continue to assume perfect coherence throughout this chapter and the chapters that follow. Coherence is a very rich topic and is essential to the overall understanding of wave and quantum optics [9, 10]. However, outside of what we have already mentioned here, it is beyond the scope of our work here.

Fringe patterns can also be thought of as 2D projections of what is called the wavefront. The wavefront is defined as a surface of constant phase in a coherent electric field. When coherent light is reflected or scattered off of a surface, the shape of its wavefront is deformed by that interaction. This deformation encodes information about the shape of the surface into the field's wavefront. The altered wavefront then propagates away from the surface. The exact mechanics of this reflection or scattering from the surface depend on the shape of the incident wavefront, the surface's material properties and 3D structure, as well as the frequency of illumination [6]. We assume some of the light leaving the surface will propagate through some medium and be captured by an optical system's pupil and then be measured by the system's photodetector, otherwise we would have nothing to study.

Photodetectors are used to measure the energy of a field distribution and thus cannot directly measure phase, due to the elimination of the phase term by the modulus-squared operation shown in Eq. (1.2). In fact, directly measuring the intensity of a coherent field which was incident on a surface in the far field only supplies information about the autocorrelation of the fields

in the plane of that surface. However, if one manages to recover the phase and amplitude of a coherent scalar field in one transverse spatial plane, e.g. the pupil plane, then one has access to all information present in the field. If, additionally, the index of refraction of the propagation medium can be assumed constant, then the distribution of the field in any transverse plane can be recovered [7]. In many applications, including coherent imaging, this is a primary reason for the use of interferometry. It allows for the straightforward calculation of phase and amplitude of measured coherent fields by interfering them with another field. When two fields are interfered and their intensity is measured, their phase and amplitude information both survive within the modulation term of Eq. (1.3).

Interferometric techniques fall broadly into two categories, homodyne and heterodyne. The former measures a field which has been interfered with one or more copies of itself, the copies having been in some way altered, e.g. a temporal delay or spatial modulation. The latter interferes an unknown field with one or more different known fields. Both types of techniques can be used to determine the phase and amplitude information of a field, though heterodyne detection is typically a more straightforward approach when imaging real-world objects which are not generally smooth.

1.2 Turbulence Mitigation in Range-compressed Digital Holography

In the real world, random temperature differentials in the atmosphere result in a generally non-uniform index of refraction in the region between a target and an optical system's pupil. These random temperature differentials are known as volumetric turbulence [9, 11, 12]. Numerically propagating electric fields through turbulence, under the assumption that the index of refraction is uniform, distributes energy incorrectly in the output plane. Accordingly, fields measured in the pupil of long-range imaging systems often give rise to images with degraded quality because no correction has compensated for the effects of the volumetric turbulence. These effects are often so severe that these atmospheric perturbations are the limiting factor of these imaging systems [11, 12]. In this dissertation much of our work explores techniques used to compensate for these effects in range-compressed holography systems used at long standoff ranges (≥ 1 km).

Traditional holography dates back to Dennis Gabor, the Hungarian physicist who pioneered X-ray holography as a byproduct of work in electron microscopy. After the invention of the laser, Yuri Denisyuk in the Soviet Union and Emmett Leith and Juris Upatnieks at the University of Michigan pioneered early hologram recordings in the optical regime. The invention in the mid 1970's of the charge-coupled device (CCD) image sensor along with its later commercialization gave rise to the creation of a new sub-field

of holography, known as digital holography (DH) [13, 14]. The optical frequencies used in DH offer greater potential resolution than other actively illuminated remote imaging systems such as synthetic aperture radar (SAR), which uses much lower frequencies in the radio bands. DH is a heterodyne technique where a beam with known phase and amplitude (known as the reference beam or local oscillator) is interfered with a field whose phase and amplitude are unknown [14]. The method for extracting the phase and amplitude of the hologram varies by holography technique. The type of holography used in the following chapters extracts the desired information in a re-imaged pupil plane. Extracted field data can then be numerically propagated back to the plane of the target in order to form an image of the target. As an additional advantage, the scaling of the desired field by the amplitude of the reference beam is often used to boost the SNR of a weak signal with respect to the read noise [15]. This boost is known as the heterodyne gain.

If a number of digital holograms are recorded at multiple, evenly spaced frequencies of illumination and the same relative orientation to the target, we can treat the collection of 2D hologram arrays as a three-dimensional array with axes x, y, and ν, respectively. This array can be mathematically transformed to generate a 3D intensity image. This is accomplished by conducting a Fourier transform with respect to the frequency axis of the array. The Fourier conjugate of the frequency in the array is proportional to time delay of light propagating to the target surface and then returning to the pupil [16, 17]. Since the time delay is also proportional to relative displacement

along the optical axis, we can discern depth information about the target surface. We call this Fourier transformation of the array with respect to frequency axis range-compression. Chapter 2 details necessary background information on the topics of field propagation, turbulence simulation, and range-compressed holography.

Unfortunately, 3D intensity images are difficult to display. However, 3D intensity arrays can readily be converted into a 2D range maps where the two axes correspond to the transverse dimensions in the 3D array and the color-coding of the pixels corresponds to the relative depth of the highest intensity pixel along the z-axis [18]. Below, Fig. 1.2 shows two sample range images. Fig. 1.2 (a) was generated from range-compressed hologram fields which were propagated through a medium with constant index of refraction, while Fig. 1.2 (b) was generated from fields which were propagated through simulated turbulence after scattering from the target. We see that there is a demonstrable loss in image quality in Fig. 1.2 (b) as compared to Fig. 1.2 (a). Mitigating these deleterious effects is the goal of the work in Chapters 3 and 4.

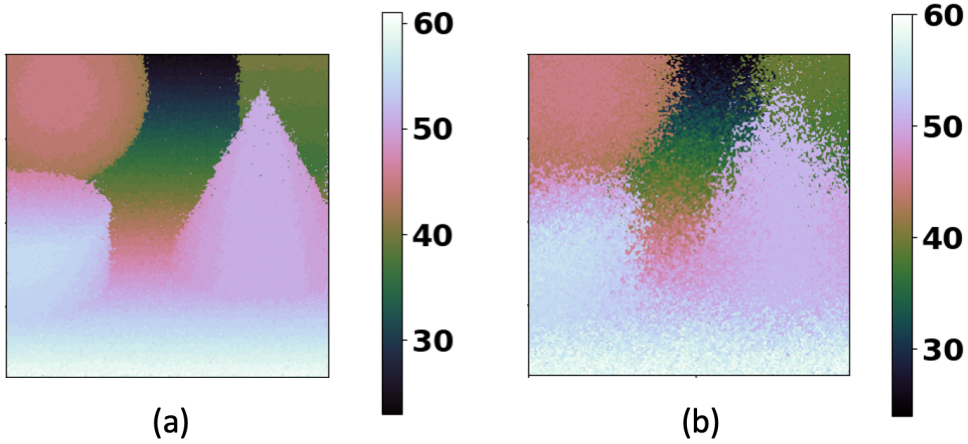

Figure 1.2: (a) Range map generated from RCH fields propagated through uniform index of refraction. (b) Range map generated from RCH fields propagated through non-uniform index of refraction resulting from simulated volume turbulence.

1.3 Phase Retrieval in Generalized Optical Interferometry

The generalized optical interferometer (GOI) builds on more fundamental types of interferometers (we chose to use the Mach-Zehnder [19]), by adding one or more generalized phase operators (GPOs) to the instrument [20]. Each GPO can just be thought of as a cylindrical lens whose radius of curvature we can change between measurements. In the lab, this is typically accomplished using spatial light modulators [21] which work by encoding phase on either a transmissive element or on a reflective surface. Additionally, the spatially resolving photodetector of the Mach-Zehnder is replaced with a detector that

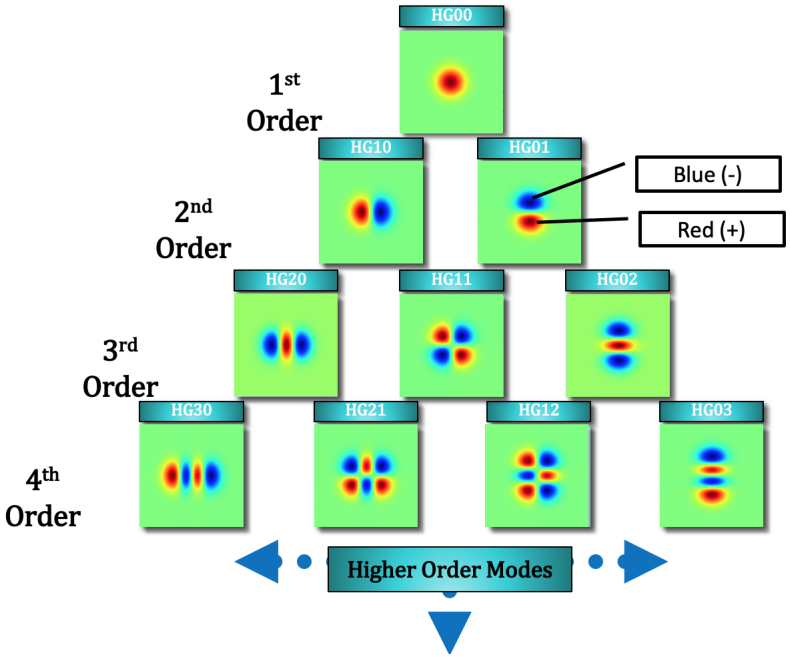

Figure 1.3: 1st through 4th order HG mode field distributions.

only measures the total energy incident on its photosensitive region, also known as a "bucket" detector. By collecting a number of interference measurements, each measurement having a unique combination of GPO radii of curvature cylindrical lenses, the interferometer is able to determine how much of the input field's energy content came from each member of a certain transverse basis set modes. In our work, we concern ourselves with the Hermite Gaussian (HG) modes [22]. In Fig. 1.3 a few sample HG mode intensities are shown. This ability to extract spatially varying field amplitude coefficients associated with each mode without any spatial measurement is quite novel and has potential application in imaging and laser communications. Generally, interferometers are useful because they allow for the measurement of

both phase and amplitude. Unfortunately, using a non-spatial detector in the GOI makes it impossible to extract the phase information using traditional interferometric techniques. In this dissertation, we examine the use of a technique known as phase retrieval to reliably estimate this phase. This topic is the focus of Chapter 5.

Phase retrieval is a prototypical example of what is called an inverse problem. It can be described as the process of using observations to determine the causal factors that produced said observations. There are many paths light could have traveled in order to make a particular energy distribution on a photodetector. Phase retrieval seeks to use known problem constraints about an optical system and signal to build a physically valid mathematical model of the energy distribution on the detector as a function of phase. Phase parameters are then iteratively estimated by minimizing the difference between photodetector measurements and the output of the physical model. During each iteration, the optimization algorithm determines new values of phase which result in a smaller net difference between the output of the physical model and the measured intensity. This process is repeated until no further improvement can be made to the estimated phase parameters with the available data. This technique has found use in many different applications including including optical metrology [23] and referenceless coherent imaging [24]. In Chapter 5, We explored two different algorithms to retrieve phase in a GOI, one which uses array detector data to retrieve the phase and one which uses bucket detector data.

1.4 Chapter Outline

Chapter 2 discusses the background information needed to understand and conduct numerical field propagations, simulate atmospheric turbulence, and create range compressed holograms and range images.

In Chapter 3, we discuss our range-compressed holography turbulence mitigation simulations. We begin by describing our physical model and algorithm in detail. We then go on to discuss our detailed Monte Carlo study of turbulence mitigation efforts in range-compressed hologram fields which were aberrated by multiple planes of simulated turbulence. We show that we did indeed achieve the desired outcome with these simulations, significant mitigation of simulated multiplane turbulence in RCHs. As expected, the turbulence mitigation greatly improved the output range images. We conclude the chapter by discussing how the ability to mitigate turbulence and improve range images relates to the relative strength of turbulence and shot noise.

Chapter 4 contains our RCH laboratory work. We begin by discussing the construction of our laboratory apparatus. This includes descriptions of the hardware, targets and turbulence screens. We then discuss the collection of our range compressed holography data sets then show our results where we reconstructed two RCH images with varying degrees of turbulence induced aberration, showing increased quality over the aberrated images. Finally, we

conclude our turbulence mitigation work by discussing significant improvements made to our laboratory data sets and the corresponding improvement of their output range images for significant phase screen turbulence profiles.

Chapter 5 covers our efforts to retrieve phase in the generalized optical interferometry systems. We first describe in detail the basic amplitude recovery function of a GOI. Next we cover the mathematical and physical relationships which allow the instrument to determine their constituent HG mode amplitudes without spatial measurement. This is followed by the discussion of our two novel phase retrieval algorithms. The description of each algorithm is followed by discussion of the results from each algorithm's simulations. We show that with both spatial detector data and single-pixel detector data, retrieval of modal phase information is possible.

Finally, Chapter 6 summarizes our results from all Chapters and discusses future work.

Bibliography

[1] T. Young, "II. The Bakerian lecture. On the theory of light and colours," Philos. T. R. Soc. A. pp. 12–48 (1802).

[2] A. A. Michelson and E. W. Morley, "On the relative motion of the earth and of the luminiferous ether," Sidereal Messenger, vol. 6, pp. 306-310 **6**, 306–310 (1887).

[3] A. A. Michelson and F. G. Pease, "Measurement of the diameter of alpha-orionis by the interferometer," P. Natl. Acad. Sci. USA **7**, 143 (1921).

[4] LIGO Scientific Collaboration and Virgo Collaboration, "Observation of gravitational waves from a binary black hole merger," Phys. Rev. Lett. **116**, 061102 (2016).

[5] K. Creath and J. C. Wyant, "Absolute measurement of surface roughness," Appl. Opt. **29**, 3823–3827 (1990).

[6] C. A. Balanis, *Advanced engineering electromagnetics* (John Wiley & Sons, 1999).

[7] J. W. Goodman, *Introduction to Fourier optics, 4th Ed.* (W. H. Freeman, 2017).

[8] E. Hecht, *Optics*, vol. 997 (1998).

[9] J. W. Goodman, *Statistical optics* (John Wiley & Sons, 2015).

[10] L. Mandel and E. Wolf, *Optical coherence and quantum optics* (University Science Books, 1986).

[11] M. C. Roggemann and B. M. Welsh, *Imaging through turbulence* (CRC press, 2018).

[12] R. L. Philips and L. C. Andrews, *Laser beam propagation through random media* (SPIE, 1998).

[13] J. W. Goodman and R. W. Lawrence, "Digital image formation from electronically detected holograms," Appl. Phys. Lett. **11**, 77–79 (1967).

[14] U. Schnars, C. Falldorf, J. Watson, and W. Jüptner, *Digital holography and wavefront sensing.* (Springer, 2016).

[15] S. F. Jacobs, "Optical heterodyne (coherent) detection," Am. J. Phys. **56**, 235–245 (1988).

[16] J. C. Marron and K. S. Schroeder, "Three-dimensional lensless imaging using laser frequency diversity," Appl. Opt. **31**, 255–262 (1992).

[17] J. C. Marron and K. S. Schroeder, "Holographic laser radar," Opt. Lett. **18**, 385–387 (1993).

[18] J. W. Stafford, B. D. Duncan, and D. J. Rabb, "Holographic aperture ladar with range compression," J. Opt. Soc. Am. A **34**, A1–A9 (2017).

[19] B. E. A. Saleh and M. C. Teich, *Fundamentals of photonics* (John Wiley & Sons, 2019).

[20] A. F. Abouraddy, T. M. Yarnall, and B. E. A. Saleh, "Generalized optical interferometry for modal analysis in arbitrary degrees of freedom," Opt. Lett. **37**, 2889–2891 (2012).

[21] T. Malhotra, W. E. Farriss, J. Hassett, A. F. Abouraddy, J. R. Fienup, and A. N. Vamivakas, "Interferometric spatial mode analyzer with a bucket detector," Opt. Express **26**, 8719–8728 (2018).

[22] A. E. Siegman, *Lasers* (University Science Books, 1986).

[23] G. R. Brady and J. R. Fienup, "Measurement range of phase retrieval in optical surface and wavefront metrology," Appl. Opt. **48**, 442–449 (2009).

[24] J. R. Fienup, "Direct-detection synthetic-aperture coherent imaging by phase retrieval," Opt. Eng. **56**, 113111 (2017).

Chapter 2

Scalar Optical Propagation, Turbulence in Imaging, and Range-compressed Holography

2.1 Numerical Propagation of Scalar Electric Fields

2.1.1 Coherent Wave Propagation

The ability to simulate the propagation of scalar electric fields between arbitrary transverse planes along the optical axis is essential to the turbulence mitigation work in this dissertation. To that end we discuss the pertinent propagation theory and generate an overall strategy for propagation, which contributes to algorithm design in the chapters that follow. Throughout this work, heavy use will be made of a number of mathematical expressions and

Parameters/Operator	Equivalence/Eq. No.	Description
c	-	Speed of light
$\vec{r}_{\perp \mathrm{i}}$	$x_{\mathrm{i}}\hat{x} + y_{\mathrm{i}}\hat{y}$	Transverse position vector in the i-th plane
$\Delta\vec{r}_{\perp \mathrm{ij}}$	$\vec{r}_{\perp \mathrm{j}} - \vec{r}_{\perp \mathrm{i}}$	Difference vector of transverse position vectors in i-th and j-th transverse planes, respectively
z_{ij}	$z_{\mathrm{j}} - z_{\mathrm{i}}$	Distance between planes i and j along the optical axis
$\vec{\alpha}^{\,2}$	$\vec{\alpha} \cdot \vec{\alpha}$	Dot product of $\vec{\alpha}$ with itself
$\mathscr{F}_{\vec{\alpha}\to\vec{\beta}}\{f(\vec{\alpha})\} = \tilde{f}(\beta)$	$\int_{-\infty}^{\infty} \exp\left(-i2\pi\vec{\alpha}\cdot\vec{\beta}\right) f(\vec{\alpha}) d\vec{\alpha}$	Fourier transform
$\mathscr{F}^{-1}_{\vec{\beta}\to\vec{\alpha}}\{f(\vec{\beta})\} = f(\vec{\alpha})$	$\int_{-\infty}^{\infty} \exp\left(i2\pi\vec{\alpha}\cdot\vec{\beta}\right) \tilde{f}(\vec{\beta}) d\vec{\beta}$	Inverse Fourier transform
$f(\vec{\alpha}) * g(\vec{\alpha})$	$\int_{-\infty}^{\infty} f(\vec{\beta}) \cdot g(\vec{\alpha} - \vec{\beta}) d\vec{\beta}$	Convolution
$f(\vec{\alpha}) \star g(\vec{\alpha})$	$\int_{-\infty}^{\infty} f(\vec{\beta}) \cdot g^*(\vec{\beta} - \vec{\alpha}) d\vec{\beta}$	Cross-Correlation
$f(\vec{\alpha}) \star f(\vec{\alpha})$	$\int_{-\infty}^{\infty} f(\vec{\beta}) \cdot f^*(\vec{\beta} - \vec{\alpha}) d\vec{\beta}$	Autocorrelation
$\mathscr{P}_{\mathrm{i}\to\mathrm{j}}\{f_{\mathrm{i}}(\vec{r}_{\perp \mathrm{i}}, z_{\mathrm{i}}; \nu)\}$	$f_{\mathrm{j}}(\vec{r}_{\perp \mathrm{j}}, z_{\mathrm{j}}; \nu)$; Eq. (2.3)	Fresnel propagation from transverse plane i → j
$\mathscr{P}^{-1}_{\mathrm{j}\to\mathrm{i}}\{f_{\mathrm{j}}(\vec{r}_{\perp \mathrm{j}}, z_{\mathrm{j}}; \nu)\}$	$f_{\mathrm{i}}(\vec{r}_{\perp \mathrm{i}}, z_{\mathrm{i}}; \nu)$; Eq. (2.3)	Inverse Fresnel propagation from transverse plane j → i
D_{i}	-	Transverse extent of propagation trapezoid in i-th transverse plane
ν	$\frac{c}{\lambda}$	Frequency of illumination; equal to speed of light divided by the wavelength of illumination
K_{ij}	$\frac{\nu D_{\mathrm{i}} D_{\mathrm{j}}}{c z_{\mathrm{ij}}}$	Fresnel region 1D space-bandwidth product or the number of speckles across one transverse output space dimension
m_{ij}	$\frac{\delta_{\mathrm{j}}}{\delta_{\mathrm{i}}} = \frac{D_{\mathrm{j}}/N_{\mathrm{j}}}{D_{\mathrm{i}}/N_{\mathrm{i}}}$	Linear magnification; scaling of output sampling with respect to input sampling
N_{i}	-	Number of samples in the i-*th* transverse plane
N_p	-	Number of samples across one fully zero-padded transverse dimension of an array
Q_{i}	$\frac{N_p}{N_{\mathrm{i}}}$	Ratio of zero-padded array length to unpadded array length in i-*th* transverse plane
$\Delta\nu$	-	Illumination frequency sampling interval
B_ν	-	Illumination frequency bandwidth

Table 2.1: Table of commonly used parameters and operations

variables. Table 2.1 defines commonly used expressions and variables. Provided that we can discount the effects of polarization, classical light can be expressed as scalar fields. Accordingly, waves which are scalar in nature must then satisfy the scalar wave equation

$$\nabla^2 U(P,t) = \frac{n^2}{c^2}\frac{\partial^2 U(P,t)}{\partial t^2}, \tag{2.1}$$

which is derived from Maxwell's equations [1–5] where P is an arbitrary point in 3D-space, t is the time, n is the index of refraction of the medium, and c is the speed of light. Furthermore, if the light can be assumed to be monochromatic, Eq. (2.1) can be replaced with the Helmholtz Equation [3]

$$\left[\nabla^2 + \left(\frac{2\pi n\nu}{c}\right)^2\right] U(P,t) = 0, \tag{2.2}$$

where ν is the frequency of the monochromatic illumination. Huygens-Fresnel propagation shows that under these conditions, the expression for a scalar monochromatic electric field in any transverse plane along the optical axis can be reliably calculated, provided one knows the electric field in another transverse plane. In the non-paraxial regime, the full Rayleigh-Sommerfeld (RS) diffraction formula can be used. However, these full propagation integrals are often very computationally demanding. In the paraxial regime, significant simplifications can be made to the RS diffraction formula. These

simplifications yield the Fresnel approximation [1, 3–5]

$$U_{\mathrm{j}}(\vec{r}_{\perp\mathrm{j}}, z_{\mathrm{j}}; \nu) = \exp\left(\frac{i2\pi\nu z_{\mathrm{ij}}}{c}\right) \int_{-\infty}^{\infty} U_{\mathrm{i}}(\vec{r}_{\perp\mathrm{i}}, z_{\mathrm{i}}; \nu) \exp\left(\frac{i\pi\nu}{c z_{\mathrm{ij}}} \Delta r_{\perp\mathrm{ij}}^2\right) d\vec{r}_{\perp\mathrm{i}}, \quad (2.3)$$

to within a constant multiple. We use the index i to represent coordinates in the input plane and the index j to represent coordinates in the output plane.

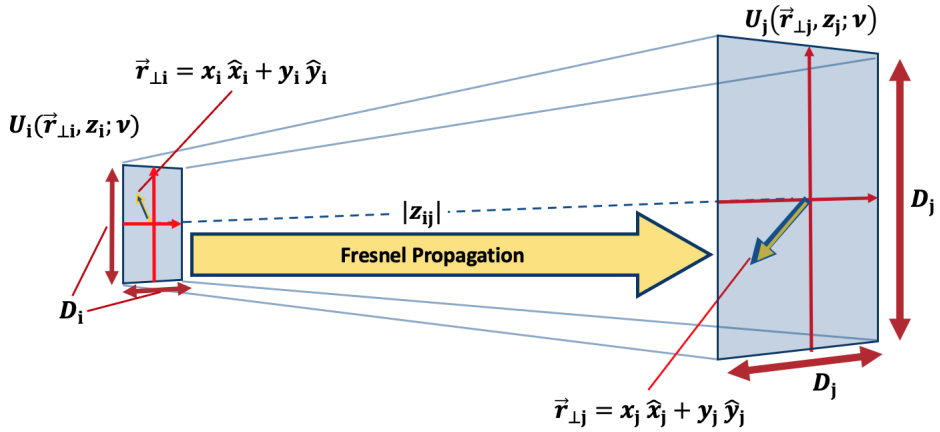

Figure 2.1: Plane-to-plane propagation space, coordinates, physical parameters

2.1.2 Scalable Angular Spectrum Propagation, K-formalism, and Sampling Constraints

The Fresnel approximation can be manipulated into an equivalent form which allows for the arbitrary input and output scaling (transverse linear magnifi-

cation) as [1, Eq. (6.67)]

$$\begin{aligned} U_{\mathrm{j}}(\vec{r}_{\perp\mathrm{j}}, z_{\mathrm{j}}; \nu) =& \frac{1}{m_{\mathrm{ij}}} \exp\left(\frac{i2\pi\nu z_{\mathrm{ij}}}{c}\right) \exp\left[\frac{i\pi\nu}{c z_{\mathrm{ij}}}\left(\frac{m_{\mathrm{ij}}-1}{m_{\mathrm{ij}}}\right)\vec{r}_{\perp\mathrm{j}}^{\,2}\right] \\ & \times \int_{-\infty}^{\infty} \exp\left(\frac{-i\pi c z_{\mathrm{ij}}}{\nu m_{\mathrm{ij}}}\vec{f}_{\perp\mathrm{i}}^{\,2}\right) \exp\left[i2\pi\left(\frac{\vec{f}_{\perp\mathrm{i}}}{m_{\mathrm{ij}}}\cdot\vec{r}_{\perp\mathrm{j}}\right)\right] \\ & \times \int_{-\infty}^{\infty} \exp\left[\frac{i\pi\nu}{c z_{\mathrm{ij}}}\left(1-m_{\mathrm{ij}}\right)\vec{r}_{\perp\mathrm{i}}^{\,2}\right] \exp\left[-i2\pi\left(\vec{f}_{\perp\mathrm{i}}\cdot\vec{r}_{\perp\mathrm{i}}\right)\right] \\ & \times U_{\mathrm{i}}(\vec{r}_{\perp\mathrm{i}}, z_{\mathrm{i}}; \nu) d\vec{r}_{\perp\mathrm{i}} d\vec{f}_{\perp\mathrm{i}}. \end{aligned} \tag{2.4}$$

Note that we have added the leading piston phase term, omitted by [1]. A full derivation of this propagator starting from the Fresnel transform Eq. (2.3) can be found in [1, Eqs.(6.31)-(6.67)] . This propagator is known as the scalable angular spectrum (SAS). This is identical to Sziklas and Siegman's propagation [6], which reduced the computational burden of their problem by using different sample spacing in the input and output plane, instead of using a traditional angular spectrum where sample spacing must be the same in the input and output planes. It is of note to mention that there is also a non-paraxial angular spectrum propagator. Work has been done to adapt the non-paraxial angular spectrum for use in cases with transverse magnification [7]. Though only marginally more computationally complex than the SAS, non-paraxial sampling criteria are less intuitive and concise. Moreover, there is no appreciable loss of accuracy using paraxial propagators for the long-range propagations which we conduct here. We have thus chosen to use the Fresnel-based paraxial SAS propagator in this work.

When conducting numerical wave propagation, it is essential that ample consideration is given to the discretization process. Schmidt [1] derives sampling constraints which are needed to support the numerical computation of propagated fields using the SAS propagator [1]. Sampling and zero-padding constraints are necessary in order to ensure physically realistic, unaliased wave propagation calculations. The constraints presented in [1, 8] are correct for most situations and served as a foundation for our propagation work. Unfortunately, the parameterization of sampling constraints in [1] is labyrinthine. Here, we ease this burden and make the constraints easier to understand and implement through an alternative parameterization based on unitless parameters. We call this reparameterization "K-formalism."

Our K-formalism uses the unitless quantities K_{ij}, the Fresnel region 1D space-bandwidth product (SBWP) defined within the trapezoidal region between the i and the j planes and m_{ij}, the ratio of the physical output sample spacing to the input sample spacing, i.e., transverse linear magnification. Note that this expression is not a valid SBWP for short range propagations. However, all propagations in this work are such that the K parameter, as defined in Table 2.1, is the SBWP.

When computing numerical Fresnel propagations, K_{ij} and m_{ij} are the only independent parameters needed to propagate fields in the i-th plane to the j-th plane, to within a piston phase term. Assuming square arrays, the only dependent parameters needed for computation are N_{i} and N_{j}, the

sampled pixel extent of the input and output arrays, respectively, and N_p, the total pixel extent of both the input and output arrays after the addition of zero padding. And save for that piston phase term, propagations which have the same space-bandwidth product, magnification, and number of sampled input and output array pixels are mathematically identical, regardless of differences in actual physical parameters. In order to convert Eq. (2.4) to K-formalism, we must convert each of the phase terms. For the innermost quadratic phase term, we substitute sampled coordinates $\delta_{\mathrm{i}}^2(v^2+w^2)$ for $\vec{r}_{\perp\mathrm{i}}^{\,2}$ and group terms to put in terms of SBWP and magnification

$$\begin{aligned}\exp\left[\frac{i\pi\nu}{cz_{\mathrm{ij}}}(1-m_{\mathrm{ij}})\delta_i^2(v^2+w^2)\right] &\rightarrow \exp\left[\frac{i\pi\nu}{cz_{\mathrm{ij}}}(1-m_{\mathrm{ij}})\frac{N_{\mathrm{i}}N_{\mathrm{j}}\delta_{\mathrm{j}}}{N_{\mathrm{i}}N_{\mathrm{j}}\delta_{\mathrm{j}}}\delta_i^2(v^2+w^2)\right] \\ &= \exp\left[i\pi(1-m_{\mathrm{ij}})\frac{\nu D_{\mathrm{i}}D_{\mathrm{j}}}{cz_{\mathrm{ij}}}\frac{\delta_{\mathrm{i}}}{\delta_{\mathrm{j}}}\frac{1}{N_{\mathrm{i}}N_{\mathrm{j}}}(v^2+w^2)\right] \\ &= \exp\left[\frac{i\pi(1-m_{\mathrm{ij}})K_{\mathrm{ij}}}{m_{\mathrm{ij}}N_{\mathrm{i}}N_{\mathrm{j}}}(v^2+w^2)\right],\end{aligned} \tag{2.5}$$

where $\delta_{\mathrm{i}}, \delta_{\mathrm{j}}$ are the input and output plane sample spacing, respectively, v, w are the input plane pixel indices, and $N_{\mathrm{i}}, N_{\mathrm{j}}$ are the number of pixels in the input and output domain, respectively. Similarly, the inverse Fourier transform kernel in Eq. (2.4) is discretized by substituting $\delta_f(p\hat{x}+q\hat{y})$ for

$\vec{f}_{\perp \mathrm{i}}$ and $\delta_{\mathrm{j}}(t\hat{x} + u\hat{y})$ for $\vec{r}_{\perp \mathrm{j}}$

$$\begin{aligned} \exp\left[i2\pi \frac{\vec{f}_{\perp \mathrm{i}}}{m_{\mathrm{ij}}} \cdot \vec{r}_{\perp \mathrm{j}}\right] &\rightarrow \exp\left[\frac{i2\pi}{m_{\mathrm{ij}}} \delta_f \delta_{\mathrm{j}} (tp + uq)\right] \\ &= \exp\left[\frac{i2\pi}{m_{\mathrm{ij}}} \frac{\delta_{\mathrm{j}}}{N_p \delta_{\mathrm{i}}} (tp + uq)\right] \\ &= \exp\left[i \frac{2\pi}{N_p} (tp + uq)\right], \end{aligned} \tag{2.6}$$

where $\delta_f = 1/(\delta_{\mathrm{i}} N_p)$ is the sampling in the frequency domain, p, q are the pixel indices in the frequency domain, t, u are the output plane pixel indices, and N_p is the fully padded array extent. Other phase terms are converted in a similar fashion using substitution and grouping of terms to achieve K-formalism. In addition to the conversion of phase terms, the integral operators are replaced with summation over discretized coordinates.

The SAS with leading is expressed in K-formalism as

$$\begin{aligned} U_{\mathrm{j}}(t,u) =& \frac{1}{m_{\mathrm{ij}}} \exp\left(\frac{i2\pi\nu z_{\mathrm{ij}}}{c}\right) \exp\left[i \frac{\pi (m_{\mathrm{ij}} - 1) K_{\mathrm{ij}}}{N_{\mathrm{i}} N_{\mathrm{j}}} (t^2 + u^2)\right] \\ &\times \sum_{p,q=1}^{N_p} \exp\left[-i \frac{\pi N_{\mathrm{i}} N_{\mathrm{j}}}{K_{\mathrm{ij}} N_p^2} (p^2 + q^2)\right] \exp\left[i \frac{2\pi}{N_p} (tp + uq)\right] \\ &\times \sum_{v,w=1}^{N_p} \exp\left[-i \frac{2\pi}{N_p} (vp + wq)\right] \\ &\times \exp\left[\frac{i\pi (1 - m_{\mathrm{ij}}) K_{\mathrm{ij}}}{m_{\mathrm{ij}} N_{\mathrm{i}} N_{\mathrm{j}}} (v^2 + w^2)\right] U_{\mathrm{i}}(v,w). \end{aligned} \tag{2.7}$$

As illustrated in Fig. 2.1, if the input region, e.g., the pupil, of our propagation is a $D_{\mathrm{i}} \times D_{\mathrm{i}}$ area region propagated to a square $D_{\mathrm{j}} \times D_{\mathrm{j}}$ output field

area, the propagation medium is bounded by the faces of a truncated square pyramid whose base and top are defined by these two areas $D_\mathrm{i} \times D_\mathrm{i}$ and $D_\mathrm{j} \times D_\mathrm{j}$, respectively. Thus, a 2D cut-through where either x_i and $x_\mathrm{j} = 0$ or y_i and $y_\mathrm{j} = 0$, has a trapezoidal shape. Note that $D_\mathrm{i} = N_\mathrm{i}\delta_\mathrm{i}$, where δ_i is the sample spacing in the i-th plane.

2.1.3 Propagation Constraints

We begin discussion of propagation constraints by defining quantities which are relevant to determining sampling requirements. These quantities are the extent of the input plane in length units, D_i, the extent of the output plane in length units, D_j, the difference between the input and output z-coordinates, z_ij, the frequency of illumination in inverse time units, ν, and the desired sample spacing of the input and output planes in length units, δ_i and δ_j, respectively. From these quantities, we can calculate the SBWP and the magnification associated with our propagation. We must determine the sampling constraints on N_i and N_j which ensure that the resulting output signal is unaliased and has the required fidelity. Note that there is often little flexibility in N_i and N_j due to hardware or system constraints. In these cases, other parameters may need to flex, such as the physical extent of the pupil or target plane field-of-view (FOV). In addition to the input and output sampling parameters, we must also ensure that each array has the necessary zero-padding to avoid aliasing.

As stated earlier, we chose to identify the constraints on N_i and N_j first, followed by the array padding constraints which yield the required overall array size, N_p. This is not a mandatory order of operations. For example, one could choose to start with a known total array size and from there determine the constraints on N_i, N_j, and then go on to determine the best problem geometry to support the array sizes and sample spacings. The choice of how to go about determining propagation sampling is ultimately dependent upon the desired end-state and user preference.

We reparameterized each of the constraints in [1, 8] into unitless K-formalism. The first sampling constraint, based on problem geometry, is [1, Eq. (7.14)]

$$|K_{\mathrm{ij}}| \leq \frac{N_\mathrm{i} N_\mathrm{j}}{N_\mathrm{i} + N_\mathrm{j}}, \tag{2.8}$$

which describes the SBWP which can be supported give an array with N_i input pixels and N_j output pixels. Note that this sampling constraint is specific to intensities, not fields, and can be considered fairly conservative under certain circumstances, such as very short distance propagations [8]. It is, however, a sufficient condition for our work. The next sampling constraint begins with [1, Eq. (7.53)]. We discard the $\Delta z/R$ term in this equation which is designed to compensate for additional quadratic phase from a converging or diverging wavefront. We also substitute c/ν instead of λ. So [1, Eq. (7.53)] becomes

$$\delta_\mathrm{i} - \frac{c z_{\mathrm{ij}}}{\nu D_\mathrm{i}} \leq \delta_\mathrm{j} \leq \delta_\mathrm{i} + \frac{c z_{\mathrm{ij}}}{\nu D_\mathrm{i}}. \tag{2.9}$$

We subtract δ_{i} from both sides and change to a single inequality with absolute values

$$|\delta_{\mathrm{j}} - \delta_{\mathrm{i}}| \leq \left| \frac{c z_{\mathrm{ij}}}{\nu D_{\mathrm{i}}} \right|. \tag{2.10}$$

Dividing both sides by δ_{j} yields

$$\begin{aligned} \left| 1 - \frac{1}{m_{\mathrm{ij}}} \right| &\leq \left| \frac{c z_{\mathrm{ij}}}{\nu \delta_{\mathrm{j}} D_{\mathrm{i}}} \right| \\ &\leq \left| \frac{N_{\mathrm{j}}}{K_{\mathrm{ij}}} \right|. \end{aligned} \tag{2.11}$$

We then rearrange terms to find the constraint on N_{j} as a function of magnification and SBWP

$$N_{\mathrm{j}} \geq |K_{\mathrm{ij}}| \left| \frac{1 - m_{\mathrm{ij}}}{m_{\mathrm{ij}}} \right| (1 + \gamma), \tag{2.12}$$

where the added γ parameter accounts for the presence of additional quadratic phase in the beam or in turbulence screens (which we discuss shortly). This constraint ensures the innermost quadratic phase of the SAS does not alias. The constraint is derived by limiting the absolute value of pixel-to-pixel phase jumps in the quadratic phase term to less than π radians, or equivalently, that the local spatial frequency at the edges of our quadratic phase satisfies the Nyquist criterion. The constraint gives the SBWP that can be supported for a given magnification and unpadded output array extent, without aliasing the input quadratic phase. The parameter γ is an empirically determined nonnegative number which should be used to account for any additional

quadratic phase present in the input (or output) plane of the propagation, such as quadratic phase content in atmospheric turbulence in the input plane (which we will discuss in an upcoming section).

Schmidt [1] does not explicitly derive a constraint for the external quadratic phase. However, it is very similar to Eq. (2.12) with the i, j indices reversed

$$\begin{aligned} N_{\mathrm{i}} &\geq |K_{\mathrm{ij}}|\,|m_{\mathrm{ij}} - 1|\,(1+\gamma) \\ &\geq |K_{\mathrm{ji}}|\left|\frac{1-m_{\mathrm{ji}}}{m_{\mathrm{ji}}}\right|(1+\gamma)\,. \end{aligned} \tag{2.13}$$

This describes the SBWP that the propagation can support, without aliasing the outermost quadratic phase of Eq. (2.7), given a certain pixel extent in the unpadded input array. The second line in Eq. (2.13) shows that our quadratic phase sampling constraints are consistent. That is to say, if we were propagating from plane j to plane i, instead of from plane i to plane j, the quadratic phase constraint would be identical to Eq. (2.12) with swapped indices. Again the chosen parameter γ is the same as in Eq. (2.12). When propagating fields scattered from optically rough objects, setting $\gamma = 0$ is generally appropriate provided there is negligible atmospheric turbulence. This is because the field phases have been randomized by interaction with the object's rough surface and are very unlikely to have appreciable quadratic phase content after this interaction.

The first constraint on the total array size [1, Eq. (7.20)]

$$N_p \geq \frac{N_\mathrm{i} + N_\mathrm{j}}{2} + \frac{N_\mathrm{i} N_\mathrm{j}}{2|K_\mathrm{ij}|}, \tag{2.14}$$

ensures there is ample array padding to support the highest spatial frequencies present in the output plane. It does however permit wrap-around energy to encroach into the zero padding region of the array. This is because energy that wraps around and encroaches into the region of zero padding array will not adversely affect the area of the image $D_\mathrm{j} \times D_\mathrm{j}$ which one is concerned with. The second, and final, constraint on total array size originates from [1, Eq. (7.59)]. In order to realize the K-formalism format for this constraint, the right side of the inequality is multiplied by $N_\mathrm{i} N_\mathrm{j} / N_\mathrm{i} N_\mathrm{j}$. Since $D_\mathrm{i} = N_\mathrm{i} \delta_\mathrm{i}$ and $D_\mathrm{j} = N_\mathrm{j} \delta_\mathrm{j}$, K_ij can then be substituted into the denominator of the right side of [1, Eq. (7.59)] leaving only $N_\mathrm{i} N_\mathrm{j}$ in the numerator:

$$N_p \geq \frac{N_\mathrm{i} N_\mathrm{j}}{|K_\mathrm{ij}|}. \tag{2.15}$$

This describes the zero-padding necessary to avoid aliasing the transfer function of the SAS propagator.

If one is only concerned with the 2D intensity of the output of the SAS, then constraint Eq. (2.13) can be ignored. Additionally, once a field has been limited in extent in both the input and output planes, Eq. (2.14) can also be ignored. That is to say, if all energy is contained within the trapezoidal

region between the input and output system.

For our work, it was acceptable to choose an equal number of input and output pixels in our propagation arrays. Doing this further simplifies constraints. We find Eq. (2.8) becomes

$$N \geq 2|K_{\mathrm{ij}}|. \tag{2.16}$$

In the event we illuminate an optically rough target with a coherent square illumination pattern, this constraint mandates a minimum of 2 samples per speckle diameter in the pupil plane [9]. Next, Eqs. (2.12) and (2.13), respectively, become

$$N \geq (1+\gamma)|K_{\mathrm{ij}}||m_{\mathrm{ij(ji)}} - 1|. \tag{2.17}$$

The geometric zero-padding constraint, Eq. (2.14), becomes

$$N_p \geq N\left(1 + \frac{N}{2|K_{\mathrm{ij}}|}\right). \tag{2.18}$$

Plugging Eq. (2.16) into Eq. (2.18) we find

$$N_p \geq 2N, \tag{2.19}$$

which indicates an array padding ratio

$$Q \geq 2. \tag{2.20}$$

Finally, the transfer function padding constraint of Eq. (2.15) becomes

$$N_p \geq \frac{N^2}{|K_{\text{ij}}|}. \tag{2.21}$$

Dividing both sides by N and plugging Eq. (2.16) into Eq. (2.21), we find an equivalent array padding ratio to Eq. (2.20).

As an example, assume one desires to propagate a field whose sampled input and output regions have have an equal number of pixels, where we are only concerned with the output intensity, $\gamma = 1$, and $m_{\text{ij}} = 2/3$, and one is only concerned with the intensity of the output field, they would need $N = 2|K_{\text{ij}}|$ pixels in the input/output planes and $Q = 2$ zero-padding. This results in a full array with $N_p = 4|K_{\text{ij}}|$.

2.1.4 Chain Propagation

Any SAS propagation between two transverse planes can be broken up into an arbitrary number of successive SAS partial propagations, creating a chain of propagations between a pupil and target. This "chain propagation" produces output fields equivalent to those produced using a single SAS propagation between pupil and target [1]. This propagation method using partial propagation allows us to include the effects of phase screens, which we will discuss in the next subsection, that are located in planes between the target and pupil of a system.

Note that the innermost quadratic phase in Eq. (2.7) is only necessary during the first partial propagation in a chain. Similarly, the outer most quadratic phase is only necessary after the last propagation in a chain (if then). This is because each inner quadratic phase after the first in the propagation chain has a complimentary outer phase which is its complex conjugate, making the product of the two phases unity [1].

Chain propagations use slightly modified sampling constraints. Each partial propagation needs to satisfy the transfer function sampling constraint, Eq. (2.15), since this constraint prevents aliasing in each partial propagation's transfer function. However, being able to spread the transfer function requirements across multiple partial propagations can often reduce overall array size requirements. The signal sampling constraint, Eq. (2.8), need only be satisfied with respect to the overall space-bandwidth product of the target-to-pupil system, not the SBWP of each partial propagation. This is because Eq. (2.8) is a constraint resulting from the geometry and spatial frequency content of the overall optical system, which will still be valid for the chain propagation. Similarly, the geometric array-padding constraint, Eq. (2.14), and both non-transfer function quadratic phase constraints, Eqs. (2.12) and (2.13), need only be satisfied for the SBWP of the entire system because of the canceling of quadratic phases detailed in the previous paragraph.

We describe a chain propagation as a series of S partial propagations

from a transverse pupil plane with index 0 to a transverse target plane with index S, where S is an integer greater than or equal to 2. For simplicity, we chose all of our chain propagations in the following work to have a uniform number of samples in all planes, i.e., $N_s = N,\ \forall\ s \in [0, S]$. For the duration of Chapters 2, 3, and 4, we will continue to use this 0 to S indexing scheme.

In cases where one desires to have a uniform number of pixels in all planes, but does not have a number of detector pixels equal to the number of pixels one wishes across the output region, the detector array data can be upsampled (note this will not increase resolution) or zero-padded until the array has the same number of pixels as one desires across the output plane. In the former case, the upsampling may introduce undesirable artifacts. In the latter case, the SBWP must be adjusted to reflect the new input and output plane extents, and multiple SBWPs and magnifications will be needed to conduct chain propagation. There is a relation between the total SBWP (from pupil plane to target plane) and the partial propagations which can be used to conduct an equivalent propagation. If, for example, one conducts a chain propagation from a pupil plane 0 to a middle plane with index s, then a second propagation from s to a target plane S, as in Fig. 2.2, two SBWP values will be needed, K_{0s} and K_{sS}, as well as two magnification terms, m_{0s} and m_{sS}.

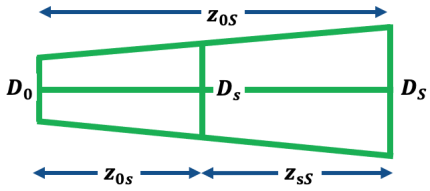

Figure 2.2: SBWP conservation geometry.

From the propagation geometry, we see that

$$\begin{aligned} D_s &= D_0 + \left(\frac{z_{0s}}{z_{0S}}\right)(D_S - D_0) \\ &= \left(\frac{z_{0S}D_0 - z_{0s}D_0 + z_{0s}D_S}{z_{0S}}\right) \\ &= \left(\frac{z_{0s}D_0 + z_{sS}D_0 - z_{0s}D_0 + z_{0s}D_S}{z_{0S}}\right) \\ &= \left(\frac{z_{sS}D_0 + z_{0s}D_S}{z_{0S}}\right). \end{aligned} \tag{2.22}$$

Now, consider the sum of the reciprocal of each partial propagation's SBWP

$$\begin{aligned} \frac{1}{K_{0s}} + \frac{1}{K_{sS}} &= \frac{cz_{0s}}{\nu D_s D_0} + \frac{cz_{sS}}{\nu D_s D_S} \\ &= \frac{c}{\nu D_s}\left(\frac{z_{0s}}{D_0} + \frac{z_{sS}}{D_S}\right) \\ &= \frac{c}{\nu}\left(\frac{z_{0S}}{z_{sS}D_0 + z_{0s}D_S}\right)\left(\frac{z_{0s}}{D_0} + \frac{z_{sS}}{D_S}\right) \\ &= \frac{c}{\nu}\left(\frac{z_{0S}}{z_{sS}D_0 + z_{0s}D_S}\right)\left(\frac{z_{sS}D_0 + z_{0s}D_S}{D_0 D_S}\right) \\ &= \frac{c}{\nu}\left(\frac{z_{0S}}{D_0 D_S}\right) \\ &= \frac{1}{K_{0S}}. \end{aligned} \tag{2.23}$$

This equation can be used as many times as necessary, splitting each of the two smaller trapezoids into additional trapezoids, allowing us to determine the SBWPs of all partial propagations. Thus, for a chain propagation consisting of S total partial propagations from input plane 0 to output plane S

$$\frac{1}{K_{0S}} = \sum_{s=1}^{S} \frac{1}{K_{s-1,s}}. \tag{2.24}$$

Magnifications are recovered in similar fashion, using Eq. (2.22), but instead substituting the physical extent of a plane, D_x, with the sample spacing of that arbitrary plane, δ_x. Recall that magnifications will be the ratio of each partial propagation's output sample spacing to that partial propagation's input sample spacing.

2.1.5 Simulating and Propagating through Atmospheric Turbulence

The freespace propagations we have defined thus far assume a uniform index of refraction equal to unity. This is a good estimate as the average index of refraction of air is very close to unity. However, pressure and temperature fluctuations in the atmosphere cause random nonuniformities in the index of refraction. Over long distances these generally small nonuniformities aggregate, causing substantial phase errors or aberrations. These pockets which occur in the propagation medium due to pressure and temperature fluctuations are called turbulent eddies. These aberrations frequently cause

degradation in images generated from these fields.

Turbulent effects are separated into two types of aberrations in the Rytov approximation of turbulence: phase aberrations and log-amplitude aberrations (scintillation) [10, 11]. We assume signal detection that are short enough that the turbulence profile has negligible transverse movement. Phase aberrations redirect photons in undesirable ways similar to surface aberrations in a system of optical elements, causing blurring and warping of target images. The second type of aberration, log-amplitude aberrations, cause changes in the illuminated target as a result of the illumination beam passing through turbulence. Additionally, log-amplitude aberrations cause intensity fluctuations in the pupil due to image beam passing through turbulence on its way to the pupil [11]. In this work, though, we considered only phase aberrations. This choice simplified our simulations, allowing us to effectively isolate and address one type of the two types of Rytov model turbulence. Phase aberrations alone are responsible for the majority of the overall image degradation caused by turbulence and, as such, correcting only for phase aberration has the potential to significantly improve image quality.

Modeling propagation through an entire volume of turbulence would be very computationally intensive and, and so would be poorly suited to real-time/near real-time imaging applications. For our purposes, volumetric phase turbulence can be well approximated by a series of two dimensional phase objects [1, 10–13]. These will be referred to as turbulence screens.

We used two parameters to discuss the strength of our system's turbulence, Fried's parameter, known as r_0, and the isoplanatic angle, known as θ_0. Fried's parameter r_0, also known the atmospheric coherence diameter [14] or "seeing cell" size, is commonly understood as the diameter of the largest aperture in a plane which would yield near-diffraction-limited results [10, 13]. An effective r_0 for the system as measured in the pupil plane can be calculated from the collection of all turbulent phase screen contributions [1, Eq. (9.72)]

$$r_0 = \left[\sum_{s=0}^{S-1} r_{0,s}^{-5/3} \left(\frac{z_{0S} - z_{0s}}{z_{0S}} \right)^{(5/3)} \right]^{(-3/5)}, \tag{2.25}$$

where $r_{0,s}$ is the r_0 of the $(s+1)$-st turbulence screen located in the plane s, which is a distance of $|z_{0s}|$ away from the pupil plane. We will refer to this aggregate quantity in the pupil as r_0 and the seeing cell diameter for a single turbulence screen s as $r_{0,s}$, in what follows. Note, this assumes a phase screen also exists in the pupil, at plane index 0. The unitless measure of turbulence strength in the pupil experienced by a particular system will be referred to as D/r_0. This quantity is calculated by dividing the diameter of the pupil, D_0, by r_0. It can be interpreted as the number of diffraction-limited apertures that fit across the diameter of the system's pupil.

Our second measurement of turbulence strength, also first described by Fried [15], is the isoplanatic angle, θ_0, is a turbulent effect that causes the OTF/PSF to lose shift invariance. Roggemann states [10, pp 171] that, "(t)he

isoplanatic angle is qualitatively the maximum angular separation between the object and beacon, such that the turbulence induced wave front deformation for the object and beacon wave fronts are still reasonably similar." Roggemann speaks about this in the context of the angle between a beacon and an astronomical object in an adaptive optics system. The analogue in our work is the maximum angular separation between two different points in the target plane that see reasonably similar turbulence. The isoplanatic angle is formally defined [10, Eq. (5.1)]

$$\theta_0 = 58.1 \times 10^{-3} \left(\frac{c}{\nu}\right)^{\frac{6}{5}} \left[\int_0^L dz C_n^2(z) z^{5/3}\right]^{-3/5}, \tag{2.26}$$

where L is the pathlength through turbulence, and $C_n^2(z)$ is the empirically determined turbulence structure constant. When using the phase screen approximation, the continuous integral can be converted into a discrete summation using [10, Eq. (3.40)]

$$\theta_0 = 58.1 \times 10^{-3} \left(\frac{c}{\nu}\right)^{6/5} \left[\sum_{s=0}^{S-1} C_{n,s}^2 z_{0s}^{5/3}\right]^{-3/5}. \tag{2.27}$$

Also when approximating turbulence as discrete phase screens, the isoplanatic angle can be described in terms of the r_0 values of all screens and their locations. To do so, we solve [10, Eq. (3.57)] for $C_{n,s}^2$ and plug it into Eq. (2.27) yielding

$$\theta_0 = 0.314 \left[\sum_{s=0}^{S-1} \left(\frac{z_{0s}}{r_{0,s}}\right)^{(5/3)}\right]^{(-3/5)}. \tag{2.28}$$

Observe from Eq. (2.28) that turbulence screens in the pupil offer no contribution to anisoplanatism. Note also, that r_0, and therefore θ_0, is defined by its relationship to C_n^2, the refractive index structure function parameter.

Images of a target with points that have angular subtense larger than θ_0 are said to suffer from anisoplanatism. We desired a measure of anisoplanatism which was relative to a given optical system, much like D/r_0, is relative to the diameter of a system's pupil. We chose to use the parameter κ as the unitless descriptor of the degree of anisoplanatism in an optical system, where

$$\kappa = \frac{\text{AFOV}}{\theta_0} = \frac{D_S}{\theta_0 |z_{0S}|}. \tag{2.29}$$

A value of $\kappa = 1$ would mean the images captured with the system would be very unlikely to experience the effects of anisoplanatism, whereas a value of, say, 50 would represent a system where images would be very likely to exhibit severe anisoplanatic effects.

In the following chapters, we used turbulence screens which were generated using the technique from Lane, Glindemann, and Dainty [16]. Though there are a number of other methods which could be used to simulate atmospheric turbulence [1, 10, 17]. The method we used is based on the Kolmogorov power spectrum which describes the statistical abundance and size of turbulent eddies as a function of spatial frequency. Turbulence screens were created by first creating arrays with twice the pixel extent of the desired turbulence screen. These oversized arrays were then populated with

complex circular Gaussian random numbers. Random number arrays were then multiplied by the square root of the discrete Kolmogorov power spectrum with the desired r_0

$$\Phi^{\frac{1}{2}}(p,q) = \sqrt{0.023}\left(\frac{2D}{r_0}\right)^{\frac{5}{6}}\left(p^2+q^2\right)^{\frac{-11}{6}}, \tag{2.30}$$

where p, q are array indices and D is the spatial extent of the desired turbulence screen in length. Array products of complex-valued random numbers and Kolmogorov spectra were then cosine transformed and cropped to the desired pixel extent, removing wrap-around continuity resulting from the cosine transform. At this stage, the spatial frequencies represented in the turbulence screens were limited to harmonics of the discretely sampled power spectra, i.e., spatial frequencies were greater than or equal to $1/(2D)$. Sub-harmonic turbulence content was added to the overall phase to ensure physically realistic turbulence simulations. Addition of these sub-harmonics is an essential part of the turbulence simulation, since much of total spectral-power of atmospheric turbulence is contained in these bands. Six sets of sub-harmonics were added to each of the turbulence screens we used in the following Chapters in accordance with best practices outlined in [16]. The final real-valued turbulence screen in the s plane was denoted $\phi_s(t,u)$.

Statistically, simulated phase screens require ≥ 3 pixels per r_0 to be adequately sampled and thereby represent their highest spatial frequency content with good fidelity [18]. All of the screens in our work contained ≥ 5

samples per r_0. This yielded a margin of safety, ensuring adequate sampling of the array products of turbulence screens with propagating fields.

To apply simulated turbulence to our fields, each real-valued Kolmogorov phase, $\phi_s(t,u)$, was then inserted as the phase of a complex exponential function

$$\Psi_s(t,u) = \exp\left[i\phi_s(t,u)\right], \tag{2.31}$$

and placed in a plane s along the path of optical propagation. The term $\Psi_s(t,u)$ is known as the turbulence screen transmittance function.

Note that the transverse extents of our turbulence screens were chosen so that they contained almost all unaberrated beam energy and had additional extent added as a function of the turbulence strength in order to compensate for energy redirected by the turbulence. When a field was propagated through turbulence screens, the extent of the phase screens needed to be larger in order to accommodate the spatially expanded energy distributions. In the following section, we describe how to calculate increased input and output plane extents which, when used in conjunction with Eq.(2.22), allowed for the determination of larger turbulence screen extents.

2.1.6 Overall Propagation Strategy

SAS propagations in the following chapters were conducted back and forth between a pupil plane and a target plane. Each propagation between these

planes used a series of consecutive partial propagations to intermediate planes which contained phase screens, as shown in Fig. 2.3. We used an indexing scheme where, again, the 0 index denotes the pupil plane and the S index denotes the target plane. Thus, our phase screens had indices in the range $[0, S-1]$. Note that a phase screen placed in the plane of the target has no blurring effect on the image, as implied by Eq. (2.25). Accordingly, no turbulence screens were included in the target plane. The full propagation

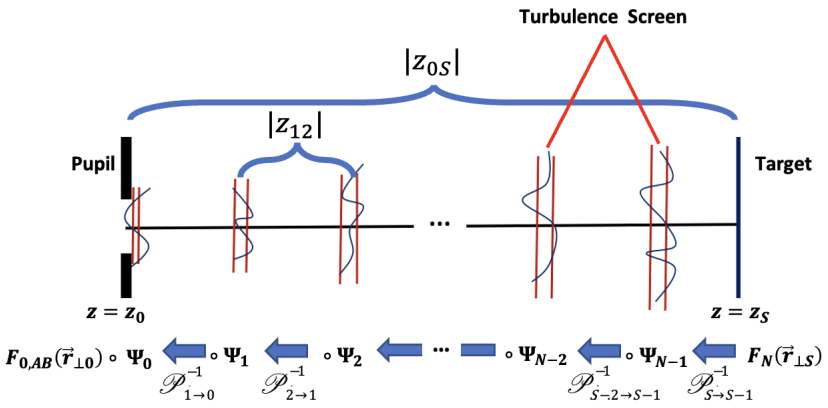

Figure 2.3: Target-to-pupil propagation through simulated turbulence screens.

process we used to simulate turbulence-based aberrations in our fields started with the ideal target fields $U_S(\vec{r}_{\perp S}; \nu)$. Next, a total of S partial propagations were conducted using the discrete SAS propagator. After each partial propagation, the Hadamard product of the field with the co-located turbulence screen transmittance function was propagated to the next plane. This pro-

cess was repeated until reaching the pupil plane. Once in the pupil plane, the final turbulence screen transmittance function was multiplied by the output field of the final partial propagation, resulting in the fully aberrated pupil field, $U_{0,AB}(\vec{r}_{\perp 0};\nu)$. Aberrated images of target fields were observed by propagating aberrated pupil fields directly back to the target plane from the pupil without any phase screens.

As previously stated, propagating through random turbulence redirects energy in unpredictable ways. The region containing relevant energy for a propagation is often not contained in the volume of the truncated square pyramids formed between the desired $D_0 \times D_0$ region and $D_S \times D_S$. Some higher spatial frequency information that would not have otherwise have been captured by the pupil and, conversely, some lower spatial frequency information we expect to see captured by the pupil being lost outside of the pupil. In order to ensure fields aberrated by turbulence are simulated in a physically realistic manor, we used the technique in [1, Eqs. (9.84), (9.85)] to calculate expanded input and output extents. This model is based on a simple model of single slit diffraction. It yields larger, adjusted values for D_0 and D_S based r_0

$$\begin{aligned} D_0' &= D_0 + \xi \frac{c|z_{0S}|}{\nu r_0} \\ D_S' &= D_S + \xi \frac{c|z_{0S}|}{\nu r_{0,REV}}, \end{aligned} \tag{2.32}$$

where $r_{0,REV}$ is the r_0 of all system phase screens as if calculated in the

reverse direction from Eq. (2.25), and ξ is an adjustable parameter indicating the sensitivity of the model to turbulence. The value of ξ typically ranges between 2, which captures $\sim$ 97% of the energy, and 4, which typically captures $\sim$ 99% [19].

2.2 Digital Holography and Multiwavelength Range-Compressed Holographic Lidar

2.2.1 Digital Holography

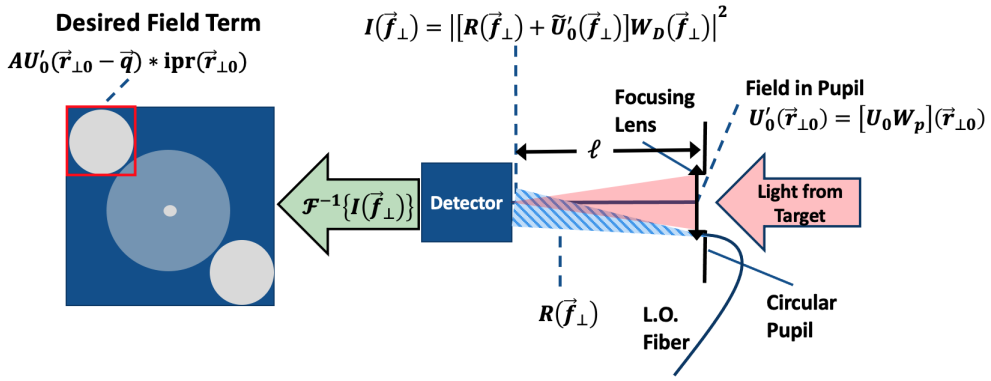

Figure 2.4: Diagram of focal plane recording geometry digital holography mixing and field extraction.

As mentioned in Chapter 1 our work uses a coherent spatial heterodyne technique known as digital holography (DH). Specifically, we use an off-axis focal plane (image plane) DH recording geometry (image plane recording geometry) as shown in Fig. 2.4. We assume a reference beam or local oscillator

(LO) field in the plane of the detector,

$$\begin{aligned} R\left(\vec{f}_{\perp}\right) &= A\exp\left(-ik\ell\right)\exp\left(-i2\pi q f_{\perp}\right)\exp\left[\frac{i\pi c\ell}{\nu}(\vec{f}_{\perp}\cdot\vec{f}_{\perp})\right] \\ &= A\exp\left(-ik\ell\right)\exp\left[-i2\sqrt{2}\pi q(f_x+f_y)\right]\exp\left[\frac{i\pi c\ell}{\nu}(f_x^2+f_y^2)\right], \end{aligned} \tag{2.33}$$

as a tilted spherical wave, A is the real-valued amplitude, ℓ is the focal length of the focusing lens immediately behind the pupil, q is the magnitude of the tip and tilt components of the reference beam, $f_{\perp} = \hat{f}_{\perp}\cdot\vec{f}_{\perp} = \sqrt{2}(\hat{x}+\hat{y})\cdot(f_x\hat{x}+f_y\hat{y}) = \sqrt{2}(f_x+f_y)$, $\vec{f}_{\perp}\cdot\vec{f}_{\perp} = f_x^2+f_y^2$, $f_x = \nu x_0/c\ell$, and $f_y = \nu y_0/c\ell$, where $\vec{r}_0 = x_0\hat{x}+y_0\hat{y}$, is the pupil-plane coordinate. Tip and tilt is purposely applied to the reference beam in order to separate the desired complex-valued pupil field terms from the on-axis autocorrelation terms for pupil field extraction. The field exiting the pupil is

$$U_0'\left(\vec{r}_{\perp 0}\right) = W_p(\vec{r}_{\perp 0})U_0(\vec{r}_{\perp 0}), \tag{2.34}$$

where $W_p(\vec{r}_{\perp 0})$ is the binary window function associated with the pupil and $U_0(\vec{r}_{\perp 0})$ is the field scattered off of the target and propagated to the pupil-plane. We see from Fig. 2.4 that the intensity on the detector is

$$I_{det}\left(\vec{f}_{\perp}\right) = \left|\left[\tilde{U}_0'\left(\vec{f}_{\perp}\right) + R\left(\vec{f}_{\perp}\right)\right]W_D\left(\vec{f}_{\perp}\right)\right|^2 \tag{2.35}$$

where $\tilde{U}_0'\left(\vec{f}_\perp\right)$ is the Fourier transform of $U_0'(\vec{r}_{\perp 0})$, and

$$W_D\left(\vec{f}_\perp\right) = \Pi\left[\frac{c\ell}{\nu D} f_x\right] \Pi\left[\frac{c\ell}{\nu D} f_y\right], \tag{2.36}$$

is the square binary window function of the detector, where $\Pi(\cdot)$ is a rectangle function, D is the extent of the detector window, ν is the frequency of illumination, and ℓ again is the focal length of the focusing lens. Note we assume here that the LO has been pathlength-matched to the focused pupil field, ensuring optimal coherence. Additionally, the quadratic phases are matched in the reference and focused pupil fields. To extract the pupil field, the inverse Fourier transform of the intensity on the detector is first calculated

$$\begin{aligned}
\mathscr{F}^{-1}_{\vec{f}_\perp \to \vec{r}_{\perp 0}}\left\{I_{det}\left(\vec{f}_\perp\right)\right\} &= \mathscr{F}_{\vec{f}_\perp \to \vec{r}_{\perp 0}}\left\{\left[\left|\tilde{U}_0'\left(\vec{f}_\perp\right) + R\left(\vec{f}_\perp\right)\right] W_D\left(\vec{f}_\perp\right)\right|^2\right\} \\
&\propto \Big[U_0'\left(\vec{r}_{\perp 0}\right) \star U_0'\left(\vec{r}_{\perp 0}\right) + A^2|\delta(\vec{r}_{\perp 0})| \\
&\quad + AU_0'(\vec{r}_{\perp 0} - \vec{q}) + AU_0'^{*}\left(\vec{r}_{\perp 0} + \vec{q}\right)\Big] \\
&\quad * \left[\operatorname{sinc}\left(\frac{\nu D}{c\ell} x_0\right) \operatorname{sinc}\left(\frac{\nu D}{c\ell} y_0\right)\right],
\end{aligned} \tag{2.37}$$

where the $\star$ operator denotes autocorrelation, $\vec{r}_{\perp 0} = x\hat{x}_0 + y\hat{y}_0$, again are pupil-plane coordinates, $\vec{q}$ is the displacement vector in the hologram plane cause by the tip and tilt of the reference beam, and A is the reference beam amplitude. We extract and recenter the desired pupil field to the optical axis.

Finally, we multiply by a binary pupil window function, $W_p(\vec{r}_{\perp 0})$, in order to suppress energy from the autocorrelation and delta-function terms

$$U_0''(\vec{r}_{\perp 0}) = W_p(\vec{r}_{\perp 0}) \left\{ A U_0'(\vec{r}_{\perp 0}) * \left[\operatorname{sinc}\left(\frac{\nu D}{c\ell} x_0 \right) \operatorname{sinc}\left(\frac{\nu D}{c\ell} y_0 \right) \right] \right\}. \quad (2.38)$$

2.2.2 Wave Optics Depth of Field

Consider again the Fresnel transform. For a single monochromatic illumination frequency in the absence of aberrations, it maps a transverse pupil plane, $U_0(\vec{r}_{\perp 0}, z_0; \nu)$, to a target field, $U_S(\vec{r}_{\perp S}, z_S; \nu)$

$$\begin{aligned} U_S(\vec{r}_{\perp S}, z_S; \nu) \propto \exp\left(\frac{i 2\pi \nu z_{0S}}{c} \right) \int_{-\infty}^{\infty} U_0(\vec{r}_{\perp 0}, z_0; \nu) \\ \times \exp\left(\frac{i\pi\nu}{c z_{0S}} \Delta \vec{r}^{\,2}_{\perp,0S} \right) d\vec{r}_{\perp S}. \end{aligned} \quad (2.39)$$

It is reasonable to ask how 2D intensity images of 3D objects can be created if the Fresnel propagator is defined to map only from a single transverse plane to another transverse plane. One way this can be understood is to realize that there is an effective wave-optics corollary to the geometric optics concept depth-of-field (DOF). DOF functionally extends the plane-to-plane mapping to include other target planes up to a distance of $\pm\delta_z$, where $2\delta_z$ = DOF, length units from the z-axis location used in the Fresnel transform without appreciable loss of accuracy (save for the leading piston phase term).

We can see from Eq. (2.39) that the external piston phase term oscillates at a much higher rate than the internal quadratic phase for remote sensing scenarios where $|\Delta r_\perp|_{\max} \ll |z_{0S}|$. We desire to describe the DOF range where the change in the internal quadratic phase term is negligible and the change in the external piston phase term remains significant, and in doing so, establish constraint expressions for $|\delta_z|$.

Since the SAS propagator is an equivalent operation to the Fresnel transform Eq. (2.39) in a continuous space, it is reasonable to derive constraints for the quadratic phase terms in Eq. (2.4). Constraining the quadratic phase terms in the SAS also allows us to avoid an indeterminate phase constraint for discretely sampled systems with unity magnification, which occurs when constraining the discretely sampled Fresnel's quadratic phase.

We define $|\delta_z|$ as the maximum distance from a designated output location along the z-axis which results in a phase difference magnitude of less than $\pi/2$ radians (a quarter wave) in each of the three SAS quadratic phase terms. Note that we use a quarter wave maximum phase difference as sufficient for our work imaging distant targets. Certain applications may require a stricter or allow for a less strict quadratic phase error tolerance.

The condition derived from the outermost quadratic phase on line 1 of

Eq. (2.4) is

$$\begin{aligned} \frac{\pi\nu}{c}\left|\left(\frac{m_{0S}-1}{m_{0S}}\right)\left(\frac{1}{z_{0S}}-\frac{1}{z_{0S}\pm\delta_z}\right)\left(r_{\perp 0}^2\right)\right| &< \frac{\pi}{2} \\ \frac{\nu}{c}\left|\left(\frac{m_{0S}-1}{m_{0S}}\right)\left[\frac{\pm\delta_z}{(z_{0S}\pm\delta_z)z_{0S}}\right]\left(r_{\perp 0}^2\right)\right| &< \frac{1}{2}. \end{aligned} \tag{2.40}$$

We assume (for all three constraints) $\delta_z \ll z_{0S}$, so we can say that the Eq. (2.40) is approximately the same as

$$\frac{\nu}{c}\left|\frac{m_{0S}-1}{m_{0S}}\right|\left(\frac{\delta_z}{z_{0S}^2}\right)\left\{r_{\perp 0}^2\right\}_{\max} < \frac{1}{2}. \tag{2.41}$$

Solving Eq. (2.41) for δ_z gives

$$\delta_z < \frac{c}{2\nu}\left|\frac{m_{0S}}{m_{0S}-1}\right|\left(\frac{z_{0S}^2}{\left\{r_{\perp 0}^2\right\}_{\max}}\right), \tag{2.42}$$

where the expression $\{\cdot\}_{\max}$ signifies the maximum value of the argument. In this case, the expression represents the largest squared magnitude of the transverse coordinate vector in the pupil plane which is of interest. Translating into a discretized space and using K-formalism, the constraint becomes

$$\delta_z < \frac{2z_{0S}}{K_{0S}}\left|\frac{1}{m_{0S}-1}\right|. \tag{2.43}$$

Similarly, we find the condition resulting from the innermost quadratic

phase on line 3 Eq. (2.4) to be

$$\delta_z < \frac{c}{2\nu}\left|\frac{1}{1-m_{0S}}\right|\left(\frac{z_{0S}^2}{\{r_{\perp S}^2\}_{\max}}\right). \tag{2.44}$$

In K-formalism the innermost quadratic phase condition is

$$\delta_z < \frac{2z_{0S}}{K_{0S}}\left|\frac{m_{0S}}{1-m_{0S}}\right|. \tag{2.45}$$

And lastly, the constraint derived from the transfer function phase on line 2 of Eq. (2.4) is

$$\delta_z < \frac{\nu m_{0S}}{2c}\left(\frac{1}{\{f_{\perp S}^2\}_{\max}}\right). \tag{2.46}$$

We know that the maximum spatial frequency of a paraxial system with target extent D_S, pupil extent D_0, frequency ν, and input-to-output plane distance z_{0S} is [1, Eqs. (7.5),(7.6), (7.12)]

$$f_{\max} = \frac{\nu(D_S m_{0S} + D_0)}{2cz_{0S}}. \tag{2.47}$$

Substituting Eq. (2.47) into Eq. (2.46) we find

$$\delta_z < \frac{2cm_{0S}}{\nu}\left(\frac{z_{0S}}{D_S m_{0S} + D_0}\right)^2. \tag{2.48}$$

As a brief aside, if we are, say, interested in observing the psf of a

system with unity magnification, propagating from the pupil to the focus of the system, and under the assumption that $D_0 \ll D_S$, we find

$$\delta_z < \frac{2cz_{0S}^2}{\nu D_S^2}, \tag{2.49}$$

yielding a DOF region

$$\text{DOF} = \frac{4cz_{0S}^2}{\nu D_S^2}. \tag{2.50}$$

This is exactly the geometric optics DOF for a system with a square pupil.

In K-formalism, the transfer function condition is

$$\delta_z < \frac{2z_{0S}}{K_{0S}} \left[\frac{m_{0S}}{m_{0S}+1} \right]^2. \tag{2.51}$$

We see from Eq. (2.4) that in systems with a unity magnification, only the transfer function condition is relevant, since the innermost and outermost quadratic phase terms become unity. Thus, when magnification is unity, the only constraint is

$$\delta_z < \frac{z_{0S}}{2K_{0S}}. \tag{2.52}$$

So the DOF in this case is simply the ratio of the propagation distance to the SBWP.

As an example, assume a system with pupil extent $D_0 = 0.25$ m, transverse target extent (TFOV) $D_S = 4$ m, propagation distance $z_{0S} = 10^4$ m, a wavelength $\lambda = c/\nu = 10^{-6}$ m, and a unity magnification. We find $\delta_z \leq 50$

m. This illustrates the poor range (axial) resolution that can be had by depth of focus (or triangulation over the width of the aperture), as compared with the fine range resolution that can be had using optical frequency diversity as discussed later in this section. We can see from this DOF example how, in spite of using a propagator defined exclusively for plane-to-plane mapping, we are able to generate 2D intensity images where all parts of a 3D target object appear in-focus.

2.2.3 Range-compressed holography

To begin our derivation of range-compressed holography (RCH), assume that a target is illuminated by a coherent source co-located with the pupil. Our development of the wave optics DOF concept allows us to approximate the 3D complex reflectance of this target as the product of a 2D transverse reflectance function and and an accompanying delta function containing the depth location of the target surface for each location in the transverse S plane

$$U_S(\vec{r}_{\perp S}, z) \approx U_{S\perp}(\vec{r}_{\perp S})\delta[z - Z(\vec{r}_{\perp S})], \tag{2.53}$$

where $Z(\vec{r}_{\perp S})$ is the distance from the pupil to the target's surface as a function of the transverse coordinates in the target plane.

Next, we conduct a Fresnel propagation of Eq. (2.53) from the various planes in our DOF region to the pupil plane. We use z_S as the known z-axis

value for the internal phase portion of the Fresnel integral shown in Eq. (2.39). Since we have stated that the target is contained within the DOF region as described above, the transverse Fresnel phase term is approximately constant in this region with respect to changes in z and ν. We choose the z_S plane to indicate the approximate halfway point in the region of the z axis which contains the target object. The Fresnel propagation of the all the planes in our target DOF region

$$\begin{aligned} U_0(\vec{r}_{\perp 0}, z_0; \nu) \propto & \int_{z_S-\delta_z}^{z_S+\delta_z} \exp\left[\frac{-i4\pi\nu(z-z_0)}{c}\right] \int_{-\infty}^{\infty} U_{S\perp}(\vec{r}_{\perp S})\delta[z - Z(\vec{r}_{\perp S})] \\ & \times \exp\left(\frac{-i\pi\nu}{cz_{0S}}\Delta\vec{r}^{\,2}_{\perp 0S}\right) d\vec{r}_{\perp S} dz \\ = & \int_{-\infty}^{\infty} U_{S\perp}(\vec{r}_{\perp S}) \exp\left\{\frac{-i4\pi\nu[Z(\vec{r}_{\perp S}) - z_0]}{c}\right\} \\ & \times \exp\left[\frac{-i\pi\nu}{cz_{0S}}\Delta\vec{r}^{\,2}_{\perp 0S}\right] d\vec{r}_{\perp S}, \end{aligned} \tag{2.54}$$

where the additional factor of two in the linear phase term is the result of illumination traveling from the pupil/transmitter plane to the target object and back again (double pass). Fresnel transforming both sides of Eq. (2.54) from the pupil plane to the nominal target plane z_S

$$\begin{aligned} & \exp\left(\frac{i2\pi\nu z_{0S}}{c}\right) \int_{-\infty}^{\infty} U_0(\vec{r}_{\perp 0}, z_0; \nu) \exp\left(\frac{i\pi\nu}{cz_{0S}}\Delta\vec{r}^{\,2}_{\perp,0S}\right) d\vec{r}_{\perp 0} \\ & \qquad \propto \exp\left(\frac{i2\pi\nu z_{0S}}{c}\right) U_{S\perp}(\vec{r}_{\perp S}) \exp\left\{\frac{-i4\pi\nu[Z(\vec{r}_{\perp S}) - z_0]}{c}\right\}. \end{aligned} \tag{2.55}$$

Conducting a change of coordinates, substituting $z' = z - z_S$ and therefore also $Z'(\vec{r}_{\perp S}) = Z(\vec{r}_{\perp S}) - z_S$ gives

$$\begin{aligned} \exp\left(\frac{i2\pi\nu z_{0S}}{c}\right) \int_{-\infty}^{\infty} U_0(\vec{r}_{\perp 0}, z_0; \nu) \exp\left(\frac{i\pi\nu_0}{cz_{0S}} \Delta\vec{r}^{\,2}_{\perp,0S}\right) d\vec{r}_{\perp 0} \\ \propto \exp\left(\frac{i2\pi\nu z_{0S}}{c}\right) U_{S\perp}(\vec{r}_{\perp S}) \exp\left\{\frac{-i4\pi\nu[Z'(\vec{r}_{\perp S}) + z_{0S}]}{c}\right\} \\ = \exp\left(\frac{-i2\pi\nu z_{0S}}{c}\right) U_{S\perp}(\vec{r}_{\perp S}) \exp\left\{\frac{-i4\pi\nu Z'(\vec{r}_{\perp S})}{c}\right\}, \end{aligned} \tag{2.56}$$

which offers an intuitive result with the right hand side of the equation now only having a single pass worth of piston phase between the pupil and target. Propagating fields from the pupil back to the nominal target plane added an opposite piston phase term over a distance z_{0S}, leaving only the piston phase accrued during the illumination beam's initial propagation to the target. Additionally, the complex reflectance information detailing surface features of the target is still represented by the phase term which varies transversely as a function of $Z'(\vec{r}_{\perp S})$. Next, multiplying both sides by $\exp(i2\pi\nu z_{0S}/c)$

$$\begin{aligned} \exp\left(\frac{i2\pi\nu z_{0S}}{c}\right) \int_{-\infty}^{\infty} U_0(\vec{r}_{\perp 0}, z_0; \nu) \exp\left(\frac{i\pi\nu_0}{cz_{0S}} \Delta\vec{r}^{\,2}_{\perp,0S}\right) d\vec{r}_{\perp 0} \\ = U_{S\perp}(\vec{r}_{\perp S}) \exp\left\{\frac{-i4\pi\nu Z'(\vec{r}_{\perp S})}{c}\right\}, \end{aligned} \tag{2.57}$$

the right hand side becomes the 2D complex reflectance function and a linear phase term which is a function of both illumination frequency and depth as a function of transverse coordinates.

Instead of creating a digital hologram using only a single frequency of

monochromatic illumination, suppose we collected a continuum of illumination frequencies. Inverse Fourier transforming Eq. (2.57) with respect to $2\nu/c$ gives

$$\mathscr{F}^{-1}_{\frac{2\nu}{c}\to z'}\left\{U_{S\perp}(\vec{r}_{\perp S})\exp\left[\frac{-i4\pi\nu Z'(\vec{r}_{\perp S})}{c}\right]\right\} = U_{S\perp}(\vec{r}_{\perp S})\delta[z' - Z'(\vec{r}_{\perp S})]$$
$$= U_{S\perp}(\vec{r}_{\perp S})\delta[z - Z(\vec{r}_{\perp S})], \tag{2.58}$$

the original approximated 3D complex reflectance function from Eq. (2.53). This final inverse Fourier transform step is known as a range-compression. We see that RCH enables us to recover the target's depth information in addition to its 2D field and intensity image. Note that this equation assumes infinite spatial-frequency bandwidth and transverse extent, as well as an infinite range of ν. We address this topic in the next subsection. Note that we chose to use the continuous space Fresnel propagation integral in this derivation as its notation is more compact than the equivalent SAS, whose discrete form we use in Chapters 3 and 4. We could just as easily used the SAS to achieve the same result. Our use of an inverse Fourier transform to perform the range-compression in Eq. (2.58), was the result of the propagation convention we used which assumed that "forward" Fresnel propagations started at the pupil plane and ended at the target plane, as shown in Eq. (2.39). Had we chosen the opposite convention where "forward" propagation went from the target plane to the pupil plane, range-compression would have used

a Fourier transform to recover the depth information-containing delta function. Finally, in spite of our opaque target assumption, the range-compressed holography technique that derived here will work to varying degrees for non-opaque objects as well.

2.2.4 Illumination Bandwidth, Range Resolution, and Spectrum Sampling

In order to produce the necessary laser bandwidth to conduct RCH, a number of laser technologies can be used. We consider a linear frequency-modulated (LFM), or chirp-pulse, laser as the physical model we will use to explain the generation of this bandwidth [4]. The means by which a frequency-chirped signal generates illumination bandwidth, how this limited bandwidth affects RCH range resolution, and how discretely sampling the tunable spectrum affects range measurements are discussed. Often, these topics are treated using more traditional techniques from the radar/SAR world involving matched filters [20, 21]. However, we discuss them in more traditional wave optics parlance. The physics, however, is the same. We begin with a paraxial chirp illumination beam [4]

$$U_{\text{chirp}}(\vec{r}_\perp, z; t) \propto \exp\left[\frac{i2\pi\nu(t)z}{c}\right] \exp\left[\frac{(\vec{r}_\perp \cdot \vec{r}_\perp)}{w(z)}\right] \exp\left[\frac{i\pi\nu(t)}{cz}(\vec{r}_\perp \cdot \vec{r}_\perp)\right] \times \exp\left[i\varphi(z)\right] b(t) \exp\left(-\alpha t^2\right) \exp\left(i2\pi\nu_0 t + i\beta t^2\right), \tag{2.59}$$

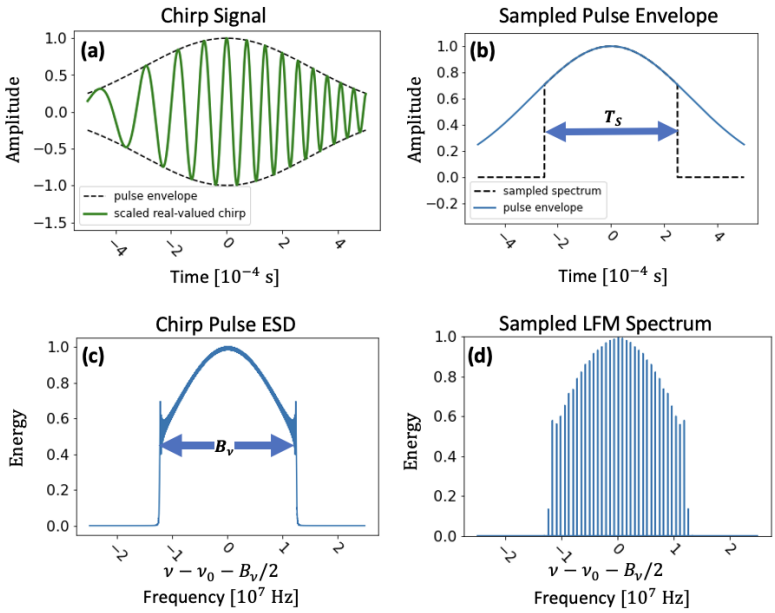

Figure 2.5: a) Gaussian pulse envelope and real-valued chirped signal component, with the base frequency and chirp constant scaled to aid visibility of pulse compression. (b) Windowed signal showing the FWHM region of Gaussian envelope from which illumination frequencies will be sampled, $T_S = T_P = 5 \times 10^{-4}$ s. (c) Windowed LFM pulse energy spectral density (ESD) indicating bandwidth $B_\nu = 25$ MHz. (d) Discretely sampled LFM ESD showing sampled frequencies.

where $\nu_0 t + (\beta/\pi)t^2$ is known as a frequency chirp. The instantaneous frequency function, $\nu(t)$, is proportional to the rate of change of the chirp

$$\nu(t) = \frac{d}{dt}\left\{\nu_0 t + \frac{\beta}{2\pi}t^2\right\} = \nu_0 + \frac{\beta}{\pi}t, \tag{2.60}$$

$\exp\{[i\pi\nu(t)/cz]\,(x^2+y^2)\}$ is the spatial quadratic phase component of a Gaussian beam, $\exp[(\vec{r}_\perp \cdot \vec{r}_\perp)/w(z)]$ is the transverse Gaussian amplitude, $w(z)$ is the radius at which beam amplitude falls off to $1/e$ its peak value,

$\varphi(z)$ is the Guoy phase, $\exp(-\alpha t^2)$ is the temporal Gaussian amplitude envelope, where $\alpha = 2\ln 2/T_P^2$ and T_P is the FWHM of the Gaussian intensity in time, and $b(t)$ is a non-negative real-valued window function

$$b(t) = \text{rect}(t/T_S), \tag{2.61}$$

which selects the desired portion of the signal. The variable β is a known chirp rate constant in Hz·radians/s units such that $\beta t/\pi \ll \nu_0$, ν_0 is the base illumination frequency, and T_S represents the span of time during the chirp we can detect (used to generate RCH frequencies) where, typically, $T_S \leq T_P$. The truncated temporal Gaussian envelope was chosen because of its ability to suppress unwanted sidelobes resulting from the Fresnel integral.

The window, temporal amplitude envelope, and chirp phase of Eq. (2.59)

$$U_{\text{chirp}}(t) = b(t)\exp\left(-\alpha t^2\right)\exp\left(i2\pi\nu_0 t + i\beta t^2\right). \tag{2.62}$$

are of primary concern when describing coherent range imaging. Observe that Eqs. (2.60) and (2.61) imply an overall illumination bandwidth

$$B_\nu = \frac{\beta}{\pi}T_S. \tag{2.63}$$

To understand the energy distribution of the chirp signal and confirm the bandwidth of the spectrum, we calculate the energy spectral density (ESD)

[13, Eq. (3.3-5)] of the selected time period, T_S, in the chirp signal Eq. (2.62)

$$\begin{aligned} \mathrm{ESD}(\nu) &= \left|\mathscr{F}_{t\to\nu}\left\{U_{\mathrm{chirp}}(t)\right\}\right|^2 \\ &\propto \left|\mathrm{sinc}\left(T_S\nu\right) * \exp\left[\frac{\pi^2\left(\nu-\nu_0\right)^2}{\Gamma}\right]\right|^2, \end{aligned} \tag{2.64}$$

where $\Gamma = \alpha - i\beta$.

The process of calculating the ESD of our windowed beam is a 1D time analogue of the intensity produced by a 2D Fresnel propagation in the Fraunhofer regime [22, Eq. (4.3-12)]. The quadratic phase term in Eq. (2.64), caused by the signal chirp, has the capability to produce significant bandwidth broadening in the ESD. Shown in Fig. 2.5 (c) is the ESD of a sample function with $T_S = 5 \times 10^{-4}$ s selected period and $\beta = 50$ MHz/s. The bandwidth of the chirp signal is calculated to be 25 MHz, which is shown in Fig. 2.5 (c). Alternatively, we can easily tell that the ESD of a non-chirped signal would have a width equivalent to that of the main lobe of the Fourier transform of $b(t)$, or roughly 10^3 Hz.

In practice, illumination bandwidth is finite, not infinite as was assumed in Eq. (2.58). To limit our bandwidth in this case, we multiply the argument of the Fourier transform on the left hand side of Eq. (2.58) by $b(\nu) = \mathrm{rect}[(1/B_\nu)(\nu - \nu_0 - B_\nu/2)]$, which is our window function changed to a function of frequency using Eq. (2.60). The introduction of this window

$b(\nu)$ to Eq. (2.58) gives

$$\begin{aligned}\mathscr{F}^{-1}_{\frac{2\nu}{c}\to z'}&\left\{b(\nu)\exp\left[\frac{-i4\pi\nu Z'(\vec{r}_{\perp S})}{c}\right]U_{S\perp}(\vec{r}_{\perp S})\right\}\\&=\mathscr{F}^{-1}_{\frac{2\nu}{c}\to z'}\{b(\nu)\}*\mathscr{F}^{-1}_{\frac{2\nu}{c}\to z'}\left\{\exp\left[\frac{-i4\pi\nu Z'(\vec{r}_{\perp S})}{c}\right]U_{S\perp}(\vec{r}_{\perp S})\right\},\end{aligned} \tag{2.65}$$

which to within a linear phase shift in z' is equivalent to

$$\begin{aligned}\operatorname{sinc}\left(\frac{2B_\nu}{c}z'\right)&*\{U_{S\perp}(\vec{r}_{\perp S})\delta[z'-Z'(\vec{r}_{\perp S})]\}\\&=U_{S\perp}(\vec{r}_{\perp S})\operatorname{sinc}\left\{\frac{2B_\nu}{c}[z'-Z'(\vec{r}_{\perp S})]\right\}.\end{aligned} \tag{2.66}$$

We use the width from the center to the first null of the sinc's main lobe to define our range resolution using the Rayleigh two-point criterion

$$\rho_z=\frac{c}{2B_\nu}. \tag{2.67}$$

Note that Eq. (2.65) assumes infinite spatial-frequency bandwidth and a continuous spatial domain, for the sake of conciseness in our expressions. Limitations due to finite spatial-frequency bandwidth and discrete spatial sampling are well known and can be found in [1], or more generally, discretely sampled Fourier transforms are studied in depth in [23].

Analogous to discrete spatial sampling conducted by detector pixels, the range component of RCH also relies on discretization of illumination frequen-

cies. With the chirp laser we have assumed, different frequencies are attained by sampling the signal at different times by collecting a number of detector frames during each chirp. Since we trigger the detector to isolate certain frequencies, it effectively multiplies the spectrum by small rect functions, $1/M$ times the width of $b(t)$, where M is the number of discrete frequencies we sample. Again, we can think of calculating the ESD for each sampled frequency as analogous to conducting a Fresnel propagation [22, pp 132]. It will be a convolution of the the Fourier transform of the rect function that samples the pulse in time (a sinc function) and the Fourier transform of the quadratic chirp term. Whereas earlier when the convolution of the quadratic chirp term and the sinc resulting from the Fourier transform of $b(t)$ yielded Fig. 2.5 (c), a very broad ESD as compared to the spectrum of the sinc, the convolution of the wider sinc resulting from the more narrow frequency sampling rect function will result in an ESD that is many times narrower than Fig. 2.5 (c). This is because while the convolution of the chirp phase with the narrower sinc function is analogous to a propagation in the Fraunhofer regime, thus broadening the distribution, the convolution of the chirp with the wider sinc resulting from an individual frequency sample is analogous to a near-field Fresnel where the spectrum is much closer in width to that of the convolved sinc. This indicates that the spectral purity of each sampled frequency increases with the total number of sampled frequencies.

Similar to the discrete sampling of spatial coordinates by a detector, the discrete sampling of the illumination by the comb function produces a

wrap-around ambiguity in range determination. Multiplication by a comb-function, $\text{Ш}_{2\Delta\nu/c}(2\nu/c) = \sum_{k=-\infty}^{\infty}\delta(2\nu/c - 2k\Delta\nu/c)$, in the illumination frequency space becomes a convolution with the Fourier transform of the sampling comb function (another comb function) in the range-compressed space

$$\mathscr{F}_{\frac{2\nu}{c}\to z'}\left\{\exp\left[\frac{-i4\pi\nu Z'(\vec{r}_{\perp S})}{c}\right]\text{Ш}_{\frac{2\Delta\nu}{c}}\left(\frac{2\nu}{c}\right)\right\} \propto \delta[z' - Z'(\vec{r}_{\perp S})] * \text{Ш}_{\frac{c}{2\Delta\nu}}(z'), \tag{2.68}$$

where $\Delta\nu$ is the frequency sampling interval, i.e., the distance between peaks in Fig. (2.5) (d). This convolution creates periodic copies of the original sampled function's Fourier transform. It is these copies which create the aforementioned range ambiguity

$$R_{\text{amb}} = \frac{c}{2\Delta\nu}, \tag{2.69}$$

as it is impossible, generally, to distinguish between the true range to target given by $\delta[z' - Z'(\vec{r}_{\perp S})]$, and $\delta[z' - Z'(\vec{r}_{\perp S}) - nR_{\text{amb}}]$, where n is an integer. So in RCH, range can only be measured modulo R_{amb}.

Range Map Formation

After computing the 3D intensity array, $U_S(\vec{r}_{\perp S}, z')$ in Eq. (2.58), one needs to determine the best way to visualize the 3D data as a 2D figure. We

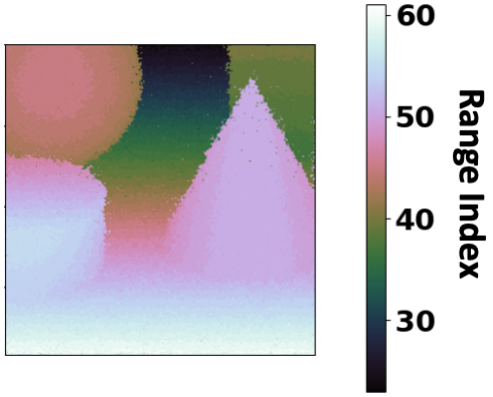

Figure 2.6: Sample range map created from RCH fields using Eq. (2.70)

.

estimate the correct location of a scatterer in the target in z' for all of $\vec{r}_{\perp S}$ by choosing the location with the largest intensity [24]

$$R(\vec{r}_{\perp S}) = \underset{z'}{\operatorname{argmax}} \left\{ \left| U_S(\vec{r}_{\perp S}, z') \right|^2 \right\}. \tag{2.70}$$

An example range map created from simulated RCH fields is shown in Fig. 2.6. The resulting 2D function has spatial information along the transverse dimensions as with a typical intensity image, but instead of intensity data the color-coded pixel value, range data is instead the color-coded pixel value. In practice, discrete arrays often exhibit multiple intensity peaks along z'-axis of a single cross-range pixel. This is typically due to pixels containing edge information or objects being semi-transparent. As an example, if an object in the foreground only takes up part of the target plane cross-sectional area

of a pixel, some of the photon flux of the illuminator will travel past the part of the object in the foreground and possibly reflect off an object further away in range. This leads to two distinct intensity peaks along the z'-axis for a single cross-range pixel. In this case, choosing the highest intensity discards the information from either the first or second peak, depending on which had the highest intensity. Note that there are peak-finding techniques from traditional radar and SAR which are likely helpful in this situation, but we have not explored application of these techniques to RCH.

Bibliography

[1] J. D. Schmidt, *Numerical simulation of optical wave propagation with examples in MATLAB* (SPIE, 2010).

[2] C. A. Balanis, *Advanced engineering electromagnetics* (John Wiley & Sons, 1999).

[3] J. W. Goodman, *Introduction to Fourier optics, 4th Ed.* (W. H. Freeman, 2017).

[4] A. E. Siegman, *Lasers* (University Science Books, 1986).

[5] M. Born and E. Wolf, *Principles of optics, 7th Ed. (expanded)* (Press Syndicate of the University of Cambridge, 1999).

[6] E. A. Sziklas and A. E. Siegman, "Diffraction calculations using fast Fourier transform methods," P. IEEE **62**, 410–412 (1974).

[7] X. Yu, T. Xiahui, Q. Yingxiong, P. Hao, and W. Wei, "Band-limited angular spectrum numerical propagation method with selective scaling of observation window size and sample number," J. Opt. Soc. Am. A **29**, 2415–2420 (2012).

[8] J. D. Mansell, R. Praus, and S. Coy, "Determining wave-optics mesh parameters for complex optical systems," Proc. SPIE **6675**, 66750H (2007).

[9] J. W. Goodman, *Speckle phenomena in optics: theory and applications* (Roberts and Company Publishers, 2007).

[10] M. C. Roggemann and B. M. Welsh, *Imaging through turbulence* (CRC press, 2018).

[11] R. L. Philips and L. C. Andrews, *Laser beam propagation through random media* (SPIE, 1998).

[12] H. G. Booker, J. A. Ratcliffe, and D. H. Shinn, "Diffraction from an irregular screen with applications to ionospheric problems," Philos. T. R. Soc. S-A **242**, 579–607 (1950).

[13] J. W. Goodman, *Statistical optics* (John Wiley & Sons, 2015).

[14] D. L. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," J. Opt. Soc. Am. **56**, 1372–1379 (1966).

[15] D. L. Fried, "Anisoplanatism in adaptive optics," J. Opt. Soc. Am. **72**, 52–61 (1982).

[16] R. G. Lane, A. Glindemann, and J. C. Dainty, "Simulation of a Kolmogorov phase screen," Waves Random Media **2**, 209–224 (1992).

[17] Z. Chen, D. Zhang, C. Xiao, and M. Qin, "Precision analysis of turbulence phase screens and its influence on simulation of gaussian-beam

propagating in the turbulent atmosphere," Appl. Opt. **59**, 3726–3735 (2020).

[18] R. A. Johnston and R. G. Lane, "Modeling scintillation from an aperiodic Kolmogorov phase screen," Appl. Opt. **39**, 4761–4769 (2000).

[19] S. Coy, "Choosing mesh spacings and mesh dimensions for wave optics simulation," Proc. SPIE **5894**, 589405 (2005).

[20] M. Soumekh, *Synthetic aperture radar signal processing*, vol. 7 (New York: Wiley, 1999).

[21] C. E. Cook, "Pulse compression-key to more efficient radar transmission," Proc. IRE **48**, 310–316 (1960).

[22] B. E. A. Saleh and M. C. Teich, *Fundamentals of photonics* (John Wiley & Sons, 1991).

[23] R. N. Bracewell, *The Fourier transform and its applications* (McGraw-Hill New York, 1986).

[24] J. W. Stafford, B. D. Duncan, and D. J. Rabb, "Phase gradient algorithm method for three-dimensional holographic ladar imaging," Appl. Opt. **55**, 4611–4620 (2016).

Chapter 3

Multiplane Turbulence Mitigation Simulations Using Iterative Sharpness Maximization

In Section 2.1.5 we described how scalar field propagations are altered in the presence of volumetric turbulence, the parameters used to describe that turbulence, and the methods by which we approximate and simulate its effects. In Section 2.2.3 we described imaging using range-compressed holography (RCH) and the process by which 2D range maps are created from the resulting 3D intensity images. Here we consider the intersection of these two topics. Since RCH is dependent upon scalar wave propagation, it is vulnerable to volumetric turbulence and image quality can be heavily degraded by the resulting aberrations. In this chapter, we describe our novel algorithm we developed and implemented to mitigate turbulent effects on range-compressed holograms. We detail the simulations we conducted in order to study the efficacy of our algorithm and analyze the results of these studies.

The algorithm we used to mitigate simulated turbulence is known as iterative sharpness maximization (ISM). It estimates phase planes along the path of optical propagation so as to maximize image sharpness, i.e., energy concentration intensity arrays. It has been used in the past to mitigate turbulence in a number of different imaging applications including astronomical imaging [1], synthetic aperture radar [2], and digital holography [3, 4]. In past work, we have used ISM to correct aberrations caused by a single plane of simulated atmospheric turbulence in the pupil [5]. We expand on this work here by employing multiple phase plane estimates to mitigate the turbulent effects on RCH fields.

3.1 Algorithm optimization model

The most convenient way to begin thinking about our optimization model is to consider it as the inverse of the process described in Chapter 2.1.6 and shown in Fig. 2.3. Since this plane-to-plane propagation through turbulence screens caused the aberrations in the pupil plane fields (or to the extent that it is a good approximation of the causal process), it stands to reason that the inverse process would serve to remove these aberrations. Provided we know the exact turbulence screens that aberrated the fields measured in the pupil plane, the majority of turbulent effects on the image can be corrected.

We therefore designed our algorithm to estimate the turbulence screens.

The algorithm assumed that turbulence-aberrated intensity images of a target are less sharp than intensity images of targets where the fields have not propagated through turbulence. Sharpness was quantified using a measure of energy concentration in the 3D RCH intensity image. Specifically, we used an exponential sharpness metric [3]

$$\mathcal{J}(\hat{\mathbf{\Phi}}) = \operatorname{sgn}(\alpha - 1) \sum_{\vec{r}_{\perp S}, z'} I^{\alpha}_{\text{avg}}(\vec{r}_{\perp S}; z; \hat{\mathbf{\Phi}}), \tag{3.1}$$

where $I^{\alpha}_{\text{avg}}(\vec{r}_{\perp S}; z'; \hat{\mathbf{\Phi}})$ is the range-compressed, incoherently averaged, target intensity which has been propagated through our estimated point-by-point phase screens, $\hat{\mathbf{\Phi}}$, α is the power to which the intensity is taken, and $\operatorname{sgn}(\cdot)$ is the sign function. We maximized this metric with respect to $\hat{\mathbf{\Phi}}$ and in doing so estimated the turbulence screens.

To maximize the sharpness as a function of the estimated phase screens, we constructed a physical optimization model. The first part of the model propagated fields in the pupil to a series of successive intermediate planes between the target and pupil that contained our estimated phase screen transmittance functions. We keep consistent with notation used in Chapter 2, where we defined propagation from the pupil to the target plane as forward Fresnel propagation denoted $\mathscr{P}_{s \to s+1}\{\cdot\}$ and propagation from the pupil to the target plane as "inverse" Fresnel propagation denoted $\mathscr{P}^{-1}_{s+1 \to s}\{\cdot\}$. Each plane, excluding the target plane but including the pupil, contained an estimated phase screen transmittance function, $\Psi_s(\vec{r}_{\perp s}; \hat{\Phi}_s)$. Prior to beginning

each propagation, the field in that plane was multiplied by the co-located transmittance function. This is expressed as

$$\begin{aligned} U_S(\vec{r}_{\perp S},\nu,\eta;\hat{\mathbf{\Phi}}) = \mathscr{P}_{S-1\to S}\Big\{ & \Psi_{S-1}(\vec{r}_{\perp S-1};\hat{\Phi}_{S-1}) \\ & \circ \mathscr{P}_{S-2\to S-1}\Big\{ \Psi_{S-2}(\vec{r}_{\perp S-2};\hat{\Phi}_{S-2})\dots \\ & \circ \mathscr{P}_{1\to 2}\Big\{ \Psi_1(\vec{r}_{\perp 1};\hat{\Phi}_1) \\ & \circ \mathscr{P}_{0\to 1}\Big\{ \Psi_0(\vec{r}_{\perp 0};\hat{\Phi}_0)\circ U_0(\vec{r}_{\perp 0},\nu,\eta)\Big\}\Big\}\dots\Big\}\Big\} \end{aligned} \tag{3.2}$$

where η represents one of H total unique speckle realization. The symbol " $\circ$ " represents a Hadamard product, $U_0(\vec{r}_{\perp 0},\nu,\eta)$ is again the three-dimensional array of frequency-diverse pupil-plane fields, and $\Psi_s(\vec{r}_{\perp s};\hat{\Phi}_s)$ is one of S complex-valued phase screen transmittance functions which are functions dependent on the phase estimate $\hat{\Phi}_s$. Employing intuition developed in [4, 6], we used the method of sieves (MoS) technique [7] to improve our optimizations. This technique convolved estimated phase screens with a Gaussian kernel function, creating a phase screen transmittance function

$$\Psi_s(\vec{r}_{\perp s},\hat{\Phi}_s) = \exp\Big\{ i\hat{\Phi}_s(\vec{r}_{\perp s}) * G_s(\vec{r}_{\perp s};\sigma_s)\Big\}, \tag{3.3}$$

and can be constructed from elementary operations in Table 7.5 in Appendix 7.1.1. The function $G_s(\vec{r}_{\perp s};\sigma_s)$ is the s-th Gaussian kernel with a standard

deviation σ_s:

$$G_s(\vec{r}_{\perp s};\sigma_s) = \frac{1}{\sigma_s^2 2\pi}\exp\left(\frac{\vec{r}_{\perp s}\cdot\vec{r}_{\perp s}}{2\sigma_s^2}\right), \tag{3.4}$$

and the $*$ symbol denotes convolution. The MoS promoted accumulation of smooth, lower spatial frequency phase early in the optimization before allowing the algorithm to fit higher spatial frequency phase information which is allowed to occur by decreasing the standard deviation of the Gaussian.

In the nominal target plane, the fields were range-compressed, and the exponential sharpness of the incoherently-averaged, range-compressed intensity is calculated

$$\mathcal{J}(\hat{\mathbf{\Phi}}) = \frac{\operatorname{sgn}(\alpha-1)}{\mathcal{J}_{\text{init}}H}\sum_{\vec{r}_{\perp S},z'}\left(\sum_{\eta}\left|\tilde{U}_S(\vec{r}_{\perp S};z',\eta;\hat{\mathbf{\Phi}})\right|^2\right)^{\alpha} \tag{3.5}$$

where

$$\tilde{U}_S(\vec{r}_{\perp S},z,\eta;\hat{\mathbf{\Phi}}) = \mathscr{F}^{-1}_{\frac{2\nu}{c}\to z'}\{U_S(\vec{r}_{\perp S},\nu,\eta;\hat{\mathbf{\Phi}})\} \tag{3.6}$$

is the range-compression operation from Eq. (2.58).

Previous work [4, 8] used multiple speckle realizations, averaging them together incoherently before computing sharpness. Work in [3] showed that having too few speckle realizations negatively impacted sharpness algorithm performance. Though we include the function argument for speckle realization in our optimization model for later use in Chapter 4, here we assume the much more difficult case of image sharpening using a single speckle realization. We expected that the additional information provided by frequency

diversity and the comparative sparsity of range-compressed intensity images would more than compensate for the robustness lost by using only a single speckle realization. The optional $\mathcal{I}_{\text{init}}$ divisor

$$\mathcal{I}_{\text{init}} = \sum_{\vec{r}_{\perp S}, z'_{,}} \left(\sum_{\eta} \left| \mathscr{F}^{-1}_{\frac{2\nu}{c} \to z'} \left\{ U_S(\vec{r}_{\perp S}, \nu, \eta) \right\} \right|^2 \right)^{\alpha}, \tag{3.7}$$

is the initial sharpness of the RCH without phase correction. Since there is no phase correction included, this term is a constant independent of $\hat{\mathbf{\Phi}}$ and behaves as a constant in gradient of this optimization model. The $\mathcal{I}_{\text{init}}$ was used to normalize the initial magnitude of the sharpness metric to unity. We observed early on that optimizations where sharpness metric values were allowed to become very large often resulted in sub-optimal results, likely due to assumptions regarding values of parameters within the optimizer. This normalization both eliminated these difficulties and allowed us to more easily balance the weighting of the sharpness portion of our objective function with the penalty terms in our objective function, which we will discuss shortly. Note also that the change in image quality as a function of the sharpness metric value, though generally positively correlated when properly constrained, is unpredictable and nonlinear.

In our research, we have encountered failure modes which commonly affect the usefulness of the power law sharpness metric used for turbulence mitigation. One is an afocal telescoping effect often by opposing quadratic phase screens which have the effect of shrinking the transverse extent of

the target energy distribution and erroneously increasing the value of the sharpness metric [4]. Similarly, we have observed other arrangements of phase screen estimates which, in aggregate, cause accumulation of energy in the target plane into smaller regions than is physically realistic, also causing an undesirable increase in the sharpness metric which does not correlate with increased image quality. To help address these concerns, two penalty terms are assessed in addition to the sharpness metric.

The first penalty term in our model penalized an overabundance of defocus phase in any of the phase screens, which often lead to the afocal telescoping and can often also contribute to the second, more general, oversharpening-based failure mode. Specifically, this metric was designed to penalize the relative amount of defocus in the overall estimated phase of a given phase screen, as opposed to the absolute amount of defocus phase, ideally creating a penalty term whose weighting parameter, λ_D, would be more independent of changes in the strength of the turbulence. This penalty term is expressed

$$p_D(\hat{\mathbf{\Phi}}) = \frac{\lambda_D}{S} \sum_s \frac{\left[\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s(\vec{r}_{\perp s}) \circ Z_4(\vec{r}_{\perp s})\right]^2}{\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s^2(\vec{r}_{\perp s}) + \beta}, \tag{3.8}$$

where $Z_4(\vec{r}_{\perp s})$ (using the Zernike numbering scheme in [9]) is the defocus Zernike term and the sum in the numerator is a projection of the phase onto Z_4, which is normalized by the denominator, and the square is summed over all the phase screens. This term penalizes an overabundance of defocus phase in each of the phase screens, which otherwise often lead to the afocal tele-

scoping and can often also contribute to the second, more general, failure mode. Specifically, this metric was designed to penalize the relative amount of defocus in the overall estimated phase, as opposed to the absolute amount of defocus phase, ideally creating a penalty term whose weighting parameter, λ_D , would be more independent of changes in the strength of the turbulence The constant λ_D is the weighting coefficient of the penalty term. As shown in Appendix 7.2, the maximum possible value of this penalty term is λ_D, enabling us to have more precise control on the scale of the $p_D(\hat{\mathbf{\Phi}})$ term, avoiding scaling-related algorithm instability resulting from this penalty term. Also of concern when designing this penalty term was that early in each optimization when the total estimated phase in each screen is small, $p_D(\hat{\mathbf{\Phi}})$ becomes unstable due the small denominator and causes an undesirable, out-sized effect on the gradient. A small nonnegative value of the regularization term β helped to ensure a well-behaved gradient early in the optimization process when the accrued phase in each screen is small. Note that $p_D(\hat{\mathbf{\Phi}})$ term is an alternative to the penalty term described in [4].

We know from the similarity theorem in [10] that a decrease in energy distribution extent in the spatial domain will cause an increase in the corresponding distribution width in the Fourier domain. Our second penalty term and fifth and final optimization model component therefore penalizes nonphysical energy distribution outside of the pupil after fields have been corrected by phase screens, and is similar in effect to the penalty term in [4]. This offers a penalty for any type of phase screen arrangements which would

demagnify some or all of the target energy distribution without having to choose explicitly which aberrations to penalize and allowing our optimization model to better represent physical reality:

$$p_C(\hat{\mathbf{\Phi}}) = -\frac{\lambda_C}{\epsilon_{init}} \sum_{\vec{r}_{\perp 0},\nu,\eta} \mathscr{M}(\vec{r}_{\perp 0}) \circ \left| \mathscr{P}_{S\to 0}^{-1} \left\{ U_S(\vec{r}_{\perp S},\nu,\eta;\hat{\mathbf{\Phi}}) \right\} \right|^2, \tag{3.9}$$

where the propagation is without aberrations in intermediate planes and where λ_C is another empirically determined weighting coefficient and

$$\epsilon_{init} = \sum_{\vec{r}_{\perp 0},\nu,\eta} |U_0(\vec{r}_{\perp 0},\nu,\eta)|^2, \tag{3.10}$$

is a normalizing parameter, i.e., the sum of all energy at all frequencies in the aberrated pupil fields. The $\mathscr{M}(\vec{r}_{\perp 0})$ term is pupil mask whose value was unity in most of the region outside of the pupil and zero inside the pupil region. The transition between these regions was a radially symmetric Gaussian roll-off. This roll-off region served two purposes. First, it ensured that an abrupt change from unity to zero resulting from a binary mask did not cause wild swings in the gradient. Second, and most importantly, the roll-off region allowed our model to account for energy which was captured by the pupil that, if not for turbulence, would have in fact fallen outside of our pupil [11]. Thus, the best-case corrected system will have a small but non-trivial amount of energy outside of the pupil region if one were to propagate the corrected fields back to the pupil plane without passing through phase screens. The

unity transmittance region allowed for the strong penalization of solutions that resulted in non-physical amounts of energy outside of the pupil, while the Gaussian roll-off region allowed solutions which caused some energy to occupy the area immediately adjacent to the pupil boundary. The extent of this roll-off region was determined by the change in energy distribution extent in the pupil plane as discussed in [11, Section 9.4].

Finally, we combine Eqs. (3.5), (3.8), and (3.9) to form the complete objective function which we maximized as a function of $\hat{\mathbf{\Phi}}$

$$f(\hat{\mathbf{\Phi}}) = \mathcal{I}(\hat{\mathbf{\Phi}}) - p_C(\hat{\mathbf{\Phi}}) - p_D(\hat{\mathbf{\Phi}}). \tag{3.11}$$

In addition to this optimization model, we have also included the analytic gradient of our optimization model as Appendix 7.1.2 and the algorithmic differentiation methodology used to derive that gradient as Appendix 7.1.1, which we have referenced above.

3.2 Noiseless Simulations

3.2.1 Simulation setup

The encounter geometry for these simulations is shown in Fig 3.1. Once our optimization model and its gradient were established, we conducted a

study of noiseless, aberrated range-compressed fields in order to explore ISM performance under ideal conditions. The study used ISM to correct various uniquely aberrated versions of the same target scene at each of five different D/r_0, where D/r_0 represents the strength of turbulence in the pupil plane resulting from all turbulence along the path of propagation. A target was simulated using a scene containing 3D rendered facet models of four geometric solids resting on a flat, off-angle plane. This target scene was converted into two-dimensional range and reflectivity maps, each with 2320×2320 pixels. Gaussian illumination beams propagations at 32 frequencies were simulated from a transmitter in the pupil plane to the target region 8900 m away. Beams had a base frequency $\nu_0 = 194$ THz (1.545μm) and interfrequency spacing of $\Delta\nu = 23.9$ MHz. After reaching the target region, each propagated Gaussian beam was then multiplied by the target's reflectivity map. These fields then were multiplied by complex circular Gaussian random numbers with zero mean and variance $\sigma^2 = 0.5$ to simulate reflection from an optically rough surface which would in turn produce a speckled im-

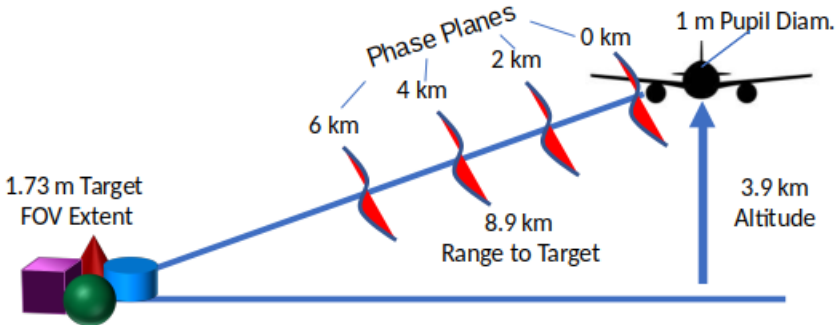

Figure 3.1: Simulation encounter geometry.

age [3]. Note that for all of the 32 illumination frequencies, only a single speckle realization is simulated and therefore no speckle averaging occurred in any of the subsequent optimizations. This is realistic due to the small fractional bandwidth of $\nu/\nu_0 \approx 0.125 \times 10^{-6}$. Following the application of surface roughness, the resulting fields were then multiplied by an array of target depth-dependent phases containing the relative range information about the target scene. We then sinc-downsampled our fields to a size of 580×580 pixels to ensure proper speckle statistics.

The effects of volumetric turbulence on our fields were simulated by propagating to the pupil from the target plane through simulated turbulence screens, in accordance with the plane-to-plane propagation protocol described in Section 3.1. After each of the four partial propagations, our fields were multiplied by a Kolmogorov turbulence screen [12]. Turbulence screens were located at 2 km, 4 km, and 6 km from the pupil, as well as in the pupil plane. Each of the four turbulence screens had the same $D/r_{0,i}$ and yielded D/r_0 values shown in Table 3.1. We characterized the relative severity of anisoplanatism using the ratio

$$\kappa = \frac{\text{AFOV}}{\theta_0}, \tag{3.12}$$

where AFOV is the angular field-of-view of the system and θ_0 is the isoplanatic angle [9]. The κ values for each of our turbulence cases are also shown in Table 3.1. As mentioned in Section 3.1, turbulence causes some photons to

Table 3.1: **Simulated Turbulence Characteristics**

Case	**1**	**2**	**3**	**4**	**5**
$D/r_{0,i}$	5	10	15	20	25
D/r_0	7	14	22	29	36
κ	18	35	54	73	91

be captured by the pupil which would otherwise be outside the spatial bandwidth limit of the system and, conversely, also causes some photons which would have been captured by the pupil to be lost. In order to make sure that higher spatial frequency information was available to be redirected into the pupil as the result of turbulence, simulated fields were propagated through the turbulence screens with a larger physical pupil, 580 pixel diameter, and target field of view, 580 pixel extent, than we used in our optimizations which had 384 pixel pupil diameter and target FOV extents, respectively. Sample sizes, however, were unchanged throughout this process. The pupil had a pixel pitch of 2.6 mm and a final diameter of 1 m, providing a sampling rate of 10 samples per speckle area or about 3.16 samples per speckle in each transverse dimension. The 3D intensity image had a FOV extent of 1.7 m after the aberration process and a cross-range resolution $\rho_x = \rho_y = 1.7$ cm. The range ambiguity interval of the compressed image was 6.27 m. Fields were zero padded in range at $Q = 2$. Range resolution was $\rho_z = 19.6$ cm.

Each of our optimizations was conducted in eight rounds. Unless otherwise noted, optimization phase screens were placed in locations as shown in Fig. 3.1. We used the SciPy implementation of the L-BFGS optimizer [13]

to maximize our objective function Eq. (3.11). Each round of optimization used the method of sieves [7] (MoS) technique which, during the computation of the metric, convolved the point-by-point phase estimate with a Gaussian kernel given by Eq. (3.4). The standard deviations $\{\sigma_s\}$, and therefore the size of the Gaussian kernels, varied by phase plane and round of optimization. Initial MoS σ_s values decreased as proximity to the target increased. The initial MoS kernel σ_s values decreased with each round of optimization. Kernel σ_s values varied each round, as shown in Table 3.2. Empirical metric parameters were set at $\alpha = 0.5$, $\lambda_D = 0.1$, and $\lambda_C = 0.07$. Optimizations used the exact analytic gradients as derived in Appendix 7.1.2. The optimizer was always was always run for 80 iterations. We did not expect local minima to cause difficulty with this algorithm because the MoS has a smoothing effect on the search space, likely eliminating some, if not all, local minima. Best-case stopping points will be discussed in the results section.

Table 3.2: **Method of Sieves σ_s values**

Round	1	2	3	4	5	6	7	8
Screen Loc.	σ (pixels)							
$z = 0$ m (pupil)	16	14	12	10	8	6	4	2
$z = 2000$ m	12	11	9	8	6	6	3	2
$z = 4000$ m	8	7	6	5	5	4	3	2
$z = 6000$ m	4	4	3	3	3	3	2	2

3.2.2 Noiseless Simulation Results

After simulations were conducted in accordance with the procedures outlined in the previous section, we analyzed our results. To quantify our results, we calculated the root-mean-square-error (RMSE) of the range maps created from fields corrected by our algorithm with respect to the best-case range maps. Best-case range maps were constructed by propagating aberrated pupil fields back through the exact phase screens which were used to aberrate them, range-compressing those fields, taking their intensity, and then applying Eq. (2.70) to the 3D intensity array. We report RMSE values in centimeters (cm). Note also that due to range ambiguity wrap-around, each pixel error included in our RMSE values was restricted to a magnitude between 0 and $K/2$, where K is the total number of range bins in our 3D array. Prior to calculation of the RMSE, both aberrated and sharpened range maps were registered to within one-hundredth of a pixel accuracy [14] with respect to the best-case range maps, and shifted with sub-pixel accuracy into best alignment using Fourier linear phase translation. Fig.3.2 shows example images of range map results from our simulation study. Aberrated range maps are shown in the left-most column, range maps generated from single plane corrected fields are in the center-left column, range maps generated from fields corrected using four phase planes are shown in the center-right column, and best-case range maps are displayed in the right-most column. From Fig. 3.2 we first observe that range imaging using RCH, as we have executed it in our study, appears

to be generally somewhat robust to the effects of volumetric turbulence even without sharpening. We attribute this robustness to both the information diversity afforded through the use of many wavelengths of illumination and the localization of perturbations to cross-range dimensions.

As mentioned in Section 3.2.1, our study consisted of optimizations which estimated four phase screens in order to correct for turbulent effects. The results in the center-left column of Fig 3.2, however, are taken from simulations where a single phase plane in the pupil was estimated in order to mitigate turbulent effects. These single plane results were included specifically to display the added utility in optimizing over multiple phase planes, which are able to compensate for the effects of anisoplanatism. Multiple phase plane estimates were also used in [4] for 2D images. However, our work here estimated phase planes as a means of sharpening sparse 3D intensity data, which is new. Both the single plane and the multiple phase plane results are assumed to be optimal after the completion of the seventh round of optimization since they appeared to have the best quality. RMSE range error values in centimeters are listed under each image. Although both single and multiple phase screen estimations showed vast improvement over the original aberrated cases, multiple phase screen cases almost always showed greater improvement than single screen cases. In all but the $D/r_0 = 36$ case, multiple phase screen estimations are almost indistinguishable from best-case images, although RMSE values still show some differences. Note that while RMSE is useful for trending improvement, it is not a clear indicator of overall perceived

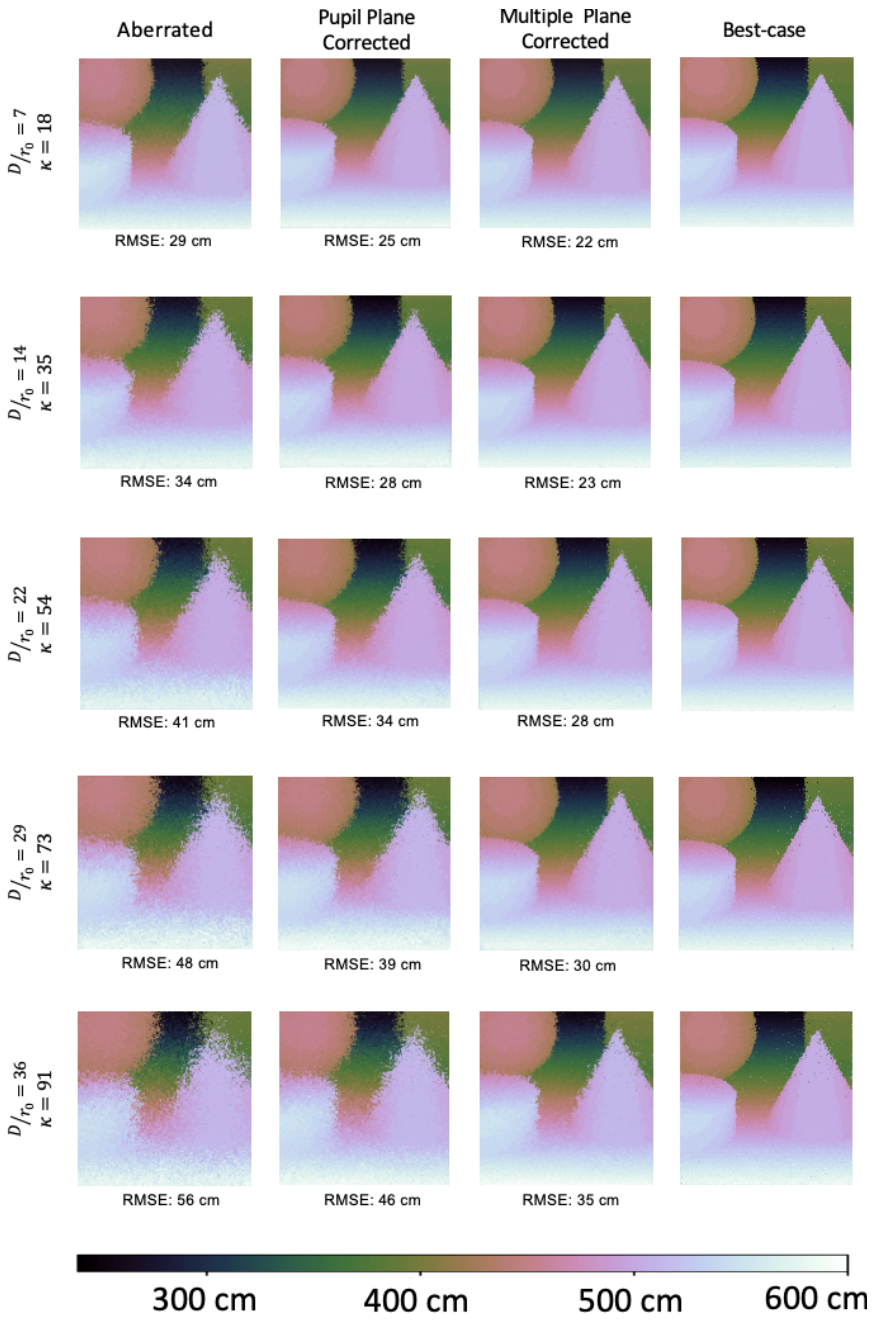

Figure 3.2: Range maps for various strengths of turbulence using no correction (left-most column), single phase plane correction (center-left column), four phase plane correction (center-right column), and best-case exact correction (right-most column).

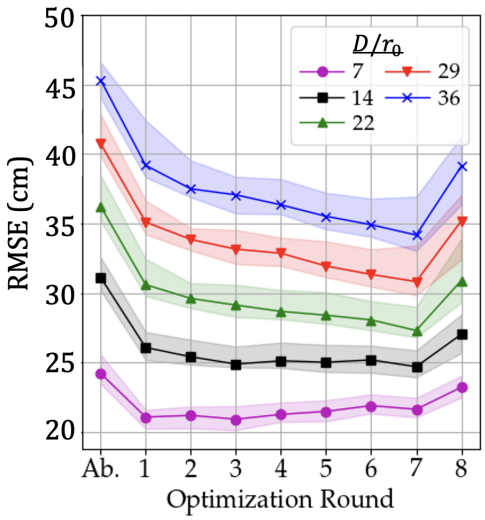

Figure 3.3: Comparison of median RMSE in centimeters. The "Ab." column details the original aberrated RMSE with no correction.

image quality. Fig. 3.3 shows aggregate results of all twenty optimizations at each of the five turbulence strengths. Shown are median RMSE value trends using solid colored lines, while shaded color fills show the 80% confidence interval for all twenty cases. For all turbulence strengths, the greatest single-round improvement in RMSE was seen after completing the first round of optimization. We see that in the weakest turbulence case, $D/r_0 = 7$, best results were achieved after 3 rounds of optimization. In the $D/r_0 = 14$ case, RMSE values stayed relatively steady and at a minimum from round 3 through round 7. Stronger turbulence cases, $D/r_0 = 22, 29,$ and 36, were clearly best in round 7. Uniformly poor performance was observed during round 8 of optimization. We attribute this degradation of quality in round 8 to Gaussian kernel decreasing to a size where the presence of high spatial

frequency phase allowed for over-fitting.

Overall we see that the RMSE range errors observed in the case studies shown in Fig. 3.2 are consistent with the trends we see in Fig. 3.3. We note that our case study at $D/r_0 = 14$ in Fig. 3.2 did fall outside of the 80% confidence interval in Fig. 3.3 at round 7, but is not enough of an outlier to cause concern. From Fig. 3.3 we can conclude that it might have been better to choose the our example result for Fig. 3.2 after round three of optimization for our case study at $D/r_0 = 7$. However, given the high quality result shown for the multiple phase screen case at $D/r_0 = 7$, it is unlikely that we would have been able to visually discern a significant difference between the round seven and the round three corrected range images.

We observe that for all strengths of turbulence, our algorithm is generally effective in improving the quality of aberrated range images for the problem geometry and turbulence profiles used. We do note from the $D/r_0 = 36$ case in Fig. 3.2 that there seemed to some a gradual decrease in the algorithms ability to enhance images as D/r_0 increased (given our current algorithm configuration). We see that even after the multiple plane correction, significant visual differences between the multiple plane correction and best-case correction remain for this large value of D/r_0. Examining our log files for this optimization showed that our objective function values behaved similarly to those in lower turbulence cases, with the net change in objective function value during the tenth iteration of each round having gradually de-

creased to a small fraction of the change in objective function during first iteration. These results only showed evidence of overfitting after the eighth round of optimization. This indicates that the lack of correction was not the result of an insufficient number of optimizer iterations. We postulate that this failure to attain the image quality of the images aberrated by lower turbulence strengths may have been the result MoS kernel size transitions being to coarse from round to round. If this is correct, more total rounds of optimization may be necessary to improve results in higher turbulence cases. It is also possible that, in spite of our best efforts, our penalty terms may need to be adjusted for different strengths of turbulence in order to achieve optimal results. This would also fit our results where the magnitude of the gradient has become very small prior to reaching the point of optimum image correction. Moreover, there is no evidence to support the supposition that sub-optimal algorithm performance in stronger turbulence is caused by only one of these factors. It may very well be caused by a number of different contributing factors.

3.3 Shot-Noise-Limited Simulations

3.3.1 Noise Model

In order to better understand the limitations of the sharpness maximization algorithm as a function of signal-to-noise ratio (SNR), we conducted trials using fields which were aberrated and then were further modified by the application of shot noise. We first derived a noise model where Poisson (shot) noise is the dominant SNR concern in our simulations. A focal plane recording geometry, where the field incident on the detector is the sum of the local oscillator (i.e., the holographic reference beam) and an aberrated image field (the Fourier transform of the fields incident on the system's pupil [15]), was simulated. In the development of our noise model, we drew heavily from noise model development in [8, 16–18]. The local oscillator term $R(j_D, k_D)$, where (j_D, k_D) are discrete detector plane pixel coordinates, was assumed to have much greater energy than the image fields, $U_D(j_D, k_D)$, i.e., $|R(j_D, k_D)|^2 \gg |U_D(j_D, k_D)|^2$.

The initial detection noise model was

$$\begin{aligned} H(j_D, k_D) = \text{Poisson}\left[|R(j_D, k_D) + U_D(j_D, k_D)|^2\right] \\ + \sigma_r \mathscr{N}_{0,1;r}(j_D, k_D) + \mathscr{U}_{0,\sigma_q^2}(j_D, k_D), \end{aligned} \tag{3.13}$$

where $H(j_D, k_D)$ is the hologram intensity in the detector plane, the operator

Poisson $[\cdot]$ represents the Poisson noise operator, σ_r was the RMS read noise, $\mathscr{N}_{0,1;r}(j_D, k_D)$ is the zero-mean, unity-variance random Gaussian variable associated with our read noise, and $\mathscr{U}_{0,\sigma_q^2}(j_D, k_D)$ is a zero-mean, σ_q^2-variance uniformly-distributed random variable representing analog-to-digital quantization error [19]. A unity quantum efficiency was assumed so that expressions in photon and photo-electron units were interchangeable. We thus could consider intensity on the detector in units of photo-electrons (e^-). The strong local oscillator model allowed the Poisson portion of the detected intensity to be split into two additive components, giving [16]

$$\begin{aligned} H(j_D, k_D) = |R(j_D, k_D) + U_D(j_D, k_D)|^2 + h(j_D, k_D) \\ + \sigma_r \mathscr{N}_{0,1;r}(j_D, k_D) + \mathscr{U}_{0,\sigma_q^2}(j_D, k_D) \end{aligned} \tag{3.14}$$

where $\langle H(j_D, k_D)\rangle$ is mean number of photo-electrons in the detector signal and $h(j_D, k_D)$ is an IID zero-mean Gaussian random variable [16, 18]

$$\sigma_s^2 = \langle h^2(j_D, k_D)\rangle = |R(j_D, k_D) + U_D(j_D, k_D)|^2, \tag{3.15}$$

in units of photo-electron. We expand Eq. (3.14)

$$\begin{aligned} H(j_D, k_D) = |R(j_D, k_D)|^2 + |U_D(j_D, k_D)|^2 \\ + 2|R(j_D, k_D)||U_D(j_D, k_D)| \\ \times \cos\left[\phi_U(j_D, k_D) - \phi_R(j_D, k_D)\right] \\ + h(j_D, k_D) + \sigma_r \mathscr{N}_{0,1;r}(j_D, k_D) + \mathscr{U}_{0,\sigma_q^2}(j_D, k_D), \end{aligned} \tag{3.16}$$

where $\phi_U(j_D, k_D)$, $\phi_R(j_D, k_D)$ are the phases of the focused pupil fields and the tilted reference wave, respectively. Additionally, in Eq. (3.16) we assumed the local oscillator $R(j_D, k_D)$ supplied Σ photon-electrons per detector pixel, where $\sqrt{\Sigma} \gg 1$ and also that the pupil field energy on the detector, $|U_D(j_D, k_D)|^2 \ll \sqrt{\Sigma}$, allowing us to assume negligible the $|U_D(j_D, k_D)|^2$ image intensity term and yielding

$$\begin{aligned} H(j_D, k_D) \approx \Sigma &+ 2\sqrt{\Sigma}|U_D(j_D, k_D)| \cos\left[\phi_U(j_D, k_D) - \phi_R(j_D, k_D)\right] \\ &+ \sqrt{\Sigma}\mathscr{N}_{0,1;p}(j_D, k_D) + \sigma_r \mathscr{N}_{0,1;r}(j_D, k_D) \\ &+ \mathscr{U}_{0,\sigma_q^2}(j_D, k_D), \end{aligned} \tag{3.17}$$

where $\mathscr{N}_{0,1;p}(j_D, k_D)$ is zero-mean, unity variance Gaussian random variable associated with our approximated Poisson noise contribution. We considered a notional photodetector whose array consisted of 1536×1536 detector pixels, 16-bit digitization (65536 gray levels), RMS read-noise $\sigma_r \leq 5\ e^-$, and a well-depth $8 \times 10^4\ e^-$. We assumed our local-oscillator-supplied a mean of photo-electrons $\Sigma \approx 6.8 \times 10^4\ e^-$, equivalent to 85% of our detector's well-depth. So $\sigma_r \ll \sqrt{\Sigma}$, allowing us to neglect read noise. The variance of our additive quantization noise model is $\sigma_q^2 \approx (1/12)\text{LSB}^2$, where LSB is the least significant bit of our detector's binary readout number. In photo-electron units then, LSB$= (5 \times 10^4\ e^-)/(65536$ gray levels$) \approx 1.22\ e^-$. Thus, $\mathscr{U}_{0,\sigma_q^2}(j_D, k_D)$ was also negligible because because $\sigma_q \ll \sqrt{\Sigma}$. The final

expression for our noisy detector plane measurement was

$$H(j_D, k_D) \approx \Sigma + 2\sqrt{\Sigma}|U_D(j_D, k_D)| \cos\left[\phi_U(j_D, k_D) - \phi_R(j_D, k_D)\right] + \sqrt{\Sigma}\mathscr{N}_{0,1;p}(j_D, k_D). \tag{3.18}$$

To extract the pupil fields, we Fourier transformed the simulated detector readouts shown in Eq. (3.18) with no additional zero-padding of the array ($Q = 1$). We then isolated the desired pupil term, which had a diameter of 384 pixels, by multiplying the transformed field by an offset binary circular pupil mask. The offset in the pupil data was generated via the tip/tilt term included in our local oscillator that allowed for the separation of the desired pupil term from its twin image and from the DC term. Finally, we translated the pupil field so that the extracted fields were centered at the origin. We move forward ignoring the noise associated with $\sqrt{\Sigma}|U|$, using the justification above. We can see then that all of the noise in our model originates from the Poisson noise term on the right-hand side of Eq. (3.13) whose variance is Σ. The expression for the extracted field was

$$U_{0,E}(j_0, k_0) = \sqrt{\Sigma}P(j_0, k_0)U_0(j_0, k_0) + \sqrt{\Sigma}P(j_0, k_0)\mathscr{N}_{0,1;C}(j_0, k_0), \tag{3.19}$$

where (j_0, k_0) are discrete pupil plane coordinates, $P(j_0, k_0)$ is a binary circular pupil mask representing the aperture and $\mathscr{N}_{0,1;C}(j_0, k_0)$ is an IID zero-mean, unity variance circular complex Gaussian random noise. Hence, the

SNR in the pupil plane over the area of the pupil is then

$$\mathrm{SNR}^2[|U_0(j_0,k_0)|] = \frac{|U_0(j_0,k_0)|^2}{E[|\mathscr{N}_{C,0,1}(j_0,k_0)|^2]} = |U_0(j_0,k_0)|^2, \tag{3.20}$$

the square root of the average number of target photons/photo-electrons per pixel in the pupil. Note that this is the RMS SNR as opposed to the power SNR, which is the square of Eq. (3.20) and is equal to $|U_0(j_0,k_0)|^2$, the number of photons per pixel from the object field in the pupil plane. Since noise on the detector was uniformly distributed IID Gaussian random noise, it followed that noise in the Fourier transform plane was also uniformly distributed and Gaussian. Due to Parseval's theorem, we are only left with 20% of the initial noise on the detector since we window out the pupil in the extraction plane which only accounts for roughly 20% of the total area in the $Q = 1$ extraction plane. This results in compression gain of about $5\times$ in our image. Since there are $\sim$10 pixels per speckle area our simulations, we report shot noise levels in the more fundamental measure of signal photons per speckle.

3.3.2 Simulation with noise and Results

For our study of ISM with noise, we randomly selected five out of the twenty aberrated fields at each of the five turbulence strengths that we simulated in Section 3.2.1and applied noise to their hologram intensities in the detector

plane. Simulated shot noise levels were 10^{-2}, 10^{-1}, $1, 10$, and 10^{2} target photons per speckle.

Comparing the RMS error of corrected range images made from noisy data with respect to the best case noiseless range images would not be the whole story, since the reconstructed range images will be degraded on account of both the measurement noise and the residual turbulence due to the algorithm's inability to perfectly correct for the turbulence.

We wanted to assess how well the algorithm mitigated aberrations for fields exposed to the exact same turbulence, but which contained different levels of shot noise. However, we needed to remove the noise from the comparison to make sure we did not bias our comparison in favor of range images made from fields with lower shot noise. So we used the estimated phase screen results from each level of shot noise and applied them to noiseless fields propagated through identical turbulence.

After noiseless aberrated fields were propagated through our phase screen estimates to the target plane, they were range-compressed and range maps were generated. These range maps were registered to a within a hundredth of a pixel accuracy with their respective best-case range maps. Again, best-case range maps were created from the exact phase screens used to aberrate the fields. The center 384×384 pixel square (the image region) of each range map was cropped out. The RMSE between these cropped regions was calculated.

Our noisy simulation results are shown in Fig. 3.4. Best performance

in noisy cases was uniformly seen in the same round of optimization as with their noiseless counterparts shown in Fig. 3.3. We again plotted the median RMSE versus SNR for each of the five turbulence-strength profiles as the bold line. The shaded color fills again showed the 80% confidence interval of the RMSE values. We see that our algorithm is generally unable to make improvements with a signal strength of 10^{-2} target photons per speckle; the results are near identical to the uncorrected case. This meant there were $\sim$150 target photons per detector integration over the entire array, so it is not surprising given most pixels had no photons at all. But it does create a worst-case scenario baseline for comparison purposes. Median RMSE and confidence intervals at this level of shot noise are very similar to the aberrated cases with no correction. These estimated phase screens accrued very little phase at all. The 10^{-1} target photons per speckle level shows the algorithm made significant corrections but that the algorithm's effectiveness was limited due to SNR. At an average of one target photon per speckle, we see RMSE values are within $\sim$5% of noiseless (infinite SNR) RMSE cases shown on the far right gridline of our plot except in the $D/r_0 = 36$ case. At the 10 and 10^2 target photon per speckle levels, we have achieved similar RMSE to the noiseless case for all turbulence strengths.

In summary, we can clearly see that performance improves as a function of SNR. For the turbulence strengths and noise profiles studied here, between 1 and 10 target photons per speckle appear to be needed in order to achieve optimal results, and higher numbers of target photons yield little additional

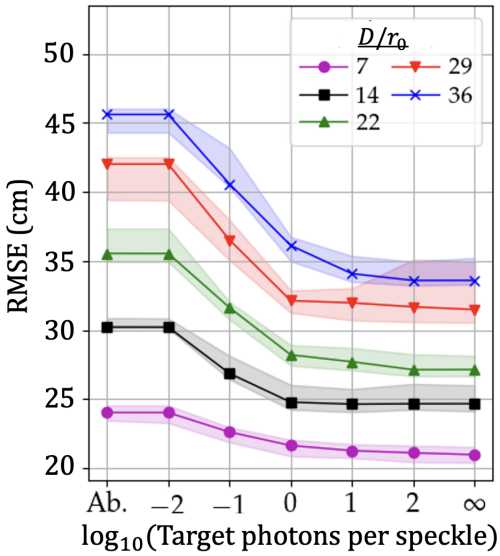

Figure 3.4: Comparison of median RMSE for shot noise simulations. The "Ab." label in the far left column represents the original aberrated RMSE values with infinite SNR and no correction. The right-most horizontal axis column denoted by "∞" represents the noiseless ISM corrected results from Section 3.2.

benefit.

Bibliography

[1] R. A. Muller and A. Buffington, "Real-time correction of atmospherically degraded telescope images through image sharpening," J. Opt. Soc. Am. **64**, 1200–1210 (1974).

[2] J. R. Fienup and J. J. Miller, "Aberration correction by maximizing generalized sharpness metrics," J. Opt. Soc. Am. A **20**, 609–620 (2003).

[3] S. T. Thurman and J. R. Fienup, "Phase-error correction in digital holography," J. Opt. Soc. Am. A. **25**, 983–994 (2008).

[4] A. E. Tippie and J. R. Fienup, "Multiple-plane anisoplanatic phase correction in a laboratory digital holography experiment," Opt. Lett. **35**, 3291–3293 (2010).

[5] W. E. Farriss, J. R. Fienup, J. W. Stafford, and N. J. Miller, "Sharpness-based correction methods in holographic aperture ladar (HAL)," Proc. SPIE **10772**, 107720K (2018).

[6] A. E. Tippie, A. Kumar, and J. R. Fienup, "High-resolution synthetic-aperture digital holography with digital phase and pupil correction," Opt. Express **19**, 12027–12038 (2011).

[7] U. Grenander, *Abstract inference* (Taylor & Francis, 1981).

[8] A. E. Tippie, *Aberration correction in digital holography* (PhD. thesis, University of Rochester, 2012).

[9] M. C. Roggemann and B. M. Welsh, *Imaging through turbulence* (CRC press, 2018).

[10] J. W. Goodman, *Introduction to Fourier optics, 4th Ed.* (W. H. Freeman, 2017).

[11] J. D. Schmidt, *Numerical simulation of optical wave propagation with examples in MATLAB* (SPIE, 2010).

[12] R. G. Lane, A. Glindemann, and J. C. Dainty, "Simulation of a Kolmogorov phase screen," Waves Random Media **2**, 209–224 (1992).

[13] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, "Scipy 1.0: Fundamental algorithms for scientific computing in Python," Nat. Methods **17**, 261–272 (2020).

[14] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," Opt. Lett. **33**, 156–158 (2008).

[15] J. W. Stafford, B. D. Duncan, and M. P. Dierking, "Experimental demonstration of a stripmap holographic aperture ladar system," Appl. Opt. **49**, 2262–2270 (2010).

[16] F. Verpillat, F. Joud, M. Atlan, and M. Gross, "Digital holography at shot noise level," J. Disp. Technol. **6**, 455–464 (2010).

[17] Y.-q. Li, D. Guzun, and M. Xiao, "Sub-shot-noise-limited optical heterodyne detection using an amplitude-squeezed local oscillator," Phys. Rev. Lett. **82**, 5225 (1999).

[18] R. W. Boyd, *Radiometry and the detection of optical radiation* (Wiley & Sons, 1983).

[19] R. M. Gray and D. L. Neuhoff, "Quantization," IEEE T. Inform. Theory **44**, 2325–2383 (1998).

Chapter 4

Experimental Turbulence Mitigation in Range-compressed Holography

4.1 RCH Lab Apparatus

4.1.1 System Layout

In order to test our ISM turbulence correction algorithm described in Chapter 3 on laboratory data, we constructed a RCH system in which 2 or 3 Lexitek psuedo-random turbulence phase plates were placed along the path of optical propagation between the target and the pupil. A diagram and photo of our system are shown in Figs. 4.1 and 4.2, respectively. The system's flood illumination transmitter (TX) was located 5 inches from the system pupil in the same transverse plane. This transmitter offset allowed for target illumination which did not pass through the phase screens on its way to the

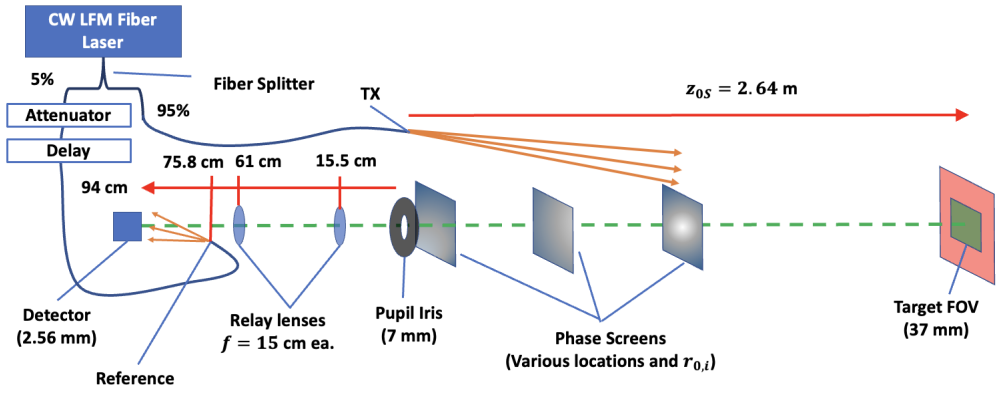

Figure 4.1: Laboratory RCH system diagram

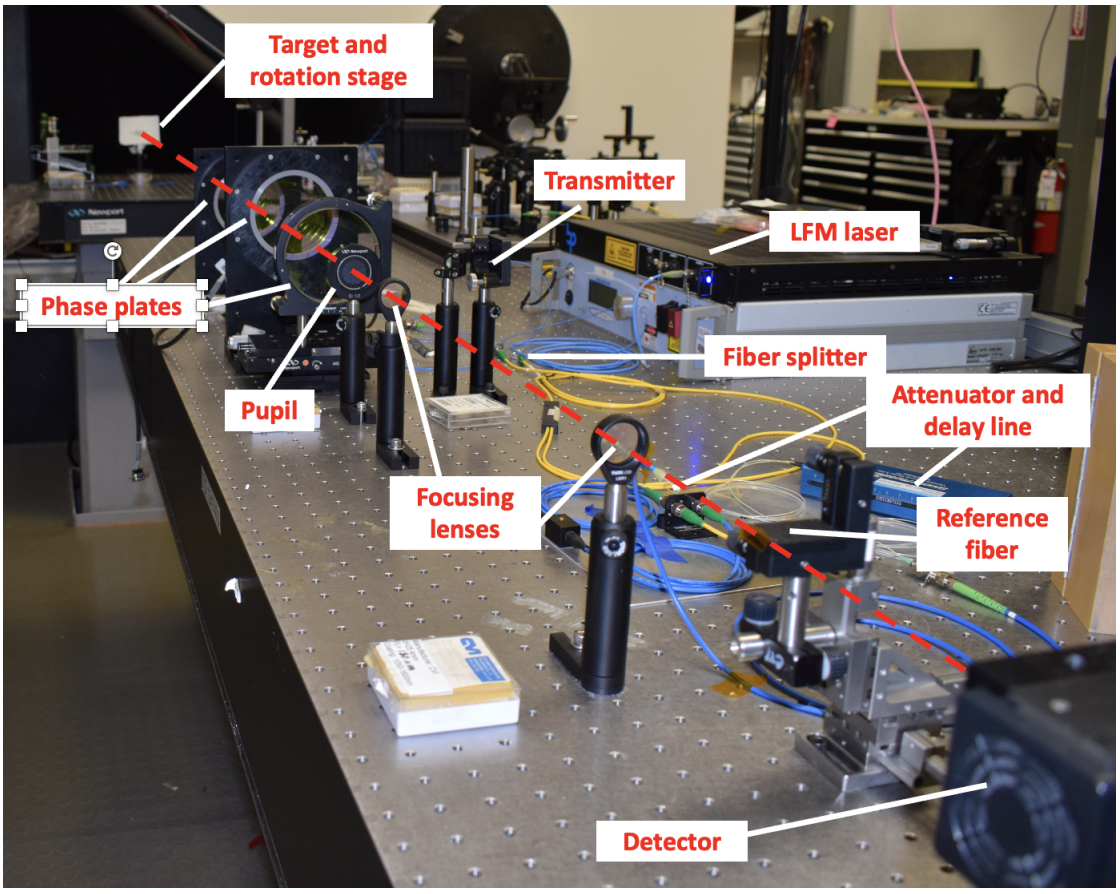

Figure 4.2: Laboratory RCH system setup.

target. Thus, our target data exhibited the effects of phase aberrations while avoiding the effects of log-amplitude (scintillation) aberrations, as mentioned in Section 2.1.5. The transmitter was placed as close as possible to the pupil without the transmitter beam being clipped by the mounts holding our phase plates. This, along with a convenient target orientation, helped minimize the amount of shadowing resulting from our bi-static configuration.

Our target was located 2.64 m from the pupil on a precision rotation stage. The rotation stage allowed us to capture multiple independent speckle realizations of target without having to manually re-position the target between each captured frame of data. After the flood illumination was scattered by the target, some of the illumination passed through the phase plates and was collected by our receiver optics.

The Lexitek phase plates were constructed from N-BK7 glass with an index of refraction equal to 1.5 at a frequency of 1.55 μm and a thickness of approximately 2 cm. This meant with two phase screens, the effective propagation distance from pupil to target increased to 2.66 m, and with three phase screens the effective propagation distance became 2.67 m. Placement of our phase screens was constrained to the region closer to the pupil in order to ensure our transmit illumination was not clipped by the phase screen mounts.

The first element in the receiver imaging system was an iris diaphragm that formed the entrance pupil of our receiver optical system. The iris was followed by two biconvex relay lenses, 15.5 cm and 61 cm from the iris,

respectively. Each lens had a focal length of 15 cm and they were positioned to mimic the effect of a single focusing lens with an 18.2 cm focal length placed directly behind the pupil iris.

The reference beam transmitter was positioned 75.8 cm behind the iris and, like the flood illuminator. The reference beam transmitter also possessed the necessary transverse offset and tip/tilt orientation to separate the pupil data from its complex conjugate during pupil field extraction, as described in Chapter 2 Eqs. (2.33) and (2.37) and shown in Fig. 2.4. The reference beam and focused pupil fields were interfered on the detector array, which was placed 94 cm behind the pupil. The reference beam transmitter's exact location relative to the detector was calibrated to match the quadratic wavefront curvature of the focused pupil fields and thereby avoid unwanted defocusing of our extracted pupil data.

4.1.2 System Hardware

We used a custom Bridger Photonics fiber laser which provided a linearly frequency modulated (LFM) chirped waveform with a center wavelength, $\lambda = 1.55\mu$m. This laser supplied the spectral bandwidth necessary to conduct RCH imaging. The rising sawtooth frequency chirp waveform had a slope, $\beta/\pi = 117.89$ THz/s and a 500 Hz repetition rate, where β is the chirp rate [1]. Due to modulation instability at the sawtooth resets, the frequency modulation was only linear over approximately 80% of the period of the chirp.

While the laser was continuous-wave (CW), it exhibited a marked decrease in output power during each linear frequency chirp. By time averaging over many chirps, the laser power was measured at approximately 150 mW.

The linearly polarized fiber laser output was delivered via single-mode, polarization-maintaining (PM) fiber, to a fiber splitter. In total, 95% of the input laser power was directed through PM fiber to our flood illumination transmitter where a fiber collimator reduced the divergence angle of our Gaussian beam to roughly 20 mrads.

The remaining 5% of the laser power from the beamsplitter acted as the reference beam (local oscillator). The reference passed through both a variable fiber attenuator (in order to set the optical power ratio of the reference and target return beams) and variable fiber delay line (in order to match OPL with the received signal). The cleaved end of the reference beam fiber was placed in the plane of the exit pupil. The reference fiber was displaced from the optic axis, in order to introduce a tip/tilt aberration which allowed for the separation of the pupil field distributions in the extraction plane.

The target object, shown in Fig. 4.3, was composed of 2 metal lapel pins and a metal badge stacked on top of one another, epoxied to one another, and then secured to a thick card stock backing. It was painted with matte white paint to create surfaces which exhibited a close approximation to ideal Lambertian reflectivity and which possessed surface roughness necessary to

Figure 4.3: Photo of laboratory target.

generate speckled images. The target was then affixed to a post mounted on our precision rotation stage. Notice first the most prominent feature in our target, the letters "US" in the foreground, as they will be very important target features to look for in our images. Additionally, the sword tip, handle, and flower petals of the lapel pin behind the "US" insignia are also important features to note in our target. Finally, on the last item in our target, pay attention to the general support of the wreath and the gaps between the wreath and the bayonet where illumination will pass through and scatter off of the thick card stock backing. This target was constructed because of the complex features of various scales and orientations and at various ranges inside of our range ambiguity.

As stated in the last subsection, in order to simulate atmospheric turbulence in our lab data, one or more Lexitek pseudo-random phase plates were placed between the target and the pupil. These phase plates possessed random optical-path-length (OPL) across the transverse plane with spatial

frequency content that approximated a modified von-Karman spectrum[2]. The strength of each phase plate was characterized by its coherence length parameter, $r_{0,\mathrm{i}}$, discussed in Chapter 2.

Our detector was a 640×512 pixel InGaAs photodetector with 20 micron pixel pitch and near unity fill factor. This camera was based on a Xenics Cheetah camera. With the camera windowed to a 128×128 region of interest (ROI), the frame read out time was about 42.8 μs. With the integration time set at 18μs, the measured frame rate was approximately $1/(42.8\ \mu\mathrm{s} + 18\ \mu\mathrm{s}) = 16.45$ kHz. By further reducing the effective pixel count or integration time, higher frame rates would be possible. The custom read-out acquired a burst sequence of frames at 8 bit resolution and stored them in DRAM, limiting the burst sequence to about 8000 frames at the 128×128 format. As we will see, this camera was extremely noisy, which was a reasonable trade-off to achieve our desired frame rate.

4.1.3 Calibration and Testing

The image plane at 2.64 m was located by placing a point object in the target plane and dithering the axial location of the detector array to minimize the point distribution's diameter. The extent of the detector array as well as the choice of imaging lens focal length determined the 37 mm target FOV while the 7 mm pupil regulated spatial frequency content of the image and therefore controlled the transverse resolution of system, $\rho_x = \rho_y = 578\ \mu$m

in object space, using the Rayleigh criterion.

The unsynchronized camera and laser were observed to operate at stable repetition rates. We recorded roughly 32.8 frames per chirp period while the laser was free-running and the camera was recording at 16.45 kHz. The frequency step size between frames was $\Delta\nu \approx 7.17$ GHz. Only the frames from the linear portion of the chirp were retained for our holograms. The number of recorded frequencies was 26. The range ambiguity and range resolution, from Eqs. (2.69) and (2.67), respectively, were $R_{amb} = 20.9$ mm and $\rho_z = 0.8$ mm.

Fig. 4.4 (a) shows an interference intensity recorded during calibration and testing. The target in this case was a highly reflective featureless Spectralon panel. The reference field intensity on the detector plane had a Gaussian spatial profile whose $1/e^2$ diameter was roughly 36 mm compared to the 2.56 mm detector plane extent, making for a nearly uniform reference intensity. Fig. 4.4 (a) displays the large, spatially varying noise profile of our camera. In order to help mitigate this detector noise, the target signal was blocked and reference-only data frames were recorded. A mean reference-only frame was generated, averaging over multiple frames to reduce shot noise. This mean reference-only frame was then subtracted from the noisy interference intensity, which is shown in Fig. 4.4 (a), resulting in the intensity pattern shown in 4.4 (b). This process removed much of the spatially varying camera noise. In the 4.4 (d) the Fourier transform of the interference

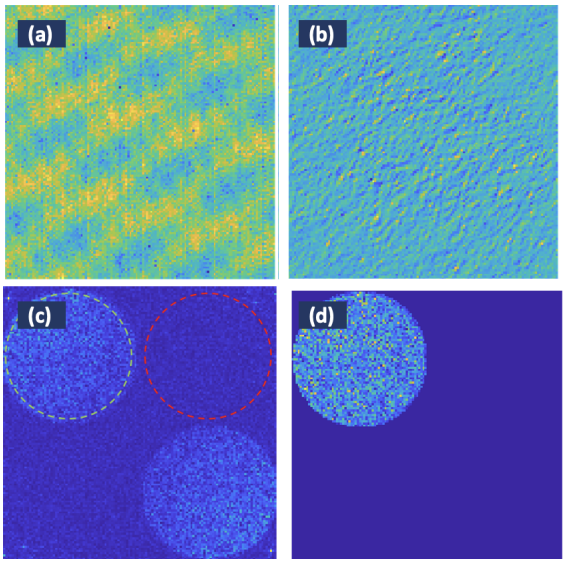

Figure 4.4: (a) The raw hologram intensity distribution on the detector. (b) Hologram intensity after reference beam subtraction (detector intensity shown). (c) Discrete Fourier transform of (b); green and red circled regions used in the calculation of carrier-to-noise ratio. (d) Extracted pupil field (intensity shown).

pattern is shown. As expected, the autocorrelation term fell below the noise floor due to the strong reference beam. The lack of autocorrelation term can be observed in Fig. 4.4 (c).

The fidelity of this coherent field measurement depended on the signal-to-noise ratio (SNR). The SNR was experimentally estimated by computing the mean power within the pupil (shown by the green circle in Fig. 4.4 (c)) and outside the pupil (shown by the red circle 4.4 (c)). Assuming the shot noise of the reference field intensity was uniformly distributed over Fourier space, the experimentally determined SNR was

$$\mathrm{SNR} = \frac{\frac{1}{N_g}\sum_{\mathrm{i}} I_g(x_{\mathrm{i}}, y_{\mathrm{i}})}{\frac{1}{N_r}\sum_{\mathrm{j}} I_r(x_{\mathrm{j}}, y_{\mathrm{j}})} - 1, \tag{4.1}$$

where N_g and N_r are the total number of pixels in the green and red circles of Fig. 4.4 (c), respectively, and $I_g(x_{\mathrm{i}}, y_{\mathrm{i}})$ and $I_r(x_{\mathrm{j}}, y_{\mathrm{j}})$ are the intensities of the pixels in the same green and red circles, respectively. Much of the noise present throughout pupil extraction plane is suppressed by the applying an offeset pupil shaped mask to the data in the extraction plane, leaving the desired pupil field with noise superimposed on it, as seen in Fig. 4.4 (d).

The quality of our images depended on adequate and consistent fringe contrast in our holograms at each frequency. This required an accurate OPL match between the reference and transmitter/object legs. This was accomplished through variable delay in the reference beam. The fiber delay line consisted of an input and an output fiber coupling, each with a fiber collima-

tor, and an optical trombone (a retroreflector on a 1D translation stage). A change in the relative position of the retroreflector either added or subtracted OPL to or from our reference beam.

In order to collect sufficient signal photons, the camera had to integrate over an appreciable period of time. Whereas mismatches in path length in interferometric systems typically results in lower contrast fringes which are constant in time, significant path length mismatch in our system resulted in temporally oscillating intensity measurements. These oscillations occurred as a result of our laser and the camera not being synchronized, i.e. we collected approximately 32.8 frames per chirp. This meant that the portion of the chirp interference integrated by each camera frame changed in time, and thus different time windows had different intensities. As path mismatch becomes worse, we have consistently low contrast which oscillates between peaks and nulls as a function of OPD. If our camera had been synchronized to our detector, measuring the same portions of each chirp in time, intensities would have varied frame to frame and been low generally due to poor contrast, but they would have been temporally consistent from chirp to chirp. It is likely that there were some temporal oscillations in our path matched data due to synchronization issues. They were likely just not noticeable/significant.

On a related note, our camera's finite integration period temporally averaged the interference intensity resulting in a loss of mixing efficiency, i.e. a loss in fringe contrast. The fiber delay line was adjusted prior to recording

each set of measurements in order to maximize the experimentally measurable signal-to-noise ratio (SNR). The temporal mixing efficiency, η_{temp}, is expressed analytically as [3, 4]

$$\eta_{temp} = \frac{T_{\text{int}}}{T_{\text{frame}}} \operatorname{sinc}^2\left(\frac{2\beta_{\text{chirp}} T_{\text{int}}}{c} \cdot \text{OPD}\right), \tag{4.2}$$

where T_{int} represents the time span of the time window when the detector is collecting photons, and T_{frame} is the total time the frame lasts, such that $T_{\text{int}} < T_{\text{frame}}$.

Ideally, SNR should be proportional to η_{temp}. So as a check on the calibration of our system, we verified that our scaled SNR values showed good agreement with the theoretical temporal mixing efficiency. We measured SNR as a function of range offset for the case $T_{\text{int}} = 28$ μs, $T_{\text{frame}} = 70.8$ μs, and the case $T_{\text{int}} = 50$ μs, $T_{\text{frame}} = 92.8$ μs, and plotted these data points against theoretical values. The SNR values were scaled so that the peak SNR matched the theoretical mixing efficiency. This comparison of scaled SNR and η_{temp} are shown in Fig. 4.5. Target data was typically captured with the camera integration time set at 18 μsec, which allowed us to collect a sufficient number of wavelengths per chirp for a range ambiguity that exceeds the depth of our target, even at off-angles. The short integration time and high camera read noise resulted in low SNR images. To improve SNR, we coherently summed complex-valued fields from multiple chirps. This coherent frame stacking required target return fields which did not change between chirps.

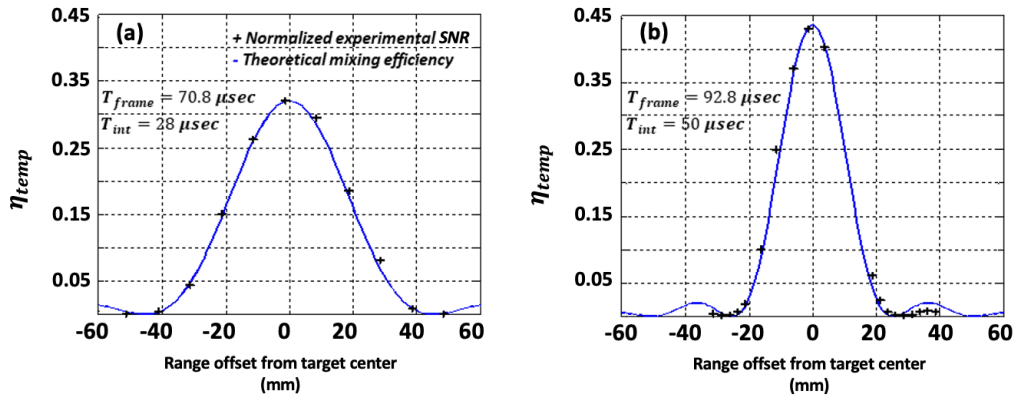

Figure 4.5: Theoretical temporal mixing efficiency η_{temp} plotted against scaled experimental SNR as function of range offset for (a) $T_{\text{int}} = 28$ μs, $T_{\text{frame}} = 70.8$ μs, and (b) $T_{\text{int}} = 50$ μs, $T_{\text{frame}} = 92.8$ μs.

This required a static target, transceiver, propagation medium, and waveform between chirps. Typically, these are not advantages that one would have outside of the laboratory. In our lab system, the target, transceiver, and propagation medium variations were not significant from chirp to chirp. Recorded chirp waveforms, however, differed primarily because the camera was not synchronized to the laser. This chirp-to-chirp difference manifested itself primarily as a spatially uniform piston phase difference between the recorded fields. The mean phase difference between chirps was measured and removed prior to coherent field summation. The measured SNR improvement that resulted from coherent frame stacking is evident in Fig. 4.6. The exact number of stacked frames we used to achieve good SNR was generally between 10 and 20. Our images were created from pupil plane data in accordance with the SAS propagation, range-compression, and range image

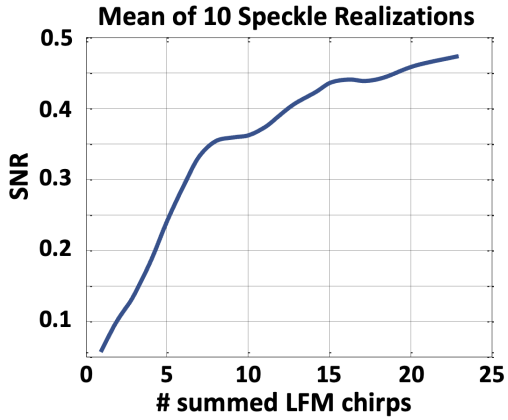

Figure 4.6: SNR as a function of number of coherently summed measurement frames.

generation methods outlined in Chapter 2. The extracted pupil data had a diameter of 64 pixels and required a total padded array size of 256×256 pixels in order to avoid aliasing the innermost quadratic phase in the first partial propagation's quadratic phase, shown in Eq. (2.7). The output fields of the SAS of the pupil in the target plane were 256×256 pixel arrays. Our image plane recording geometry dictated that our target FOV was 128 samples across, so the SAS output images were effectively upsampled by a factor of two. Accordingly, output images shown in this chapter have been downsampled to 128×128 transverse pixels. To enhance range image precision, range images are produced from fields which were zero padded at Q = 20 along the range dimension prior to range-compression.

All energy-based images that follow in this chapter are normalized to the brightest pixel in the frame. The energy-based images shown in this chapter

are RMS amplitudes, $A_{RMS}(x,y) = \sqrt{\sum_{\mathrm{i,j}} |F_{\mathrm{ij}}(x,y)|^2/(IJ)}$, where i denotes a discrete frequency of illumination, j denotes a speckle realization, I is the total number of discrete frequencies, and J is the total number of speckle realizations.

Note, though, that if one only desires to view the unaberrated intensity image(s) resulting from pupil data and did not want to propagate from plane-to-plane, an FFT of the pupil plane data will provide a near ideal intensity image of the target. From this intensity image a near optimal range image can be generated. The SAS was used here exclusively as a means to conduct plane-to-plane propagations between the target and pupil.

Experimental range precision was computed by averaging range profiles over a relatively flat target area. Fig. 4.7 (b) is the mean range profile spatially averaged over the red rectangle area in the amplitude and range images, Figs. 4.7 (a) and (b), respectively. The mean range profile peak exhibits the expected 0.8 mm full-width-half-maximum (FWHM).

Once the SNR of the coherent field measurements was sufficient, the three-dimensional data quality was improved by capturing target data with independent speckle realizations.

To image our target shown in Fig. 4.3, we collected 26 different measurements (counting each coherent sum across numerous chirps as one measurement), one for each frequency. This constituted one speckle realization of our 3D data. After a speckle realization was collected, the precision stage

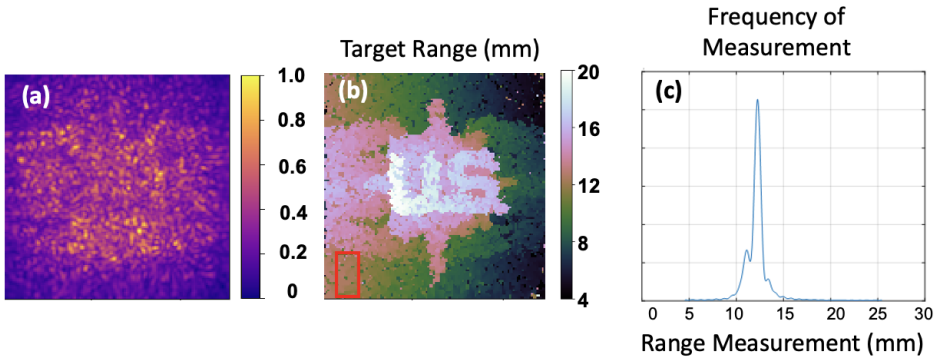

Figure 4.7: (a) Single speckle realization $A_{RMS}(x, y)$ averaged over 26 frequencies and one speckle realization. (b) Range map created from range-compressed intensity containing 26 frequencies and a single speckle realization. (c) Frequency plot of range values over the flat target area within the red rectangle showing ~ 0.8 mm FWHM.

rotated the target slightly such that the target reflected speckle field swept through one full diameter of the entrance pupil, allowing the system to collect an independent speckle realization data cube. This process was repeated, obtaining ten total independent speckle realizations. For our system parameters, we could capture at least ten independent speckle realizations with single-axis target rotations without significantly altering the range or transverse measurement of the points on the rotating edge of our target FOV. The improvement in the target intensity data fidelity, and the resulting range image, from the incoherent summation of ten independent speckle realizations is evident by comparing Fig. 4.7 (a) and (b) with Fig. 4.8 (a) and (b).

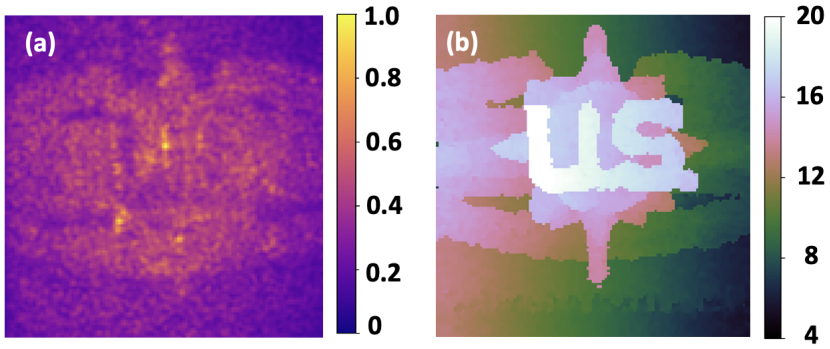

Figure 4.8: (a) Unaberrated $A_{RMS}(x, y)$ averaged over 10 speckle realization and 26 discrete frequencies. (b) Unaberrated range map created from incoherent sum of 10 range-compressed speckle intensities.

4.2 ISM Turbulence Mitigation

In this section we discuss our employment of the ISM algorithm, described in Section 3.1, to two RCH datasets aberrated using Lexitek phase plates. The first aberrated dataset had the weaker of the two turbulence profiles. It used two total phase plates, the first of which had an $r_0 = 1$ mm placed as close to the pupil plane as possible (~ 4 cm from the pupil) and a second with $r_0 = 2.6$ mm placed 0.74 m from the pupil. This resulted in an overall turbulence profile with a $D/r_0 = 7$ and a $\kappa = 20$, which we consider a mild to moderate turbulence profile with regard to coherent imaging. For each speckle realization we collected 12 pupil fields at each of 26 discrete frequencies, coherently summing them together after removing the global piston phase term from each coherent pupil field. In total, ten speckle realizations were collected. The aberrated amplitude and range images for this case are

shown in Figs. 4.9 (a) and 4.10 (a), respectively. We see that the general outline of the target can still be observed in amplitude, but it is difficult to discern any additional target details using only this image. Though one can still recognize the target in the aberrated range image, it is significantly blurred. Additionally, our κ value indicates that there are likely significant anisoplanatic effects present in the aberrated images.

The optimization used the objective function defined by Eq. (3.11) containing an exponential sharpness metric [5] with a coefficient $\alpha = 1.1$ as discussed in Chapter 3, a defocus penalty term, with $\lambda_D = 0.1$ and a pupil containment penalty term, also with $\lambda_C = 0.1$. The pupil containment penalty term also used a mask with diameter of 70 pixels where the mask's value was zero, after this 70 pixel diameter, mask values increased gradually through a soft Gaussian boundary region until reaching a value of unity at a diameter of approximately 76 pixels.

We estimated phase screens at two locations, one coincident with the pupil, the approximate location of our first phase screen, and one at 0.74 m, the location of our second phase screen. We estimated point-by-point phase screens, using the method of sieves [6] (MoS) (described in Chapter 3) on the second phase screen located at 0.74 m. The first phase screen was an unaltered point-by-point estimate during the entire optimization. The MoS standard deviation for the second screen, whose size regulated the spatial frequency content of the phase accrued during each round, started at 3 pixels

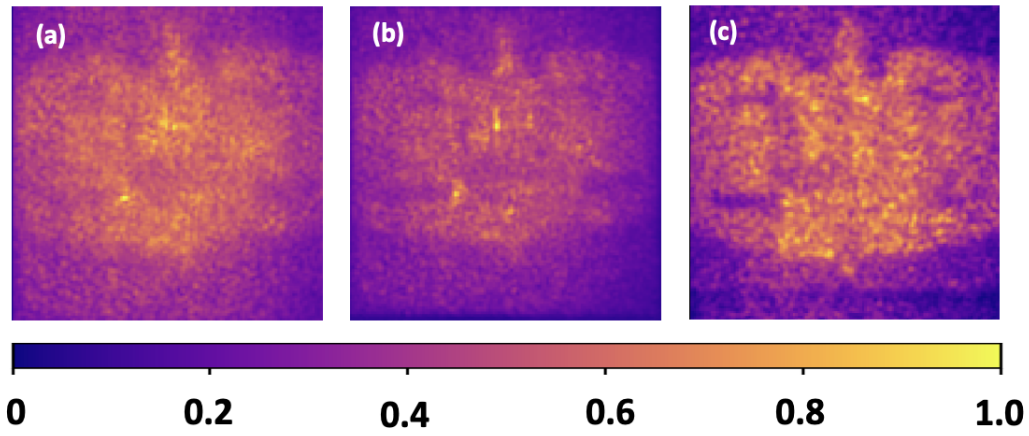

Figure 4.9: Turbulence case 1: $D/r_0 = 7$, $\kappa = 20$ speckle and frequency averaged amplitudes where (a) is the aberrated amplitude, (b) is the sharpened amplitude, and (c) is an unaberrated amplitude from a very similar data collection. Amplitude colormap scaled to brightest pixel.

and decreased each round by 0.5 pixels, resulting in a final standard deviation of 0.5 pixels during the 6th and final round of optimization.

The first three rounds of optimization consisted of six optimizer iterations each and the latter three rounds each consisted of ten optimizer iterations. The change in the number of iterations was based on our observation that when MoS kernel sizes were larger, fewer optimizer iterations were necessary to fit the majority of the useful phase to our screen, i.e. phase which sharpens the image and improves image quality. Fitting this higher spatial frequency content required more optimizer iterations than were needed when kernels were comparatively large. Our results for the weak turbulence case are shown in Figs. 4.9 and 4.10, where images in the former are amplitudes

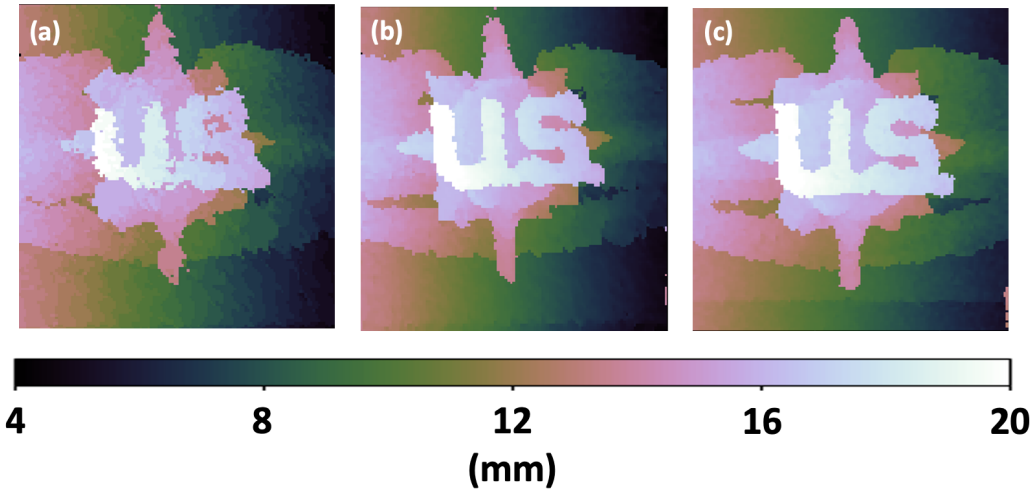

Figure 4.10: Turbulence case 1: $D/r_0 = 7$, $\kappa = 20$ range images where (a) is the aberrated amplitude, (b) is the sharpened amplitude, and (c) is an unaberrated amplitude from a very similar data collection.

downsampled by a factor of 2 incoherently summed over all ten speckle realizations and 26 frequencies in range-compressed space, and the latter are range images created from incoherently summed intensities whose transverse information has been downsampled by a factor of two and sinc upsampled by a factor of 20 in range. Image (a) in both figures is the uncorrected aberrated image of each respective type, amplitude and range. Image (b) in both figures has been corrected by our ISM algorithm. Image (c), which was obtained from unaberrated data, has been included for comparison purposes. Note that unaberrated images are not the exact same image, so differences unrelated to turbulence are present between (b) and (c) images. For example, Figs. 4.9 (b) and 4.10 (b) have a small but noticeable counterclockwise rotation relative to their corresponding (c) images. The result of our opti-

mization in Figs. 4.9 (b) and 4.10 (b) are near ideal. We also observe a small, yet perceivable demagnification in the sharpened images. In the Fig 4.9, we observe a consistency between bright pixels in the aberrated, sharpened, and ideal images. In Fig. 4.10 (b), one small recess near the top left of the U in US has not been fully recovered.

Our second aberrated dataset, which had the stronger of the two turbulence profiles, had a phase screen with $r_0 = 0.6$ mm placed as close to the pupil plane as possible, a phase screen with $r_0 = 1$ mm placed 0.23 m from the pupil, and a third phase screen with $r_0 = 1$ mm placed 0.6 m from the pupil. This resulted in an overall turbulence profile with a $D/r_0 = 16$ in the pupil and a $\kappa = 42$, which as we will see is a severe turbulence profile with significant anisoplanatism. This dataset consisted of 18 coherently summed pupil fields with piston offset removed. We coherently summed a larger number of pupil fields than the previous case in hopes that a higher SNR would aid in our efforts to sharpen the aberrated image, given the increased severity of the turbulence applied. Again 26 discrete frequencies and 10 speckle realizations were used. The aberrated amplitude and range images for this case are shown in Figs. 4.11 (a) and 4.12 (a), respectively. We see that in both amplitude and range the target is unrecognizable, the former being indistinguishable from a roughly square illumination beam incident on a flat (but still optically rough) background.

Again, an exponential sharpness metric with a coefficient of $\alpha = 1.1$ was

used, along with a defocus penalty term with coefficient $\lambda_D = 0.5$, and pupil containment penalty term, with coefficient $\lambda_C = 0.5$. The pupil containment penalty term also used a mask with diameter of 70 pixels with a Gaussian boundary thereafter extending to a diameter of approximately 76 pixels.

Estimated screens were located in the pupil plane, at a distance 0.23 m from the pupil, and at a distance of 0.6 m from the pupil. We executed a total of 10 rounds of optimization in this case. The number of optimizer iterations during each round was 5, 6, 10, 10, 10, 10, 20, 50, 50, and 50 iterations, in that order. All three screens used the MoS. Each screen started with a MoS kernel standard deviation of 6 pixels, which decreased by just over 0.61 pixels per round of optimization, with all screens having a kernel standard deviation of 0.5 pixels in the final round of optimization. The optimization had a total of 6 rounds.

Our ISM results for the severe turbulence case are shown in Figs. 4.9 (b) and 4.10 (b). Again, image (a) in both figures is the uncorrected aberrated image of each respective type, amplitude and range and image (c) is an ideal, unaberrated image included for comparison. Whereas our results in this case are not pristine corrections of the target, they show very profound improvement over the aberrated image. There is some minimal warping in the sharpened images, though it is not severe enough to affect the identification of the target. We see in the sharpened amplitude image that the support of the target is largely intact and is very similar to the support of the unaberrated

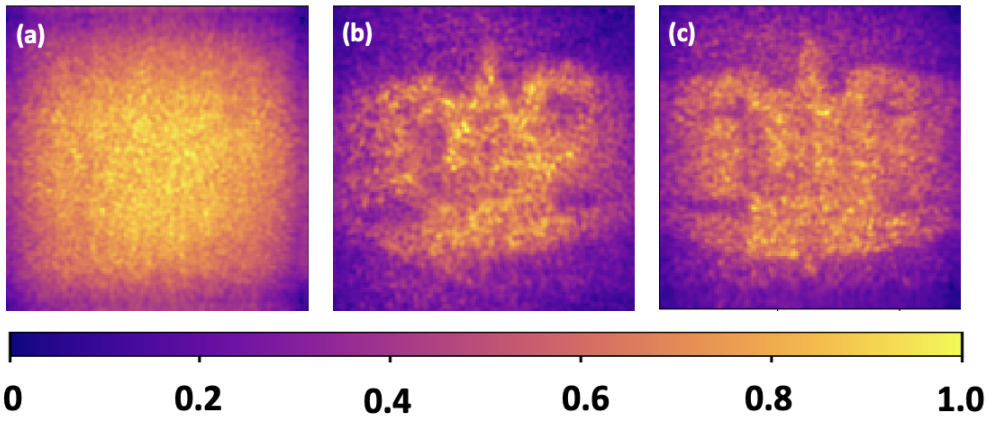

Figure 4.11: Turbulence case 2: $D/r_0 = 16$, $\kappa = 42$ speckle and frequency averaged amplitudes where (a) is the aberrated amplitude, (b) is the sharpened amplitude, and (c) is an unaberrated amplitude from a very similar data collection. Colormap scaled to brightest pixel.

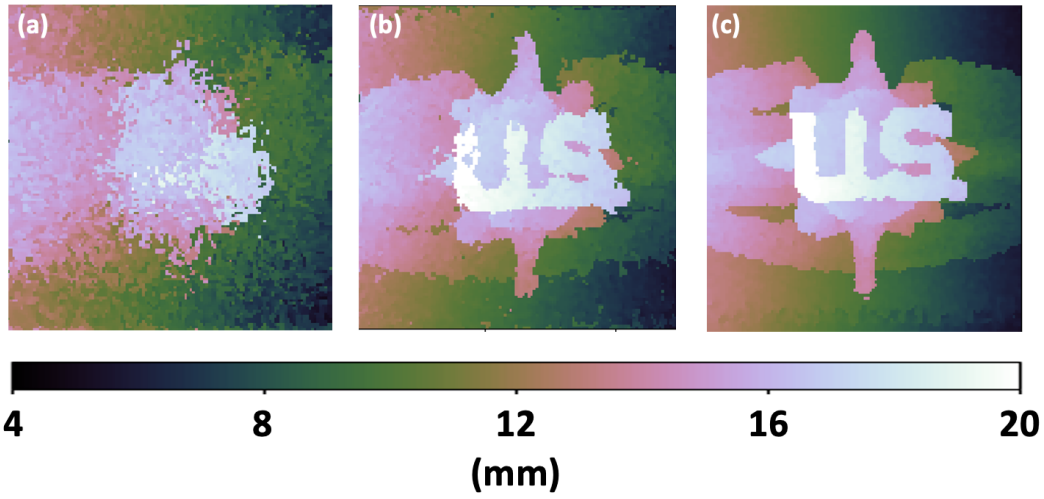

Figure 4.12: Turbulence case 2: $D/r_0 = 16$, $\kappa = 42$ range images where (a) is the aberrated amplitude, (b) is the sharpened amplitude, and (c) is an unaberrated amplitude from a very similar data set.

amplitude. We can also make out the "U" in "US," at a level of detail similar to the unaberrated amplitude.

Bibliography

[1] A. E. Siegman, *Lasers* (University Science Books, 1986).

[2] M. C. Roggemann and B. M. Welsh, *Imaging through turbulence* (CRC press, 2018).

[3] M. I. Skolnik, "Introduction to radar systems," New York (1980).

[4] C. E. Cook, "Pulse compression-key to more efficient radar transmission," P. IRE **48**, 310–316 (1960).

[5] S. T. Thurman and J. R. Fienup, "Phase-error correction in digital holography," J. Opt. Soc. Am. A. **25**, 983–994 (2008).

[6] U. Grenander, *Abstract inference* (Taylor & Francis, 1981).

Chapter 5

GOI Phase Retrieval

5.1 Introduction

Methods which analyze modal coefficients of fields have been studied using a variety of different techniques. Historically, much of the modal analysis work has been associated with sorting fields propagated through multimodal fibers for use in optical communication systems [1]. Recently however, there has been a great deal of focus on analyzing propagated freespace fields into a transverse basis set for use in both satellite and terrestrial-based communication systems [2–4]. Experimental recovery of messages encoded in the relative phase between two superpositioned freespace modes has also been conducted over long propagation distances through atmospheric turbulence using machine learning techniques [5].

In this chapter, we describe a phase retrieval algorithm for a novel in-

terferometry system known as a generalized optical interferometer (GOI). Note that we will use the abbreviation GOI to refer to both the general technique of generalized optical interferometry as well as the generalized optical interferometer instrument. It will be clear from context to which we are referring. A GOI enables one to analyze a monochromatic, scalar optical field propagated through free-space. In a GOI, GPOs are added in one or more arms of an interferometer (in our work it is specifically a Mach-Zehnder type interferometer). Mathematically, GPOs take the form of a transform kernel associated with the manipulation of a physical property of the optical field. These properties include phase, polarization, temporal delay, parity (flipping of the field as in a dove prism), etc. [6]. For cases of concern here, this interferometer contains two generalized phase operators (GPOs), one in either arm of the interferometer. Each GPO conducts a fractional Fourier transform (fFT) with respect a different transverse dimensions of the input field. The optical elements that conduct the fFTs balance the beam parameters so that the transverse scaling of the system's output field distributions are unchanged from the input fields [7]. Moreover, this is accomplished without physically repositioning any optical elements in the system. By considering the field transiting through the interferometer as a weighted superposition of Hermite Gaussian (HG) modes, the system is able to recover the amplitudes of the generally complex-valued weighting coefficients [6, 8].

HG modes are defined at the beam waist in two dimensions as

$$HG_{mn}(x,y) = H_m\left(\frac{\sqrt{2}x}{w_0}\right) H_n\left(\frac{\sqrt{2}y}{w_0}\right) \exp\left[\frac{-\left(x^2+y^2\right)}{w_0^2}\right], \tag{5.1}$$

where H_m and H_n are Hermite polynomials of order m and n, respecively, and w_0 is the radius of the beam waist. They are well known solutions to the paraxial wave equation [9]. Throughout this paper HG mode superpositions will be refererred to by order. The ℓ^{th} order of HG modes is inclusive of all modes such that the indices m and n in a given mode, HG_{mn}, satisfy the condition $m+n \leq \ell$.

In Section 5.2 we describe in detail the theoretical underpinnings of GOI, briefly describe the system configuration which we use in our work, and briefly describe the recovery of these amplitudes. An additional approach to amplitude recovery can also be found in [10]. After the amplitude of each coefficient is recovered, we wish to recover the phase in order to complete the characterization of the field. Here, we developed and implemented two phase retrieval algorithms which accomplish this.

Phase retrieval techniques have been used in a variety of applications including wave-front sensing [11], metrology [12], crystallography [13], microscopy [14], astronomical speckle imaging [15, 16], x-ray diffraction imaging [17], and decomposition of TEM modes in multimodal fibers [18]. Sections 5.3 and 5.4 describe our novel nonlinear optimization-based phase retrieval algorithms, the first using measurements from an array detector and the

second using measurements from a bucket detector. We share results from monte carlo studies of these algorithms and provide best practices for their use.

5.2 Generalized Optical Interferometry Theory

Consider a GPO whose only effect on a field is to add a known phase to that field. Each term in the linear combination that describes the original field is a weighted member of the previously mentioned HG basis set. When HG modes are operated on by a GPO, the GPO produces phase-only eigenvalues. We have included a detailed proof of the eigenfunction relationship HG modes share with the fractional Fourier transform as Appendix 7.3. So, given a transverse scalar field

$$U(x,y) = \sum_{mn} c_{mn} HG_{mn}(x,y), \tag{5.2}$$

a GPO operator, Λ_α, is defined

$$\Lambda_\alpha \{U(x,y)\} = \sum_{mn} c_{mn} e^{\frac{im\pi\alpha}{2}} HG_{mn}(x,y), \tag{5.3}$$

where α is the transform parameter of the GPO.

The fFT is a linear cannonical transform whose effect on the phase-space distribution of a transverse field distribution is a rotation of the rectangular

region whose area is the SBWP of a single dimension [19, 20]. A $\pi/2$ radian rotation of this rectangular region in phase space corresponds to a Fourier transform and a $-\pi/2$ radian rotation corresponds to an inverse Fourier transform. Operations which cause phase space rotations that are not discrete multiples of $\pi/2$ radians are "fractions" of a full Fourier transform and are thus referred to as fractional Fourier transforms. The order of an fFT, α, corresponds to the amount of rotation in phase space imparted by the transform. An fFT can also be thought of as a propagator similar to a Fresnel transform in that it can be used to conduct a propagation from the exit pupil of the system to any other plane with the addition of an additional quadratic phase term [21]. In this paper, if θ is the phase space rotation in radians, the fFT will be represented as $\mathscr{F}_\alpha\{\cdot\}$, where $\alpha = 2\theta/\pi$. So the first Fourier transform plane is at $\alpha = 1$, which corresponds to a $\pi/2$ rotation in phase space. The fFT is a periodic operator in the sense that the operation performed at α is the same as the operation performed at $\alpha + 4p$ where p is any integer. The symbols $\mathscr{F}\{\cdot\}$ and $\mathscr{F}^{-1}\{\cdot\}$, without an α value, will represent a standard Fourier transform and inverse Fourier transform, respectively.

5.2.1 Generalized Mach-Zehnder Interferometry

In this work, we used a Mach-Zehnder interferometer (MZI) configuration GOI, as shown in Fig. 5.1. Each arm of the interferometer is understood to contain a GPO as described by Eq. (5.3), which performs an equivalent

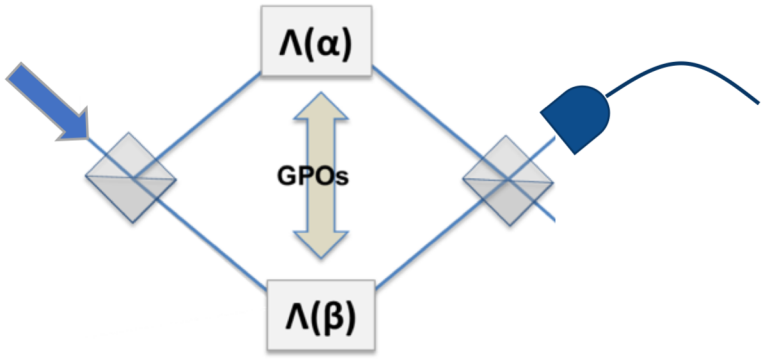

Figure 5.1: Mach-Zehnder GOI configuration used for phase retrieval simulations. The array detector placed at the upper port of the figure also functions as a bucket detector for GOI amplitude recovery by integrating over all pixels.

function to a cylindrical lens with an arbitrary radius of curvature in one transverse dimension. Each GPO performs fFTs with respect to a different transverse dimension. Our first phase retrieval algorithm relies on explicitly measured spatial information and therefore requires an array detector, whereas our second algorithm has been designed for use with data collected using a bucket detector.

Using Eq. (5.3), the field exiting the upper output port of the interfer-

ometer is expressed

$$\begin{aligned} U_{out}(x,y;\alpha,\beta) &= \frac{1}{\sqrt{2}}\left[\Lambda_\alpha + \Lambda_\beta\right] U_{in}(x,y) = \frac{1}{\sqrt{2}}\left[\Lambda_\alpha + \Lambda_\beta\right] \sum_{m,n} c_{mn} HG_{mn}(x,y) \\ &= \frac{1}{\sqrt{2}} \sum_{m,n} c_{mn} \left[e^{\frac{im\pi\alpha}{2}} HG_{mn}(x,y) + e^{\frac{in\pi\beta}{2}} HG_{mn}(x,y) \right] \\ &= \frac{1}{\sqrt{2}} \sum_{m,n} c_{mn} HG_{mn}(x,y) \left(e^{\frac{im\pi\alpha}{2}} + e^{\frac{in\pi\beta}{2}} \right), \end{aligned} \tag{5.4}$$

where Λ_α and Λ_β are the GPOs in the upper and lower arms of the interferometer, respectively.

The intensity at the output port of the MZI in Fig. 5.1 detector is

$$\begin{aligned} I_{out}(x,y;\alpha,\beta) &= \left| \frac{1}{\sqrt{2}} \sum_{mn} c_{mn} HG_{mn}(x,y) \left(e^{\frac{im\pi\alpha}{2}} + e^{\frac{in\pi\beta}{2}} \right) \right|^2 \\ &= \frac{1}{2} \Bigg\{ \sum_{m,n} |c_{mn}|^2 HG_{mn}^2(x,y) \left[2 + 2\cos\left(\frac{m\pi\alpha}{2} - \frac{n\pi\beta}{2} \right) \right] \\ &\qquad + \Bigg[\sum_{\substack{m,n;m',n' \\ m,n\neq m',n'}} c_{mn} c_{m'n'}^* HG_{mn}(x,y) HG_{m'n'}(x,y) \\ &\qquad\qquad \times \left(e^{\frac{im\pi\alpha}{2}} + e^{\frac{in\pi\beta}{2}} \right) \left(e^{-\frac{im'\pi\alpha}{2}} + e^{-\frac{in'\pi\beta}{2}} \right) \Bigg] \Bigg\}. \end{aligned} \tag{5.5}$$

During the amplitude recovery process, the entire transverse spatial extent of the detector is integrated over. If we are using an array detector, we do this integration in post-processing. If we are using a bucket detector, this

integration is done implicitly.

Using the orthogonality relation of HG modes,

$$\iint_{\infty}^{\infty} dxdyHG_{mn}(x,y)HG_{m'n'}(x,y) = S(m,n)\delta_{mm'}\delta_{nn'}, \tag{5.6}$$

where δ is the Kroenecker delta function, the signal measured after spatial integration is

$$\begin{aligned} P_{out}(\alpha,\beta) &= \iint dxdyI_{out}(x,y;\alpha,\beta) \\ &= \sum_{m,n} |c_{mn}|^2 \left[1 + \cos\left(\frac{m\pi\alpha}{2} - \frac{n\pi\beta}{2}\right)\right] S(m,n). \end{aligned} \tag{5.7}$$

For convenience, we assume HG mode energy is unit normalized, i.e. $S(m,n) = 1$. The bias term can be removed from the interferogram. Since $P_{out}(0,0) = 2\sum_{mn}|c_{mn}|^2$, its measurement can be used in a well calibrated system to remove the bias term as follows:

$$\begin{aligned} P'_{out}(\alpha,\beta) &= P_{out}(\alpha,\beta) - \sum_{mn} |c_{mn}|^2 \\ &= \sum_{mn} |c_{mn}|^2 \cos\left(\frac{m\pi\alpha}{2} - \frac{n\pi\beta}{2}\right). \end{aligned} \tag{5.8}$$

For simplicity, assume that we are able to sample $P_{out}(\alpha,\beta)$ finely enough in α and β to approximate a continuous function space. The continuous Fourier

transform of $P'_{out}(\alpha,\beta)$ is

$$\begin{aligned}\tilde{P'}_{out}(\mu,\eta) &= \mathscr{F}\{P'_{out}(\alpha,\beta)\}_{\{\alpha,\beta\}\to\{\mu,\eta\}} \\ &= \sum_{m,n}|c_{mn}|^2\mathscr{F}\left\{\cos\left(\frac{m\pi\alpha}{2}-\frac{n\pi\beta}{2}\right)\right\} \\ &= \sum_{m,n}|c_{mn}|^2\left[\delta\left(\mu-\frac{m}{4},\eta+\frac{n}{4}\right)+\delta\left(\mu+\frac{m}{4},\eta-\frac{n}{4}\right)\right].\end{aligned} \tag{5.9}$$

So the recovery of the amplitude coefficients is conceptually just a matter of measuring the intensities of the δ-function peaks in $\tilde{P}'_{out}(\mu,\eta)$. The δ-function offset as a factor of $1/4$ is the result of the frequency modulation terms $\alpha\pi/2$ and $\beta\pi/2$ in the cosine function. If one so desires, a simple change of variables will make this offset into an integer value. The δ-function would instead peak at integer values if these terms were instead $2\alpha\pi$ and $2\beta\pi$. The $P'_{out}(\alpha,\beta)$ interferogram from a sample 2^{nd} order HG superposition is shown in Fig. 5.2 and its Fourier transform, $\tilde{P}'_{out}(\mu,\eta)$, is shown in Fig. 5.3. The values of the nonzero pixels in Fig. 5.3 are indicative of the relative contribution of each HG mode to the overall energy of the sample field.

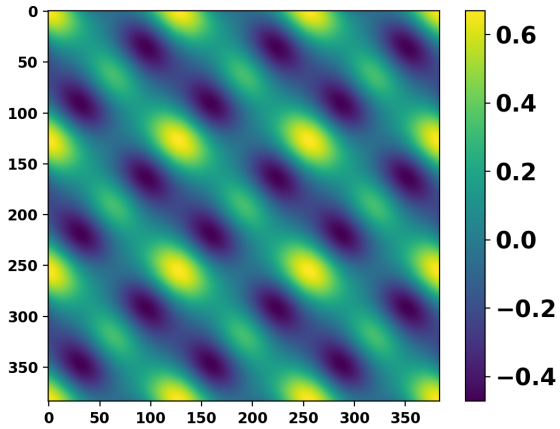

Figure 5.2: $P'_{out}(\alpha, \beta)$ measurement for a sample field created from a 2^{nd} order HG superposition.

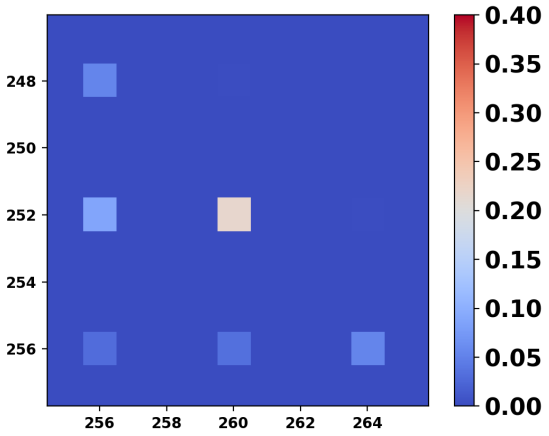

Figure 5.3: The upper right quadrant of $\tilde{P}'_{out}(\mu, \eta)$, the Fourier transform of the interferogram in Fig. 5.2. Non-zero pixel values represent the relative strength of a particular HG mode's amplitudes in a sample field, save for bright pixel in the lower left corner representing twice the relative contribution of the HG_{00} mode's amplitude.

5.3 Phase Retrieval in a Generalized Mach-Zehnder Interferometer Using an Array Detector

5.3.1 Algorithm Design

To fully characterize a monochromatic scalar field described as a linear superposition of HG modes, both the phase and amplitude of the complex-valued weighting coefficients, $c_{mn} = |c_{mn}|e^{i\phi_{mn}}$, must be determined. We assume the amplitude information of the field, $|\vec{c}| = \{c_{mn}\}$, has been successfully recovered using the method in Section 5.2. Now $\vec{\phi}$ must be recovered. Here, we implemented an algorithm which recovers $\vec{\phi}$ using array detection.

Given the intensity distribution in Eq. (5.5) and previously recovered amplitude coefficients, we seek to solve the inverse problem of retrieving the phases, $\vec{\phi} = \{\phi_{mn}\}$, associated with each amplitude coefficient. This can be cast as a nonlinear optimization problem where the error metric to be minimized is

$$E = \sum_{x,y} \left| I_{\text{est}}\left(x, y; \alpha, \beta; \hat{\vec{\phi}}\right) - I_{\text{d}}\left(x, y; \alpha, \beta\right) \right|^2, \tag{5.10}$$

where $I_{\text{d}}(x, y)$ is the measured detector intensity in the upper output of the

interferometer and

$$I_{\text{est}}\left(x, y; \alpha, \beta; \hat{\vec{\phi}}\right) = \frac{1}{2}\left|\sum_{mn} |c_{mn}| HG_{mn}(x, y) e^{i\hat{\phi}_{mn}} \left(e^{\frac{im\pi\alpha}{2}} + e^{\frac{in\pi\beta}{2}}\right)\right|^2, \quad (5.11)$$

is an estimated intensity distribution computed from $\hat{\vec{\phi}}$, an estimate of $\vec{\phi}$.

Note that, based on the intensity model in Eq. (5.11), there exist values where certain modes will be suppressed from the output intensity of the interferometer. Any time that $\exp(im\pi\alpha/2) + \exp(in\pi\beta/2) = 0$, the mode HG_{mn} will not contribute to the output intensity. For example, in the case of an inverted image transform, $\alpha = \beta = 2$, any modes where $m+n$ is an odd whole number will result in an intensity plane that contains no contribution from those modes. Hence, fFT order pairs resulting in suppressed modes do not aid in retrieving the phases of those suppressed modes because we cannot modulate the output intensity of the instrument by altering those phases. Fortunately, values of α and β can be chosen that do not suppress modes. Moreover, multiple measured intensity planes, each with a different α, β pair, can be used to add robustness to our phase retrieval and aid in overcoming mode suppression.

To retrieve $\vec{\phi}$ we employed the LBFGS [22] algorithm from scipy.optimize package [23]. The LBFGS was supplied an initial phase estimate, $\hat{\vec{\phi}}_{\text{init}}$, drawn from a uniform random distribution and was allowed to iterate until the error metric value reached a minimum. The LBFGS is a gradient-based

optimizer. So in order to help the optimizer converge to a minima much faster, we supplied an analytic gradient of our physical model, as a funciton of $\hat{\vec{\phi}}$. It is worth noting that the ability to express propagation as a trivially parallelizable series of scalar multiplications, instead of a traditional DFT (discrete Fourier transform)-based transform (or FFT), means that it has a very low computational cost when compared to most other phase retrieval algorithms that rely on DFTs (or FFTs) [24].

The optimization process described above does not always yield the desired phase values. There are two common failure modes for this process. The first failure mode occurs when the gradient information guides the optimizer into a local minimum of the error metric. The second common failure mode occurs when using a single output intensity plane, due to what we call a "twin image" problem [25, 26], although the term "twin image" is a misnomer in this work. This failure case is the result of a degeneracy resulting from the modulus operation in Eq. (5.10). One of these global minimums will yield the correct relative phases. The other global minimum, generally, will not.

Twin image convergence failures are eliminated by adding information diversity to the optimization in the form of measurements of one or more additional planes of intensities, each with a different α, β pair. This is akin to information-diverse phase retrieval techniques such as defocus-diverse phase retrieval [11] and transverse-translation-diverse phase retrieval [17]. The er-

ror metric function for one or more output intensity planes is

$$E = \sum_{\{\alpha,\beta\}} E_{\alpha\beta} = \sum_{\{\alpha,\beta\}} \sum_{x,y} \left| I_{est}\left(x, y; \alpha, \beta; \hat{\vec{\phi}}\right) - I_d(x, y; \alpha, \beta) \right|^2 . \tag{5.12}$$

So in multiple-plane cases, the "twin-image" is eliminated because the degenerate global minimum of the error metric with respect to $\vec{\phi}$ at one α, β intensity plane will not be a global minimum for intensity planes with different α, β pairs. Furthermore, adding a greater diversity of information to a nonlinear optimization-based phase retrieval algorithm has been shown to increase the rate of successful retrieval in terms of convergence to the global minimum and resistance to measurement noise. Simulation experiments were conducted to test the performance of this approach. Simulations were always conducted over many different fields (i.e. the fields were generated using different random $|\vec{c}|$ and $\vec{\phi}$ values) to ensure that success or failure was agnostic of the particular field being retrieved. Intensitites were generated as superpositions of unit-normalized HG modes, centered at the beam waist (HG modes are real valued at the beam waist). Modes were well oversampled in 256×256 pixel arrays with pixel spacings $\delta x,\ \delta y = w_0/32$, where w_0 is the radius of the beam at the waist. Coefficient amplitudes were assumed to be known in advance of initiating the phase retrieval process, since they would be available from the analysis of $\tilde{P'}_{out}(\mu, \eta)$ data. Unless otherwise stated explicitly, all simulations were conducted for superpositions of 2nd order (3 modes) through 6th order (21 modes) HG modes. The condition for success

in our trials was an error of less than error$<$ 0.07 waves RMS of retrieved relative phase values with respect to the true phase values. As in wave-front sensing phase retrieval, only the relative phases can be recovered. So, the recovered phase values likely will differ from the actual phase values by a global piston phase.

5.3.2 Single-Plane Phase Retrieval

The first test of the phase retrieval algorithm was to see if it was possible to retrieve phase from only a single spatially distributed intensity in the $\alpha = \beta = 0$ plane. This plane is representative of the intensity of the field in the interferometer without any operation being executed on the field. Twenty random fields were generated for each order ℓ. Fifty retrieval attempts were made for each of those twenty fields (a total of one thousand retrieval attempts at each order). All 50 retrieval attemps for a single field used a different random starting guess, $\hat{\vec{\phi}}_{init}$. Note that the random starting points were seeded and each one of the twenty fields used the same 50 random starting points. As seen in Table 5.1, a greater number of modes corresponded to a smaller fraction of successes, as one would expect. The algorithm converged to “twin-image" solutions about as often as it converged to the true solution. All non-“twin-image" failures resulted from optimizer stagnation at a local minima.

Next, we studied phase retrievals were whose intensity distributions re-

Table 5.1: **Success of single $\alpha = \beta = 0$ plane GI phase retrieval**

HG order	# Modes	% Success	% Twin Image	HG order	# Modes	% Success	% Twin Image
$\ell = 2$	3	46.7	44.6	$\ell = 7$	28	18.1	19.1
$\ell = 3$	6	37.1	34.5	$\ell = 8$	36	14.3	14.5
$\ell = 4$	10	31.4	27.4	$\ell = 9$	45	10.9	10.4
$\ell = 5$	15	23.1	22.8	$\ell = 10$	55	10.6	10.0
$\ell = 6$	21	22.6	24.4	$\ell = 11$	66	9.8	10.0

sulted from a variety of α, β pairs assigned to the GPOs. Parameters α and β were sampled in twenty-five equally spaced increments from 0 to 2. Ten random fields were generated for each order ℓ and α, β pair. Retrieval of each field was attempted with up to five random $\hat{\vec{\phi}}_{init}$ starting guesses. With the exception of planes where retrieval failed due to the aforementioned mode suppression, no trend is seen that indicates a certain range of α, β values will yield a significantly greater number of successes than any other. Success rates for single-plane phase retrieval where α and β are varied do not change demonstrably from success rates where $\alpha = \beta = 0$. Single-plane success rates are shown in the bottom curve of Fig. 5.4, which treats the results of each $\hat{\vec{\phi}}_{init}$ separately. Given the twin image problem, around 50% is generally the highest success rate possible for the single-plane phase retrieval with a single $\hat{\vec{\phi}}_{init}$. However, when we allowed up to five different $\hat{\vec{\phi}}_{init}$ guesses for a single-plane phase retrieval, the success rates increased to 94% for 2nd order, 83% for 3rd order, 78% for 4th order, and 77% for 5th and 6th orders.

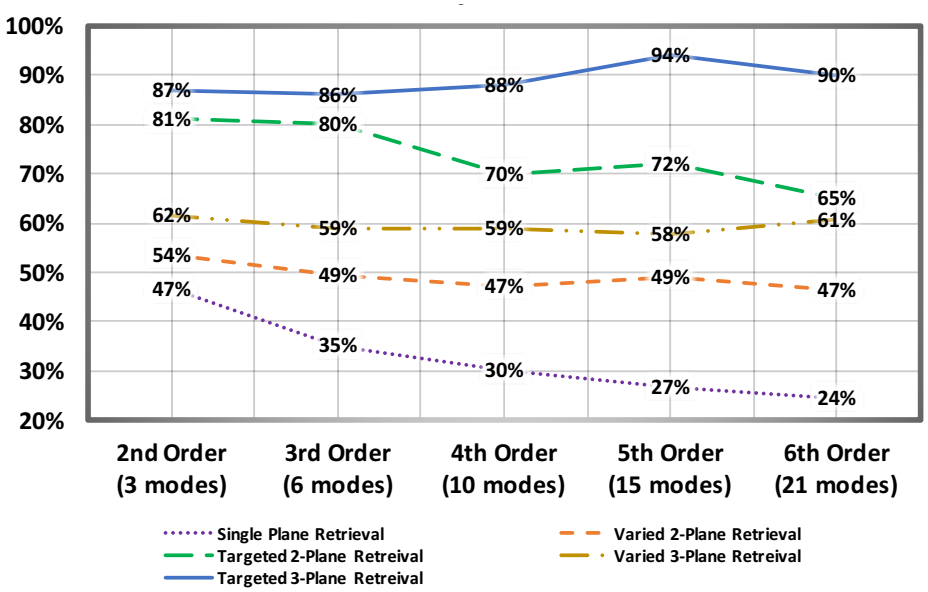

Figure 5.4: Chart compares percent of successful retrievals for single-plane, two-plane, two-plane targeted, three-plane, and three-plane targeted phase retrieval techniques using only a single set of $\hat{\vec{\phi}}_{init}$ starting values.

5.3.3 Two-Plane Phase Retrieval

In the next round of simulations, additional planes of intensity information were added in order to further constrain the nonlinear optimization, which includes the ability to overcome the twin image problem. These monte carlo studies were conducted to determine if any specific combinations of α, β intensity planes yielded significantly improved success rates for retrieval. For simplicity, we held $\alpha = \beta$, in order to keep the overall size of the monte carlo study manageable.

Runs were conducted for HG superpositions consisting of modes for or-

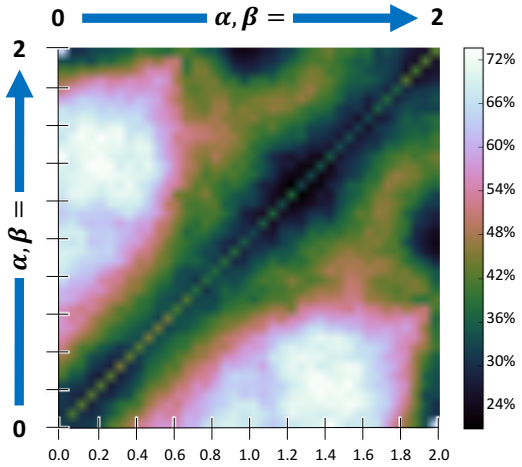

Figure 5.5: Heatmap of successful retrieval rates as a function of α, β values for two-plane GI phase retrieval with 10 modes (4th order).

ders from $\ell = 2$ through $\ell = 6$. Forty evenly-spaced fFT order values (α, β) with values between $[0, 2]$ were used. Phase retrievals were conducted using twenty different GOI input fields for every pair of α, β values. Retrieval of each of these twenty fields was allowed up to fifty different $\hat{\vec{\phi}}_{init}$ starting guesses. A heat map of the success rates for order $\ell = 4$ (10 superpositioned modes) is shown in Fig. 5.5. A preference is seen for α, β pairs where one value is between 0.1 and 0.3 and the other value is between 1.4 and 1.6. Similar preferred plane combinations were seen for all HG orders tested. We suspect this preference might be due to planes in these regions having the greatest amount of unique information to constrain the problem and containing the fewest suppressed modes. Having located preferential α, β ranges, we conducted tests to determine the performance of the algorithm for these preferred values. Monte carlo simulations were run for α, β

planes in the targeted regions for 50 additional randomly generated fields. In each optimization, $\alpha,\ \beta$ values were chosen randomly from the ranges in column 4 of Table 5.2. Using a single random $\hat{\vec{\phi}}_{init}$, the percentage of phase vectors retrieved successfully is shown in column 3 of Table 5.2. With two well-chosen planes, a similar fraction of successful retrievals were obtained for field superpositions of up to 21 modes (6^{th} order).

Table 5.2: **Success rates of two intensity plane GOI phase retrieval**

HG order	# Modes	% Success	$(\alpha,\ \beta)$ Ranges
$\ell = 2$	3	81	(0.1, 0.3), (1.4, 1.6)
$\ell = 3$	6	80	(0.1, 0.3), (1.2, 1.4)
$\ell = 4$	10	70	(0.1, 0.3), (1.2, 1.4)
$\ell = 5$	15	72	(0.1, 0.3), (1.4, 1.6)
$\ell = 6$	21	65	(0.1, 0.3), (1.2, 1.4)

5.3.4 Three-Plane Phase Retrieval

With the increased success of the two-plane phase retrievals, a third plane was added in an attempt to further improve rate of successful retrieval with a single $\hat{\vec{\phi}}_{init}$ starting guess. To simulate this, the range of the fFT orders was limited to $[0.2-0.4]$ for the first plane and $[1.4-1.6]$ for the second plane, each in range sampled in 4 equally-spaced increments. α, β pair selections for the third intensity plane were allowed to vary from order $\alpha = \beta = 0$ to $\alpha = \beta = 2$,

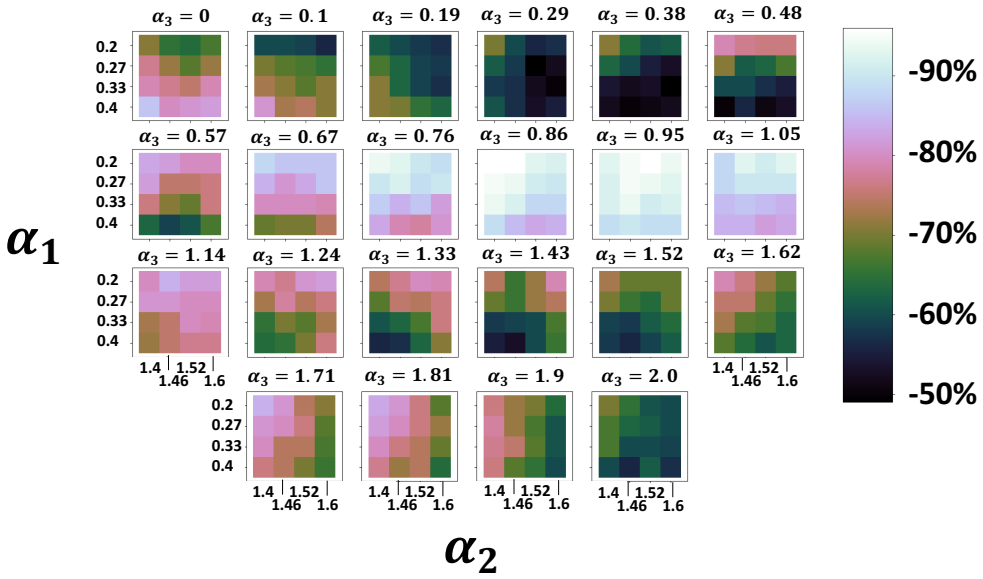

Figure 5.6: Heatmaps detail success of three-plane phase retrieval of a 6^{th} order superposition with 22 equally-spaced values of α in the third plane.

in 22 equally spaced intervals. The percentage of successful retrievals with respect to the α-value of three intensity planes used is shown in Fig. 5.6. The best success rates are seen for a third fFT plane whose α, β values are in the range $[0.72, 1.04]$. This range is roughly equidistant from the regions of the first two planes. This would be in line with our earlier speculation that the preferred planes will occur in the regions that provide the optimizer with the most unique information. The rate of successful retrieval with intensity plane combinations in this range is consistently greater than 90% for a single $\hat{\vec{\phi}}_{init}$ starting point. No obvious trend was observed, with respect to amplitude and phase values, that would imply that certain values of coefficient phases and amplitudes, or combinations thereof, caused more failures or successes than

any others. In Fig. 5.4 there are a few cases where a greater number of modes experienced a greater percentage of overall successes. This is most likely due to the randomness of the fields generated and the modest number of fields and $\hat{\vec{\phi}}_{init}$ values that were used. Given a large enough sample of randomly generated fields and random starting points, we expect the percentage of success will be monotonically decreasing as the number of modes present in the superposition increases. It is worth noting that the process for choosing the α, β range for the three plane superposition may not be ideal. The best way to determine the most ideal three plane combination would be to conduct a monte carlo simulation allowing the α, β pairs for all three planes to vary simultaneously. We chose not to use this process because of the time and resources needed to conduct a monte carlo simulation in this way and because the method we did use to determine preferential α, β values was more than sufficient to create a robust algorithm.

5.3.5 Final Array Detector Algorithm Performance and Discussion

With success rates consistently greater than 85% achieved in the three-plane retrievals for a single $\hat{\vec{\phi}}_{init}$, the algorithm was deemed effective and final performance runs with targeted plane selections were conducted. For each set of simulated data, the algorithm was permitted a maximum of five attempts, each with a different $\hat{\vec{\phi}}_{init}$, to retrieve the correct phase. Phase retrieval

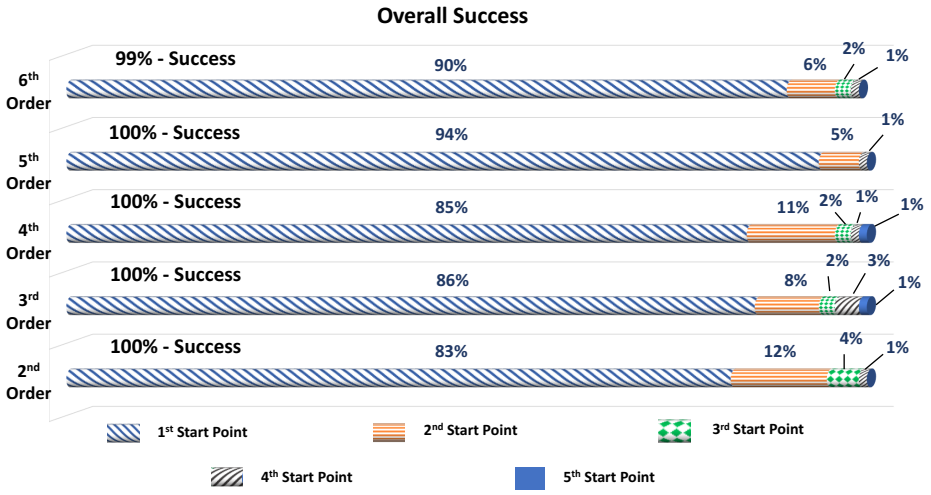

Figure 5.7: Percent successful retrievals when three targeted α planes and five random starting sets of phase values are permitted (noiseless).

was attempted using 100 random fields composed of 2^{nd} through 6^{th} order superpositions. The results of this final algorithm, with a breakdown of success by starting-guess number, are shown in Fig. 5.7. It shows that if the algorithm fails with the first random starting guess, the second starting guess was successful the majority of the time. Almost 100% of cases succeeded within the five permitted attempts. For real world applications where one cannot check the algorithm against the known phase values, the error metric was more than sufficient to identify successes and failures for the fields we retrieved. The smallest gap between the error metric of failed retrievals and the error metric of successful retrievals occurred in 4th order phase retrievals. In this noise-free group, the average error metric of a failed retrieval was greater than 10^{-5} while average error metric value for successful retrievals

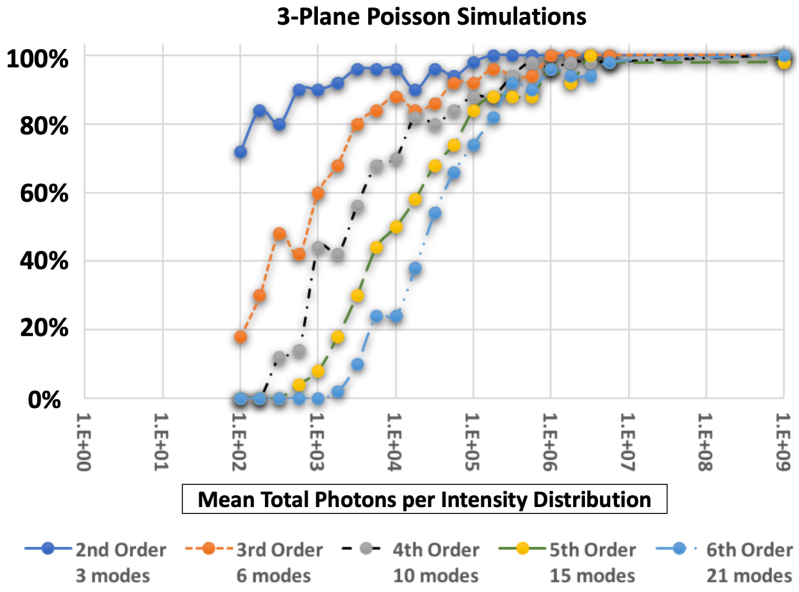

Figure 5.8: Variability of success with Poisson noise.

was less than 10^{-20}.

Additionally, runs were conducted where reference intensities reflected Poisson noise statistics. These reference intensities varied in mean total number of photons, enabling us to test the robustness of the algorithm and establish SNR requirements for good performance. Again, up to five attempts to retrieve the phase vector were permitted. Each attempted retrieval used three targeted intensity planes from the preferential α-value regions established in earlier simulations. The mean number of total photons per intensity over the 256×256 pixel array varied from 10^2 to 10^7. The exponent determining the average number of total photons varied in 21 uniformly spaced values from 2 to 7. Fig. 5.8 shows the results: Nearly all phases were correctly

retrieved at 10^7 average total photons per intensity plane and greater than 90% of all phases were retrieved at 10^6 average total photons per intensity, for up to 6^{th} order superpositions. As expected, superpositions composed of fewer modes are generally more successful, especially for low SNRs. For example, at 10^3 average total photons per intensity plane, 90% of 2^{nd} order superposition phases were recovered successfully, whereas no 6^{th} order superposition phases were recovered successfully. When the average total photons increased to 10^4, 25% of phases were recovered at 6^{th} order and 96% all 2^{nd} order superposition phases were recovered successfully.

5.4 Phase Retrieval in a Generalized Mach-Zehnder Interferometer Using a Single-Pixel Detector

Here, we again consider a Mach-Zehnder-type GOI which contains two generalized phase operators (GPOs), one in either arm as shown in Fig. 5.1. We build on our array-based phase retrieval work in [27] by designing and implementing a phase retrieval algorithm in simulation which only requires bucket detection. In Section 5.5, we describe our novel nonlinear optimization-based phase retrieval algorithm which retrieves modal coefficient phases of an input field using measurements from a bucket detector. In Section 5.6 we describe monte carlo studies conducted using this algorithm, then share the results of these studies, providing our analysis.

5.5 Algorithm Design and Performance Assessment

5.5.1 Nonlinear Optimization

Again, the intensity at the upper output port of the MZI containing the detector in Fig. 5.1 is [27]

$$\begin{aligned} I_{out}(x,y;\alpha,\beta) = &\sum_{m,n}|c_{mn}|^2 HG_{mn}^2(x,y)\left[1+\cos\left(\frac{m\pi\alpha}{2}-\frac{n\pi\beta}{2}\right)\right] \\ &+\frac{1}{2}\sum_{\substack{m,n;m',n' \\ m,n\neq m',n'}}|c_{mn}||c_{m'n'}^*|e^{i(\phi_{mn}-\phi_{m'n'})} \\ &\times HG_{mn}(x,y)HG_{m'n'}(x,y) \\ &\times\left(e^{\frac{im\pi\alpha}{2}}+e^{\frac{in\pi\beta}{2}}\right)\left(e^{-\frac{im'\pi\alpha}{2}}+e^{-\frac{in'\pi\beta}{2}}\right), \end{aligned} \tag{5.13}$$

where α and β are the orders of fFT encoded on the upper and lower GPOs, respectively. During the amplitude recovery process, the entire transverse extent of the output field is assumed to be integrated spatially. Ideally this measurement is conducted using a bucket detector because of its desirable hardware characteristics, as compared to an array detector.

In our array detector-based phase retrieval, we used a gradient-based nonlinear optimization to minimize a sum-of-squared-error objective function to estimate the desired modal piston phases [27]. This approach proved very effective and was robust to shot noise. Thus, we sought to develop a similar

phase retrieval technique that could be implemented using a bucket detector.

Observe that the second summation term on the right-hand side of Eq. (5.13) is suppressed due to HG mode orthogonality if all of the output field energy is measured by the bucket detector (i.e. integrated over all x and y). Since this second term is the only one which is dependent on the modal phases we would thus be unable to construct a useful error metric for phase retrieval.

To remedy this, we proposed a technique where we only measure only a portion of intensity distribution with each detector integration. This allowed the phase-dependent cross terms in Eq. (5.13) to be nonzero, making it possible to modulate a bucket detector signal as a function of coefficient phase. This strategy could be implemented in a laboratory apparatus by simply placing a binary mask in front the detector, allowing only a limited portion of the energy in the detector plane to be collected by the bucket. The only firm restriction on the detector plane region you select with the mask is that the change in the number of detected photons α, β and $\vec{\phi}$.

In our simulations, we chose to integrate only over the first quadrant of the plane of the detector. This meant we assumed a binary mask where all photons in the second, third, and fourth quadrants of the output port's energy distribution were not detected. The expression for the total measured

energy for a given α, β with phase-independent terms removed is

$$P(\alpha,\beta) = \sum_{\substack{m,n;m',n' \\ m,n\neq m',n'}} |c_{mn}||c_{m'n'}| e^{i\phi_{mn}} e^{-i\phi_{m'n'}} \times \left(e^{\frac{im\pi\alpha}{2}} + e^{\frac{in\pi\beta}{2}}\right)\left(e^{\frac{-im'\pi\alpha}{2}} + e^{\frac{-in'\pi\beta}{2}}\right) S_{mn;m'n'}, \tag{5.14}$$

where $S_{mn;m'n'} = \sum_{x,y} A(x,y) HG_{mn}(x,y) HG_{m'n'}(x,y)$, where $A(x,y)$ is the binary mask covering all but the first quadrant of the energy distribution, and HG modes are assumed to be energy normalized.

We chose our error metric to be

$$E = \sum_{\alpha,\beta} \left[P_{\text{est}}\left(\alpha,\beta;\hat{\vec{\phi}}\right) - P_{\text{ref}}(\alpha,\beta)\right]^2, \tag{5.15}$$

where $\hat{\vec{\phi}}$ is a vector containing $\hat{\phi}_{mn}$ for all m, n in order ℓ. The error metric was minimized with respect to $\hat{\vec{\phi}}$. $P_{\text{ref}}(\alpha,\beta)$ represents the total energy measured by the bucket detector in the first quadrant and $P_{\text{est}}(\alpha,\beta;\hat{\vec{\phi}})$ is the estimate of the measured energy as a function of our estimated phases, $\hat{\vec{\phi}}$. We used the SciPy implementation of the L-BFGS gradient-based optimizer [23] with an analytic gradient to conduct minimization.

5.6 Monte Carlo Simulations

5.6.1 Monte Carlo Study Design

After developing our phase retrieval algorithm, we conducted a monte carlo study to characterize its performance. One hundred randomized fields were simulated for each combination of parameters in the study. The study's parameter space consisted of three key elements, varied α, β parameters for each measurement, a varied number of measurements, and varied Poisson noise statistics. Each simulation attempted phase retrieval of 100 randomized fields with a unique combination of elements from these three categories.

We simulated fields using 2^{nd} through 6^{th} order HG mode superpositions. This meant our study included fields made from 3, 6, 10, 15, or 21 total HG modes. All HG modes were unit-normalized and centered at the beam waist. Simulated modes were well-sampled for their order in 32×32 pixel arrays whose transverse sample spacing was $w_0/10$, where w_0 is the radius of the beam at the waist. In our study, 100 random amplitude and phase vectors, $|\vec{c}|$ and $\vec{\phi}$ were created for every order of HG mode superposition we wished to simulate. In total, our study included 500 unique fields, before the application of shot noise statistics.

We ran our study with various numbers of α and β values determined using the following selection methodology. A given phase retrieval used the

set of α values $\{(4/R)(r-1)\}$ where r is a natural number in the range $[1, R]$ and R is the total number of α values. Similarly, β values were the set of values $\{(4/S)(s-1)\}$ where s is a natural number in the range $[1, S]$ and S was the total number of β values. Simulations were conducted for all combinations of the natural numbers R and S in the range $[1, 12]$, subject to the caveats that $S \leq R$ (to avoid redundancy in simulations) and $R/S \geq 2/3$ (to ensure comparable amounts of information diversity were delivered by each GPO). The $\alpha\beta$-pairs used in a given phase retrieval were therefore all the combinations of the α and β values in the current set totaling $R \cdot S$ pairs and therefore an $R \times S$ $P_{\text{ref}}(\alpha, \beta)$ array. Values of α or β less than zero or greater than or equal to four would have been redundant (assuming an ideal system) due to the periodic nature of the fFT [21]. This is why our α and β values are all in the range $[0, 4)$. In all, simulations were conducted with between 1 and 144 measurements per phase retrieval. Though due to multiplicity of $R \cdot S$ values, simulations were not conducted in one measurement increments. Given our constraints, total measurements followed the pattern $(1, 4, 6, 9, 12, 16, ..., 110, 121, 132, 144)$, all the products of the natural numbers between 1 and 12. One can deduce from the parameter selection criteria that $P_{\text{ref}}(\alpha, \beta)$ arrays were not generally square since the number of α and β values were not generally equal.

So to give a couple short examples of sets of α values, in the $R = 1$ case, the only α value was 0. For the $R = 3$ case, α values were 0, $4/3$, $8/3$. For the $R = 8$ case, the α values were 0, $1/2$, 1, $3/2$, 2, $5/2$, 3, and $7/2$. With respect to

β selections, in the $R = 3$ case, the only S values which fit our constraints are $S = 2$ or $S = 3$. Therefore the β values were either identical to the α values or the β values were 0 and 2, yielding either 6 total $\alpha\beta$-pairs (measurements) in the $R = 3, S = 2$ case, or 9 total $\alpha\beta$-pairs in the $(R = 3, S = 3)$ case. In the $R = 8$ case, S would have been 6, 7, or , 8, and produced β values in accordance with the previously defined relation.

To explore the fundamental noise limitations of our algorithm, phase retrieval for each intensity distribution was conducted using shot noise profiles of $10^3, 10^4, 10^5, 10^6$ and 10^7 mean signal photons per detector measurement (mspm). To be clear, this is the mean number of photons in the detector plane, not the mean number of detected photons. For comparison purposes, simulations were also conducted with no shot noise.

Coefficient amplitudes, $|\vec{c}|$, were assumed to be recovered using the technique outlined in [27], and therefore considered known prior to conducting phase retrieval. In order to avoid failure through convergence to local minima, each phase retrieval attempt was permitted up to 5 attempts to retrieve the correct relative phases. Each of these attempts used a different random starting guess for $\hat{\vec{\phi}}$. A phase retrieval was considered successful if it achieved a relative error of 0.07 waves RMS between the estimated phases, $\hat{\vec{\phi}}$ and the actual simulated phases, $\vec{\phi}$.

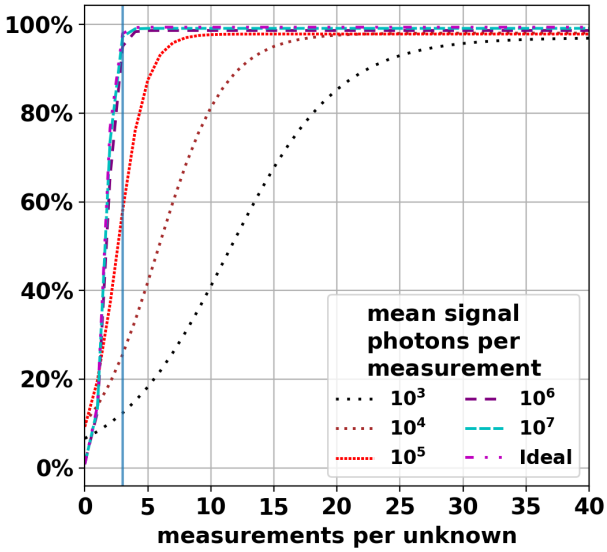

Figure 5.9: Logistic regressions of data points showing different shot noise profiles as a function of measurements per unknown phase coefficient (mpu).

5.6.2 Monte Carlo Results and Discussion

In Fig. 5.9 are logistic curves showing success as a function of measurements per unknown (mpu). Logistic regression was used to visualize data to enhance the interpretability of trends, as the placing all data points in the same figure resulted in tightly clustered and often overlapping data points which made it difficult to interpret. Though we show two of these plots individually in what follows. Each data point in our study represented the percentage of successful phase retrievals, out of 100, for a given combination of noise profile, collection of $\alpha\beta$-pairs (i.e. GPO settings for the interferometer), and order of superposition.

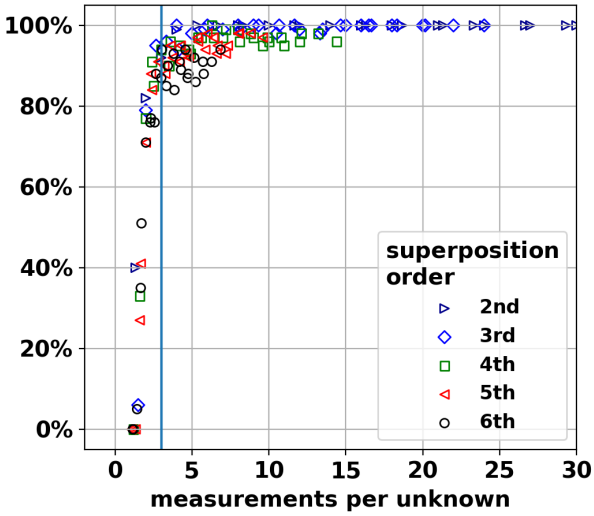

Figure 5.10: 10^7 mspm series success scatter plot as a function of mpu.

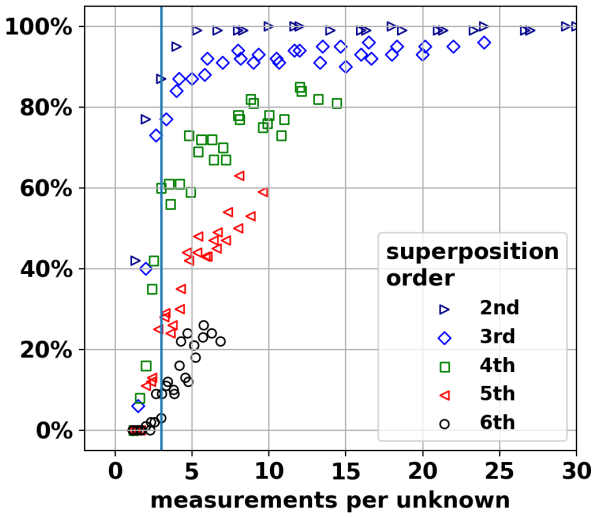

Figure 5.11: 10^4 mspm series success scatter plot as a function of mpu.

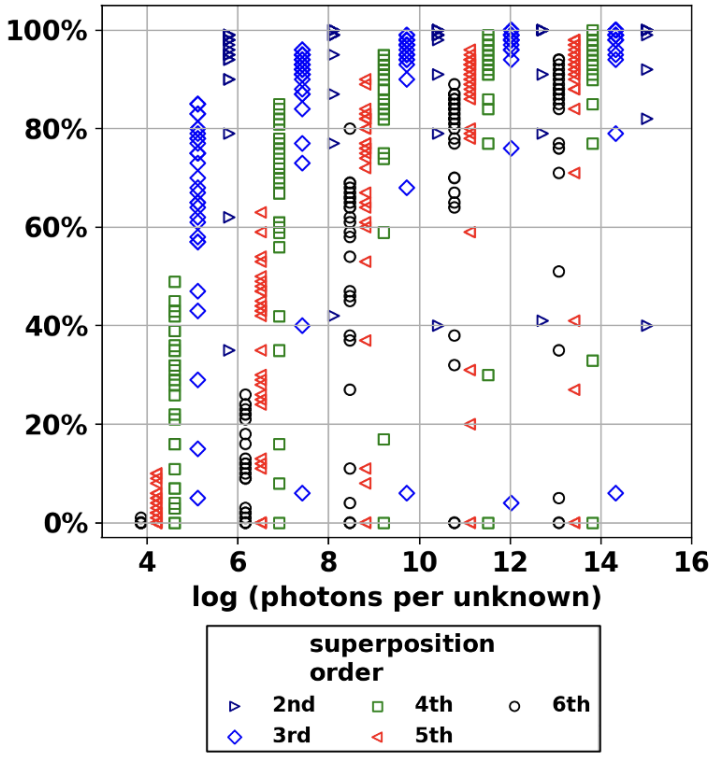

Figure 5.12: Scatter plot success rate data points of of all mspm values for each superposition order as a function of ppu phase coefficient per measurement.

Most importantly in Fig. 5.9 we observe that the phase retrieval algorithm is consistently capable of retrieving the ϕ_{mn} phase coefficients, given a sufficient number of photons and measurements. We also observe a not unexpected relationship between mspm and mpu. In short, more mspm meant fewer mpu were necessary. In order to get consistently successful results ($\geq$ 95% average success) in the noise-free (Ideal) case, as well as the 10^7 mpsm and 10^6 mpsm cases, we needed three to four mpu. This is somewhat reminiscent of the requirement in phase-shifting interferometers where three or four measurements are required to calculate a closed-form solution for the phase. We've added an additional vertical line to the chart at mpu= 3 for reference. The 10^5 mspm case achieved consistent success between 8-10 mpu, the 10^4 mspm case with between 18-20 mpu, and finally the 10^3 mspm with between 33-35 mpu.

Notice that the linear regression for 10^5, 10^4, and 10^3 mspm fits do intercept y-axis at 0% at mpu= 0. Consider as examples Figs. 5.10 and 5.11, plots of the 10^7 and 10^4 mspm series, respectively. Data points in the 10^4 mspm series are much more dispersed over the mpu axis than in the 10^7 series. We attribute this to the increasing negative effects which Poisson statistics had on the fidelity of our simulated measurements as signal photons became more scarce.

Additionally, these two figures display an interesting trend, though it is more apparent in the 10^4 mspm case. There is a discernible difference in

success for orders which have very similar mpu. Since superposition order is essentially a way of describing the total number of modes present in a superposition, we interpret this as mpu not the only contributing factor to the relative difficulty of our optimizations.

We thought perhaps this was due to a significant difference in the number of photons per unknown per measurement (ppu), which would indicate a higher likelihood of Poisson statistics deteriorating the necessary data. Said differently, the more modes that are present in a single measurement, the larger the effect each Poisson fluctuation will cause in the optimization.

In Fig. 5.12 we plot all of the data points as a function of log(ppu). Again, the plot markers represent the order of superposition, i.e. they are representative of how many phases we are trying to retrieve. All levels of shot noise can be observed in this graphic as well as order. Levels of shot noise increase from the left-most set of 2nd through 6th order data points at 10^3 mspm to right-most set of 2nd through 6th order data points at 10^7 mspm.

We see that the 10^4 mspm series, the 2nd set of 2nd through 6th order data point if moving left to right, has broad success marker spreads for the fourth, fifth, and sixth orders, than in any other mspm series, save for perhaps the 10^3 mspm series. We interpret these broad data point spreads as indicators that these optimization orders were on the cusp between having enough and too little information for the algorithm to succeed, largely as a

result of ppu. Clearly, in the 10^7 mspm series, we have very tight clustering of almost all data markers for all superposition orders, indicating that the optimizer had ample information to use in estimating coefficient phases and small Poisson fluctuations caused little adverse effect.

We would like to point out also in Fig. 5.12 that in spite of having more ppu, superpositions with more modes did not perform as well as superpositions with fewer modes. As an example, consider the 6th order cases in the 10^7 mspm series. All orders of superposition in the 10^6 mspm series, save for the 6th order, clearly displayed better performance. We cite this as a clear example of the curse of dimensionality [28], referring to the phenomena where a profound increase in calculation/optimization difficulty occurs as the dimensionality of a problem increases.

It's clear that if one has more than 10^6 mspm, and more than, say, five mpu, GOI phase retrievals will likely be consistently successful. It's also clear that success at mspm levels at or below 10^5 will be lower unless additional mpu are collected by the instrument to offset this poor signal. However, as stated, there does appear to be some minimum ppu threshold below which the chances of successful phase retrieval quickly diminish, regardless of how many measurements were collected. Finally, as is typical in problems with increasing dimensionality, the more phases we try to retrieve, the more difficult the process will become.

Bibliography

[1] N. Riesen and J. D. Love, "Design of mode-sorting asymmetric y-junctions," Appl. Opt. **51**, 2778–2783 (2012).

[2] M. Malik, M. O'Sullivan, B. Rodenburg, M. Mirhosseini, J. Leach, M. P. J. Lavery, M. J. Padgett, and R. W. Boyd, "Influence of atmospheric turbulence on optical communications using orbital angular momentum for encoding," Opt. Express **20**, 13195–13200 (2012).

[3] M. N. O'Sullivan, M. Mirhosseini, M. Malik, and R. W. Boyd, "Near-perfect sorting of orbital angular momentum and angular position states of light," Opt. Express **20**, 24444–24449 (2012).

[4] P. Villoresi, T. Jennewein, F. Tamburini, M. Aspelmeyer, C. Bonato, R. Ursin, C. Pernechele, V. Luceri, G. Bianco, A. Zeilinger *et al.*, "Experimental verification of the feasibility of a quantum channel between space and earth," New J. Phys. **10**, 033038 (2008).

[5] M. Krenn, J. Handsteiner, M. Fink, R. Fickler, and A. Zeilinger, "Twisted photon entanglement through turbulent air across vienna," P. Natl. Acad. Sci. USA **112**, 14197–14201 (2015).

[6] A. F. Abouraddy, T. M. Yarnall, and B. E. A. Saleh, "Generalized optical interferometry for modal analysis in arbitrary degrees of freedom," Opt. Lett. **37**, 2889–2891 (2012).

[7] J. A. Rodrigo, T. Alieva, and M. L. Calvo, "Programmable two-dimensional optical fractional Fourier processor," Opt. Express **17**, 4976–4983 (2009).

[8] A. F. Abouraddy, T. M. Yarnall, and B. E. Saleh, "Angular and radial mode analyzer for optical beams," Opt. Lett. **36**, 4683–4685 (2011).

[9] A. E. Siegman, *Lasers* (University Science Books, 1986).

[10] D. Mardani, A. F. Abouraddy, and G. K. Atia, "Efficient modal analysis using compressive optical interferometry," Opt. Express **23**, 28449–28458 (2015).

[11] B. H. Dean, D. L. Aronstein, J. S. Smith, R. Shiri, and D. S. Acton, "Phase retrieval algorithm for jwst flight and testbed telescope," Proc. SPIE pp. 626511–626511 (2006).

[12] D. L. Aronstein, J. S. Smith, T. P. Zielinski, R. Telfer, S. C. Tournois, D. B. Moore, and J. R. Fienup, "Wavefront-error performance characterization for the james webb space telescope (jwst) integrated science instrument module (isim) science instruments," Proc. SPIE **9904**, 990409–990409 (2016).

[13] R. P. Millane, "Phase retrieval in crystallography and optics," J. Opt. Soc. Am. A **7**, 394–411 (1990).

[14] L.-H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller, "Experimental robustness of Fourier

ptychography phase retrieval algorithms," Opt. Express **23**, 33214–33240 (2015).

[15] J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).

[16] J. R. Fienup and C. C. Dainty, "Phase retrieval and image reconstruction for astronomy," Image recovery: Theory and application pp. 231–275 (1987).

[17] M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," Opt. Express **16**, 7264–7278 (2008).

[18] O. Shapira, A. F. Abouraddy, J. D. Joannopoulos, and Y. Fink, "Complete modal decomposition for optical waveguides," Phys. Rev. Lett. **94**, 143902 (2005).

[19] B. M. Hennelly and J. T. Sheridan, "Generalizing, optimizing, and inventing numerical algorithms for the fractional Fourier, Fresnel, and linear canonical transforms," J. Opt. Soc. Am. A **22**, 917–927 (2005).

[20] M. A. Alonso, "Wigner functions in optics: describing beams as ray bundles and pulses as particle ensembles," Adv. Opt. Photonics **3**, 272–365 (2011).

[21] H. M. Ozaktas, Z. Zalevsky, and M. A. Kutay, *The Fractional Fourier*

transform: with applications in optics and signal processing (Wiley, New York, 2001), 1st ed.

[22] J. Nocedal and S. J. Wright, *Numerical optimization 2nd edition* (Springer, 2006).

[23] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, "Scipy 1.0: Fundamental algorithms for scientific computing in Python," Nat. Methods **17**, 261–272 (2020).

[24] A. S. Jurling and J. R. Fienup, "Applications of algorithmic differentiation to phase retrieval algorithms," J. Opt. Soc. Am. A **31**, 1348–1359 (2014).

[25] J. R. Fienup and C. C. Wackerman, "Phase-retrieval stagnation problems and solutions," J. Opt. Soc. Am. **3**, 1897–1907 (1986).

[26] M. Guizar-Sicairos and J. R. Fienup, "Understanding the twin-image problem in phase retrieval," J. Opt. Soc. Am. **29**, 2367–2375 (2012).

[27] W. E. Farriss, J. R. Fienup, T. Malhotra, and A. N. Vamivakas, "Phase retrieval in generalized optical interferometry systems," Opt. Express **26**, 2191–2202 (2018).

[28] R. Bellman, *Dynamic Programming* (Princeton University Press, Princeton, NJ, USA, 1957), 1st ed.

Chapter 6

Conclusions and Future Work

6.1 Key Contributions

6.1.1

In Chapter 2 we laid out the necessary conceptual framework to understand the material presented in Chapters 3 and 4. Additionally, we presented our K-formalism which we use to approach sampling and propagation using the SAS. This formalism eased the burden of understanding sampling requirements for plane-to-plane propagation and permits the user a the ability to conduct discrete Fresnel transforms using fewer discrete parameters than would otherwise required. We also presented a novel DOF-based approach used to determine the limits of separability between range-compression integration and spatial phase propagation integrals when conducting RCH imag-

ing.

6.1.2 Multiplane Turbulence Mitigation Simulations Using Iterative Sharpness Maximization

In Chapter 3 we developed and tested our novel ISM algorithm used to mitigate the effects of turbulence in RCH imaging. Our monte carlo study showed success for noiseless cases, greatly improving range image fidelity for turbulence strengths up to and including $D/r_0 = 36$ and $\kappa = 128$ using four estimated phase screens and only a single speckle realization of simulated data. Additionally, we were able to test the robustness of our ISM turbulence mitigation in the presence of shot noise, characterizing the drop-off in algorithm performance as a function mtpps. We showed that our algorithm was unable to improve images through phase plane estimation with 10^{-2} mtpps (understandably since this is very little target signal at all). We showed that the algorithm's turbulence mitigation performance increased for the 10^{-1} mtpps case and continued increasing until achieving near-parity with the noiseless case at 10^{2} mtpps level.

6.1.3 Experimental Turbulence Mitigation in Range-compressed Holography

In Chapter 4, we constructed a RCH laboratory system using a CW-LFM laser and a high speed IR detector. After calibrating this system to ensure that our datasets attained the best possible SNR for the available components and system design, we used our lab system to generate high quality 2D intensity and 3D range images. After producing these high quality results, we introduced Lexitek phase plates to the path of optical propagation, producing good approximations to turbulence aberrated fields with significant anisoplanatism. We presented two case studies showing that our ISM algorithm was able to correct the effects of phase turbulence. For a mild to moderate turbulence profile created using two phase plates to aberrate our fields, our algorithm achieved results which were nearly indistinguishable from cases with no phase turbulence effects. In the second case using three phase plates to aberrate our fields, our algorithm was able to profoundly improve the images aberrated by severe turbulence, taking an amplitude image which was nearly indistinguishable from an empty frame and making an image which was clearly very similar to the unaberrated amplitude. The corrected range image in the severe case achieved a mildly aberrated but clearly recognizable image of our target, having started from an unrecognizable series of colored blobs.

6.1.4 GOI Phase Retrieval

In our GOI phase retrieval work we have shown that, using a modest number of random starting guesses for phase values, both array detector data and bucket detector data can reliably be used to retrieve the modal coefficient phases of fields which are composed of superposed HG modes.

Our array detector phase retrieval algorithm showed robustness to shot noise, achieving near 100% success for fields composed of up to 21 superposed HG modes with as few as 10^5 mean signal photons per array detector measurement.

Our single pixel phase retrieval work showed most importantly that, in theory, a GOI can generate enough information diversity to successfully retrieve phases using only a single pixel of measurement for each unique GPO setting. Additionally, we showed that the ability to retrieve modal phases in the single pixel case is dependent on both the number of measurements per unknown (mpu) and photons per unknown (ppu). We also showed that, regardless of how many mpu or ppu are available, it gradually becomes more difficult to retrieve phases using our single pixel approach as the number of parameters we wish to estimate increases.

6.2 Future Work

This thesis has shown work that ISM algorithms can be used to mitigate the effects of atmospheric turbulence, both in simulation and experiment, and that phase retrieval can be conducted on data from generalized optical interferometry systems to enable complete characterization of the measured coherent scalar field.

That said, our work has also presented us with a myriad of ideas regarding work which could and should be conducted to improve and/or better utilize these techniques/systems.

6.2.1 Improving the ISM Algorithm

The ISM algorithm has been shown to be a useful method of mitigating the effects of phase turbulence in our work. However, application of this algorithm has always been ad hoc, tuning empirical parameters, e.g. sharpness coefficient, penalty term weighting coefficients, method of sieves kernel standard deviation size, etc., to suit the exact situation the imaging system is presented with.

The end state goal for this algorithm is to enable turbulence mitigation in long range RCH imaging systems, and to ideally execute this task in real time. In order to accomplish this, work must be completed which

makes the algorithm adaptable to changing targets and conditions. Future work should, in part, focus on discovering which empirical parameters work best in a given situation and, ideally, how these parameter should change with variables such as target contrast, turbulence severity, distance to target, target reflectivity, etc. Additionally, future development should also focus on making this algorithm a continuously updating "online" algorithm which continually improves turbulence estimates. This work should consider adapting other control theory-based online turbulence mitigation algorithms which have been developed for applications such as adaptive optics correction of turbulent effects in terrestrial-based astronomical telescopes. These algorithms have proven very effective at improving system resolution.

Finally, we believe that while it would add hardware complexity to the imaging system, the addition of a structured illuminator, i.e. an illuminator which creates some known pattern with predictable null amplitude regions on the target, should be explored. We believe this would be extremely helpful in creating additional optimization constraints which would aid not only in estimating phase aberrations, but could also prove helpful in the mitigation of log-amplitude aberrations, which we did not study in this work.

[Write about exploring degenerate phase solutions using a data consistency metric on simulated turbulence to force the creation of degenerate phase solutions and to study these solutions]

6.2.2 Improving the RCH Laboratory Apparatus

Though our lab apparatus ultimately proved adequate to show experimental proof that our algorithm could indeed mitigate turbulent effects in real world image data, improvements to the current system should be considered prior to attempting any additional turbulence mitigation studies. In order to be able to sample our chirp spectrum, we needed to use a camera with an extremely high frame rate. This resulted in a trade-off where we accepted both a smaller number of effective pixels in our detector array and considerable amount of spatially varying detector noise. Our decision to bin pixels to boost effective frame rate resulted in our data being critically sampled. Whereas this is adequate for imaging under ideal conditions, it offered very little data redundancy to help overcome turbulent effects and noise when conducting turbulence mitigation in post-processing.

Though we implemented both coherent and incoherent summing techniques to make our data more robust, these methods do reach a point of diminishing returns and are not a perfect substitute for per-frame SNR. Purchasing a camera with a 16.45 KHz frame rate (using no pixel binning) with decreased noise would likely prove very helpful, though we understand this is a non-trivial detector engineering challenge.

Aside from the camera improvements, augmenting laser power by using our laser as the seed for a master oscillator power amplifier setup would also be useful. Our need to engage in coherent summing of frames to increase

SNR would become less important if each chirp-sampled frame had a higher SNR. Moreover, this increased laser power would permit us to use a more distant target and attain more independent speckle realizations per radian of stage rotation.

6.2.3 Future Work with the GOI

Our GOI phase retrieval work has generated some interesting ideas. As very fundamental research, it is ripe with opportunity to adapt the system and techniques to address various challenges. Foremost in our mind, we consider the ability to use mode suppression, caused by the summing of GPO terms in the output field, as a mechanism to greatly expedite modal coefficient analysis by creating a small cascade of GOI interferometers, with each previous interferometer's output creating the input for two new interferometers. We think about this as an optical processing approach to dimensionality reduction in our optimization problem, with the output field of each interferometer deeper in the cascade consisting of fewer HG modes than were input into the last interferometer. We propose to use the coefficients of this expedited modal analysis as a feature set for target recognition applications, avoiding the commonly used and computationally costly practice of principal component analysis of isolating data into orthogonal basis vectors and coefficients.

There is a great deal of fundamental work which would need to be real-

ized prior to achieving this goal. First, finding the most economical/efficient arrangement of GPO settings in the interferometer cascade is essential to the process. Additionally, a preliminary study of the performance of HG mode coefficients as features in a target recognition problem needs to be conducted. This study should compare the relative worth of the HG modes as a feature set in terms of its per photon information as well as how HG modes/coefficients compare to other more traditional feature sets in terms of required computation and processing time.

We believe that a technique like the one described here could have significant SNR advantages over more traditional methods of feature extraction from images because of the presence of phase information, the ability to boost signal through heterodyne gain, and the ability to use only bucket detectors.

Chapter 7

Appendices

7.1 Analytical Gradient

7.1.1 Algorithmic Differentiation

The derivation and use of analytic gradients was necessary for timely execution of our nonlinear optimizations. We eased the burden of this often cumbersome and error-prone task of deriving an analytic gradient by employing a technique called algorithmic differentiation [1]. If what follows does not immediately strike you as compact and non-burdensome, we would encourage the reader to consider the alternative, explicit method of gradient derivation for a very similar objective function contained in [2].

Algorithmic differentiation allows one to construct an analytic gradient by using the elementary operations of the optimization model to derive their

corresponding elementary derivative pieces. After finding these derivatives, the full analytic gradient is typically straight-forward to assemble through algebraic substitution (as we will show).

This technique, at its core, is a user-friendly way of implementing the chain rule and combining different contributions to a gradient. We include five tables in this appendix, each corresponding to one of the five optimization model components described in Section 3.1. Table 7.1 shows the necessary optimization model elementary operations (left columns) needed for Eq. (3.5), Table 7.2 the components necessary for Eq. (3.9), Table 7.3 the components for Eq. (3.8), Table 7.4 the components for Eq. (3.2), and finally Table 7.5 the components for Eq. (3.3). The overbar notation signifies a partial derivative of the final output quantity of the physical model with respect to the symbol under the bar. For example in Table 7.1

$$\frac{\partial \jmath}{\partial \tilde{U}_S} = \overline{\tilde{U}_S}. \tag{7.1}$$

Note that a single elementary operation in the optimization model can result in more than one gradient operation, as seen in Table 7.3 for the q_s elementary operation. When this occurs, we have placed a dash in the optimization model column below the original optimization model component in order to provide two cells in the right column for the necessary gradient components. To promote compact notation, only terms on the left side of the equation in the table columns are accompanied by their function arguments. Opti-

Table 7.1: Sharpness component algorithmic differentiation components needed for Eqs. (3.5).

Model component	∇ component	Ref [1] Eqs.
$\mathfrak{z} = \frac{\operatorname{sgn}(\alpha-1)}{\mathfrak{z}_{\text{init}}} \sum_{\vec{r}_{\perp S},z} E$	$\overline{E} = \frac{\operatorname{sgn}(\alpha-1)}{\mathfrak{z}_{\text{init}}} \overline{\mathfrak{z}}$	(47),(51)
$E(\vec{r}_{\perp S}, z') = I_{\text{avg}}^{\alpha}$	$\overline{I_{\text{avg}}}(\vec{r}_{\perp S}, z') = \alpha I_{\text{avg}}^{\alpha-1} \overline{E}$	(52)
$I_{\text{avg}}(\vec{r}_{\perp S}, z') = (1/\eta) \sum_{\eta}^{\eta} I$	$\overline{I}(\vec{r}_{\perp S}, z') = (1/\eta) \overline{I_{\text{avg}}}$	(47),(51)
$I(\vec{r}_{\perp S}, z', \eta) = \lvert\tilde{U}_S\rvert^2$	$\overline{\tilde{U}_S}(\vec{r}_{\perp S}, z', \eta) = 2\tilde{U}_S \circ \overline{I}$	(53)
$\tilde{U}_S(\vec{r}_{\perp S}, z', \eta) = \mathscr{F}_{2\nu/c \to z}\{U_S\}$	$\overline{U_S}(\vec{r}_{\perp S}, \nu, \eta) = \mathscr{F}^{-1}_{z' \to 2\nu/c}\left\{\overline{\tilde{U}_S}\right\}$	(80)

mization model components represented by an upper-case letter (Greek or Roman) are assumed to be computationally represented by arrays with two or more dimensions, while lower-case letters represent scalar or values which are a one dimensional array with length S.

A single explicit expression for the gradient of a quantity of interest can be constructed by starting in the top row of the gradient column of a table and plugging each respective overbar expression (or expression from the optimization model in some cases) into the partial derivative component(s) which is(are) dependent upon that overbar expression in the rows beneath it. Continue substitution until only a single expression for the gradient remains.

Table 7.2: Pupil penalty term algorithmic differentiation components necessary for Eq. (3.9).

Model component	∇ component	Ref [1] Eqs.
$p_C = (\lambda_c/\epsilon_{init})j$	$\overline{j} = (\lambda_c/\epsilon_{\text{init}})\overline{p_C}$	(51)
$j = \sum_{\vec{r}_{\perp 0},\nu,\eta} K$	$\overline{K} = \overline{j}$	(47)
$K(\vec{r}_{\perp 0},\nu,\eta) = \mathscr{M}(\vec{r}_{\perp 0}) \circ L$	$\overline{L}(\vec{r}_{\perp 0}) = \mathscr{M}(\vec{r}_{\perp 0}) \circ \overline{K}$	(49)
$L(\vec{r}_{\perp 0},\nu,\eta) = \lvert\breve{U}_0\rvert^2$	$\overline{\breve{U}_0}(\vec{r}_{\perp 0},\nu,\eta) = 2\breve{U}_0 \circ \overline{L}$	(53)
$\breve{U}_0(\vec{r}_{\perp 0},\nu,\eta) = \mathscr{P}_{S\to 0}\{U_S\}$	$\overline{U_S}(\vec{r}_{\perp S},\nu,\eta) = \mathscr{P}^{-1}_{0\to S}\{\overline{\breve{U}_0}\}$	(49),(80),(81)

Table 7.3: Defocus penalty term algorithmic differentiation components needed for Eq. (3.8).

Model component	∇ component	Ref [1] Eqs.
$p_D = (\lambda_D/S)\sum_s q_s$	$\overline{q_s} = (\lambda_D/S)\overline{p_D}$	(47),(51)
$q_s = r_s s_s$	$\overline{r_s} = s_s\overline{q_s}$	(49)
-	$\overline{s_s} = r_s\overline{q_s}$	(49)
$s_s = t_s^{-1}$	$\overline{t_s} = -t_s^{-2}\overline{s_s}$	(52)
$r_s = \chi_s^2$	$\overline{\chi_s} = 2\chi_s\overline{r_s}$	(52)
$\chi_s = \sum_{\vec{r}_{\perp s}} V_s$	$\overline{V_s} = \overline{\chi_s}$	(47)
$V_s(\vec{r}_{\perp s}) = \hat{\Phi}_s \circ Z_4$	$\overline{\hat{\Phi}_{s,1}}(\vec{r}_{\perp s}) = \overline{V_s} \circ Z_4$	(49)
$t_s = \sum_{\vec{r}_{\perp s}} F_s$	$\overline{F_s} = \overline{t_s}$	(47)
$F_s(\vec{r}_{\perp s}) = \hat{\Phi}_s^2 + \beta$	$\overline{\hat{\Phi}_{s,2}}(\vec{r}_{\perp s}) = 2\hat{\Phi}_s\overline{F_s}$	(47),(52)

Table 7.4: Propagation algorithmic differentiation components necessary for Eq. (3.2).

Model component	∇ component	Ref [1] Eqs.
$U_s(\vec{r}_{\perp s}\nu,\eta) = \mathscr{P}_{s-1\to s-1}\{U'_{s-1}\}$	$\overline{U'_{s-1}}(\vec{r}_{\perp s-1},\nu,\eta) = \mathscr{P}^{-1}_{s\to s-1}\{\overline{U_s}\}$	(49),(80),(81)
$U'_{s-1}(\vec{r}_{\perp s-1}\nu,\eta) = U_{s-1}\circ\Psi_{s-1}$	$\overline{\Psi_{s-1}}(\vec{r}_{\perp s-1},\nu,\eta) = U^*_{s-1}\circ\overline{U'_{s-1}}$	(49)
-	$\overline{U_{s-1}}(\vec{r}_{\perp s-1},\nu,\eta) = \Psi^*_{s-1}\circ\overline{U'_{s-1}}$	(49)

Table 7.5: Phase screen algorithmic differentiation components needed for (3.3).

Model component	∇ component	Ref [1] Eqs.
$\Psi_s(\vec{r}_{\perp s}) = \exp(iA_s)$	$\overline{A}_s(\vec{r}_{\perp s}) = \sum_{\nu,\eta}\mathrm{Im}\{\Psi^*_s\circ\overline{\Psi_s}\}$	(57)
$A_s(\vec{r}_{\perp s}) = \hat{\Phi}_s * G_s$	$\overline{\hat{\Phi}}_s(\vec{r}_{\perp s}) = \overline{A_s} * G_s$	(49),(80),(81)

7.1.2 Explicit Gradient Expressions

As an example, consider our Table 7.3, the most complicated of our derivative tables. To construct the explicit gradient $\left[\overline{\hat{\Phi}_s}\right]_{p_D}$, we first substitute $\overline{q_s}$ into $\overline{r_s}$

$$\overline{r_s} = \frac{\lambda_D}{S}s_s = \frac{\lambda_D}{S}t_s^{-1} = \frac{\lambda_D}{S}\left(\sum_{\vec{r}_{\perp s}}U_s(\vec{r}_{\perp s})\right)^{-1} = \frac{\lambda_D}{S}\left(\sum_{\vec{r}_{\perp s}}\hat{\Phi}_s^2(\vec{r}_{\perp s})+\beta\right)^{-1}, \tag{7.2}$$

and also into $\overline{s_s}$

$$\overline{s_s} = \frac{\lambda_D}{S} r_s = \frac{\lambda_D}{S} \chi_s^2 = \frac{\lambda_D}{S} \left(\sum_{\vec{r}_{\perp s}} V_s(\vec{r}_{\perp s}) \right)^2 = \frac{\lambda_D}{S} \left(\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s(\vec{r}_{\perp s}) \circ Z_4(\vec{r}_{\perp s}) \right)^2, \tag{7.3}$$

since $\overline{p_D} = \partial p_D / \partial p_D = 1$. Next, we insert Eq. (7.2) into $\overline{\chi_s}$

$$\overline{\chi_s} = \frac{2\lambda_D}{S} \left(\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s(\vec{r}_{\perp s}) \circ Z_4(\vec{r}_{\perp s}) \right) \left(\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s^2(\vec{r}_{\perp s}) + \beta \right)^{-1}, \tag{7.4}$$

and then insert Eq. (7.3) into $\overline{t_s}$

$$\overline{t_s} = -\frac{\lambda_D}{S} \left(\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s^2(\vec{r}_{\perp s}) + \beta \right)^{-2} \left(\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s(\vec{r}_{\perp s}) \circ Z_4(\vec{r}_{\perp s}) \right)^2. \tag{7.5}$$

We observe that $\overline{\chi_s} = \overline{V_s}$ and thus we insert $\overline{\chi_s}$ directly into $\overline{\Phi_s}$

$$\overline{\hat{\Phi}_{s,1}}(\vec{r}_{\perp s}) = \frac{2\lambda_D}{S} Z_4(\vec{r}_{\perp s}) \left(\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s(\vec{r}_{\perp s}) \circ Z_4(\vec{r}_{\perp s}) \right) \left(\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s^2(\vec{r}_{\perp s}) + \beta \right)^{-1}. \tag{7.6}$$

Similarly, since $\overline{t_s} = \overline{U_s}$ we insert Eq. (7.5) into $\overline{\hat{\Phi}_s}(\vec{r}_{\perp s})$

$$\overline{\hat{\Phi}_{s,2}}(\vec{r}_{\perp s}) = \frac{-2\lambda_D}{S} \Phi_s(\vec{r}_{\perp s}) \left(\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s^2(\vec{r}_{\perp s}) + \beta \right)^{-2} \left(\sum_{\vec{r}_{\perp s}} \hat{\Phi}_s(\vec{r}_{\perp s}) \circ Z_4(\vec{r}_{\perp s}) \right)^2. \tag{7.7}$$

The overall gradient of p_D with respect to $\hat{\Phi}_s(\vec{r}_{\perp s})$ is

$$\left[\overline{\hat{\Phi}_s}(\vec{r}_{\perp s})\right]_{p_D} = \overline{\hat{\Phi}_{s,1}}(\vec{r}_{\perp s}) + \overline{\hat{\Phi}_{s,2}}(\vec{r}_{\perp s}) = \frac{2\lambda_D}{S}\left(\frac{\chi_s}{t_s^2}\right)\left[t_s Z_4(\vec{r}_{\perp s}) - \chi_s \hat{\Phi}_s(\vec{r}_{\perp s})\right], \tag{7.8}$$

where $\chi_s = \sum_{\vec{r}_{\perp s}} \hat{\Phi}_s(\vec{r}_{\perp s}) \circ Z_4(\vec{r}_{\perp s})$ and $t_s = \sum_{\vec{r}_{\perp s}} \hat{\Phi}_s^2(\vec{r}_{\perp s}) + \beta$.

Using the same substitution method from Tables 7.1, 7.4, and 7.5 the sharpness portion of our gradient is expressed

$$\begin{aligned}\left[\overline{\hat{\Phi}_s}(\vec{r}_{\perp s})\right]_{\mathfrak{z}} = \frac{2\alpha\,\mathrm{sgn}\,(\alpha-1)}{H\mathfrak{z}_{\text{init}}}\Bigg[\sum_{\nu,\eta}\mathrm{Im}\Bigg\{&\Psi_s^*(\vec{r}_{\perp s}) \circ U_s^*(\vec{r}_{\perp s},\nu,\eta) \\ &\circ \mathscr{P}_{s+1\to s}^{-1}\Bigg\{\Psi_{s+1}^*(\vec{r}_{\perp s+1}) \circ \mathscr{P}_{s+2\to s+1}^{-1}\Bigg\{\Psi_{s+2}^*(\vec{r}_{\perp s+2}) \circ \dots \\ &\circ \mathscr{P}_{S\to S-1}^{-1}\Bigg\{\mathscr{F}_{z'\to 2\nu/c}\Bigg\{\tilde{U}_S(\vec{r}_{\perp s},z',\eta) \circ I_{\text{avg}}^{\alpha-1}(\vec{r}_{\perp s},z')\Bigg\}\Bigg\}\dots\Bigg\}\Bigg\}\Bigg\}\Bigg].\end{aligned} \tag{7.9}$$

Finally, the pupil penalty term contribution to the gradient is expressed

$$\begin{aligned}\left[\overline{\hat{\Phi}_s}(\vec{r}_{\perp s})\right]_{p_C} = \frac{2\lambda_C}{\epsilon_{\text{init}}}\Bigg[\sum_{\nu,\eta}\mathrm{Im}\Bigg\{&\Psi_s^*(\vec{r}_{\perp s}) \circ U_s^*(\vec{r}_{\perp s},\nu,\eta) \\ &\circ \mathscr{P}_{s+1\to s}^{-1}\Bigg\{\Psi_{s+1}^*(\vec{r}_{\perp s+1}) \circ \mathscr{P}_{s+2\to s+1}^{-1}\Bigg\{\Psi_{s+2}^*(\vec{r}_{\perp s+2}) \circ \dots \\ &\circ \mathscr{P}_{S\to S-1}^{-1}\Bigg\{\mathscr{P}_{0\to S}\Bigg\{\breve{U}_0(\vec{r}_{\perp 0},\nu,\eta) \circ \mathscr{M}(\vec{r}_{\perp 0})\Bigg\}\Bigg\}\dots\Bigg\}\Bigg\}\Bigg\}\Bigg],\end{aligned} \tag{7.10}$$

using elementary operations in Tables 7.2, 7.4, and 7.5. The gradient of our overall objective function Eq. (3.11) is then

$$\overline{\hat{\Phi}_s}(\vec{r}_{\perp s}) = \left[\overline{\hat{\Phi}_s}(\vec{r}_{\perp s})\right]_{\jmath} - \left[\overline{\hat{\Phi}_s}(\vec{r}_{\perp s})\right]_{p_C} - \left[\overline{\hat{\Phi}_s(\vec{r}_{\perp s})}\right]_{p_D}. \tag{7.11}$$

7.2 Penalty Term Bounds

As mentioned in Section 3.1, we paid special attention to the design of our penalty terms, Eqs. (3.8) and (3.9), in order to maintain control of the scale of our overall objective function, Eq. (3.11), during optimization. To do this, we designed our penalty terms such that they were bounded between zero and their scaling coefficients, λ_D and λ_C.

The maximum value of Eq. (3.9) clearly only occurs when the combined energy of all fields ends up in the region of the mask function $\mathscr{M}(\vec{r}_{\perp 0})$ with unity transmittance. In this special case

$$p_C = \frac{\lambda_c}{\epsilon_{\text{init}}} \sum_{\vec{r}_{\perp 0},\nu} |U_0(\vec{r}_{\perp 0}, \nu)|^2. \tag{7.12}$$

We observe that the summation above is equivalent to Eq. (3.10), ϵ_{init}. Thus, the maximum possible value of $p_C = \lambda_c$.

To determine the maximum value of Eq. (3.8), we consider it's gradient expression, Eq. (7.8). The only non-trivial case where this expression equals

zero is when

$$Z_4 t_s - \hat{\Phi}_s \chi_s = 0. \tag{7.13}$$

To accomplish this, we must assume the extreme case where all of the phase in our estimated screens is defocus, Z_4 and can thus be expressed $\hat{\Phi}_s = \alpha_s Z_4$, where α_s is a real-valued, scalar weighting coefficient. We assume our stabilizing constant $\beta = 0$ for the purposes of finding the maximum value. In this case

$$\begin{aligned} Z_4 t_s - \hat{\Phi}_s \chi_s &= Z_4 \sum_{\vec{r}_{\perp s}} (\alpha_s Z_4)^2 - \alpha_s Z_4 \sum_{\vec{r}_{\perp s}} \alpha_s Z_4^2 \\ &= \alpha_s^2 Z_4 \left(\sum_{\vec{r}_{\perp s}} Z_4^2 - \sum_{\vec{r}_{\perp s}} Z_4^2 \right) \\ &= 0. \end{aligned} \tag{7.14}$$

Plugging this condition for the maximum into Eq. (3.8) and assuming orthonormalized Zernikes

$$\begin{aligned} p_D &= \frac{\lambda_D}{S} \sum_s \frac{\left(\sum_{\vec{r}_{\perp s}} \alpha_s Z_4 \right)^2}{\sum_{\vec{r}_{\perp s}} \alpha_s^2 Z_4^2} \\ &= \frac{\lambda_D}{S} \sum_s \frac{\left(\sum_{\vec{r}_{\perp s}} Z_4^2 \right)^2}{\sum_{\vec{r}_{\perp s}} Z_4^2} \\ &= \lambda_D. \end{aligned} \tag{7.15}$$

7.3 The Fractional Fourier Transform Operator Eigenvalue Relationship

Here, we explicitly derive the eigenvalue relationship between HG modes and the fFT operator. The following derivation is composed of elements in [3] and [4].

Assume that one desires to construct an operator that exhibits the following eigenfunction relationship when applied to a Hermite-Gaussian function,

$$\mathscr{F}_\alpha \left\{ e^{-\frac{x^2}{2}} H_m(x) \right\} = e^{\frac{im\pi\alpha}{2}} e^{-\frac{x^2}{2}} H_m(x), \tag{7.16}$$

where $H_m(x)$ is the Hermite polynomial of degree m and α is the order parameter of the desired operator, and x is a unitless quantity.

Since Hermite-Gaussian polynomials form a basis set, any complex-valued one dimensional function can be described using a linear superposition of weighted Hermite-Gaussian polynomials

$$f(x) = \sum_{m=0}^{\infty} c_m e^{-\frac{x^2}{2}} H_m(x), \tag{7.17}$$

where the weighting coefficient, c_m, is generally complex-valued. The results of this discussion in 1-D are generalizable to the 2-D case, which is more relevant to freespace scalar wave propagation.

The orthogonality relation of the Hermite-Gaussian modes is

$$\int_{-\infty}^{\infty} H_m(x)H_s(x)e^{-x^2}dx = \delta_{mn}\sqrt{\pi}2^m m!. \tag{7.18}$$

In order to find c_m, we use this orthogonality relation on an arbitrary function that is expressed in its HG mode decomposition

$$\begin{aligned}\int_{-\infty}^{\infty} H_m(x)e^{-\frac{x^2}{2}}f(x)dx &= \int_{-\infty}^{\infty} H_m(x)e^{-\frac{x^2}{2}}\left(\sum_{n=0}^{\infty} c_s e^{-\frac{x^2}{2}}H_s(x)\right)dx \\ &= \sum_{n=0}^{\infty} c_s \int_{-\infty}^{\infty} H_m(x)H_s(x)e^{-x^2}dx \\ &= \sum_{n=0}^{\infty} c_s \delta_{mn}\sqrt{\pi}2^n n! \\ &= c_m\sqrt{\pi}2^m m!.\end{aligned} \tag{7.19}$$

$$\therefore c_m = \frac{1}{\sqrt{\pi}2^m m!}\int_{-\infty}^{\infty} H_m(x)e^{-\frac{x^2}{2}}f(x)dx. \tag{7.20}$$

From Eqs. (7.16) and (7.17)

$$\mathscr{F}_\alpha\{f(x)\} = \sum_{m=0}^{\infty} c_m e^{\frac{im\pi\alpha}{2}}H_m(x)e^{-\frac{x^2}{2}}. \tag{7.21}$$

Now we consider Mehler's Hermite Polynomial Formula [3, 4]

$$\begin{aligned}&\sum_{m=0}^{\infty}\frac{t^m}{2^m m!\sqrt{\pi}}H_m(x)H_m(x')e^{-\left(\frac{x^2+x'^2}{2}\right)} \\ &\qquad = \frac{1}{\sqrt{\pi\left(1-t^2\right)}}\exp\left[\frac{4xx't - \left(x^2+x'^2\right)\left(1+t^2\right)}{2\left(1-t^2\right)}\right],\end{aligned} \tag{7.22}$$

where t is just a placeholder variable. Let $t = \exp(i\pi\alpha/2)$. This implies

$$\sum_{m=0}^{\infty} \frac{e^{\frac{im\pi\alpha}{2}}}{2^m m!\sqrt{\pi}} H_m(x)H_m(x')e^{-\left(\frac{x^2+x'^2}{2}\right)} = \frac{1}{\sqrt{\pi\left(1-e^{i\pi\alpha}\right)}} \exp\left[\frac{4xx'e^{\frac{i\pi\alpha}{2}} - \left(x^2+x'^2\right)\left(1+e^{i\pi\alpha}\right)}{2(1-e^{i\pi\alpha})}\right]. \tag{7.23}$$

The fractional term on the right hand side of of Eq. (7.23) can be expressed

$$\frac{1}{\sqrt{\pi\left(1-e^{i\pi\alpha}\right)}} = \frac{1}{\sqrt{-2\pi i e^{\frac{i\pi\alpha}{2}}}\sqrt{\frac{e^{\frac{i\pi\alpha}{2}}-e^{\frac{-i\pi\alpha}{2}}}{2i}}} = \frac{e^{i\frac{\pi}{4}-i\frac{\pi\alpha}{4}}}{\sqrt{2\pi\sin\left(\frac{\pi\alpha}{2}\right)}}. \tag{7.24}$$

Next, we express the exponential term on the right hand side of Eq. (7.23) as

$$\begin{aligned} &\exp\left[\frac{4xx'e^{\frac{i\pi\alpha}{2}} - \left(x^2+x'^2\right)\left(1+e^{i\pi\alpha}\right)}{2(1-e^{i\pi\alpha})}\right] \\ &= \exp\left[\frac{e^{\frac{i\pi\alpha}{2}}4xx' - 2e^{\frac{i\pi\alpha}{2}}\left(x^2+x'^2\right)\left(\frac{e^{\frac{i\pi\alpha}{2}}+e^{\frac{-i\pi\alpha}{2}}}{2}\right)}{e^{\frac{i\pi\alpha}{2}}(-4i)\left(\frac{e^{\frac{i\pi\alpha}{2}}-e^{\frac{-i\pi\alpha}{2}}}{2i}\right)}\right] \\ &= \exp\left[-ixx'\csc\left(\frac{\pi\alpha}{2}\right) + \frac{i}{2}\left(x^2+x'^2\right)\cot\left(\frac{\pi\alpha}{2}\right)\right]. \end{aligned} \tag{7.25}$$

For our chosen t we can express Eq. (7.23) as

$$\sum_{m=0}^{\infty} \frac{e^{\frac{im\pi\alpha}{2}}}{2^m m!\sqrt{\pi}} H_m(x) H_m(x') e^{-\left(\frac{x^2+x'^2}{2}\right)} = \frac{\exp i\frac{\pi}{4} - i\frac{\pi\alpha}{4}}{\sqrt{2\pi \sin\left(\frac{\pi\alpha}{2}\right)}} \exp\left[-ixx' \csc\left(\frac{\pi\alpha}{2}\right) + \frac{i}{2}\left(x^2 + x'^2\right)\cot\left(\frac{\pi\alpha}{2}\right)\right]. \tag{7.26}$$

Substituting Eq. (7.20) into Eq. (7.21)

$$\begin{aligned}
\mathscr{F}_\alpha\{f(x)\} &= \sum_{m=0}^{\infty}\left[\frac{1}{2^m m!\sqrt{\pi}}\int_{-\infty}^{\infty} H_m(x)\exp\left(\frac{-x^2}{2}\right) f(x)dx\right] \\
&\qquad \times H_m(x')\exp\left(\frac{im\pi\alpha}{2}\right)\exp\left(\frac{-x'^2}{2}\right) \\
&= \int_{-\infty}^{\infty}\left\{\sum_{m=0}^{\infty}\frac{\exp\left(\frac{im\pi\alpha}{2}\right)}{2^m m!\sqrt{\pi}} H_m(x)H_m(x') \right. \\
&\qquad \left. \times \exp\left[\frac{-\left(x^2+x'^2\right)}{2}\right]\right\} f(x)dx \\
&= \frac{\exp\left(i\frac{\pi}{4} - i\frac{\pi\alpha}{4}\right)}{\sqrt{2\pi\sin\left(\frac{\pi\alpha}{2}\right)}}\int_{-\infty}^{\infty} f(x) \\
&\qquad \times \exp\left[-ixx'\csc\left(\frac{\pi\alpha}{2}\right) + \frac{i}{2}\left(x^2+x'^2\right)\cot\left(\frac{\pi\alpha}{2}\right)\right]dx.
\end{aligned} \tag{7.27}$$

Now performing a change of variables where $\sqrt{2\pi}y = x$, $dy = dx/\sqrt{2\pi}$, and

$\sqrt{2\pi}y' = x'$, we find the definition of the fFT [5]

$$\mathscr{F}_\alpha\{f(y)\} = \frac{e^{i\frac{\pi}{4}-i\frac{\pi\alpha}{2}}}{\sqrt{\sin\left(\frac{\pi\alpha}{2}\right)}}\int_{-\infty}^{\infty} f(y)e^{\left[-iyy'\csc\left(\frac{\pi\alpha}{2}\right)+\frac{i}{2}\left(y^2+y'^2\right)\cot\left(\frac{\pi\alpha}{2}\right)\right]}dy. \quad (7.28)$$

We therefore can conclude that

$$\mathscr{F}_\alpha\{f(x)\} = \mathscr{F}_\alpha\left\{\sum_{m=0}^{\infty} c_m H_m(x)e^{-\frac{x^2}{2}}\right\} = \sum_{m=0}^{\infty} c_m e^{\frac{im\pi\alpha}{2}} H_m(x)e^{-\frac{x^2}{2}}, \quad (7.29)$$

is an alternative form of the fFT of a function $f(x)$.

Bibliography

[1] A. S. Jurling and J. R. Fienup, "Applications of algorithmic differentiation to phase retrieval algorithms," J. Opt. Soc. Am. A **31**, 1348–1359 (2014).

[2] A. E. Tippie, *Aberration correction in digital holography* (PhD. thesis, University of Rochester, 2012).

[3] N. V., "The fractional order Fourier transform and its application to quantum mechanics," IMA J. Appl. Math. **25**, 241–265 (1980).

[4] G. N. Watson, "Notes on generating functions of polynomials:(2) hermite polynomials," J. London. Math. Soc. **1**, 194–199 (1933).

[5] J. A. Rodrigo, T. Alieva, and M. L. Calvo, "Programmable two-dimensional optical fractional Fourier processor," Opt. Express **17**, 4976–4983 (2009).