COMPUTATIONAL IMAGING THROUGH ATMOSPHERIC TURBULENCE

by

Nicholas Chimitt

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Electrical and Computer Engineering

West Lafayette, Indiana

August 2023

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Stanley H. Chan, Chair

School of Electrical and Computer Engineering

Dr. Mark R. Bell

School of Electrical and Computer Engineering

Dr. Charles A. Bouman

School of Electrical and Computer Engineering

Dr. Mary L. Comer

School of Electrical and Computer Engineering

Approved by:

Dr. Milind Kulkarni

Think I'll pack it in and buy a pick-up
take it down to L.A.

ACKNOWLEDGMENTS

When I look back on my Ph.D. experience at Purdue, I'll remember a cast of brilliant, kind, and funny people. The first I must acknowledge is my advisor, Professor Stanley Chan. I am thankful to him for a number of reasons that could run on for a great while. I am not alone in this, *every* student who has joined with me feels the same. I offer my sincerest thanks to him for countless hours of discussing ideas with me, the time spent laughing and talking about nonsense, and for being an excellent advisor. I also wish to extend thanks to my committee – Professors Mark Bell, Charlie Bouman, and Mary Comer. I appreciate each of you for your insights and good humour during lectures and our interactions.

Next, I should acknowledge the members of the i2lab. I specifically think of the group that joined with me: Abhiram, Zhiyuan, and Xiangyu, as well as the group not far behind in time of Yiheng, Guanzhe, Chengxi, and Yash, later adding Kent and Xingguang, and newer members Prateek, Preston, and Harshana. I'd also like to thank Misha and Akshat, two undergraduate members of the lab. I thank each of them for their participation in this process. I specifically wish to thank Abhiram, Zhiyuan, Yash, Harshana, and Xingguang; though three of them I have never had the good fortune of writing a paper with, all of them have had a great impact on me.

To my friends I have made along the way: Manish, Isha, Timur, Sarada, Deepak, and Mint I offer many thanks for many memories and laughs! I also would thank my friends Alex, Ethan, and Marc who have no idea what it is that I'm still doing in school but love to remind me that I'm still in school. I'd also thank my friends I met at PNC, Jason, Tony, Taimoor, Carroll, and Terri as well as Professors Lizhe Tan, Jean Jiang, and Alain Togbe without whom I never would've considered graduate school.

Finally, I wish to extend my thanks and love to my parents, brother, and the rest of my family. I will be marrying my fiancé Vanessa approximately one month from writing these words, and I couldn't be any happier or feel any more fortunate. I love you. Thank you for so many great memories and the many more to come.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Imaging at range for the purposes of biometric, scientific, or militaristic applications often suffer due to degradations by the atmosphere. These degradations, due to the non-uniformity of the atmospheric medium, can be modeled as being caused by turbulence. Dating back to the days of Kolmogorov in the 1940s, the field has had many successes in modeling and some in mitigating the effects of turbulence in images. Today, modern restoration methods are often in the form of learning-based solutions which require a large amount of training data. This places atmospheric turbulence mitigation at an interesting point in its history; simulators which accurately capture the effects of the atmosphere were developed without any consideration of deep learning methods and are often missing critical requirements for todays solutions.

In this work, we describe a simulator which is not only fast and accurate but has the additional property of being end-to-end differentiable, allowing for end-to-end training with a reconstruction network. This simulation, which we refer to as Zernike-based simulation, performs at a similar level of accuracy as its purely optics-based simulation counterparts while being up to 1000x faster. To achieve this we combine theoretical developments, engineering efforts, and learning-based solutions. Our Zernike-based simulation not only aids in the application of modern solutions to this classical problem but also opens the field to new possibilities with what we refer to as computational image formation.

1. INTRODUCTION

The world is a dynamic place filled with nearly constant motion. Many of these motions that we observe in our day to day life have been suitably described by the many great physicists before us. There is, however, one type of motion that continues to evade our mathematical capabilities despite our best efforts. This is the motion of **turbulent** fluid.

This is not a dissertation about *just* turbulence. It is about imaging through **atmospheric turbulence**. In modern life we may have some familiarity with these concepts through the look of a car on the road far ahead of us on a hot day, but outside of these extreme cases, it is not something we often think about. This means that this thesis describes modeling images which are taken through a phenomena which is both unfamiliar to us and not captured by the classical laws of physics. What, then, could possibly be said in this dissertation?

1.1 Context of this dissertation

I have been talking about **turbulence** for a while now and the more people I speak with, the more I realize that this word "turbulence" carries many different meanings to many different people. In fact, when people are shown images taken through **atmospheric turbulence** their intuition sometimes goes out the window: am I taking pictures through a tornado, during a hurricane, while falling off my chair? Therefore, it is important for me to tell you *exactly* what it is I mean when talking about turbulence, why it applies to the atmosphere, and why it affects images the way it does.

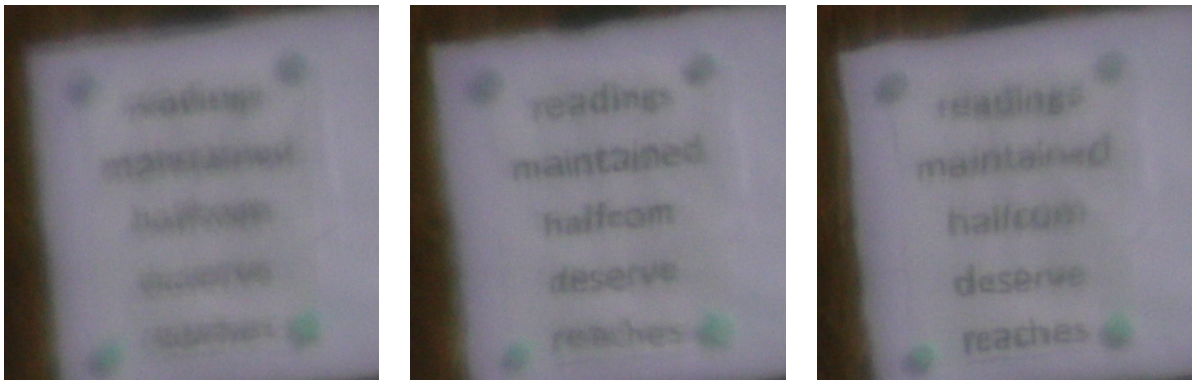


Figure 1.1. Images where we can observe degradation by the atmosphere.

In Figure 1.1, three images taken back-to-back are shown which I would describe as being degraded by atmospheric turbulence. What has happened to these images? You may wonder, under what *conditions* were these images taken? Were you in a desert when you took them? What was wrong with the camera?

The images in Figure 1.1 were taken on a regular weekday in West Lafayette, Indiana at a distance of approximately 500 meters. There are no tricks in the camera, this is truly what we observe when imaging over a long distance. The appearance of these images may remind you of looking at a heat vent releasing air into the cold winter or maybe just above the road on a hot day while driving. That is because, in essence, it is the same thing!

When air moves around on Earth, how do we think about its movement? Does it have a smooth motion with the wind moving in a constant direction or is it more *swirly*? If it were moving in a consistent direction we would say it exhibits **laminar** motion, otherwise, we would say it is **turbulent** motion. Van Gogh certainly had an opinion of this in his painting *The Starry Night* (Figure 1.2) in which he visualizes the calm night sky as having this swirling pattern to it. Looking at Van Gogh's painting, we can "feel" the air swirling around us, however, it is not chaotic or rapid or aggressive. This is because turbulence can be calm and slow-moving. Therefore, when I speak of turbulence, I only mean that the air dances in a somewhat swirling pattern; the speed and intensity at which it happens is of little concern.[1] The air is in a near-constant state of turbulent mixing, therefore, we can say the atmosphere is turbulent.

Even if we can accept the motion of the air between the object and camera which produced Figure 1.1 as turbulent, this still does not answer *why* the images look so unusual. The mathematical reason is well beyond the reach of this introduction, so how can we understand it more simply? In Figure 1.3, we can see an example of **refraction**. Refraction bends light according to a change in the **index of refraction**, with glass having an index of refraction $n \approx 1.33$ and air as approximately $n \approx 1$. However, the air's index of refraction isn't *exactly* 1 and it isn't *exactly* uniform. For a short distance we can ignore these details perfectly fine as an approximation. However, when a long distance is involved, it begins to matter much more than the case shown in Figure 1.3.

[1] ↑Of course, later I'll be more mathematically precise, but this line of thinking will still be useful!

Figure 1.2. Van Gogh, *The Starry Night.* [Source]

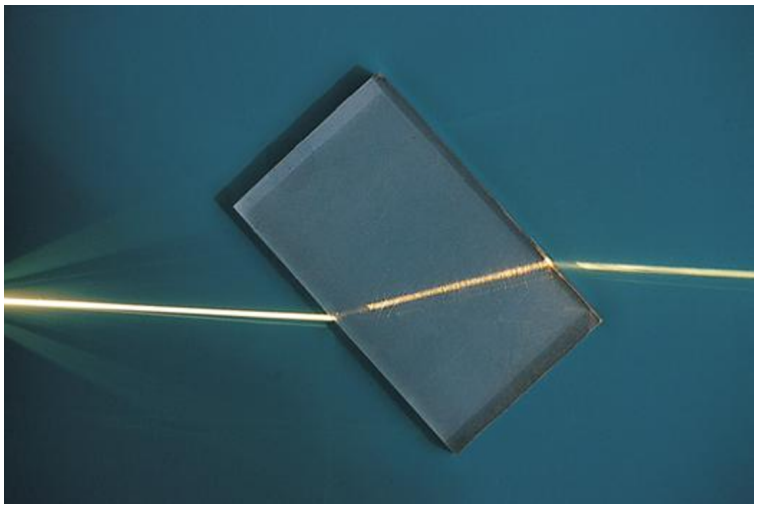


Figure 1.3. Refraction of light through glass. [Source]

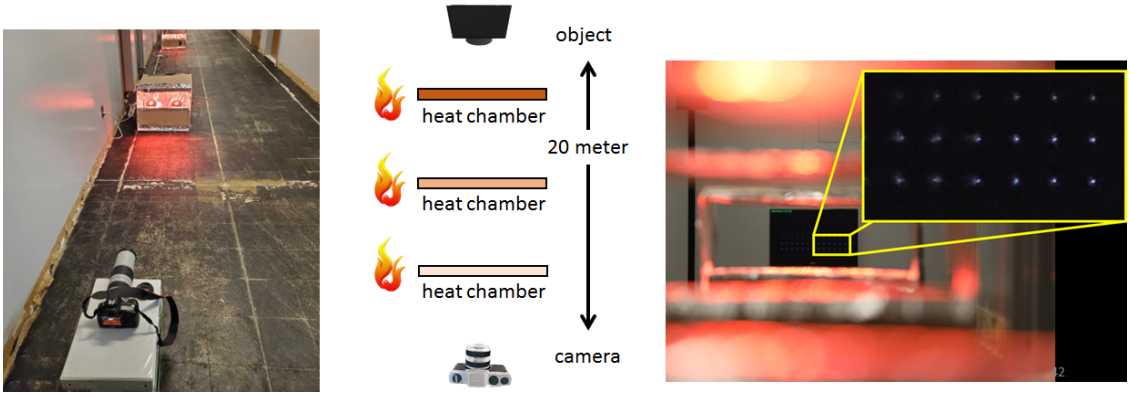


Figure 1.4. Heat chamber for turbulent imaging at close range.

Because of this turbulent mixing of the atmosphere, there will be collisions of particles in some places and particles moving away from other locations. There will also be temperature fluctuations, differences in the types of molecules which make up the air, and so on. All of these properties contribute to the index of refraction (of the air!) being non-uniform. The scale of these fluctuations in the atmosphere's index of refraction is so small, however, that they effectively don't bend the light at all. They instead act as roadblocks that make the light from a single point arrive at our camera at different times across the camera aperture. Believe it or not, this difference in arrival time is all it takes to mess the image up and give us Figure 1.1.

So why don't we see these effects all the time? If the air between you and the words of this text are in a state of turbulent mixing, why doesn't it appear distorted? These small differences in the index of refraction are simply not large enough to seriously impact the light emerging from an object. However, if we were to turn up the heat and *really* make the air turbulent, we'll increase these fluctuations in the index of refraction and thus these words would be distorted. This is why we see these effects near a heat vent or just above the road. In fact, we can exploit this to image turbulence at close range using a heat chamber as shown in Figure 1.4.

1.2 The Goal of This Thesis

The aim of this thesis is to make images like Figure 1.1 look better. This means we wish to **reconstruct** what they were supposed to look like. Of the many subdomains of Electrical Engineering, **computational imaging** uses **models** to describe the **forward process**, that is, how can we represent the process through which the image came to look as it does? To solve this problem, we must understand some of the physics involved. As it turns out, there is a lot to do for representing this forward process in a convenient manner for today's reconstruction tools.

In the last decade, data-driven machine learning algorithms have made great strides in solving problems old and new. The advantage of these models is that there are many tools to help develop and deploy them quickly as well as their effectiveness. A typical operating principle is to train them using a pair of data: the input is the degraded observation, the output is what it is supposed to look like. Once we learn this mapping from degraded input to corrected output, we can apply it to data which is similar but has not been included in our training process. At this point, the machine learning algorithm performs inference to estimate the correct output image, in other words it makes an educated guess.

However, *data-driven* implies exactly what these methods require: data. When faced with something like imaging through the atmosphere, we face a challenging problem. We almost can never have ground truth data – we cannot "turn off" the turbulence to see what the image was supposed to look like, it is a fact of nature we cannot avoid. Therefore, in order to teach the algorithm, we have to generate accurate data by means of **simulation**.

This work is focusing on simulating imaging through the atmosphere. There has been plenty of work done in this direction, too, therefore we may wonder what is left to be done? These previous efforts have largely been driven to understand new physical properties, or to verify certain mathematical assumptions often used to simplify analysis. These goals are very different than making images look good. Therefore, the type of simulation we are focused on in this work has a few significant properties which differ from typical physics-minded simulation:

1. **Fast.** In order to generate a large amount of data to feed our data-hungry algorithms, we need to have speed to generate gigabytes of data in a reasonable amount of time.

2. **Accurate.** If our data is inaccurate, we will have taught our algorithm how to invert a process which does not reflect nature, therefore, our simulation must mimick nature as accurately as possible.

3. **Differentiable.** Data-driven **deep learning** models use gradients to optimize their behavior, thus they are **differentiable**. If we want to truly integrate our physical model into our deep learning model it, too, will require this property.

1.2.1 Summarized Contributions

The contributions of this dissertation is what we refer to as **Zernike-based simulation** of atmospheric turbulent imaging. In certain situations, which we shall quantify in this thesis, the simulation is 1000× faster than its traditional optical counterpart.

The way that this type of simulation has been enabled required a number of tricks; theoretical, engineering, learning-based concepts have gone into developing a simulator which meets the three qualities of fast, accurate, and differentiable. Learning-based reconstruction methods trained with data from our Zernike-based simulation perform better than those trained on cheaper alternatives. This suggests a more close match with nature, not to mention the other benefit of differentiability which allows one to use perform simulation-in-the-loop training.

We present an overview of the simulation method described in this thesis in Figure 1.5. While the details are beyond our current discussion, we wish to highlight a few key details. First, the **Zernike space**, the main theoretical contribution of our simulator. The Zernike space allows us to describe the atmospheric effects using a **random vector field**. We can pull samples from the Zernike space using standard sampling techniques combined with an approximation we have proposed.

The image is formed through spatially varying convolution, which we approximate as a sum of weighted invariant convolutions. Our learning-based solution, known as the **Phase-**

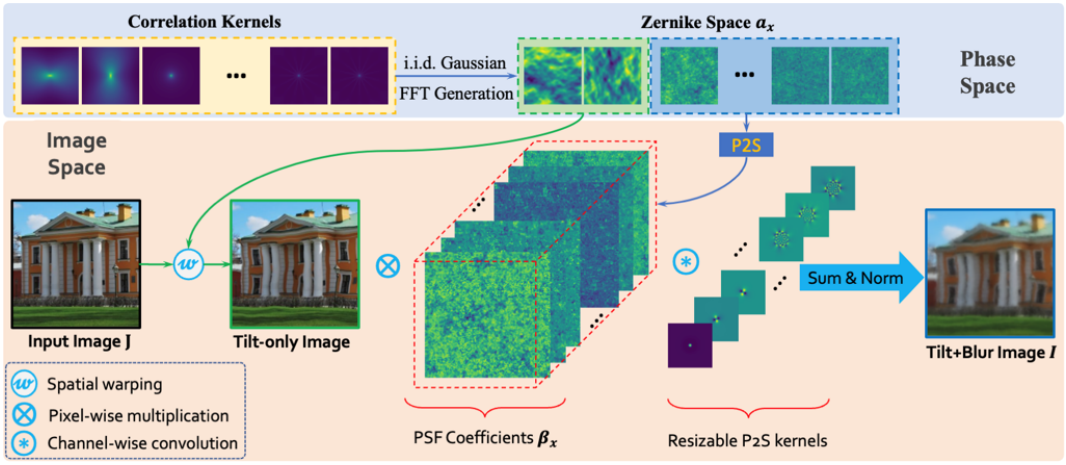


Figure 1.5. An overview of Zernike-based simulation.

to-Space (P2S) transform converts from the Zernike space to the spatial convolution weights.

At a high level, the computational complexity boils down to: Generating a collection of random fields at the size of the image, sending the coefficients through the P2S transform, and performing multiple convolutions of the image with different kernels which are then added together. We believe this to be near the optimum in simulating imaging through turbulence.

Finally, we would note that reconstruction methods trained with our data perform better than data from other, cheaper methods which use a less sophisticated physics approach in simulation. This suggests to us that (i) the generalization gap between our data and real data is minimal and (ii) our simulator can be applied to real world turbulence problems.

1.2.2 Funding acknowledgement

The work described in this dissertation has been supported by the following agencies:

- National Science Foundation under the grants CCF-1763896, CCF-1718007, IIS-2133032, and ECCS-2030570,
- Air Force Research Lab (public release approval number 88ABW 2020-0292),
- The Intelligence Advanced Research Projects Activity (IARPA) under Contract No. 2022-2110210000. The views and conclusions contained herein are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

2. MODELING IMAGE FORMATION

At the core of **imaging** is the exploitation of the transmission of **light** from one point to another. Light carries energy, and its distribution can be purely chaotic such as the background light that is not within the visible band (such as the WiFi signal probably bouncing around wherever you are right now) or it may instead be organized such that it's your favorite movie when decoded by your eye and brain[1]. Light and its movement is a major factor in how we perceive the world around us and, more importantly for this thesis, how we take images.

Light is a physical phenomenon, and nature is quite content carrying on with knowing its secrets while we are left to guess as to how it operates. The way we "guess" is through **mathematical models**. In this Chapter, we consider different mathematical models for modeling light and imaging. For this thesis, I do not wish to simply repeat the great efforts by Goodman [1], Born and Wolf [2], Hecht [3], or many others before my time. I instead would like to speak a bit more directly to folks concerned with an interest in **inverse problems**. That being said, I will not dispense with these great efforts. Therefore, I will present them as background material and interject my own experience and viewpoints when I feel it adds to the discussion.

For this thesis, we need a model which can easily incorporate atmospheric turbulence in a way that does not cause us too much grief. Therefore, we must consider which model is correct, of which many exist. I am particularly partial to the approach by Joseph Goodman [1], however, I'd like to show you some alternatives and options. We eventually will land on the model chosen for this thesis.

This chapter will progress in stages. First, we must consider how light can be modeled. There will be two models of particular use to us here. Next, how should we model an imaging system? Should we do ray tracing or a simplification which captures a majority of the effects we need? Finally, how should we extend this model in anticipation of the turbulence model we adopt in this work? These questions will be answered in this Chapter, with the later Chapters making full use of the chosen model.

[1]↑The great irony in this, however, is that the "chaotic" WiFi signal is being decoded by your streaming device, converted, and emitted from a display in such a way that it you recognize it as your favorite movie. Therefore, there is a great deal of order in what we would directly observe as chaos.

2.1 Modeling light

We first consider the model for light. Throughout history, the debate between light as a particle/ray or a wave was energetically sustained. The answer as it seems today is not so simple – it appears to be both! We will not concern ourselves with this quantum reality in this thesis, and we will stop our consideration of light modeling right *before* this advancement in thinking.

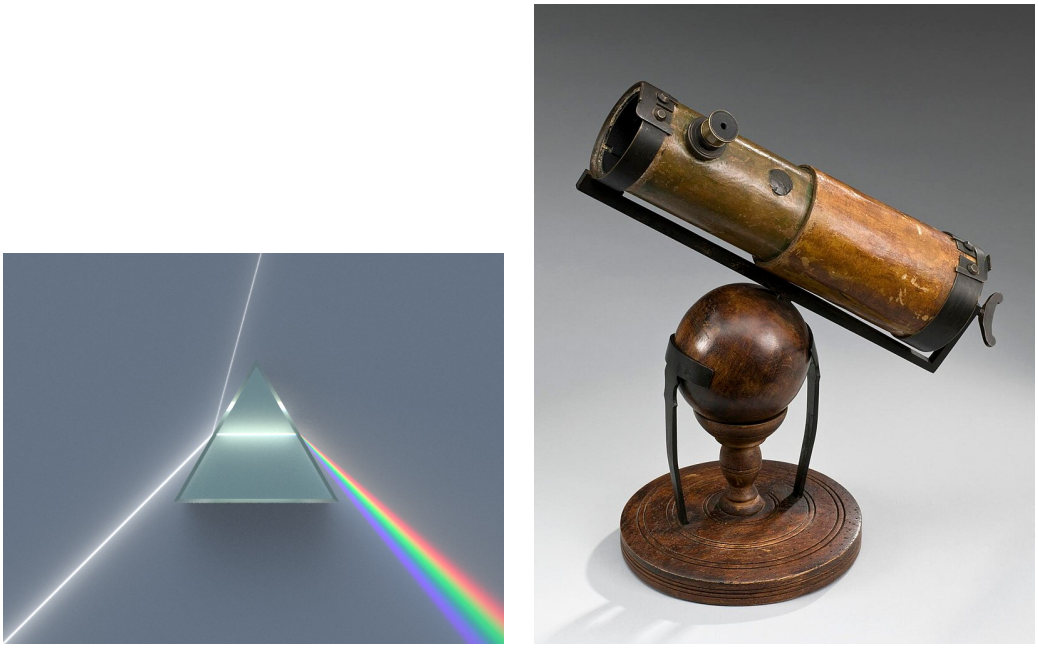
2.1.1 A history of understanding light

Though the history of studying of light dates back to the B.C. era, I will pick up around the time of Isaac Newton (i.e., the 1600's). At this time, there are two predominant models of light supported by two figures: The model that light is a particle/ray by Newton and the model that light is a wave by Christiaan Huygens. I'd like to discuss some of their experimental procedures and writings to detail how each one came to their conclusions. Interestingly, they *both* have their place!

Newton's experiments led him to believe that light acted as particles. A famous experiment of his is the splitting of **white light** into different colors. A famous example of this is through a dispersive prism which separates light into its constituent colors. We show this in Figure 2.1(a). Newton attributed this effect to the poor quality of **refractive lenses** (i.e., lenses that bend light). Because each color bent differently, you would have a blurring now known as **chromatic aberration**. Newton instead proposed the first reflective telescope of which Figure 2.1(b) presents a replica of.

Huygens championed a different model for light – that light acts as a wave. Huygens imagined that light acted as a repeated emission of **spherical waves** and that light traveled slower in "denser" media. This captured Newton's refractive concepts in a totally different manner. Huygens' model came to be more in favor than Newton's model, a trend which would only increase over the next century.

Thomas Young demonstrated the wavelike behavior of light in the early 1800's when he had two nearby pinholes in what is know known as the double slit experiment Figure 2.2. The double slit experiment proves that light does indeed **interfere** with itself. Young noticed



(a) Dispersive Prism Replica of Newton's telescope

Figure 2.1. (a) A view of the chromatic nature of light [Source] and (b) a replica of Newton's reflective telescope [Source]

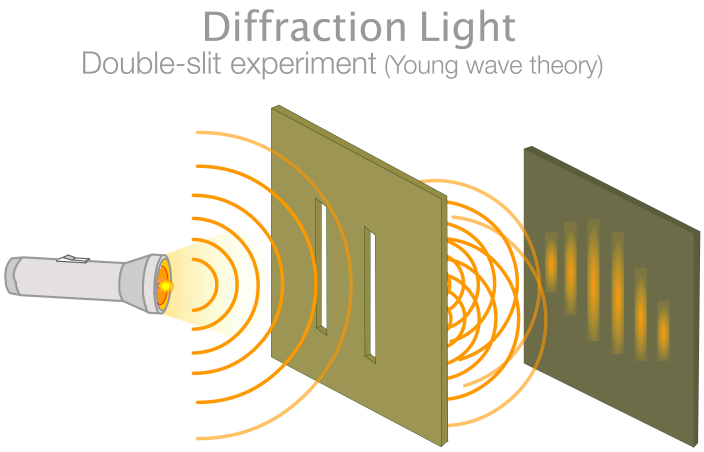


Figure 2.2. Young's double slit experiment in which interference patterns arise [Source]

if he covered the right pinhole, the pattern that arose would be the same as if the left pinhole was covered, just with a difference of a small shift. When both were open, however, the pattern was different than the sum of the two. Therefore, he concluded that light intereferes in a process which can only be described by wave motion.

In the mid 1800's James Clerk Maxwell then mathematically proved the existence of **electromagnetic waves** which propagated at the known speed of light. He concluded that light itself must be an electromagnetic wave. In fact, Maxwell's equations were *so good*, that it drove many to wonder why they seemed incompatible with Newton's equations. Of the many physicists at this time, Albert Einstein reconciled this conflict through special relativity – a modification of Newton's equations to live "in the same universe" as Maxwell's equations. This speaks to the power of Maxwell's equations and how light has influenced our development of the laws of nature (even in places where it seems it shouldn't – such as in the case of Newton's mechanics).

This may seem as if the final word had been said and thus light is a wave. If Maxwell can mathematically prove it and this proof was so influential to modify Newton's equations, what more could come to change our minds? Einstein and his study of the photoelectric effect and Planck with black body radiation and its incompatibility with Maxwell's equations suggested the reality we are still living in today: quantum mechanics. This loops us back to the time of Newton; light seems to be a particle, but interestingly enough, one that acts like

Figure 2.3. A demonstration of refraction where we can see the image of the straw is bent when viewed through water. [Source]

a wave! While this resulting duality is interesting and strange, it is not a point which will be discussed further in this thesis.

With a brief history presented, we now move to a model which still has its place today: light as a ray. Although the ultimate model we will choose in this thesis will be the wave model, this will introduce many important concepts along the way and prove useful for some points later in this Chapter.

2.1.2 Light as a ray

Ray optics or **geometrical optics**, two ways of referring to this classification of the ray model for light, views light as "arrows" which propagate through space. The behavior of light in this model includes the assumption that light travels in straight lens and bends at a change in **medium**. By a medium, we mean the thing that the light moves through. This bending of light is known as **refraction**. We present a visualization of this bending in Figure 2.3.

To quantify the amount of bending which light undergoes at the interface between two media we can define the index of refraction n_1 and n_2. These If we additionally define the angle at which it enters as θ_1 and exits at θ_2, we may write **Snell's law**,

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{2.1}$$

The interpretation of the index of refraction is the attenuation in speed by the ray. That is, if light travels at a speed c in a vacuum, it travels at a speed c/n_1 in a medium with index of refraction n_1.

Thus, the relationship between the input angle and the output angle is non-linear. We often encounter this equation in something like an undergraduate physics course. This non-linearity is the main reason for the complications that come from this model. To remedy this, there are two primary solutions: (i) to numerically perform **ray tracing** (ii) to use the **paraxial approximation**.

Ray tracing

Suppose a ray at position p_0 emits with a directional vector $\mathbf{v}$. For a few types of surfaces, a closed form solution exists (a plane being one such surface). However, in optical devices our surfaces will be more complex that just a simple plane. Therefore, we often must employ ray tracing in order to find the points of intersection. Ray tracing follows a simple procedure:

1. **Find points of intersection**. A ray at position p_0 is propagated along its directional vector $\mathbf{v}$ until it reaches a surface $\mathcal{S}$. This is often found through methods such as **Newton's method** – an iterative approach to finding roots of a function. Therefore, the intersections we find are through iterative, computational processes.

2. **Compute Snell's law**. With the point of intersection found, we know the direction the ray is traveling and compute the **normal** vector of $\mathcal{S}$ at the point of intersection. From this, we can apply Snell's law to find where the ray should travel on the *other* side of the surface (or if it should reflect).

This simple algorithmic approach can be parallelized and made incredibly fast, therefore it is perfect for computational analysis and rendering.

Paraxial ray tracing

The adoption of the paraxial approximation is more related to mathematical analysis. The paraxial approximation assumes the following:

$$\sin\theta \approx \theta, \quad \tan\theta \approx \theta, \quad \cos\theta \approx 1. \tag{2.2}$$

Put simply, we are assuming that we are primarily working with small angles! Amazingly, through this approximation, a moderate amount of surfaces can be analyzed purely analytically. This means we do not need to find points of intersection through computation, but through simple geometry. In this context, a **lens** can be described suitably by a 2×2 matrix known as the **ray transfer matrix** (sometimes called **ABCD matrix analysis**,

$$\begin{bmatrix} x_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x_1 \\ \theta_1 \end{bmatrix} \tag{2.3}$$

with

$$A = \left.\frac{x_2}{x_1}\right|_{\theta_1=0} \qquad B = \left.\frac{x_2}{\theta_1}\right|_{x_1=0}$$
$$C = \left.\frac{\theta_2}{x_1}\right|_{\theta_1=0} \qquad D = \left.\frac{\theta_2}{\theta_1}\right|_{x_1=0}$$

Furthermore, an array of lenses can be analyzed by subsequent applications of the corresponding ray transfer matrices.

Having a fair bit of familiarity with computer vision (CV) people through experience and the more optics ways of thinking through reading, I would like to make a particular distinction here which I will later reinforce when discussing camera models. It is in my opinion that two people, each from one of the fields mentioned, can be talking to each other about the same thing using the same language but mean completely different things. They will only realize it

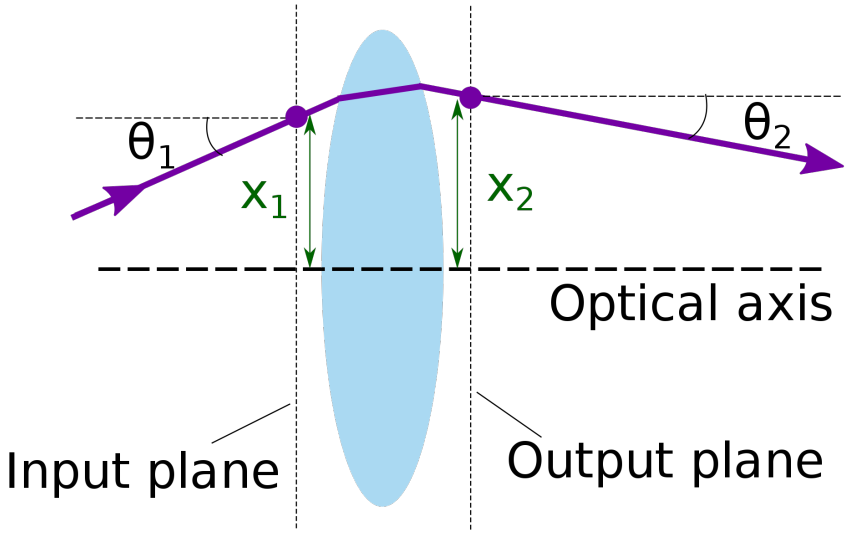


Figure 2.4. Some notational definitions for the ray transfer matrix analysis [Source]

once they get deeper into the problem. To cut to the core of this difference, I would ask them when they imagine an object emitting light, is it emitting one ray per point, or multiple?

The physicist will envision an object which emits *multiple* rays emitting from each point on the object. While they may not all hit the camera, the fact that it is observed by our camera or not is of little difference. Nature emits multiple rays, therefore the physicist thinks of it as such. The CV person will instead visualize a single ray emitting from each point on the object (they may even regard them as pixel locations!). This clearly is very different from the physicist, we are working with infinitely less rays! I will describe why I believe this difference exists between the two communities when discussing camera models.

Finally, light as a ray will not be the primary model adopted in this thesis. However, many ideas that arise from it will be useful for getting concepts across. One such place will be for the modeling of cameras. But, in general, it will be used sparingly.

2.1.3 Light as an electromagnetic wave

In the late 1800's James Clerk Maxwell developed a set of equations now known as **Maxwell's equations**. These equations captured the known behavior of electricity and magnetism. Maxwell made an additional interesting discovery through these equations: manipulating them suggested the presence of **electromagnetic waves**. However, he did not stop with mathematics but instead continued with physical reasoning, ultimately suggesting that light is one such electromagnetic (EM) wave. As it turns out, for the most part, he was suitably correct!

Maxwell's equations describe **vector calculus** relationships between different components of the **electric field** $\mathbf{E}$ and the **magnetic field** $\mathbf{B}$. I would like to make the distinction here that, very briefly, I am considering the fields to be **vector fields** – that is, fields which for every point $\mathbf{x}$

$$\mathbf{E}(\mathbf{x}) = [E_x(\mathbf{x}), E_y(\mathbf{x}), E_z(\mathbf{x})]^T, \tag{2.4}$$

where $E_a(\mathbf{x})$ is the a-component of the vector field at $\mathbf{x}$. Each of these components constitute a **scalar field**, a concept we will return to shortly. Typically, this dependence on location is dropped for a bit of simplicity in notation, therefore I will follow suit.

Maxwell's equations in differential form can be written as:

$$\begin{aligned} \boldsymbol{\nabla} \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, & \boldsymbol{\nabla} \times \mathbf{B} &= \mu_0 \left(\epsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \mathbf{J} \right), \\ \boldsymbol{\nabla} \cdot \mathbf{E} &= \rho / \epsilon_0, & \boldsymbol{\nabla} \cdot \mathbf{B} &= 0. \end{aligned} \tag{2.5}$$

It would be useful at this point to develop the EM wave equation. As a first step, we restrict our analysis to a vacuum. Combining this with the fact $\epsilon_0 \mu_0 = 1/c^2$ allows our equations (2.5) to take the form

$$\begin{aligned} \boldsymbol{\nabla} \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, & \boldsymbol{\nabla} \times \mathbf{B} &= \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}, \\ \boldsymbol{\nabla} \cdot \mathbf{E} &= 0, & \boldsymbol{\nabla} \cdot \mathbf{B} &= 0, \end{aligned} \tag{2.6}$$

where c is the speed of light. Through some vector calculus tricks (triple scalar product and BAC-CAB [4] rule), we can obtain the wave equation. Starting with taking the curl of the first equation in (2.6),

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{E}) = \boldsymbol{\nabla} \times \left(-\frac{\partial B}{\partial t} \right) \tag{2.7}$$

$$\boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \mathbf{E}) - \boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \mathbf{E} = -\frac{\partial}{\partial t} (\boldsymbol{\nabla} \times \mathbf{B}) \tag{2.8}$$

$$\nabla^2 \mathbf{E} = \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{2.9}$$

By symmetry, we could say the same for magnetic field $\mathbf{B}$. This motivates to write the following:

Theorem 2.1.1 (**Electromagnetic Wave Equation**). *The homogeneous, 3-dimensional form of the electromagnetic wave equation is given by*

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0. \tag{2.10}$$

This equation predicts an electromagnetic wave that travels at the speed of light c.

The reason for presenting this derivation is that light *undeniably* acts as a wave in the framework of Maxwell's equations. As highlighted by quantum mechanics, this isn't exactly true but we will not require any of those considerations for this thesis. We will now make some simplifications to gain us some traction in making use of this result.

The scalar wave equation

The EM wave equation (2.10) is a major result of physics and optics. It further serves as the starting place for our chosen model for light. At this point, however, we would like to make the assumption that the components of the vector field $\mathbf{E}$ were independent of one another. This assumes that there is no coupling between the EM field components. Because of this, we can analyze each component independently [1, 2].

Because each component can be analyzed independently, we need only work with *scalar* fields and not vector fields. This leads us to define f as such a scalar field which satisfies

$$\nabla^2 f(\mathbf{x}, t) - \frac{n^2}{c^2} \frac{\partial^2 f(\mathbf{x}, t)}{\partial t^2} = 0, \tag{2.11}$$

which is known as the **scalar wave equation**. Here I have defined v to be the speed of the wave which, in the case of optics, may be written as $v = c/n$ where c is the speed of light and n is the index of refraction. This gives us a bit more insight on the index of refraction, it is the dimensionless constant that attenuates the speed of light when a wave is traveling through a non-vacuum medium. Furthermore, I would again emphasize the fact that we are working with a scalar field f and no longer the vector field $\mathbf{E}$.

In the study of differential equations, it is common to use a set of elementary solutions to either (i) build insights into the equation or (ii) build more complicated solutions by virtue of linearity. Perhaps it is not a surprise that an elementary solution is a sinusoidal *wave*,

$$f(\mathbf{x}, t) = A(\mathbf{x}) \cos(2\pi\nu t + \theta(\mathbf{x})), \tag{2.12}$$

with spatially varying amplitude $A(\mathbf{x})$ and phase $\theta(\mathbf{x})$. Although this is valid, we will typically consider a **complex wave function**:

$$u(\mathbf{x}, t) = U(\mathbf{x})e^{-j2\pi\nu t}. \tag{2.13}$$

The function $U(\mathbf{x})$ represents the complex envelope, or **phasor** [1, 3], which is then modulated by the time-evolving term $e^{-j2\pi\nu t}$. This form in Equation (2.13) is particularly useful because it separates the spatial and time components. As a result, most problems of interest to us will involve finding the proper $U(\mathbf{x})$, then adding in the time behavior through multiplication with the modulation term $e^{-j2\pi\nu t}$.

We now wish to investigate a few important properties of our complex envelope. The first is the concept of a **wavefronts**. A wavefront can be mathematically described by the set $\{\mathbf{x} \mid \angle U(\mathbf{x}) = \varphi\}$ where φ is some constant $\varphi \in [0, 2\pi)$. The distance from one wavefront to another (along the normal direction of the wavefront) is referred to as the **wavelength**, denoted as λ [m]. With this we can also define the **optical frequency** ν [s^{-1}] to be

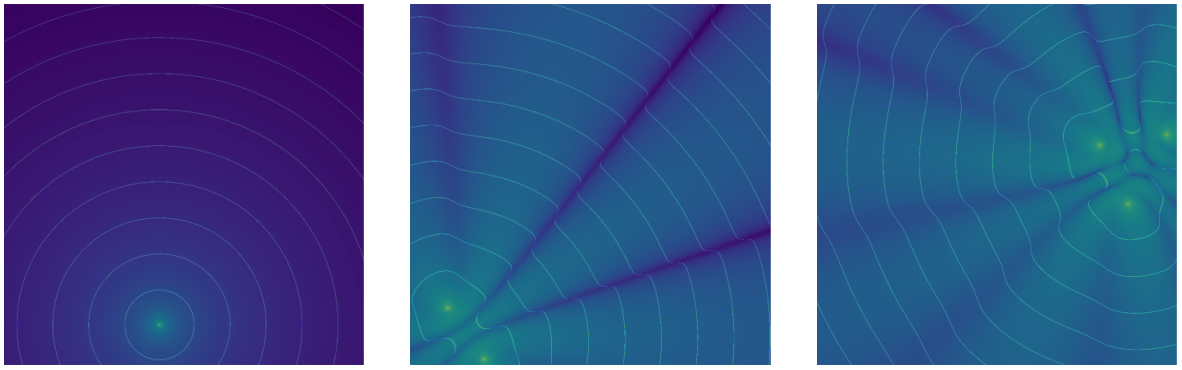
$$\nu = \frac{v}{\lambda}, \tag{2.14}$$

where v is the velocity of the wave [m/s]. The optical frequency describes the rate at which the wave oscillates as it arrives.

A mathematical description of wavefronts is perhaps precise, but it does not capture the visual that comes along with it. We therefore present Figure 2.5, where we show the following field:

$$V(\mathbf{x}) = \left| \sum_{i=1}^{N} U(\mathbf{x}; \mathbf{x}_i) \right| = \left| \sum_{i=1}^{N} \frac{\exp\{-j(2\pi/\lambda)|\mathbf{x} - \mathbf{x}_i|\}}{|\mathbf{x} - \mathbf{x}_i|} \right|,$$

where we have chosen the amplitude function $A(\mathbf{x}; \mathbf{x}_i) = 1/|\mathbf{x} - \mathbf{x}_i|$ and phase function $\phi(\mathbf{x}; \mathbf{x}_i) = (2\pi/\lambda)|\mathbf{x} - \mathbf{x}_i|$. In other words, there are sources of waves at locations $\mathbf{x}_i$ known as **point sources**, whose amplitude and phase vary as a function of distance from the point $\mathbf{x}_i$. In addition to this magnitude, we also show the wavefronts. We can see that with more point sources and the more closely they are packed together, the wavefronts become increasingly complicated.



(a) One point (b) Two points (c) Three points

Figure 2.5. The amplitudes and wavefronts created by point sources.

With the intent of using our form of the complex wave function $u(\mathbf{x}, t)$ within the scalar wave equation of Equation (2.11), we present the following pair of derivatives:

$$\nabla^2 u(\mathbf{x}, t) = \left\{\nabla^2 U(\mathbf{x})\right\} e^{-j2\pi\nu t}, \tag{2.15}$$

$$\frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} = (j2\pi\nu)^2 U(\mathbf{x}) e^{-j2\pi\nu t}. \tag{2.16}$$

Substituting Equation (2.15) and Equation (2.16) into Equation (2.11), we can write the scalar wave equation in terms of the phasor,

$$\nabla^2 U(\mathbf{x}) = -\left(2\pi\nu\right)^2 \frac{n^2}{c^2} U(\mathbf{x}).$$

Letting $k = 2\pi\nu n/c = 2\pi/\lambda_\nu$ be the **wave number** (where $\lambda_\nu = c/(\nu n)$ is the wavelength), we arrive at the **Helmholtz equation**.

Definition 2.1.1 (**Helmholtz Equation**). *With $u(\mathbf{x}, t) = U(\mathbf{x})e^{-j2\pi\nu t}$ where $\mathbf{x} = [\mathbf{x}, z]^T \in \mathbb{R}^3$ and ν , the Helmholtz equation of the scalar field $U(\mathbf{x})$ is given by*

$$\nabla^2 U(\mathbf{x}) + k^2 U(\mathbf{x}) = 0, \tag{2.17}$$

where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the Laplacian operator, and $k = 2\pi/\lambda$ is the wave number.

Types of Waves

With the Helmholtz equation Equation (2.17) presented, we now wish to show a few elementary solutions of this equation. The first is the planar wave

$$U(\mathbf{x}) = A_0 e^{-j\mathbf{k}^T\mathbf{x}} = A_0 e^{-j(k_x x + k_y y + k_z z)}, \tag{2.18}$$

where $\mathbf{k} = [k_x, k_y, k_z]^T$ is the **wave vector** such that $k_x^2 + k_y^2 + k_z^2 = k^2$ and $k = 2\pi/\lambda$ is the wave number and A_0 is the amplitude. The plane wave is at the core of **Fourier optics** in which a complex wave distribution is decomposed by a basis of plane waves (in the very same

manner as the Fourier transform). Since the rules of propagating plane waves are so simply (i.e, follow the wave vector!) this method is particularly useful for computing propagations. We will present this method when we are dealing with simulating wave propagation.

The next example is a **spherical wave** which is given as

$$U(r) = \frac{A_0}{r} e^{-jkr}, \tag{2.19}$$

where $r = |\mathbf{x}| = \sqrt{x^2 + y^2 + z^2}$ is the radius between $\mathbf{x}$ and the origin. The spherical wave has a specified origin point which we refer to as a point source. The final wave we will introduce is the **paraboloidal wave** given by

$$U(\mathbf{x}) = \frac{A_0 e^{-jkz}}{z} e^{-jk\left(\frac{|\mathbf{x}|^2}{2z}\right)}. \tag{2.20}$$

All three of the forms we have presented here (spherical, paraboloidal, or planar) satisfy the Helmholtz equation.

Suppose that we have a spherical wave emitting energy as shown in Figure 2.7. As the wave propagates out further and further, we can justify a particular approximation. At a certain distance, we may state that the wave is well approximated by a parabolic wave. Further away we may even claim that it is approximated well by a single planar wave. This conceptualization of a spherical wave turning into a parabolic then planar wave is an extremely important tool that we will use in analyzing light propagation in the next Section.

2.1.4 The Huygens-Fresnel principle

With the model of light presented, I now wish to present a particularly interesting way of thinking about wave propagation that I find useful when imagining how these things work in my own mind. This is the **Huygens-Fresnel principle (HFP)**. The HFP states the following:

- Every point on a wavefront is a source of secondary spherical wavelets;
- These secondary wavelets interfere;
- The sum of these wavelets produce a new wavefront.

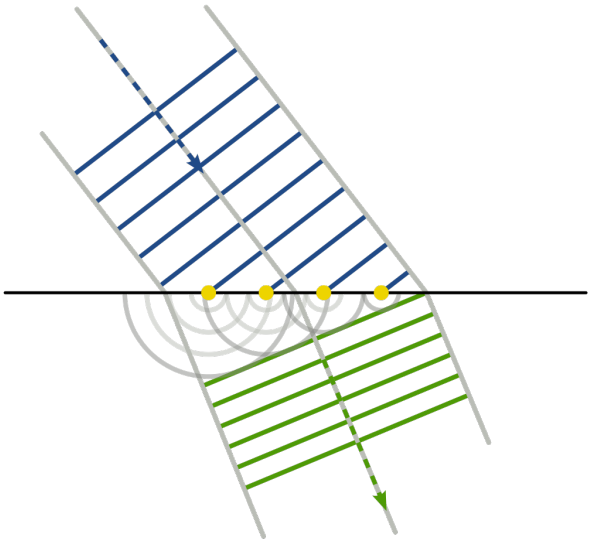


Figure 2.6. A visualization of the HFP through a change in medium [Source]

Although the principle is simply stated, it may be challenging to appreciate the beauty and elegance of such a summation. I would encourage the reader to seek out visualizations of the Huygens-Fresnel principle (of which there exists many videos and GIFs online) to get a real sense as to what it implies.

Huygens proposed his model for the wave-like behavior of light in a fashion to the set of rules as above. However, they have been tweaked over time to match what we now know. It is interesting to know that in the following Section, we shall introduce a model for light propagation that, for all intents and purposes, mirrors the rules described here. There will be some mathematical details which our above rules don't tell us, but we will feel the similarity. Therefore, when solving problems of this variety, the HFP is a fantastic guiding concept that can help us visually interpret the problem.

For an example of its utility, the HFP, when combined with the fact that the speed of light is attenuated in a different medium, can even explain why light refracts! To this end, I present Figure 2.6 in which it is shown *why* light bends. One thing I have not mentioned yet, to tie this explanation together, is that there is a connection between the wave model and the ray model for light. The rays travel in a direction that is normal to the wavefront. Therefore, in Figure 2.6 also shows arrows which correspond to the direction of rays.

2.1.5 The model of light for this thesis

With a great deal of effort dedicated to the wave model, it likely does not come as a great surprise that in this thesis I will consider light to be wave. We will make more simplifications as necessary to simplify our analysis when necessary, but we will indeed treat light as a wave. For example, one such simplification is that we will never truly need to solve a wave equation to model an imaging system with the simplifications we will outline in this Chapter. Furthermore, when describing turbulence, the route we will take will *again* not use any wave propagation.

It may be unclear *why* then we would adopt the wave model. If I say light is a wave, then why wouldn't we ever do wave propagation? The answer is subtle, and will be made more obvious over the course of this thesis, but I do wish to give some insight as to why. The answer is that we will primarily think of light as a wave which has a phase $\phi(\boldsymbol{\xi})$ and amplitude $A(\boldsymbol{\xi})$ which vary across its $\boldsymbol{\xi}$. The model for turbulence can be placed into both the phase and amplitude, though we shall primarily focus on the phase. Therefore, with ray optics it will be somewhat unclear where to put the turbulence model. Therefore, we must take light as a wave if we are to model turbulence properly.

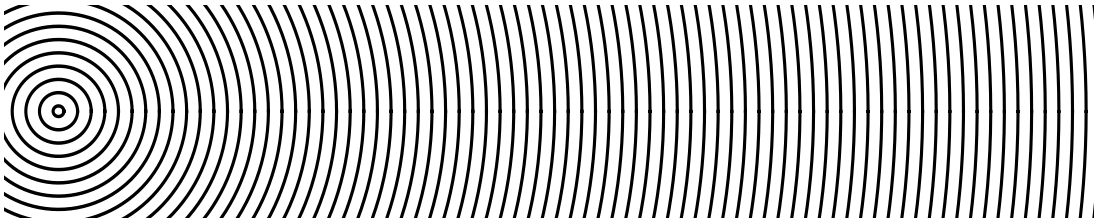
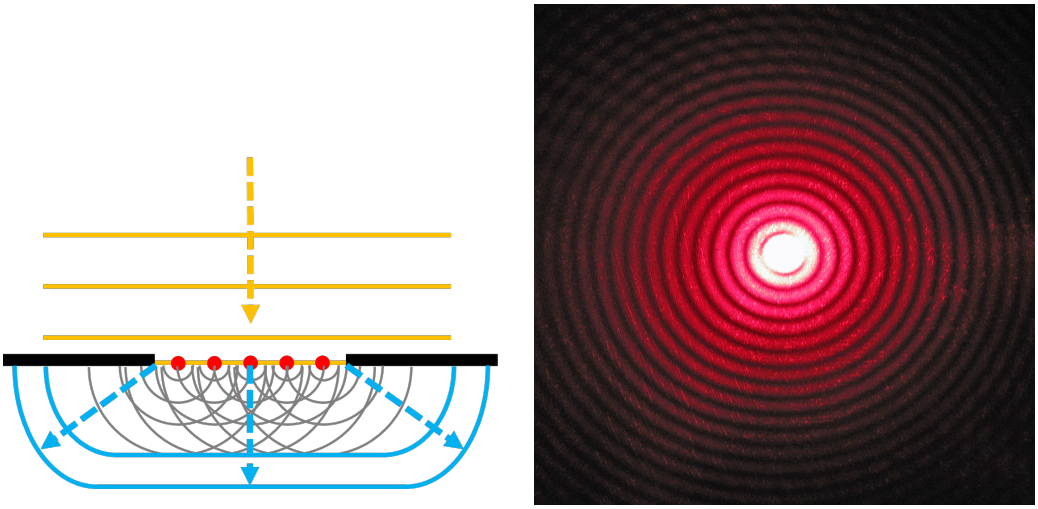


Figure 2.7. As a spherical propagates outwards, at a sufficient distance it becomes suitably paraboloidal (or parabolic), and further yet planar. We note that one can approximate a planar wave with a parabolic one. These ranges are not mutually exclusive, but rather the most appropriate approximation (in most cases) for the associated wave at a distance. Image inspired by Hecht [3].

2.2 Diffraction

Adopting the wave model comes with a few particular difficulties. One is the fundamental limit that is placed on resolution of our images by his model. This limit is caused by the



(a) (b)

Figure 2.8. (a) Visualizations of the "leaking" of light around the edges and (b) a diffraction pattern by a red laser [Source]

effect known as **diffraction**. Diffraction comes from the fact that light acts as a wave. In Figure 2.8(a) we show that a planar wave which is incident upon a non-transmitting screen with a hole has a sort of "leaking" effect in which the wave extends beyond the shadow of the screen. This leaking we will mathematically describe in this Section. The result may be a pattern as in Figure 2.8(b) which we refer to as a **diffraction pattern**.

Throughout modern history, a great deal of effort has been spent in analyzing the deceptively simple situation in Figure 2.8(a). One of the more prevailing models was the **Fresnel-Kirchhoff diffraction integral**, which primarily solves the problem through some assumptions which were later found to be inconsistent (to a moderate degree). The most accurate approach, which is in accordance with Maxwell's equations is known as **Rayleigh-Sommerfeld diffraction**. The Rayleigh-Sommerfeld equation utilizes a great deal of insight and mathematical rigor to properly analyze a diffracting screen. The problem is easily stated, yet the details involved are beyond the scope of this thesis. We would suggest the interested reader to Goodman [1] for an easy-to-follow treatment, and those who are interested in the rigorous development to Born and Wolf [2].

Seeing as this derivation takes a number of pages and concepts which we have not introduced to present, We instead will provide the theorem directly without proof:

Theorem 2.2.1 (**Rayleigh-Sommerfeld diffraction**). *The wave observed at a point* $\mathbf{x}$ *after diffracting through a surface* Σ *is described by*

$$U(\mathbf{x}) = \frac{1}{j\lambda} \iint_{\Sigma} U(\boldsymbol{\xi}) \frac{\exp\{jkr\}}{r} \cos\theta \, d\boldsymbol{\xi}, \tag{2.21}$$

where Σ *is the source surface (which we will call the* ***aperture****) and* θ *as the angle between the surface normal and the vector* $\mathbf{r}$ *pointing from* $\boldsymbol{\xi}$ *to* $\mathbf{x}$ *(with* $r = |\mathbf{r}|$*), and* $k = 2\pi/\lambda$ *as the wavenumber.*

This theorem can be understood by Figure 2.9. Each point on the opening of the diffracting screen will contribute to the phasor at a location $\mathbf{x}$. The way in which they will contribute will be through secondary spherical wavelets, which propagate starting from Σ. In other words – it as a mathematical formulation of the HFP!

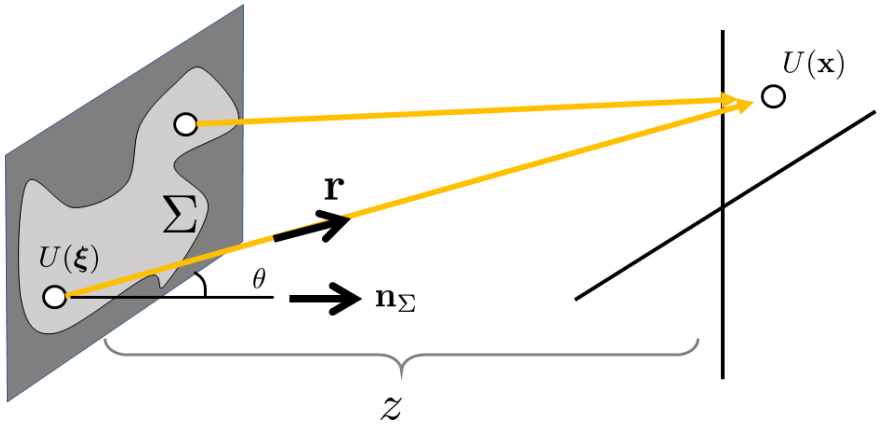


Figure 2.9. When there are more than two points in the object plane, the observed field at position $\mathbf{x}$ will be the superposition of the fields emitting from the source plane Σ.

The trouble with Rayleigh-Sommerfield diffraction

Now that we have the equation, it may seem as if we are done and can finally start solving problems. However, there is an issue with the Rayleigh-Sommerfeld diffraction theorem – it

is challenging to get closed form solutions for nearly *any* problem of interest [1]. The reason for this is that it is *so* accurate, that the cost to be paid for accuracy is difficulty.

If we recall Figure 2.5, we can see that with multiple point sources, the behavior of the wavefront will take on more dramatic and complicated forms. The Rayleigh-Sommerfeld equation *will* capture this behavior no matter how complicated the field distribution on the aperture is. I would liken this situation to the problem facing Newton's $F = ma$. Just because we have the equation does not mean that we can solve every problem in the world with it (to which I would suggest the three-body problem as an example). Therefore, in order to solve some problems with it, we will need to make a few approximations.

2.2.1 Fresnel and Fraunhofer Diffraction

A crucial concept to grasp in the previous discussion is that the complex wave on the opposite side of a diffracting screen is a superposition of secondary spherical wavelets. This, in fact, was more or less the theory proposed by Huygens (and refined by Fresnel), thus we will often refer to the Rayleigh-Sommerfeld equation as the Huygens-Fresnel equation. Huygens, of course, was not even aware of Maxwell's equations, coming roughly 200 years earlier. This speaks to the insight of Huygens' formulation.

Because the secondary wavelets emit as spherical waves, we can use our idea of approximating spherical waves as parabolic or planar waves at suitable distances. It may not appear clear as to *why* we would want to do so, but this will become very apparent once we introduce lenses. The approximation by parabolic waves was done by Fresnel, thus leading us to present the following theorem:

Theorem 2.2.2. *The* ***Fresnel diffraction integral*** *is defined as*

$$U(\mathbf{x}) = \frac{e^{jkz}}{j\lambda z} e^{j\frac{k}{2z}|\mathbf{x}|^2} \iint_{-\infty}^{\infty} \left\{ U_\Sigma(\boldsymbol{\xi}) e^{j\frac{k}{2z}|\boldsymbol{\xi}|^2} \right\} e^{-j\frac{2\pi}{\lambda z}\mathbf{x}^T\boldsymbol{\xi}} d\boldsymbol{\xi}, \tag{2.22}$$

where $U_\Sigma(\boldsymbol{\xi})$ is the incident field passing through a finite aperture Σ. This may also be written as

$$U(\mathbf{x}) = \frac{e^{jkz}}{j\lambda z} e^{j\frac{k}{2z}|\mathbf{x}|^2} \mathfrak{Fourier}\left\{ U_\Sigma(\boldsymbol{\xi}) e^{j\frac{k}{2z}|\boldsymbol{\xi}|^2} \right\}\Big|_{\mathbf{f}=\frac{\mathbf{x}}{\lambda z}}, \tag{2.23}$$

where the Fourier frequencies are evaluated at $\mathbf{f} = \mathbf{x}/(\lambda z)$.

This means that within a suitable distance away from the screen, the effects are related to a modified Fourier transform. If we go even further away to where planar waves suitably approximate the wavelets, we arrive at the following theorem:

Theorem 2.2.3. *The* ***Fraunhofer diffraction integral*** *is defined as*

$$U(\mathbf{x}) = \frac{e^{jkz} e^{j\frac{k}{2z}|\mathbf{x}|^2}}{j\lambda z} \iint_{-\infty}^{\infty} U_\Sigma(\boldsymbol{\xi}) e^{-j\frac{2\pi}{\lambda z}\mathbf{x}^T\boldsymbol{\xi}} d\boldsymbol{\xi}, \tag{2.24}$$

or equivalently, in terms of Fourier transform:

$$U(\mathbf{x}) = \frac{e^{jkz} e^{j\frac{k}{2z}|\mathbf{x}|^2}}{j\lambda z} \mathfrak{Fourier}\{U_\Sigma(\boldsymbol{\xi})\}\Big|_{\mathbf{f}=\frac{\mathbf{x}}{\lambda z}}. \tag{2.25}$$

Both of these theorems have their places in this thesis. The Fresnel diffraction integral will be critical in propagating waves from plane-to-plane without lenses, while the Fraunhofer diffraction integral will be particularly useful when lenses are in the mix.

Why make these approximations?

We can now describe waves which propagate through an aperture at some distance away from the screen. In the case of the Fraunhofer pattern, the waves need to be far away for the results to apply. This may beg the question: Why would we want this? If we are modeling an imaging system, the length of the camera body is short, so why would be want these approximations?

The answer to this is that we have not yet introduced a lens into the mix. When we do, we will see that whatever was going to happen at infinity will now happen at a finite distance. A computer vision concept known as **homogeneous coordinates** [5] can shed some

light onto this property. Let us consider where two rays which are traveling in parallel would intersect. According to this theory, it would be at the line at infinity. In our light example, this is equivalent to rays passing through the aperture with no lens.

When we add a lens into the mix, something interesting happens. The rays which *would have* intersected at infinity now are intersecting somewhere at a finite distance. Therefore, a lens can be viewed as a transformation which forces an intersection at the line at infinity to happen at a finite distance. This is significant for Fraunhofer diffraction, which we know happens infinitely far away. The ultimate conclusion, we shall see, is that the **Fraunhofer pattern** occurs at the focal plane of the imaging system.

2.2.2 Diffraction Patterns

At this point we have already shown what are known as **diffraction patterns**, however, we have not precisely defined them. A diffraction pattern, to me, is what we would observe if we had a sensor to view the wave. We wouldn't see the complex wave, instead, we would see the power per unit area.

Because the waves are sinusoidal in nature, the power at a position $\mathbf{x}$ can be written as

$$\begin{aligned} P(\mathbf{x}) &\propto \mathbb{E}[|A(\mathbf{x})\cos(2\pi\nu t + \theta(\mathbf{x}))|^2], \\ &\propto |A(\mathbf{x})|^2/2, \end{aligned}$$

where the proportionality comes from a variety of properties of the camera such as quantum efficiency and $\mathbb{E}[\cdot]$ denotes probabilistic expectation. We can define a function known as the **intensity** of a wave to be proportional to the power [1]. The intensity is given as

$$\begin{aligned} I(\mathbf{x}) &= \mathbb{E}[|U(\mathbf{x})\exp\{-j2\pi\nu t\}|^2], \\ &= |U(\mathbf{x})|^2. \end{aligned} \tag{2.26}$$

Instead of referring to the power directly, we typically prefer intensity for its simplicity. Furthermore, we drop the expectation in (2.26) due to the time invariance of the phasor.

To an optical engineer, it may appear as a glaring error that I have used the term intensity instead of **irradiance**. In this thesis, I will refer to the intesnity as power per unit area, thus being identical to the irradiance. This is done to be in coordination with Goodman as well as the bulk of the computer vision and image processing literature.

2.2.3 Diffraction Limit

Earlier it was mentioned that diffraction imposes a limit on the resolution of an imaging system. We can analyze this by Fraunhofer diffraction. If we consider the intensity of the field emerging from a circular aperture, we can show that

$$I(r) = |U(r)|^2 = \left(\frac{2\pi D^2}{\lambda z}\right)^2 \left[2\frac{J_1\left(\frac{2\pi Dr}{\lambda z}\right)}{\frac{2\pi Dr}{\lambda z}}\right]^2. \tag{2.27}$$

This intensity distribution $I(r)$ is called the **Airy disc**. The shape of the Airy disc is shown in Figure 2.10. The width of the central lobe is measured as

$$d = 1.22\frac{\lambda z}{D}. \tag{2.28}$$

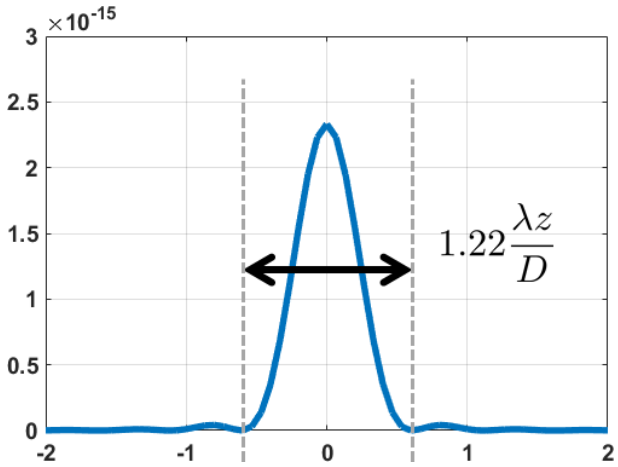


Figure 2.10. The width of the central lobe of the Airy disc is given by $d = 1.22\lambda z/D$.

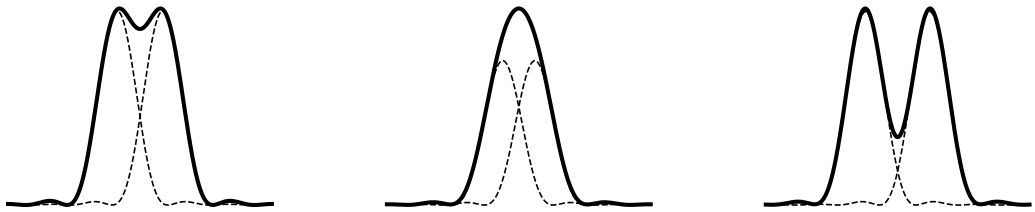
This brings us to introduce the **Rayleigh criteria**. With two pinholes in an imaging system, the emergent diffraction patterns (in this case, the Airy discs) may overlap. Because

of this, there is a minimum separation we must reach in order for the two to be resolvable. Mathematically, the minimum resolvable separation is defined at the point of half of the main lobe width,

$$\delta = 0.61\frac{\lambda z}{D}. \tag{2.29}$$

We show the three cases in Figure 2.11. Finally, these concepts are related to the **numerical aperture** (NA) as $\text{NA} = D/z$. The Rayleigh criteria may be written in terms of the numerical aperture through

$$\delta = 0.61\frac{\lambda}{\text{NA}}. \tag{2.30}$$



(a) Rayleigh criteria (b) Unresolvable (c) Resolvable

Figure 2.11. Visualizing two diffraction patterns relative to the Rayleigh criteria.

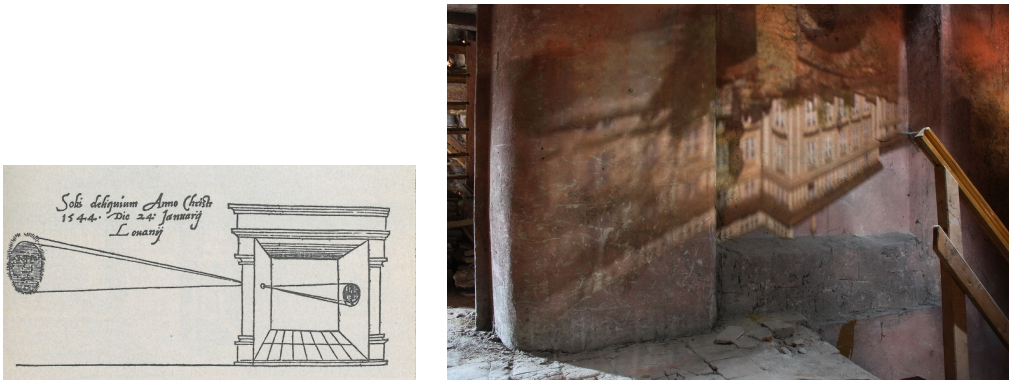
2.3 Modeling a camera

A modern **camera** is a complicated device comprised of multiple lenses, carefully designed sensor(s), mirrors, beam splitters, and so on. In Figure 2.12 we can see a Sony camera cut in half. To design this lens, experts in lens design use software to help them optimize the **lens relay**. This is because *they cannot design this complicated relay by hand.* By the same token, we cannot design a model to completely describe this lens system by hand. Therefore, we will need a simplification.

There are many ways we can model this system. One of the more sophisticated ways we may model it is by **ray tracing**, which is, in fact, the way the camera shown in Figure 2.12 is most likely designed. Commercial software such as *Zemax* or *Code V*[2] can be used to optimize certain criteria which are based around ray optics. This model, however, will be a

[2]↑Pronounced "Code Five" – don't be like me and think it's "Code vee".

Figure 2.12. A camera cut in half showing its elaborate lens design [Source]



(a) (b)

Figure 2.13. (a) A camera obscura diagram. (b) A result of a camera obscura imaging the New Royal Palace at Prague Castle, projected onto a stone wall. [Sources]

bit overkill for our purposes and will miss a crucial detail which we want to include in our formulation. To begin, I will start with the simplest model of an imaging system and our first potential candidate for our model – the **pinhole camera**.

2.3.1 The pinhole camera

The pinhole camera is the simplest choice in modeling an imaging system. A scene goes in and comes out at a certain magnification. Although there are some extensions of this idea which say otherwise[3], the pinhole camera is largely accepted to be a simple geometric transformation of the scene consisting of magnification and rotation.

The origin for the model of the pinhole camera comes from historical developments: the very first camera *actually* was a pinhole camera! Known as **camera obscura**, this particular concept for designing a "camera" is by putting a very small hole in a wall and allowing the otherwise dark room to be illuminated by it. The result is that an image is projected on the wall opposite the pinhole. In Figure 2.13 the geometry and result is shown. As you can see, an image is indeed produced, therefore it is entirely reasonable to suppose that imaging can be written as a simple geometric transformation of the scene to be imaged.

[3]↑For an example I would suggest from Prof. Avinash Kak's lecture note "Modeling the Camera" https://engineering.purdue.edu/kak/computervision/. He describes in excellent detail how to extend the rigid pinhole camera model to an extended model which can describe geometric distortions.

Mathematically, this model implies that the **object coordinate** $\hat{\mathbf{u}}$ may be transformed to the **image plane coordinate** $\hat{\mathbf{x}}$ through

$$\hat{\mathbf{x}} = \mathbf{C}\hat{\mathbf{u}}. \tag{2.31}$$

There are two important distinctions to make in the case of this model which we will only introduce here as it is relevant to the pinhole camera. The first is that by the hat notation $\hat{\mathbf{u}}$, I am implying homogeneous coordinate representation of the "real word" coordinate $\mathbf{u}$. The second is that my choice in using equality for (2.31) is slightly incorrect as it is typically written to be within a scalar constant as this is a property of the homogeneous representation. More details can be found in [5], although for this thesis I will not focus on them too much as they will not arise again.

There is a significant issue regarding this model within the context of our problem of computational imaging through atmospheric turbulence. The very first, and most significant, is that *there will be nowhere to put the turbulence model!* Of course, you will have to trust me on that for now, however, in due time we shall see that this is true. Simply put, this model is an *oversimplification* of the imaging process. It has many useful applications, even in the case of camera calibration for stereo camera design [6], however, it will serve us very little in our case.

Regarding the interpretation of ray optics from a CV researcher vs. a physicist, I would claim it is related to the camera model most often assumed by each community. The CV person often assumes a pinhole camera model for their camera model. This means that the *one* ray which they assume is emitting from the object is the exact ray which passes through this pinhole. The physicist/optical engineer assumes a far more sophisticated model for a camera. This model *needs more than one ray per point to make sense!*

The interesting thing is that in one case, the case of a **pinhole camera**, the two will be the nearly the same when it comes to describing the behavior on the other side of the pinhole. To me, this is the reason for the difference. The CV community has realized that with the pinhole model, we can visualize one ray per point while the physics community often goes

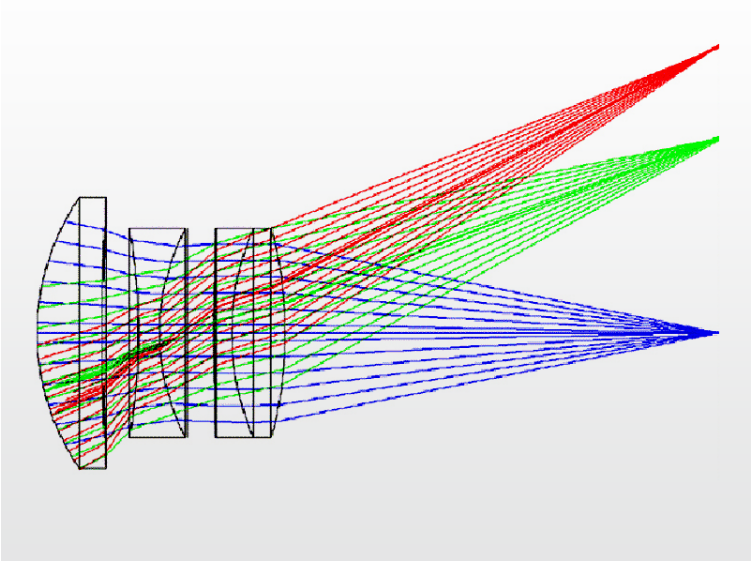


Figure 2.14. A screenshot of Zemax rendering rays passing through a lens relay.

beyond the pinhole camera, requiring multiple rays per point. In this thesis I will consider the physicist's view of multiple rays per point on the object unless mentioned otherwise.

2.3.2 Ray tracing for camera modeling

Ray tracing is *the* choice for modeling a system of lenses and other optical devices for the purposes of design. In Figure 2.14 I show an optimized lens system in Zemax where the **target function** is the spot size error; in other words, minimize the distance that the rays land from the points where they are *supposed* to land. These errors can be captured by the analysis of **aberrations**.

To truly understand aberrations, we must first present the notion of the **entrance pupil** and **exit pupil** of the camera. Hecht's book [3] great many illustrations of the entrance and exit pupil. The entrance and exit pupil are the same physical object imaged from different perspectives. They are the "maximum width" of the imaging device as seen from object and the detector side. There are far more details as to *where* they go and how to find them (based on rays which receive important names such as chief or marginal arrays), but for our purposes we will just need to know that they exist.

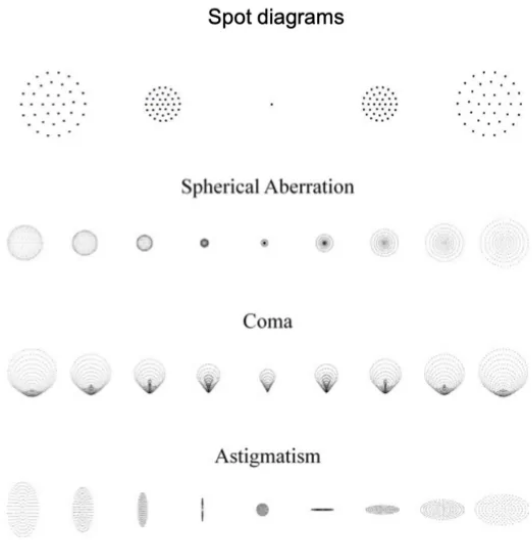


Figure 2.15. Visualization of different spot diagrams through errors in focusing. [Source]

The entrance and exit pupils, in many cases are *virtual*, meaning that they do not physically exist but are images of something which does. For us, it is important to know that the quality of an imaging system can be measured by the aberrations at the *exit* pupil. Therefore, the exit pupil represents, at least to me, the plane which we can imagine the light came from. If we can quantify the errors there, it is equivalent to describing the errors in focusing.

A common way of measuring the aberrations in a system are through a **basis representation**, one of which are the Seidel polynomials (which quantify the Seidel aberrations). This means that we can represent the aberration at the exit pupil as a sum of Seidel aberrations S_i,

$$\phi(\boldsymbol{\xi}) = \sum_i b_i S_i(\boldsymbol{\xi}), \tag{2.32}$$

where $\boldsymbol{\xi}$ is defined over the exit pupil. Different Seidel polynomials capture different types of effects, which we show some of in Figure 2.15. Here, the patterns shown are where the numerically computed rays land and do *not* capture diffractive effects.

So, why do cameras have multiple lenses? With more lenses, we have more flexibility to control higher and higher orders of aberrations! A real camera system, in even optimal

conditions, is limited by aberrations or diffraction[4], the latter which we will discuss shortly. Aberrations limit the camera due to these non-linearities in the process of the rays propagating through the lens system. The chromatic effects are exactly what drove Newton to invent the reflective telescope.

Discussing these aberrations offers us a great time to introduce they key difference between what I call the **ray tracing model** of an imaging system with the pinhole camera. For simplicity, let us assume we are imaging a flat scene which exists in a plane which $\mathbf{u}$ is defined in. Furthermore, our image will be defined on a separate plane with coordinate $\mathbf{x}$. The pinhole model would suggest that the output point $\mathbf{x}$ is related to the input point by

$$\mathbf{x} = \mathbf{u}/f \tag{2.33}$$

where f is some constant (which may be negative) which quantifies the magnification of the system. The pinhole camera has a simple philosophy for describing the way a scene is imaged: a *single* ray, which emits from $\mathbf{u}$, lands on a *single* point $\mathbf{x}$. Note this is very different from the case of diffraction.

The ray tracing model says something very different. An equation for the ray tracing model would instead be far more challenging to define, therefore we represent the lens relay as a function $h(\cdot)$ which operates in the following way.

$$\mathbf{x}_\theta = h(\mathbf{u}, \theta). \tag{2.34}$$

This equation instead says that a point $\mathbf{u}$ which emits a ray *at angle* θ will arrive at the point $\mathbf{x}_\theta$ on the imaging plane. Note that for every angle θ, we do not require that the points $\mathbf{x}_\theta$ are the same. This is one of the fundamental differences from the pinhole camera and an important step in moving towards a more realistic view of the camera system as a whole.

[4]↑One could say that it is limited by aberrations *and* diffraction, however, here I am taking this perspective that one is the limiting *factor*. That is, one is more significant than the other.

I would like to offer up my main gripe with this model: h is very irritating, or even impossible, to write out.[5] For a great example of how complicated and tedious this analysis can become, I recommend taking a peek at a few papers by Feder [7–9], a lens designer at Kodak. Note that these papers are only modeling *parabolic* surfaces of lenses – more complicated lenses would require even more equations!

I will close this discussion with a following thought which may begin to arise in your mind. Ray tracing works for the camera and ray tracing works for the scene; why can't we just use ray tracing? In truth, we could use a *modified* version of ray tracing, however, this modification would practically invoke the **wave** behavior of light. Therefore, it is ray tracing which is modified to mimic wave behavior – then why not just use the wave model? This is the goal of the next model. This will require the introduction to the wave model of light.

2.3.3 The thin lens model

The **thin lens** model, combined with the wave model of light, is the model taken for the imaging system model as described in this work. It is an idealization of the true lens model. It can be stated rather simply, after which we will provide more details. The thin lens model consists of only a simple rule for how the ray changes: the change in ray direction is related to the inverse of the thickness at the point in space of intersection. We will give a much more precise mathematical rule that determines the direction of ray propagation once we have introduced the wave model. In Figure 2.16 we show a visualization of the rays going through a true refractive system versus the thin lens model.

The thin lens model is equivalent to stating that (i) the lens transformation takes place on a 2D plane. This has many important implications. This is in contrast with a true refractive lens model, with bigger angles contributing to bigger errors. This is not true for the lens model because the transformation occurs on this 2D plane. Whether or not the ray enters in at two different angles, the change in direction will be the same.

[5]↑To quote Richard Feynman, "Geometrical optics is either very simple or else it is very complicated." I completely agree with this sentiment. To me, the former is the pinhole camera – easy! The latter is when you really try to put lenses in the mix.

If one were designing a lens relay, beyond a first guess, this model will undeniably not give optimal performance in a lens optimized according to these rules. In fact, the thin lens model suggests that one can design a one lens system that has no aberrations, a point we saw could not be true in previous discussions. While this model does not capture refractive effects, it captures diffractive effects very well. This is one of the main reasons we will choose this model.

The thin lens model can be stated simply: the thicker the lens, the more the wave is delayed in phase. Continuing with our usage of the complex representation, we present the **lens transformation**:

$$t_\ell(\boldsymbol{\xi}) = \exp\{j\phi(\boldsymbol{\xi})\} = \exp\{jk\Delta_0\}\exp\{jk(n-1)\Delta(\boldsymbol{\xi})\}, \tag{2.35}$$

where $\phi(\boldsymbol{\xi})$ represents the phase delay by the lens. I would refer the reader to [1, Chapter 5] for details on this transformation. When an incident field $U_\ell(\boldsymbol{\xi})$ reaches the lens, the lens changes its phase by multiplying $U_\ell(\boldsymbol{\xi})$ with $t_\ell(\boldsymbol{\xi})$:

$$U'_\ell(\boldsymbol{\xi}) = U_\ell(\boldsymbol{\xi})P(\boldsymbol{\xi})t_\ell(\boldsymbol{\xi}), \tag{2.36}$$

where $P(\boldsymbol{\xi})$ is the **pupil function** such that

$$P(\boldsymbol{\xi}) = \begin{cases} 1, & \boldsymbol{\xi} \text{ is inside the aperture,} \\ 0, & \text{otherwise.} \end{cases} \tag{2.37}$$

The resulting field $U'_\ell(\boldsymbol{\xi})$ is located immediately behind the lens, as shown in Figure 2.16.

As in the case of our analysis of diffraction, we are additionally interested in how the wave behaves as it moves away from the lens. For this particular problem, we are concerned with

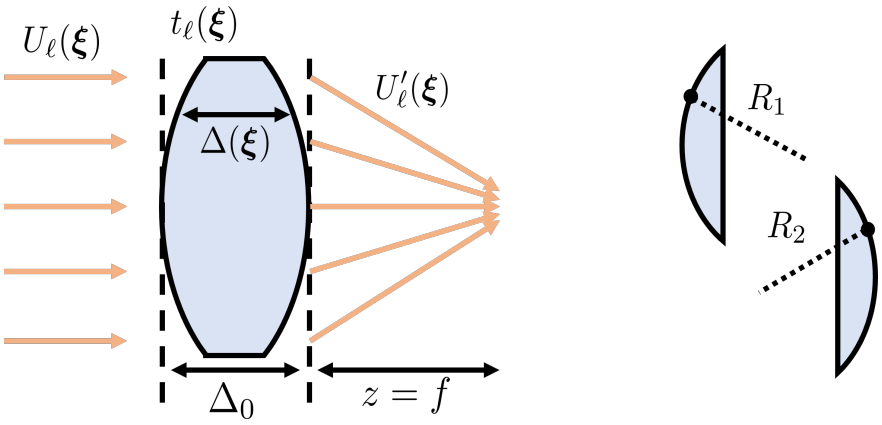


Figure 2.16. Geometry of the thin lens model. Inspired by [1]

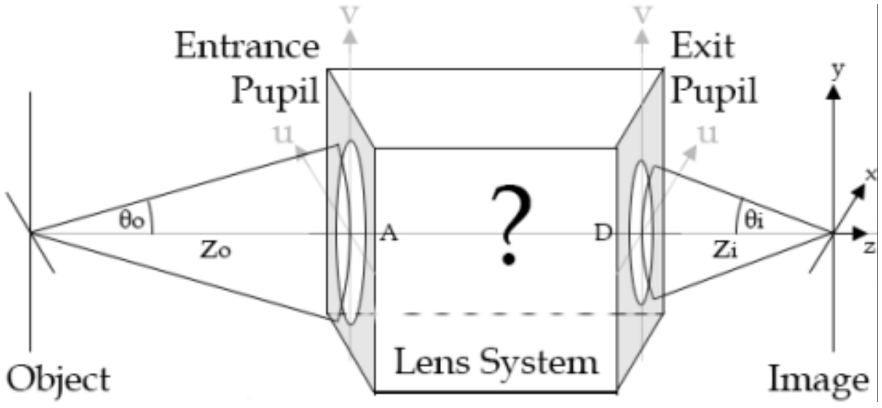


Figure 2.17. Visualization of the black box model for optics with entrance and exit pupils. [Source].

the distribution of the field at the plane of distance $z = f$. Application of Fresnel integral results in

$$\begin{aligned} U_f(\mathbf{u}) &= \frac{e^{jkf}}{j\lambda f} e^{j\frac{k}{2f}|\mathbf{u}|^2} \iint_{-\infty}^{\infty} \left\{ U'_\ell(\boldsymbol{\xi}) e^{j\frac{k}{2f}|\boldsymbol{\xi}|^2} \right\} e^{-j\frac{2\pi}{\lambda f}\boldsymbol{\xi}^T\mathbf{u}} \, d\boldsymbol{\xi} \\ &= \frac{e^{jkf}}{j\lambda f} e^{j\frac{k}{2f}|\mathbf{u}|^2} \iint_{-\infty}^{\infty} U_\ell(\boldsymbol{\xi}) P(\boldsymbol{\xi}) e^{-j\frac{2\pi}{\lambda f}\boldsymbol{\xi}^T\mathbf{u}} \, d\boldsymbol{\xi}. \end{aligned} \tag{2.38}$$

Comparing with Equation (2.24), we realize that Equation (2.38) is the Fraunhofer diffraction pattern of the windowed field incident on the lens. This is exactly the behavior we alluded to previously: with the thin lens model, we can mathematically show that the Fraunhofer pattern appears at the focal plane.

2.4 The model of image formation for this thesis

2.4.1 Imaging as a black box

Goodman describes the black box model for imaging as presented in Figure 2.17. The black box model introduces what is, to me, a nuanced point in modeling light that can likely only be appreciated with experience and exposure to the different ways of thinking about light.

First, the black box model assumes that whatever happens between the entrance and exit pupil can be described by ray optics. Everything else can be described by wave optics. This

means that the optical process can be split into three parts: (i) propagation from the object to the entrance pupil (ii) rays moving through the optical system (iii) propagation from the exit pupil to the focal plane. This means that if the Fraunhofer pattern arises at the focal plane, it has to do with whatever is happening at the *exit pupil*.

Why is this significant? (i) Our turbulence model will be described at the entrance pupil – because it has to do with whatever is between the object and camera. (ii) The thin lens model causes entrance and exit pupil to coincide – they are one the same plane. Therefore, the black box model combined with the thin lens approximation allows the rest of this thesis to exit. If we wish to use a more realistic camera model in place of the thin lens approximation, we will need to come up with some sort of new mapping.

2.4.2 Coherent and incoherent imaging

If we recall our discussion of intensity, we do not observe $U(\mathbf{x})$, but instead its corresponding intensity. This brings us to discuss the two polar opposite ends of forms of light: **coherent** and **incoherent** light. Our discussions thus far apply to both types of light, but when forming images (which are related to intensities) they will begin to diverge from one another.

When we discussed the pinhole camera, we introduced the concept of magnification. It is likely that when you imagine the image of the object, you are picturing its intensity – a magnified version of the true object's intensity at the focal plane. We will refer to this as I_g, which is the geometric image predicted by purely geometric optics (a point-to-point process). We will correspondingly define a phasor U_g such that

$$|U_g(\mathbf{x})|^2 = I_g(\mathbf{x}). \tag{2.39}$$

The concept of U_g is harder to conceptualize, but hopefully we can accept its existence.

The thin lens effectively tells us that the impulse response of a point will be the Fraunhofer pattern at the focal plane (assuming the lens system is in focus – for more details refer to [1, Chapter 5]). Because of the thin lens model, this impulse response will be **spatially invariant**. This leads us to define the **amplitude spread function (ASF)**:

Theorem 2.4.1 (**Amplitude spread function**). *The amplitude spread function of the system is*

$$h(\mathbf{x}) = \frac{A}{\lambda z_2} \iint_{-\infty}^{\infty} P(\boldsymbol{\xi}) \exp\left\{ -j \frac{2\pi}{\lambda z_2} \mathbf{x}^T \boldsymbol{\xi} \right\} \, d\boldsymbol{\xi}, \tag{2.40}$$

which is the Fourier transform of $P(\boldsymbol{\xi})$ evaluated at frequencies $(\frac{x}{\lambda z_2}, \frac{y}{\lambda z_2})$ and with $A = 1/(\lambda z_1)$, though often for notational convenience, we can set the constant $\frac{A}{\lambda z_2}$ to unity. The resulting field formed at the focal plane is

$$U_i(\mathbf{x}) = h(\mathbf{x}) \circledast U_g(\mathbf{x}), \tag{2.41}$$

where $\circledast$ denotes the 2D convolution.

I have removed a few scaling constants here, though in practise we can often assume the ASF integrates to 1. Furthermore, it is essential that h is applied to U_g with the proper sampling, so we would emphasize that U_g is a magnified version of the object's "true" phasor U_0. See Goodman [1] for more details regarding this. We will often write (2.40) as

$$h(\mathbf{x}) = \frac{A}{\lambda z_2} \mathfrak{Fourier}\{ P(\boldsymbol{\xi}) \} \Big|_{\mathbf{f} = \frac{\mathbf{x}}{\lambda z_2}}. \tag{2.42}$$

Later on in this thesis, we will typically drop this scaling as it should be understood the ASF is applied to the geometric phasor with proper sampling.

Now we must consider how to evaluate the intensity. Although it seems straightforward, the study of **coherence** makes things a bit more interesting. First, we summarize the differences in coherent and incoherent imaging as follows:

Theorem 2.4.2 (**Coherent Image Formation**). *If the light source is **coherent**, then the observed image $I_i(\mathbf{x})$ is*

$$I_i(\mathbf{x}) = \left| \widetilde{h}(\mathbf{x}) \circledast U_g(\mathbf{x}) \right|^2. \tag{2.43}$$

Theorem 2.4.3 (**Incoherent Image Formation**). *If the light source is incoherent, then the observed image* $I_i(\mathbf{x})$ *is*

$$I_i(\mathbf{x}) = \left|\widetilde{h}(\mathbf{x})\right|^2 \circledast I_g(\mathbf{x}). \tag{2.44}$$

The reason for these differences leads us down a very interesting path in the study of optics, however, one we shall not take in this thesis. I would only like to highlight a few things.

In Figure 2.18 the difference between coherent light and incoherent light is shown. Incoherent light is random in nature while coherent light is largely deterministic. Incoherent light often emerges from thermal sources whereas coherent light is from lasers. Recall that in our discussion of intensity, the statistical expectation was involved. Although we do not show it here, this randomness (and lack thereof for coherent light) is the reason for the difference in expressions for the two varieties of light.

Next I would like to comment that we will refer to $|h|^2$ as the **point spread function (PSF)**. The PSF will be at the center of this thesis as this thesis primarily focuses on incoherent light and simulation. The PSF, as currently defined, represents our impulse response of the system which is convolved with the magnified image intensity I_g.

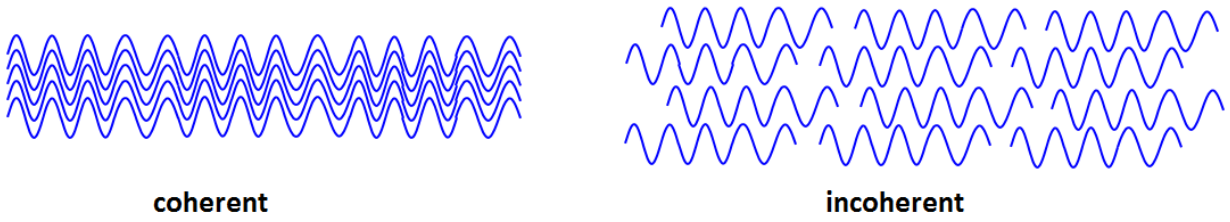


Figure 2.18. Visualizing coherent and incoherent light. [Source]

Finally, I would like to comment on a few interesting details regarding coherence. The study of coherence has been in optics for a long while, but it was often regarded as *boring*! That is until the publication of the van Cittert-Zernike theorem [10] which states that at a suitable distance, incoherent light will be imaged as coherent light. We have no need for this theorem in this thesis, but I do find it interesting to note that the community largely dispensed with the subject until a very interesting result was published. Also, Fritz Zernike,

who the theorem is partially named after, will influence a great deal of this entire thesis and will show up in a few places.

2.4.3 Putting it all together

We now have a few pieces of information to collect and apply: light is wave, the Fraunhofer pattern appears at the focal plane, and we convolve our geometric image with the point spread function. Now, it is time to put it all together:

Definition 2.4.1 (**Image formation model**). *The observed image I_i for an imaging system with aperture $P(\boldsymbol{\xi})$ and phase error $\phi(\boldsymbol{\xi})$ is related to its geometric image I_g by*

$$I_i(\mathbf{x}) = \left(\left|\mathfrak{Fourier}\{P(\boldsymbol{\xi})e^{j\phi(\boldsymbol{\xi})}\}\right|^2 \circledast I_g\right)(\mathbf{x}), \tag{2.45}$$

where we've dropped the scaling of the PSF for simplicity.

This model captures the effects we will need for the rest of this thesis in a concise, accurate, and suitably simple way. The phase function $\phi(\boldsymbol{\xi})$ is where we will eventually place the turbulence model.

2.4.4 Fourier analysis of imaging systems

Now we turn to doing a bit of a Fourier analysis for imaging systems. We'll start with coherent imaging systems and will relate them to incoherent systems. This will give us some insight into their connection as well as an alternative perspective on the diffraction limit imposed by a finite aperture. First, we define the pair of Fourier transforms

$$G_g(\mathbf{f}) \stackrel{\text{def}}{=} \mathfrak{Fourier}\{U_g(\mathbf{x})\}, \quad \text{(geometric phasor)}$$
$$G_i(\mathbf{f}) \stackrel{\text{def}}{=} \mathfrak{Fourier}\{U_i(\mathbf{x})\}, \quad \text{(observed phasor)}$$

which denote the Fourier transform of the geometric (i.e., no distortions other than magnification) and the observed phasor, though we again emphasize that we do not directly observe the phasor.

The Fourier transform of the ASF we give a special name known as

Definition 2.4.2 (**Amplitude transfer function (ATF)**). *The amplitude transfer function (ATF) is*

$$H(\mathbf{f}) \stackrel{def}{=} \mathfrak{Fourier}\{h(\mathbf{x})\}.$$

which is the Fourier transform of the amplitude spread function (ASF).

Because of the coherent image formation equation, we may write

$$G_i(\mathbf{f}) = H(\mathbf{f})G_g(\mathbf{f}). \tag{2.46}$$

However, recall that the ASF is related to the Fourier transform of the pupil (assuming no distortions). Due to repeated application of the Fourier transforming, and the appropriate scaling, the ATF can be written as

$$H(\mathbf{f}) = P(-\lambda z_2 \mathbf{f}) = P(\lambda z_2 \mathbf{f}),$$

where I've left off some constants for simplicity and utilized symmetry by a circular pupil function, i.e., $P(\boldsymbol{\xi}) = P(-\boldsymbol{\xi})$. This means the Fourier-space behavior of a coherent imaging system can be easily understood by the shape of the pupil $P(\boldsymbol{\xi})$.

We define two related quantities for incoherent imaging systems,

$$\mathcal{G}_g(\mathbf{f}) = \mathfrak{Fourier}\{I_g(\mathbf{x})\},$$
$$\mathcal{G}_i(\mathbf{f}) = \mathfrak{Fourier}\{I_i(\mathbf{x})\}.$$

We similarly define the Fourier transform of the PSF in the following definition:

Definition 2.4.3 (**Optical transfer function**). *The optical transfer function is*

$$\mathcal{H}(\mathbf{f}) = \mathfrak{Fourier}\{|h(\mathbf{x})|^2\} = \frac{\int_{-\infty}^{\infty} |h(\mathbf{x})|^2 \exp\{-j2\pi \mathbf{f}^T \mathbf{x}\} d\mathbf{x}}{\int_{-\infty}^{\infty} |h(\mathbf{x})|^2 d\mathbf{x}}.$$

The incoherent image formation equation Equation (2.44) can then be written in the Fourier domain as

$$\mathcal{G}_i(\mathbf{f}) = \mathcal{H}(\mathbf{f})\mathcal{G}_g(\mathbf{f}). \tag{2.47}$$

This leads us to now make a connection between the OTF and the ATF. For a circular aperture, the amplitude transfer function is

$$H(\mathbf{f}) = P(\lambda z_2 \mathbf{f}). \tag{2.48}$$

Through the relationship shown in 2.4.3, as elaborated by Goodman, the following relationship can be shown:

$$\mathcal{H}(\mathbf{f}) = \frac{P(\lambda z_2 \mathbf{f}) \circledast P(\lambda z_2 \mathbf{f})}{P(0,0) \circledast P(0,0)}, \tag{2.49}$$

which we can recognize as simply an autocorrelation of the scaled pupil. The overlapping region can be computed by referring to Figure 2.19. The details for the evaluation of the circular OTF may be found in Goodman [1]. Using geometry, it can be shown that

$$\mathcal{H}(f) = \begin{cases} \frac{2}{\pi}\left[\arccos\left(\frac{f}{2f_0}\right) - \frac{f}{2f_0}\sqrt{1-\left(\frac{f}{2f_0}\right)^2}\right], & f \le 2f_0, \\ 0, & \text{otherwise}, \end{cases} \tag{2.50}$$

where we defined the cutoff frequency

$$f_0 = \frac{w}{\lambda z_i}, \tag{2.51}$$

where here f_0 corresponds to the cutoff frequency of the coherent system of the same aperture, meaning the incoherent version of the same system will have a frequency cutoff that is twice

the size. The OTF obtained in Equation (2.50) is important, as it is the OTF of an incoherent **diffraction-limited** system. Assuming a perfect lens without distortions such as defocus, the system has a fundamental limit due to diffraction.

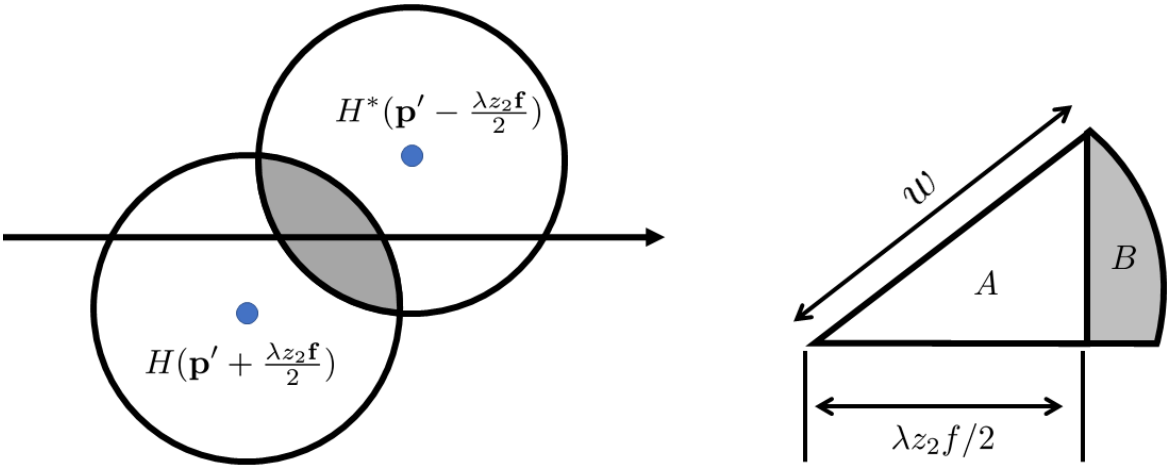


Figure 2.19. The optical transfer function $\mathcal{H}(\mathbf{f})$ is the autocorrelation function of $H(\mathbf{f})$ evaluated at position given by the two circles. This translates to computing the ratio between the overlapping area and the total area.

We wish to add a bit more discussion here and comment on an imaging system which images an object first in a coherent fashion, then in an incoherent fashion. Because of the autocorrelation property in the frequency domain, the incoherent image will have a *higher* resolution – this is because the pass band of the system is wider. Goodman [1] further elaborates on the properties of the ATF and OTF which we suggest to the interested reader for more details.

2.5 Spatially varying imaging systems

The imaging systems described by (2.45) are spatially invariant. For each point that emits light, the point spread function $|h|^2$ is identical. This is a nice equation to work with, all of the typical Fourier relationships and tricks we use to model linear systems of this variety apply here. For the sake of notational simplicity, in this Section I will use only h instead of $|h|^2$.

We now wish to extend to **spatially varying imaging systems** for two primary reasons: (i) to highlight the off-axis effects of real camera models, (ii) to build a foundation to describe turbulence. Regarding (i), as indicated by our camera modeling discussion, there will almost always be some sort of aberration in the imaging system, and it will almost certainly change as a function of angle. Furthermore, (ii) is important because, after all, we are interesting in computational imaging through atmospheric turbulence. The types of images which are produced after propagation through the atmosphere are almost always spatially varying in nature, therefore, we will undeniably need this capability in modeling.

2.5.1 Spatially varying convolution

We often assume that the systems we are working with are spatially invariant. This is often because they either are or they are *so* close that they can be suitably approximated as such. In some way, as opposed to viewing the extra lens in a camera to correct for aberrations, we can think of more lenses in the camera giving us more justification in making this approximation of spatial invariance.

Let's now introduce the possibility of the impulse response (PSF) to vary spatially. In this case, I will write the more general superposition integral

$$I(\mathbf{x}) = \int h(\mathbf{u}, \mathbf{x}) J(\mathbf{u}) d\mathbf{u}. \tag{2.52}$$

We note that the standard convolution integral is a special case of the superposition integral. Now, we wish to restrict ourselves to a particular subset of kernels that satisfy

$$h(\mathbf{u}, \mathbf{x}) = \underbrace{h(\mathbf{u}, \mathbf{x} - \mathbf{u})}_{\text{Scattering}} \qquad h(\mathbf{u}, \mathbf{x}) = \underbrace{h(\mathbf{x}, \mathbf{x} - \mathbf{u})}_{\text{Gathering}}. \tag{2.53}$$

You may be wondering *why* these two particular kernels. One reason will be apparent a bit later, but for now we note that they are *so close* to convolution. This is because it depends on the difference vector $\mathbf{x} - \mathbf{u}$, however, it also depends on the absolute position $\mathbf{u}$ or $\mathbf{x}$. For this reason, we cannot say that it is *exactly* convolution, but pretty close!

We have provided labels to (2.53) of **scattering** and **gathering**, although they cannot be appreciated fully at this point. My goal is to convince you of something that may hurt your brain if you ever learned from Oppenheim's *Signals & Systems* textbook [11]. Granted, there is nothing wrong in his textbook, however, applying the often accepted mentality towards convolution will lead you astray in the case of spatially varying convolution. Specifically, I wish to convince you of the following: *Scattering is correct for modeling imaging and gathering is wrong!* This is not to say that gathering is *always* wrong, rather that just for this situation it is incorrect.

2.5.2 Explaining scattering vs. gathering

Let us summarize the main difference in scattering versus gathering:

Proposition 2.5.1. *Scattering convolution is parameterized by the* ***source*** *location while gathering is parameterized by the* ***receiver*** *location.*

We can think of it this way, the difference vector $\mathbf{x} - \mathbf{u}$ acts as the standard translational behavior of the impulse response, while the first parameter (either $\mathbf{u}$ or $\mathbf{x}$) will dictate whether or not it depends on the input location $\mathbf{u}$ or output location $\mathbf{x}$. Since the shifting acts the same in both cases, the main difference is the parameterization based on location.

To illustrate this, let's first write out the scattering equation,

$$I(\mathbf{x}) = \int h(\mathbf{u}, \mathbf{x} - \mathbf{u}) J(\mathbf{u}) d\mathbf{u}. \tag{2.54}$$

Let's consider the case of a single point source in the input, $J(\mathbf{u}) = \delta(\mathbf{u})$. In this case equation, our output will become

$$I(\mathbf{x}) = \int h(\mathbf{u}, \mathbf{x} - \mathbf{u}) \delta(\mathbf{u}) d\mathbf{u}, \tag{2.55}$$

$$= h(\mathbf{0}, \mathbf{x}). \tag{2.56}$$

We can compare this to the gathering equation under the same conditions,

$$I(\mathbf{x}) = \int h(\mathbf{x}, \mathbf{x} - \mathbf{u}) J(\mathbf{u}) d\mathbf{u}, \tag{2.57}$$

which leads to

$$I_i(\mathbf{x}) = \int h(\mathbf{x}, \mathbf{x} - \mathbf{u}) \delta(\mathbf{u}) d\mathbf{u}, \tag{2.58}$$

$$= h(\mathbf{x}, \mathbf{x}). \tag{2.59}$$

Therefore, for a point in the *source* plane at $\mathbf{0}$ will have the corresponding output $h(\mathbf{0}, \mathbf{x})$ or $h(\mathbf{x}, \mathbf{x})$. Recall that the first position parameterizes the kernel while the latter position dictates the position. For the scattering output, the point spread function is *one* point spread function (i.e., a single parameterized PSF). Compare this to the gathering output which is the value of the $\mathbf{x}$'s PSF's $\mathbf{x}$ position (a bit of a tongue-twister).

2.5.3 Basis representations and spatially varying convolution

First, let's consider the spatially invariant convolution integral,

$$I(\mathbf{x}) = \int h(\mathbf{x} - \mathbf{u}) J(\mathbf{u}) d\mathbf{u}. \tag{2.60}$$

In many engineering applications, we often represent h through a **basis representation**, with each component of the basis set capturing a characteristic of h. Let us suppose that h lends itself to such a basis representation,

$$h(\mathbf{x}) = \sum_i a_i \varphi_i(\mathbf{x}). \tag{2.61}$$

Let's assume that the set of basis kernels $\{\varphi_i\}$ captures h exactly. In this case, we may substitute our basis representation into the convolution kernel,

$$I(\mathbf{x}) = \sum_i a_i \int \varphi_i(\mathbf{x}-\mathbf{u})J(\mathbf{u})d\mathbf{u}, \tag{2.62}$$

$$= \sum_i a_i(\varphi_i \circledast J)(\mathbf{x}). \tag{2.63}$$

In words, we may convolve the input J by our chosen basis and perform a weighted sum to get the desired output I.

A spatially varying convolutional kernel may also be decomposed by such a basis representation. In the case of a kernel parameterized by the source position, it is reasonable to state that the basis coefficients will depend on $\mathbf{u}$,

$$h(\mathbf{u}, \mathbf{x}-\mathbf{u}) = \sum_i a_{\mathbf{u},i}\varphi_i(\mathbf{x}-\mathbf{u}). \tag{2.64}$$

If we substitute this into our superposition integral, we can derive the following result:

$$I(\mathbf{x}) = \sum_i \int a_{\mathbf{u},i}\varphi_i(\mathbf{x}-\mathbf{u})J(\mathbf{u})d\mathbf{u}, \tag{2.65}$$

$$= \sum_i (\varphi_i \circledast a_{\mathbf{u},i}J)(\mathbf{x}). \tag{2.66}$$

Holding off on commenting on this for now, let's do the same with gathering convolution in which the basis coefficients dependon $\mathbf{x}$,

$$h(\mathbf{x}, \mathbf{x}-\mathbf{u}) = \sum_i a_{\mathbf{x},i}\varphi_i(\mathbf{x}-\mathbf{u}). \tag{2.67}$$

Through substitution, we find

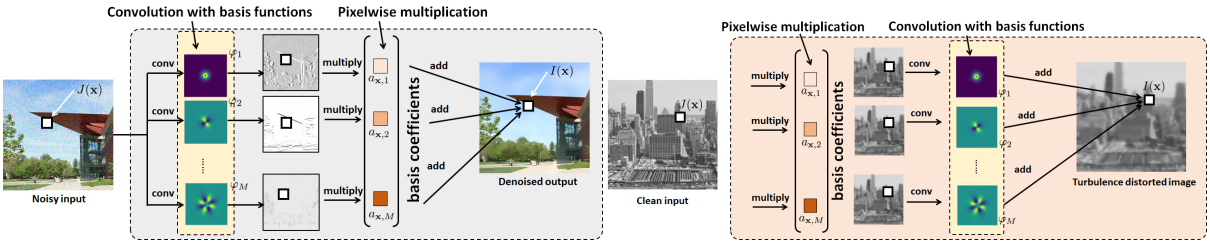
$$I(\mathbf{x}) = \sum_i a_{\mathbf{x},i} \int \varphi_i(\mathbf{x}-\mathbf{u})J(\mathbf{u})d\mathbf{u}, \tag{2.68}$$

$$= \sum_i a_{\mathbf{x},i}(\varphi_i \circledast J)(\mathbf{x}). \tag{2.69}$$

This is another way of viewing the parameterization by source or receiver.

Scattering vs. gathering at a high level

Let's now present a bit of an overview as to the differences in scattering in gathering. The first is to note that choice in model results in the order of operations changing – scattering implies basis coefficient multiplication followed by convolution and sum whereas gathering performs convolution first then a weighted sum. We can visualize this with filter banks as in Figure 2.20.



(a) $I^{\text{gather}}(\mathbf{x})$ (b) $I^{\text{scatter}}(\mathbf{x})$

Figure 2.20. Block diagram visualization of the different approximations to scattering and gathering.

Let us look at the equations more carefully through the lens of matrices and vectors. Let $\mathbf{H} \in \mathbb{R}^{N \times N}$ be the matrix representation of the spatially varying blur kernel $h(\mathbf{x}, \mathbf{u})$. We assume that there is a set of *circulant* matrices $\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_M$ representing the set of M spatially invariant basis functions $\{\varphi_1, \varphi_2, \ldots, \varphi_M\}$.

We consider two sets of diagonal matrices. For every index m, we define

$$\mathbf{D}_m^{\mathbf{x}} = \operatorname{diag}\left\{ \begin{bmatrix} a_{\mathbf{x}_1,m} \\ \vdots \\ a_{\mathbf{x}_N,m} \end{bmatrix} \right\}, \quad \mathbf{D}_m^{\mathbf{u}} = \operatorname{diag}\left\{ \begin{bmatrix} a_{\mathbf{u}_1,m} \\ \vdots \\ a_{\mathbf{u}_N,m} \end{bmatrix} \right\}$$

Then, the gathering and scattering equations are

$$\text{(Gathering)}: \qquad \mathbf{H}^{\mathbf{x}} = \sum_{m=1}^{M} \mathbf{D}_m^{\mathbf{x}} \mathbf{H}_m \tag{2.70}$$

$$\text{(Scattering)}: \qquad \mathbf{H}^{\mathbf{u}} = \sum_{m=1}^{M} \mathbf{H}_m \mathbf{D}_m^{\mathbf{u}} \tag{2.71}$$

In other words, the difference lies in how we order the diagonal matrices and the spatially invariant convolution matrices.

It is not difficult to show that the two constructions can never be the same unless $\mathbf{D}_m^{\mathbf{x}}$ and $\mathbf{D}_m^{\mathbf{u}}$ are identity matrices up to a scalar multiple. To see this, we consider the case where $M = 1$.

Theorem 2.5.1. *Let $\mathbf{H} \in \mathbb{R}^{N \times N}$ be a square matrix. Let $\mathbf{A} = diag[a_1, \ldots, a_N]$ and $\mathbf{B} = diag[b_1, \ldots, b_N]$ be two diagonal matrices. Then, $\mathbf{AH} = \mathbf{HB}$ if and only if $\mathbf{A} = \mathbf{B} = \lambda \mathbf{I}$ for some constant λ where $\mathbf{I}$ is the identity matrix.*

Proof. We just need to write out the terms. For $\mathbf{AH}$, we can show that

$$\mathbf{AH} = \begin{bmatrix} a_1 h_{11} & a_1 h_{12} & \ldots & a_1 h_{1N} \\ a_2 h_{21} & a_2 h_{22} & \ldots & a_2 h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_N h_{N1} & a_N h_{N2} & \ldots & a_N h_{NN} \end{bmatrix},$$

and for $\mathbf{HB}$, we can show that

$$\mathbf{HB} = \begin{bmatrix} b_1 h_{11} & b_2 h_{12} & \ldots & b_N h_{1N} \\ b_1 h_{21} & b_2 h_{22} & \ldots & b_N h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ b_1 h_{N1} & b_2 h_{N2} & \ldots & b_N h_{NN} \end{bmatrix},$$

By comparing terms, we can see that the only possibility for $\mathbf{AH} = \mathbf{HB}$ is to require $\mathbf{A} = \mathbf{B} = \lambda \mathbf{I}$. □

The result of the previous theorem implies that if we have a convolution matrix $\mathbf{H}_m$ (which is circulant), for the scattering and gathering operations to be equivalent, we need

$$\mathbf{D}_m^{\mathbf{x}}\mathbf{H}_m = \mathbf{H}_m\mathbf{D}_m^{\mathbf{u}}, \quad \text{for all } m.$$

Theorem 2.5.1 asserts that we need $\mathbf{D}_m^{\mathbf{x}} = \mathbf{D}_m^{\mathbf{u}} = \lambda\mathbf{I}$. But if $\mathbf{D}_m^{\mathbf{x}} = \mathbf{D}_m^{\mathbf{u}} = \lambda\mathbf{I}$, then the underlying blur must be spatially invariant.

Another observation of Theorem 2.5.1 is that in general,

$$\sum_{m=1}^{M} \mathbf{D}_m^{\mathbf{x}}\mathbf{H}_m \neq \sum_{m=1}^{M} \mathbf{H}_m\mathbf{D}_m^{\mathbf{u}}. \tag{2.72}$$

Therefore, the gathering and scattering equations are *mutually exclusive*. If we say that $h(\mathbf{x}, \mathbf{u})$ can be exactly represented by the gathering equation, then there will be an approximation error when representing $h(\mathbf{x}, \mathbf{u})$ using the scattering equation, and vice versa.

2.5.4 Do these actually matter?

The question to ask now is: Given the gathering equation and the scattering equation, does it really matter if we choose the "wrong" one? The goal of this section is to answer the question through a few examples.

Scattering Works for Optical Simulation

In the first example, we consider the problem of simulating the resulting image given a specific light source. The light source $J(\mathbf{u})$ we consider here consists of two delta functions:

$$J(\mathbf{u}) = \delta(\mathbf{u} + \Delta) + \delta(\mathbf{u} - \Delta),$$

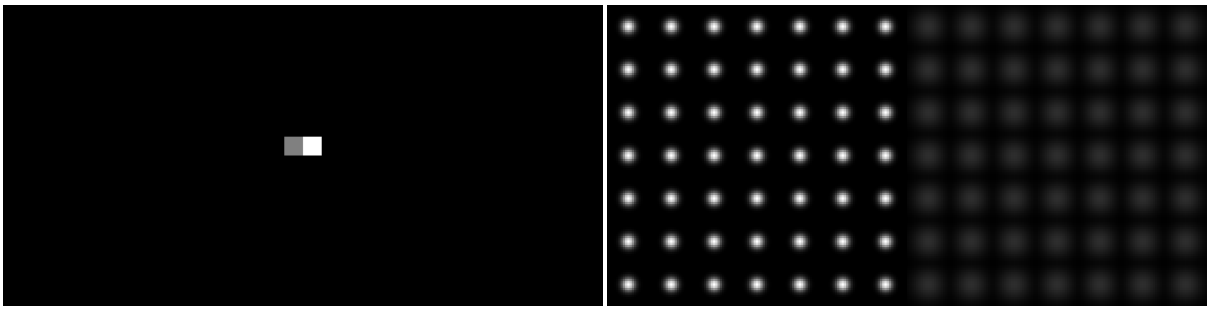
where Δ is a small displacement from $\mathbf{u}$. For convenience, we consider a plane with two halves. The separation is located at the origin; any pixel that is on the left is denoted as "$\mathbf{u} < 0$" (this notation means that the *horizontal* component of $\mathbf{u}$ is less than zero). Similarly,

any pixel that is on the right is denoted as "$\mathbf{u} \geq 0$". Therefore, $\delta(\mathbf{u} + \Delta)$ is on the left, and $\delta(\mathbf{u} - \Delta)$ is on the right.

Imagine that in front of the light source, we put two transparent sheets with different phase profiles (which can be engineered using a meta-material.) This will give us a spatially varying blur kernel $h(\mathbf{x}, \mathbf{u})$, and for simplicity, if light is emitted on the left hand side, then the blur uses a smaller radius; if the light is emitted on the right hand side, then the blur uses a larger radius. Thus, we write

$$h(\mathbf{x}, \mathbf{u}) = \begin{cases} \frac{1}{2\pi\sigma_1^2} \exp\left\{-\frac{\|\mathbf{x}-\mathbf{u}\|^2}{2\sigma_1^2}\right\} \stackrel{\text{def}}{=} \varphi_1(\mathbf{x} - \mathbf{u}), & \mathbf{u} < 0, \\ \frac{1}{2\pi\sigma_2^2} \exp\left\{-\frac{\|\mathbf{x}-\mathbf{u}\|^2}{2\sigma_2^2}\right\} \stackrel{\text{def}}{=} \varphi_2(\mathbf{x} - \mathbf{u}), & \mathbf{u} \geq 0, \end{cases} \tag{2.73}$$

where $\sigma_1 < \sigma_2$. Figure 2.21(b) illustrates these spatially varying blur kernels. For visualization purposes, we show only the PSFs at a grid of points. In reality, the PSFs is defined continuously over $\mathbf{u}$.



(a) Input $J(\mathbf{u})$ (b) Blurs $h(\mathbf{x}, \mathbf{u})$ (illustrated over a grid of points)

Figure 2.21. Visualization of an example with (a) $J(\mathbf{u})$ and (b) a grid of spatially varying blur kernels.

Before we do any calculus, we can perform a thought experiment. Figure 2.22 illustrates a hypothetical experimental setup. On the object plane there are two points emitting light through a meta surface with two different phase profiles. As the light propagates outward from the source through diffraction, the electromagnetic fields superimpose over each other. When the light reaches the aperture, the two diffraction patterns overlap. Therefore, the

resulting image $I(\mathbf{x})$, without any calculation, should be one big diffraction pattern. It is impossible to obtain a sharp cutoff and two diffraction patterns.

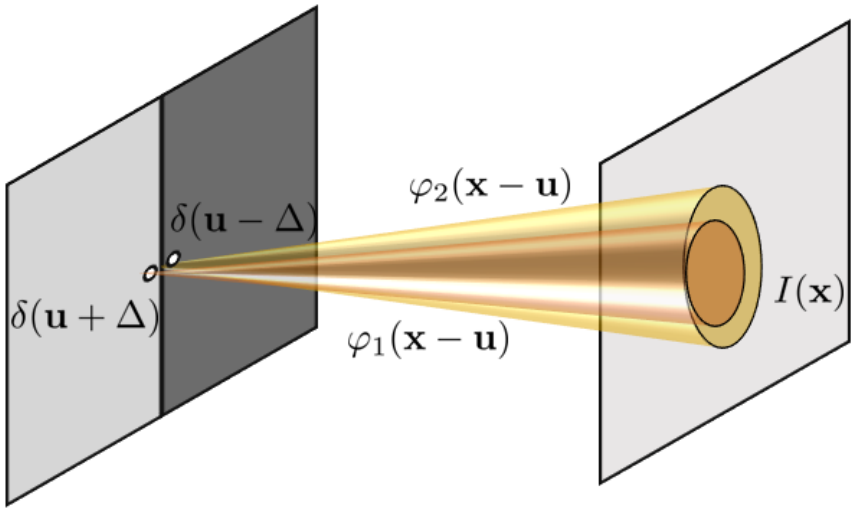


Figure 2.22. Thought experiment with two points on the object plane, diffracting through two different metasurfaces. The resulting image should, in principle, be one superimposed diffraction pattern.

With this thought experiment in mind, we can now talk about the gathering and the scattering equations. For the gathering equation, since the PSF has a simple binary structure, we can define it as

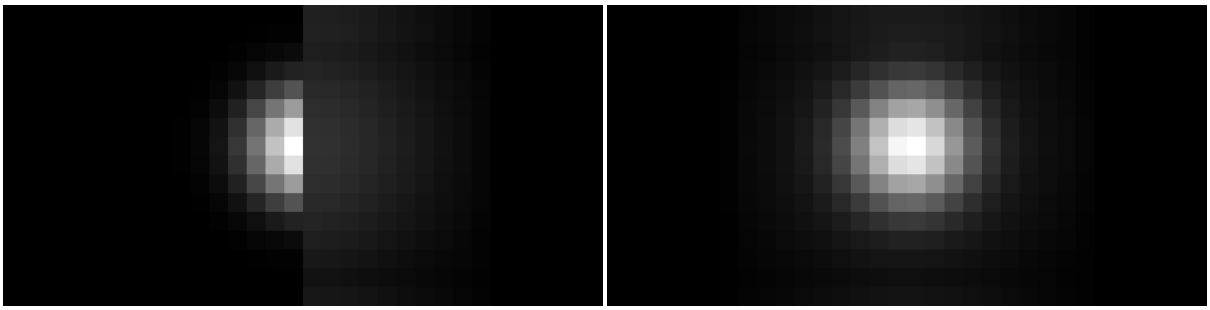
$$h^{\text{gather}}(\mathbf{x},\mathbf{u}) = \underbrace{\mathbb{I}\{\mathbf{x}\in\text{left}\}}_{=a_{\mathbf{x},1}} \times \varphi_1(\mathbf{x}-\mathbf{u}) \\ \underbrace{\mathbb{I}\{\mathbf{x}\in\text{right}\}}_{=a_{\mathbf{x},2}} \times \varphi_2(\mathbf{x}-\mathbf{u}), \tag{2.74}$$

where $\mathbb{I}\{\cdot\}$ is an indicator function For scattering, the equation takes a similar form

$$h^{\text{scatter}}(\mathbf{x},\mathbf{u}) = \underbrace{\mathbb{I}\{\mathbf{u}\in\text{left}\}}_{=a_{\mathbf{u},1}} \times \varphi_1(\mathbf{x}-\mathbf{u}) \\ \underbrace{\mathbb{I}\{\mathbf{u}\in\text{right}\}}_{=a_{\mathbf{u},2}} \times \varphi_2(\mathbf{x}-\mathbf{u}), \tag{2.75}$$

where we replaced $a_{\mathbf{x},m}$ by $a_{\mathbf{u},m}$.

By comparing the gathering equation Equation (2.74) and the scattering equation Equation (2.75) with the original spatially varying $h(\mathbf{x}, \mathbf{u})$ in Equation (2.73), it is clear that only the scattering equation will match with the original $h(\mathbf{x}, \mathbf{u})$ because they are both indexed by $\mathbf{u}$. However, to confirm that this is indeed the case, it would be helpful to look at the resulting image, as illustrated in Figure 2.23. Let us explain how these figures are obtained.



(a) $I^{\text{gather}}(\mathbf{x})$ (b) $I^{\text{scatter}}(\mathbf{x})$

Figure 2.23. Comparison between gathering and scattering for the setup in Figure 2.22. Notice that for this optical experiment, we should expect the resulting image to contain one big diffraction pattern. However, only the scattering equation demonstrates this.

The resulting image of the gathering equation can be shown as follows:

$$I^{\text{gather}}(\mathbf{x}) = \int \left(a_{\mathbf{x},1}\varphi_1(\mathbf{x}-\mathbf{u}) + a_{\mathbf{x},2}\varphi_2(\mathbf{x}-\mathbf{u})\right) J(\mathbf{u})\, d\mathbf{u}.$$

Since $a_{\mathbf{x},1} = 1$ if $\mathbf{x} \geq 0$ and $a_{\mathbf{x},1} = 0$ if $\mathbf{x} < 0$ (similarly for $a_{\mathbf{x},2}$), the coefficients $a_{\mathbf{x},1}$ and $a_{\mathbf{x},2}$ will create two cases as

$$\begin{aligned}
&I^{\text{gather}}(\mathbf{x}) \\
&= \begin{cases} \int_{-\infty}^{\infty} \varphi_1(\mathbf{x}-\mathbf{u}) J(\mathbf{u})\, d\mathbf{u}, & \mathbf{x} < 0, \\ \int_{-\infty}^{\infty} \varphi_2(\mathbf{x}-\mathbf{u}) J(\mathbf{u})\, d\mathbf{u}, & \mathbf{x} \geq 0. \end{cases} \\
&= \begin{cases} \int_{-\infty}^{\infty} \varphi_1(\mathbf{x}-\mathbf{u})(\delta(\mathbf{u}+\Delta) + \delta(\mathbf{u}-\Delta))\, d\mathbf{u}, & \mathbf{x} < 0, \\ \int_{-\infty}^{\infty} \varphi_2(\mathbf{x}-\mathbf{u})(\delta(\mathbf{u}+\Delta) + \delta(\mathbf{u}-\Delta))\, d\mathbf{u}, & \mathbf{x} \geq 0. \end{cases} \\
&= \begin{cases} \varphi_1(\mathbf{x}+\Delta) + \varphi_1(\mathbf{x}-\Delta), & \mathbf{x} < 0, \\ \varphi_2(\mathbf{x}+\Delta) + \varphi_2(\mathbf{x}-\Delta), & \mathbf{x} \geq 0. \end{cases}
\end{aligned} \tag{2.76}$$

If we draw $I^{\text{gather}}(\mathbf{x})$, we will obtain the figure shown in Figure 2.23(a).

For scattering, we can carry out the same derivation and show that

$$\begin{aligned}
I^{\text{scatter}}(\mathbf{x}) &= \int_{-\infty}^{\infty} \left(a_{\mathbf{u},1}\varphi_1(\mathbf{x}-\mathbf{u}) + a_{\mathbf{u},2}\varphi_2(\mathbf{x}-\mathbf{u})\right) J(\mathbf{u})\, d\mathbf{u} \\
&= \int_{-\infty}^{\infty} \varphi_1(\mathbf{x}-\mathbf{u})[a_{\mathbf{u},1} J(\mathbf{u})]\, d\mathbf{u} \\
&\qquad + \int_{-\infty}^{\infty} \varphi_2(\mathbf{x}-\mathbf{u})[a_{\mathbf{u},2} J(\mathbf{u})]\, d\mathbf{u} \\
&= \int_{-\infty}^{\infty} \varphi_1(\mathbf{x}-\mathbf{u})\delta(\mathbf{u}+\Delta)\, d\mathbf{u} \\
&\qquad + \int_{-\infty}^{\infty} \varphi_2(\mathbf{x}-\mathbf{u})\delta(\mathbf{u}-\Delta)\, d\mathbf{u} \\
&= \varphi_1(\mathbf{x}+\Delta) + \varphi_2(\mathbf{x}-\Delta),
\end{aligned} \tag{2.77}$$

where we used the fact that $a_{\mathbf{u},1} = \mathbb{I}\{\mathbf{u} \in \text{left}\}$ and so $a_{\mathbf{u},1}\delta(\mathbf{u}+\Delta) = \delta(\mathbf{u}+\Delta)$ whereas $a_{\mathbf{u},2}\delta(\mathbf{u}) = 0$. Similarly we have $a_{\mathbf{u},1}\delta(\mathbf{u}-\Delta) = 0$ and $a_{\mathbf{u},2}\delta(\mathbf{u}-\Delta) = \delta(\mathbf{u}-\Delta)$.

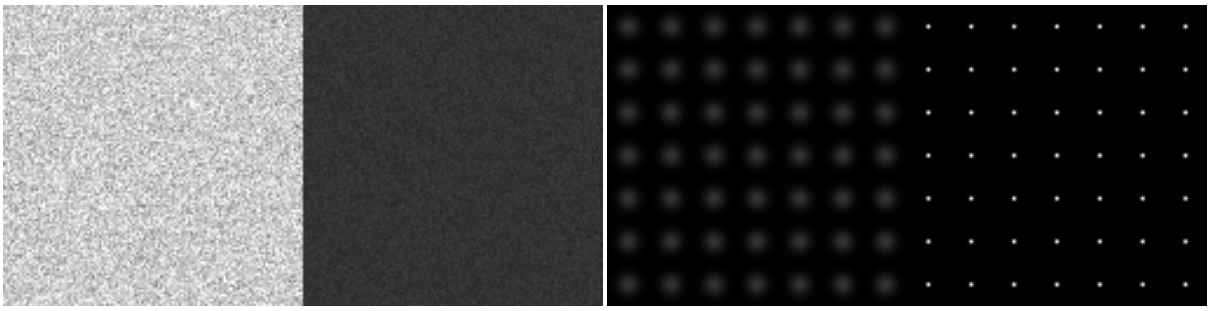
If we draw the resulting image, we will obtain the figure shown in Figure 2.23(b). This is consistent with what we expect in Figure 2.22 and the theoretical derivation in Section III.

Gathering Works for Image Filtering

In the second example, we consider the problem of *image filtering*. The scenario is that we are given a noisy image $J(\mathbf{u})$ that contains two regions:

$$\begin{aligned} J(\mathbf{u}) &= \begin{cases} \theta_1 + \text{Gauss}(0, \sigma_1^2), & \mathbf{u} < 0, \\ \theta_2 + \text{Gauss}(0, \sigma_2^2), & \mathbf{u} \geq 0. \end{cases} \\ &= (\theta_1 + W_1(\mathbf{u})) \times (1 - \text{Step}(\mathbf{u})) \\ &\qquad + (\theta_2 + W_2(\mathbf{u})) \times \text{Step}(\mathbf{u}) \end{aligned} \tag{2.78}$$

with two signal levels θ_1 and θ_2, and two noise standard deviations σ_1 and σ_2 such that $\sigma_1 > \sigma_2$. In this equation, $W_1(\mathbf{u}) \sim \text{Gauss}(0, \sigma_1^2)$ and $W_2(\mathbf{u}) \sim \text{Gauss}(0, \sigma_2^2)$ denote the Gaussian noise. The function $\text{Step}(\mathbf{u})$ denotes the horizontal step function where $\text{Step}(\mathbf{u}) = 1$ for any $\mathbf{u} \geq 0$, i.e., residing on the right, and $\text{Step}(\mathbf{u}) = 0$ for any $\mathbf{u} < 0$. For illustration, we show in Figure 2.24(a) the case where $\theta_1 = 0.8$, $\theta_2 = 0.2$, $\sigma_1 = 0.1$ and $\sigma_2 = 0.02$.



(a) Input $J(\mathbf{u})$ (b) Denoising filter $h(\mathbf{x}, \mathbf{u})$ (illustrated over a grid of points)

Figure 2.24. Thought experiment of two noisy regions in an image. To denoise this image, ideally we would want to apply to different filters with a sharp boundary at the transition.

To denoise this image, we consider the simplest approach assuming that we *knew* the partition of the two regions. Suppose that we want to denoise the left side. Since we know that the noise is stronger, we shall apply a stronger filter. As illustrated in Figure 2.25, for

this filter to be effective along the boundary, we should apply a mask *after* the filtering is done.

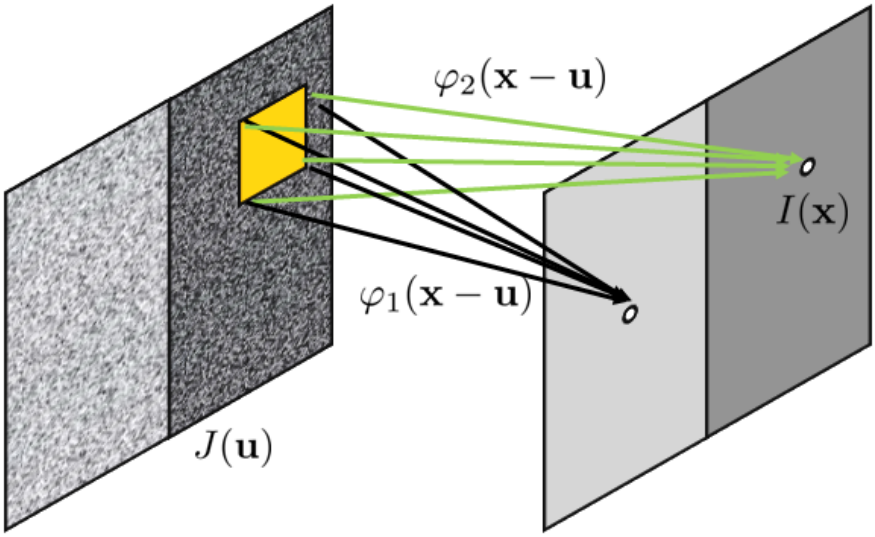


Figure 2.25. Thought experiment with two noisy half-planes. As we perform the denoising step, we would hope that the sharp boundary is preserved.

The spatially varying filter we propose here takes the form

$$h(\mathbf{x},\mathbf{u}) = \begin{cases} \frac{1}{\sqrt{2\pi s_1^2}} \exp\left\{-\frac{\|\mathbf{x}-\mathbf{u}\|^2}{2s_1^2}\right\} \stackrel{\text{def}}{=} \varphi_1(\mathbf{x}-\mathbf{u}), & \mathbf{x} < 0, \\ \frac{1}{\sqrt{2\pi s_2^2}} \exp\left\{-\frac{\|\mathbf{x}-\mathbf{u}\|^2}{2s_2^2}\right\} \stackrel{\text{def}}{=} \varphi_2(\mathbf{x}-\mathbf{u}), & \mathbf{x} \geq 0, \end{cases}$$

where we assume that $s_1 > s_2$. We are careful about the index in this equation, remarking that the conditions are applied to $\mathbf{x}$ instead of $\mathbf{u}$. We will illustrate what will happen if the conditions are applied to $\mathbf{u}$.

The gathering and the scattering equation for this example are identical to those in Equation (2.74) and Equation (2.75). Most importantly, the coefficients $a_{\mathbf{x},m}$ and $a_{\mathbf{u},m}$ are binary masks indicating whether the pixel $\mathbf{x}$ (or $\mathbf{u}$) is on the left / right hand side.

The resulting images $I^{\text{gather}}(\mathbf{x})$ and $I^{\text{scatter}}(\mathbf{x})$ follow a similar derivation as in Equation (2.76) and Equation (2.77). For the gathering equation, we recognize that

$$\begin{aligned}
I^{\text{gather}}(\mathbf{x}) &= \int_{-\infty}^{\infty} \left(a_{\mathbf{x},1}\varphi_1(\mathbf{x}-\mathbf{u}) + a_{\mathbf{x},2}\varphi_2(\mathbf{x}-\mathbf{u})\right) J(\mathbf{u})\, d\mathbf{u} \\
&= a_{\mathbf{x},1} \int_{-\infty}^{\infty} \varphi_1(\mathbf{x}-\mathbf{u}) \Big\{ \left[\theta_1 + W_1(\mathbf{u})\right] \left(1 - \text{Step}(\mathbf{u})\right) \\
&\qquad\qquad + \left[\theta_2 + W_2(\mathbf{u})\right] \text{Step}(\mathbf{u}) \Big\}\, d\mathbf{u} \\
&\quad + a_{\mathbf{x},2} \int_{-\infty}^{\infty} \varphi_2(\mathbf{x}-\mathbf{u}) \Big\{ \left[\theta_1 + W_1(\mathbf{u})\right] \left(1 - \text{Step}(\mathbf{u})\right) \\
&\qquad\qquad + \left[\theta_2 + W_2(\mathbf{u})\right] \text{Step}(\mathbf{u}) \Big\}\, d\mathbf{u} \\
&= \begin{cases}
[((\theta_1 + W_1(\mathbf{u}))(1 - \text{Step}(\mathbf{u})) \\
\qquad +(\theta_2 + W_2(\mathbf{u}))\text{Step}(\mathbf{u})) \circledast \varphi_1(\mathbf{u})](\mathbf{x}), & \mathbf{x} < 0, \\
[((\theta_1 + W_1(\mathbf{u}))(1 - \text{Step}(\mathbf{u})) \\
\qquad +(\theta_2 + W_2(\mathbf{u}))\text{Step}(\mathbf{u})) \circledast \varphi_2(\mathbf{u})](\mathbf{x}), & \mathbf{x} \geq 0,
\end{cases}
\end{aligned} \tag{2.79}$$

where the last equality holds because $a_{\mathbf{x},1}$ and $a_{\mathbf{x},2}$ are indicator functions.

For the scattering equation, we recognize that

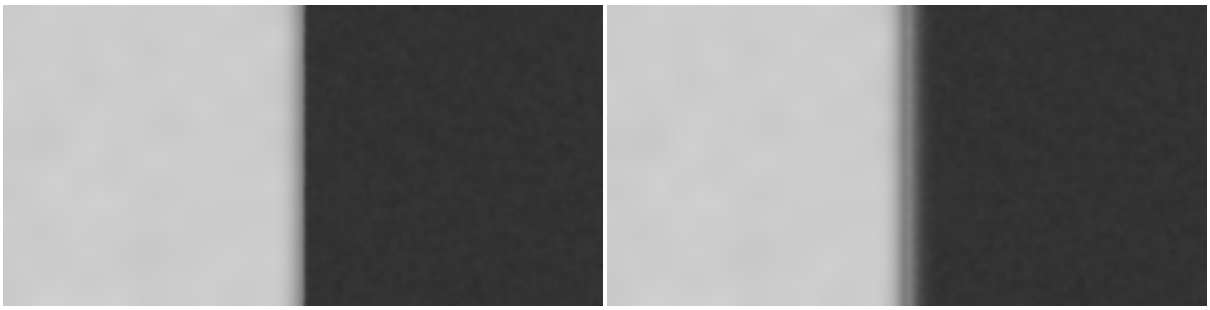
$$\begin{aligned}
I^{\text{scatter}}(\mathbf{x}) &= \int_{-\infty}^{\infty} \left(a_{\mathbf{u},1}\varphi_1(\mathbf{x}-\mathbf{u}) + a_{\mathbf{u},2}\varphi_2(\mathbf{x}-\mathbf{u})\right) J(\mathbf{u})\, d\mathbf{u} \\
&= \int_{-\infty}^{\infty} a_{\mathbf{u},1}\varphi_1(\mathbf{x}-\mathbf{u})[\theta_1 + W_1(\mathbf{u})](1 - \text{Step}(\mathbf{u}))\, d\mathbf{u} \\
&\quad + \int_{-\infty}^{\infty} a_{\mathbf{u},2}\varphi_2(\mathbf{x}-\mathbf{u})[\theta_2 + W_2(\mathbf{u})]\text{Step}(\mathbf{u})\, d\mathbf{u},
\end{aligned}$$

where we used the facts $a_{\mathbf{u},1} = (1 - \text{Step}(\mathbf{u}))$ and $a_{\mathbf{u},2} = \text{Step}(\mathbf{u})$, and hence $a_{\mathbf{u},1}\text{Step}(\mathbf{u}) = 0$ and $a_{\mathbf{u},2}(1 - \text{Step}(\mathbf{u})) = 0$. As a result, we can simplify the above equations as

$$\begin{aligned}
I^{\text{scatter}}(\mathbf{x}) &= [((\theta_1 + W_1(\mathbf{u}))(1 - \text{Step}(\mathbf{u}))) \circledast \varphi_1(\mathbf{u})](\mathbf{x}) \\
&\quad + [((\theta_2 + W_2(\mathbf{u}))\text{Step}(\mathbf{u})) \circledast \varphi_2(\mathbf{u})](\mathbf{x}).
\end{aligned} \tag{2.80}$$

So, the result is the *sum* of two convolutions with the input image $J(\mathbf{u})$.

We visualize the results in Figure 2.26. As we can see here, while both of them can offer denoising to some extent, the gathering approach handles the boundary much better because the masking is performed *after* the filtering. If the masking is performed before the filtering, then Equation (2.80) tells us that we are summing two convolutions of the same image. Therefore, the edge is blurred.



(a) $I^{\text{gather}}(\mathbf{x})$ (b) $I^{\text{scatter}}(\mathbf{x})$

Figure 2.26. Comparison between gathering and scattering for the setup in Figure 2.24. Notice that for this denoising experiment, the better method should produce a sharp transition along the boundary.

2.5.5 Spatially varying phase errors

To close this discussion on spatially varying convolution, we wish to tie it back to image formation. To do so, we will consider a particular class (of which a few exist) of spatially varing PSFs:

$$I_i(\mathbf{x}) = \left(|\mathfrak{Fourier}\{P(\boldsymbol{\xi})e^{j\phi_{\mathbf{u}}(\boldsymbol{\xi})}\}|^2 \overset{\mathbf{u}}{\circledast} I_g \right)(\mathbf{x}). \tag{2.81}$$

where we've introduced $\overset{\mathbf{u}}{\circledast}$ to represent *scattering* convolution where $\mathbf{u}$ is the coordinate over which we're allowing to parameterize the PSF. Note that (2.81) allows only the *phase* to vary spatially.

A great example of this spatially-dependent phase distortion is the case of depth-dependent camera response. In this case, $\phi_{\mathbf{u}}$ varies for every single point, with the distance of position $\mathbf{u}$ in the object having a distance $L(\mathbf{u})$. Through a proportionality constant the phase distortion can be written as

$$\phi_{\mathbf{u}}(\boldsymbol{\xi}) = \alpha|\boldsymbol{\xi}|^2(f - L(\mathbf{u})). \tag{2.82}$$

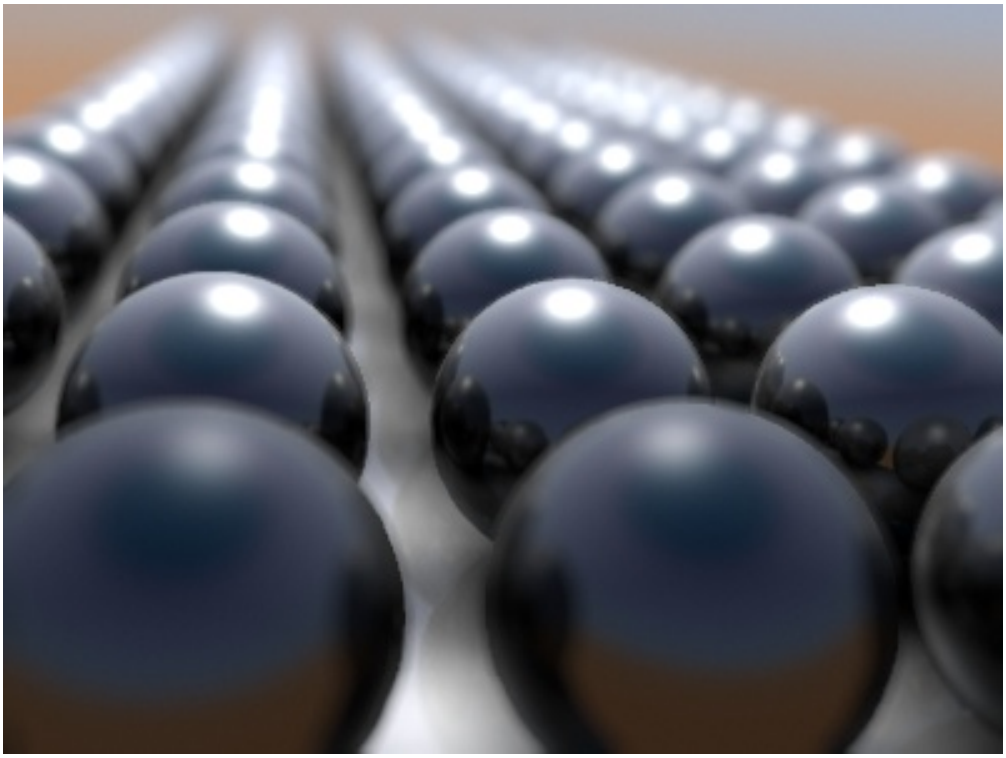


Figure 2.27. A visualization (through rendering) of defocus by depth where we can see the balls at different depths are varying in their blur. [Source]

We show an example of this depth-based defocus in Figure 2.27.

2.6 Summary

In this Chapter we have primarily discussed optics. This included ray and wave optics, with us ultimately landing on the wave model for our chosen model for the image formation process. The reasons for this have to do with turbulence – we will need the wave model to insert the turbulence into the forward model. Furthermore, we have discussed spatially varying systems, which will be of utmost importance for the case of turbulent imaging, with nearly every PSF in a wide field of view image being different.

Our model of light

Our model of light for this thesis will be the wave model of light. That being said, the focus will not be on solving partial differential equations but instead utilizing the phasor form

$$U(\boldsymbol{\xi}) = A(\boldsymbol{\xi}) \exp(j\phi(\boldsymbol{\xi})). \tag{2.83}$$

For our purposes, we will place the model for turbulence into $\phi(\boldsymbol{\xi})$ and often assume $A(\boldsymbol{\xi}) = 1$. The way we will derive the turbulent phase statistics will be through a similar trick as the thin lens approach.

The imaging model

Our imaging model for this thesis will be the following:

$$I_i(\mathbf{x}) = \left(|\mathfrak{Fourier}\{P(\boldsymbol{\xi}) e^{j\phi_{\mathbf{u}}(\boldsymbol{\xi})}\}|^2 \overset{\mathbf{u}}{\circledast} I_g \right)(\mathbf{x}). \tag{2.84}$$

where the term

$$|h_{\mathbf{u}}(\mathbf{x})|^2 = |\mathfrak{Fourier}\{P(\boldsymbol{\xi}) e^{j\phi_{\mathbf{u}}(\boldsymbol{\xi})}\}|^2 \tag{2.85}$$

is the spatially varying point spread function. Therefore, many of the tricks we have mentioned in the spatially varying section will come into play for modeling turbulence efficiently.

3. WAVE PROPAGATION THROUGH THE ATMOSPHERE

We have now presented some main results in optics and imaging, we now wish to discuss some details concerning how waves behave in the atmosphere. The approach taken in this Chapter does not represent the cutting edge of the field, but rather presents an older result which captures the effects we will focus on in this thesis. I will quantify some of these differences when we have enough background to appreciate the more modern developments and appreciate what our approach manages to explain regardless of its relative simplicity.

3.1 Random variables and processes

At the core of modeling stochastic phenomena are random variables and random processes. For example, when rolling a die, it may take on the values 1 through 6. The result of the experiment, however, cannot be considered to be deterministic but instead random. To model these types of experiments, we need probabilistic analysis. There are many excellent resources on probability, random variables, and random processes [12–14], though I will only briefly detail the key things that we will need for this thesis.

3.1.1 Random variables

In this thesis, we will only consider continuous random variables (and, perhaps, sets of them). To this end, let us define X to be a real valued random variable with **cumulative distribution function (CDF)**

$$F(t) = P\{X \leq t\}. \tag{3.1}$$

The expression $P\{X \leq t\}$ is the probability that X is less than or equal to t. The function $F(t)$ is then said to satisfy three properties: (i) monotonically increasing, (ii) be right hand continuous, (iii) have the limits $\lim_{t\to\infty} F(t) = 1$ and $\lim_{t\to-\infty} F(t) = 0$. In almost all situations of interest (and certainly the ones in this thesis) there is a corresponding **probability density function (PDF)** which is related to the CDF through

$$p(t) = \frac{\mathrm{d}F(t)}{\mathrm{d}t}. \tag{3.2}$$

For us, a commonly used distribution will be the Gaussian distribution, which is written as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \tag{3.3}$$

where

$$\mathbb{E}[X] = \mu \tag{3.4}$$

$$\mathbb{E}[(X-\mu)^2] = \sigma^2. \tag{3.5}$$

I will refer to the operation $\mathbb{E}[\cdot]$ as the **statistical expectation**, or simply just expectation defined as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xp(x)dx, \tag{3.6}$$

though one may prefer the Lebesgue-Stieltjes integral for a more formal definition. We may also extend this to the case where X is a vector, therefore, the multivariant Gaussian distribution is

$$p(x) = \frac{1}{(2\pi)^{p/2}}|R|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)^T R^{-1}(x-\mu)\right\} \tag{3.7}$$

where

$$\mathbb{E}[X] = \mu \tag{3.8}$$

$$\mathbb{E}[(X-\mu)(X-\mu)^T] = R. \tag{3.9}$$

R is called the covariance matrix which is positive definite and symmetric.

3.1.2 Random processes

A random process is a generalization of random variables where for every t in a set T, $X(t)$ is a random variable. Therefore, we refer to X as a continuous random process if each entry (for every given location t) corresponds to a random variable. In particular, we will

focus on Gaussian processes which satisfy that for every t, $X(t)$ is a Gaussian. More precisely, for any finite set of indices $t_1, \ldots, t_k$,

$$X(t_1, \ldots, t_k) = (X(t_1), \ldots, X(t_k)) \tag{3.10}$$

is a multivariate Gaussian.

For random processes, there are similar properties such as mean and autocorrelation. The autocorrelation is defined as the correlation between two points

$$\Gamma_n(\mathbf{r}_1, \mathbf{r}_2) = \mathbb{E}[X(\mathbf{r}_1)X(\mathbf{r}_2)]. \tag{3.11}$$

There are indeed requirements on the *types* of functions that are allowed to be Γ_n, which largely is the study of positive definite kernels. If we further assume homogeneity, then the following holds:

$$\Gamma_n(\mathbf{r}) = \mathbb{E}[n_1(\mathbf{r}_1)n_1(\mathbf{r}_1 - \mathbf{r})] \tag{3.12}$$

where $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. Further still we may assume isotropy,

$$\Gamma_n(r) = \mathbb{E}[n_1(\mathbf{r}_1)n_1(\mathbf{r}_1 - \mathbf{r})] \tag{3.13}$$

where $r = |\mathbf{r}_1 - \mathbf{r}_2|$. In other words, the distance of the separation between $\mathbf{r}_1$ and $\mathbf{r}_2$ is enough to describe their correlation.

3.1.3 Wiener-Khinchin theorem

When modeling physical processes, it is often that we can visualize a random field as a complicated function which is random in nature. For example, the distribution of heat in a room. This distribution lends itself to a Fourier expansion. Interestingly, it was found

that the autocorrelation function is related to the power spectral density (PSD) of a random process. Thus the pair of equations summarize the **Wiener-Khinchin theorem**

$$\Phi_X(f) = \int_{-\infty}^{\infty} \Gamma_X(\mathbf{r}) e^{j2\pi f\tau} d\tau, \tag{3.14}$$

$$\Gamma_X(\tau) = \int_{-\infty}^{\infty} \Phi_X(f) e^{-j2\pi f\tau} df. \tag{3.15}$$

We note that to arrive at this result we must assume **wide-sense stationarity**, which is equivalent to stating homogeneity (along with non spatially varying mean and that the second "moments" are finite).

3.2 Modeling turbulence in the atmosphere

Turbulence has been studied for a great deal of modern human history and even as of today does not have what many would refer to as a satisfactory answer. The situation with turbulence can be compared to the situation with knowing the equation $F = ma$. Just because we can write $F = ma$ does *not* mean that we can solve for all applications of the equation, a famous example being the three-body problem.

To the extent of my knowledge, it is generally accepted that turbulent motion is captured by the **Navier-Stokes equations** which describe fluid motion through application of Newton's equation to fluid motion (and combined with viscosity). However, just because we can write these equations does not mean that we can analyze *any* situation to predict what will happen. Turbulence is one of these phenomena that escapes suitable description by the Navier-Stokes equations. Therefore, a more common approach is to model the dynamic motion of turbulence as a stochastic process.

In 1941, in a series of papers [15, 16], Kolmogorov produced what essentially represents the foundations of turbulence as stochastic model. Statistical models for turbulence often fall into the categories of **Kolmogorov** and **non-Kolmogorov** turbulence, clearly suggesting his influence to the field. A fantastic detail of his life and contributions to turbulence and a discussion on the advancements is offered by [17].

Kolmogorov's model describes the velocity fluctuations of turbulence. A student of his by the name of Obhukov described the temperature fluctuations of a turbulent process. This is an essential step in the development of the model we will adopt because of the temperatures proportionality to the index of refraction. The statistical model for the index of refraction was then fully introduced by Tatarskii [18]. Before introducing his model and the developments from it, we first wish to provide a bit of an overview of the model for the index of refraction.

As described by Goodman [19], we break the statistical model of the index of refraction into two components:

$$n(\mathbf{r}, t, \lambda) = \underbrace{n_0(\mathbf{r}, \lambda)}_{\text{deterministic}} + \underbrace{n_1(\mathbf{r}, t)}_{\text{random fluctuations}} . \tag{3.16}$$

n_0 (neglecting the dependent variables) is the mean value or the deterministic component of the index of refraction of the atmosphere. Roughly speaking $n_0 \approx 1$. The term n_1 will contain the fluctuations, which in this thesis we will classify as the turbulent fluctuations.

3.2.1 Power spectral density of the index of refraction

Tatarskii showed that for the index of refraction, the Kolmogorov model for turbulence can be written as

$$\Phi(\mathbf{k}) = 0.033 C_n^2 |\mathbf{k}|^{-11/3}. \tag{3.17}$$

In this equation C_n^2 is known as the **structure function parameter** which can be approximately thought of as the variance of the fluctuations. The Kolmogorov model was then extended to two additional models which we will utilize in this work. The first was proposed by Tatarskii

$$\Phi(\mathbf{k}) = 0.033 C_n^2 |\mathbf{k}|^{-11/3} \exp\left(-|\mathbf{k}|^2/k_i^2\right) \tag{3.18}$$

which incorporates the inner scale through $k_i = 2\pi/\ell_0$. Furthermore, the von Kármán PSD can be written as

$$\Phi(\mathbf{k}) = \frac{0.033 C_n^2}{(|\mathbf{k}|^2 + k_0^2)^{-11/3}} \exp\left(-|\mathbf{k}|^2/k_i^2\right). \tag{3.19}$$

In Figure 3.1, we show the Kolmogorov and von Kármán PSDs plotted in the log domain.

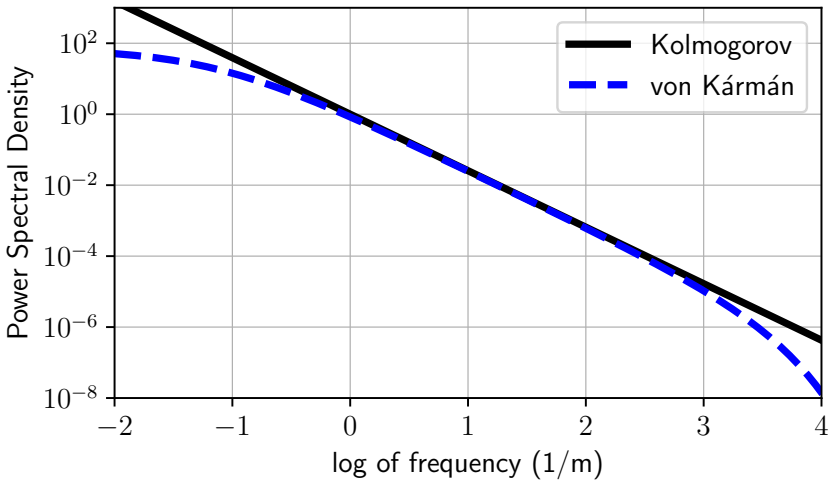


Figure 3.1. Kolmogorov and von Kármán PSDs visualized in the log space

There are more PSDs that one may choose and the way to derive them involves some fluid mechanics and other physical considerations. I have not presented them here as they do not pertain to the topics of this thesis in a direction fashion. For more details, I would refer the reader to many sources such as [20–22] for more details.

3.2.2 The structure function

The structure function is defined as [23]

$$\mathcal{D}_X(\mathbf{r}_1, \mathbf{r}_2) = \mathbb{E}\left[(X(\mathbf{r}_1) - X(\mathbf{r}_2))^2\right]. \tag{3.20}$$

Recall that a homogeneous autocorrelation function is

$$\Gamma_X(\mathbf{r}) = [X(\mathbf{r}_1)X(\mathbf{r}_1 - \mathbf{r})]. \tag{3.21}$$

Therefore, under the assumption of homogeneity,

$$\begin{aligned}
\mathcal{D}_X(\mathbf{r}_1, \mathbf{r}_2) &= \mathbb{E}\left[(X(\mathbf{r}_1) - X(\mathbf{r}_2))^2\right] && (3.22)\\
&= \mathbb{E}\left[X^2(\mathbf{r}_1) + X^2(\mathbf{r}_2) - 2X(\mathbf{r}_1)X(\mathbf{r}_2)\right] && (3.23)\\
&= 2\left(\Gamma_X(\mathbf{0}) - \Gamma_X(\mathbf{r})\right), && (3.24)
\end{aligned}$$

where we've defined $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ as before.

Recall the WK theorem which relates the PSD of a random process with its autocorrelation function via the Fourier transform. This leads us to come to understand *why* we would want to use the structure function in the first place. First, inserting the Kolmogorov PSD into the WK theorem,

$$\Gamma_n = \iiint 0.033 C_n^2 |\mathbf{k}|^{-11/3} e^{-j\mathbf{k}\cdot\mathbf{r}} d\mathbf{k} \tag{3.25}$$

we'll notice a particular difficulty – the singularity at 0. Fortunately, the structure function is defined precisely to solve this issue. We can see this by

$$\mathcal{D}_n = 2\iiint [1 - e^{-j\mathbf{k}\cdot\mathbf{r}}]\Phi_n(\mathbf{k})d\mathbf{k} \tag{3.26}$$

$$= 2\iiint [1 - \cos(\mathbf{k}\cdot\mathbf{r})]\Phi_n(\mathbf{k})d\mathbf{k}, \tag{3.27}$$

with the final step coming from the evenness of the Kolmogorov PSD. Here we can see that the PSD is attenuated near $|\mathbf{k}| = 0$. Tatarskii [18] comments that this attenuation satisfactorily controls the divergence of the Kolmogorov PSD.

This concept is even more significant when we will extend our results to a specific type of non-homogeneous and non-isotropic turbulence. Turbulence can often fluctuate in its mean value slowly over a large volume – this corresponds to the low frequency components. Due to the attenuation behavior of the PSD by the structure function, turbulence of this variety will *still* have a homogeneous and isotropic structure function. Therefore, for our problem, the structure function is absolutely essential.

If we combine the assumptions of isotropy (at least for a local region) and a constant C_n^2, the structure function for a Kolmogorov PSD becomes

$$D_n(\mathbf{r}) = C_n^2 r^{2/3}. \tag{3.28}$$

This result will be used repeatedly to develop how the structure function of the phase behaves. In short, the phase of a wave propagated through turbulence will be very much related to the statistical behavior of the index of refraction.

3.3 Waves propagated through turbulence

Truth be told, I don't often follow this approach in deriving the phase statistics. However, I do think it's important to show it in this thesis so we can appreciate a few tricks which go into solving the equations, specifically the approximations we make to reduce the complexity of the solutions. These approximations will highlight the limitations of not only this analysis

but the ones we shall take in the next Section. The reasons for such limitations, however, will be far more clear as a result of this Section.

3.3.1 The scalar inhomogeneous wave equation

Recall the standard scalar wave equation

$$\nabla^2 U(\mathbf{x}) + \frac{w^2 n^2}{c^2} U(\mathbf{x}) = 0, \tag{3.29}$$

where we've introduce $(wn/c)^2$ for a reason that will be clear in a moment. Through a series of simplifications (which are not significant to the understanding we wish to present in this thesis) the **inhomogeneous scalar wave equation** can be written as

$$\nabla^2 U(\mathbf{x}) + \frac{w^2 n(\mathbf{x})^2}{c^2} U(\mathbf{x}) = 0. \tag{3.30}$$

It should be clear that the difference is due to the positional dependence of n in the inhomogeneous wave equation.

3.3.2 The Rytov and Born approximations

We now make an interesting and extremely important simplification. First, we will utilize the model for the index of refraction, $n = n_0 + n_1$, where we are notationally the dependence on wavelength, position, etc. for some simplicity. Inserting this into the inhomogeneous wave equation, we get

$$\nabla^2 U(\mathbf{x}) + \frac{w^2 (n_0 + n_1)^2}{c^2} U(\mathbf{x}) = 0. \tag{3.31}$$

Because of the size of the fluctuations (think of how small we mentioned C_n^2 is!) we will discard n_1^2 where applicable.

We now introduce what is known as the **Born approximation**. The Born approximation follows in spirit with the splitting of n into n_0 and n_1. We will also split the phasor U into U_0 and U_1. As expected, U_0 is associated with n_0 while U_1 is associated with n_1. Since n_0 is

deterministic, then U_0 represents a wave which propagates through a completely deterministic medium, while U_1 captures the randomness. This produces two equations:

$$\nabla^2 U_0(\mathbf{x}) + \frac{w^2 n_0^2}{c^2} U_0(\mathbf{x}) = 0 \tag{3.32}$$

$$\nabla^2 U_1(\mathbf{x}) + \frac{w^2 n_0^2}{c^2} U_1(\mathbf{x}) = \frac{-2w^2 n_0 n_1}{c^2} U_0(\mathbf{x}). \tag{3.33}$$

From here, we may solve the two respective equations and obtain expressions for their amplitude and phase. This was done primarily by Tatarskii [18]. However, his approach utilized a slightly different approximation.

The Born approximation has some issues such as ignoring the case of multiple scattering [18, 23]. To remedy this, and to find a more general solution, Tatarskii used the Rytov approximation

$$\psi(\mathbf{x}) = \ln U(\mathbf{x}). \tag{3.34}$$

Here we follow the approach of Goodman [23] and do it at the end, rather than at the beginning as in the case of Tatarskii [18]. If we assume that

$$U = \exp(\psi_0(\mathbf{x}) + \psi_1(\mathbf{x})) \tag{3.35}$$

$$U_0 = \exp(\psi_0(\mathbf{x})) \tag{3.36}$$

then we may write the following

$$U/U_0 = e^{\psi_1}. \tag{3.37}$$

Note that $U_1 \neq e^{\psi_1}$, therefore we will have to do some extra work to get the fluctuations to come out of the analysis. Specifically, we will have to do the following:

$$\psi_1 = \ln\left(1 + \frac{U_1}{U_0}\right) \approx \frac{U_1}{U_0} \tag{3.38}$$

where we've utilized the fact that the magnitude of the mean value phasor dominates that of the fluctuations, i.e., $|U_1| \ll |U_0|$.

From this analysis, it can be shown that the phasors are equal to

$$U = A\exp(jS) \tag{3.39}$$

$$U_0 = A_0\exp(jS_0) \tag{3.40}$$

where A and S are the amplitude and phase of the *actual* wave while A_0 and S_0 are the same of the non-turbulent solution. With this, we may write

$$\psi_1 = \psi - \psi_0 = \ln\frac{A}{A_0} + j(S - S_0) \tag{3.41}$$

and subsequently define

$$\chi = \ln\frac{A}{A_0}, S_\delta = S - S_0. \tag{3.42}$$

The following result then constitutes the conclusion of this analysis,

$$\psi_1 = \chi + jS_\delta, \tag{3.43}$$

where the exact expressions can be found in Tatarskii or Goodman [18, 23].

It is worthwhile to mention that the statistical behavior when using the Born and Rytov approximations are different. Since we typically accept the Rytov approximation to be more physically valid, we indeed will assert that the Rytov-based solution is preferred. This tends to be justified by the empirical evidence [23]. To conclude, I would mention that the statistics for the phase are Gaussian, while the amplitude is log-Gaussian.

3.4 Structure Function of the Phase

The previous Section presented the way in which Tatarskii [18] derived the statistics for the phase and amplitude. However, since his book is a bit hard to find I would suggest Goodman's *Statistical Optics* [23] for the details (unless you want to buy Tatarskii's book!). At this point, we have two options: (i) start at the result of last Section and derive various expressions (ii) re-do the previous results in a far easier way that will give us some results

almost for free. It has been found that the *same* results can be obtained through the simpler analysis, therefore, we will go with option (ii).

This derivation is a two step process, we will first derive the statistics for a wave propagating through a *single* layer of turbulence, then we will extend it to *multiple* layers. We will use the same trick as the thin lens equation. This is a dramatic simplification over the previous Section's approach and will make things far easier for us. With a single layer, we will assume homogeneity and isotropy; extending it to multiple layers will allow us to describe a (slow-moving) inhomogeneous turbulence profile.

3.4.1 A Single Turbulence Layer

Assume a turbulent medium with thickness L to be homogeneous and isotropic Kolmogorov turbulence. A position in 3D space $\mathbf{r}$ we decompose into $\mathbf{r} = [\boldsymbol{\xi}, z]$ where $\boldsymbol{\xi} = [x, y]$. We write the phase as the accumulation of the fluctuations,

$$\phi(\boldsymbol{\xi}, L) = k \int_0^L n_1(\boldsymbol{\xi}, z)\, dz. \tag{3.44}$$

Note that we are not integrating over n, but rather n_1. We don't care about the mean value n_0 as this will delay each phase component equally, we instead wish to measure their difference which is produced by the fluctuations n_1. Assuming a thin-lens-type relationship, we may write Accordingly, the input-output relationship is given by

$$U'(\boldsymbol{\xi}) = t_\phi(\boldsymbol{\xi}), \tag{3.45}$$

$t_\phi(\boldsymbol{\xi}) = e^{j\phi(\boldsymbol{\xi})}$. In general, the distortions will have phase *and* amplitude, but we shall primarily focus on the phase.

As we are interested in the wave's phase structure function, we begin with consideration of the autocorrelation function:

$$\Gamma_{U'}(\boldsymbol{\xi}, \boldsymbol{\xi}') = \mathbb{E}[t_\phi(\boldsymbol{\xi}) t_\phi^*(\boldsymbol{\xi}')] \tag{3.46}$$

$$= \mathbb{E}\left[\exp\left\{jk\left(\int_0^L [n_1(\boldsymbol{\xi}, z) - n_1(\boldsymbol{\xi}', z)]\, dz\right)\right\}\right], \tag{3.47}$$

Here we assume (along with Tatarskii [18]) that n_1 is a zero-mean Gaussian random process. Through application of the moment generating function definition, we can show that

$$\Gamma_{t_\phi}(\boldsymbol{\xi}, \boldsymbol{\xi}') = \exp\left\{-\underbrace{\frac{1}{2}k^2\mathbb{E}\left[\left(\int_0^L [n_1(\boldsymbol{\xi}, z) - n_1(\boldsymbol{\xi}', z)]\, dz\right)^2\right]}_{\mathcal{D}_\phi(\boldsymbol{\xi},\boldsymbol{\xi}') \stackrel{\text{def}}{=} \mathbb{E}[(\phi(\boldsymbol{\xi}) - \phi(\boldsymbol{\xi}'))^2]}\right\}.$$

This gives us an interesting result: the structure function of the index of refraction appears in the exponential function! This term represents the structure function of the accumulation of the phase distortion and is referred to as the **phase structure function**.

Definition 3.4.1 (**Phase Structure Function**). *The structure function of the phase is defined as*

$$\mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}') \stackrel{def}{=} k^2\mathbb{E}\left[\left(\int_0^L [n_1(\boldsymbol{\xi}, z) - n_1(\boldsymbol{\xi}', z)]\, dz\right)^2\right], \tag{3.48}$$

with n_1 as a Gaussian random process.

This serves as a useful result, however, there is a fair amount of simplification possible. As expertly done by Fried (and with some great exposition by Roggemann and Welsh) we know that this may be simplified to

$$\mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}') = -k^2 L C_n^2 \int_{-L}^{L} \left[z^{2/3} - (|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2 + z^2)^{1/3}\right]\left(1 - \frac{|z|}{L}\right) dz$$

$$= 2.91 k^2 L C_n^2 |\boldsymbol{\xi} - \boldsymbol{\xi}'|^{5/3}. \tag{3.49}$$

Now we have a nice, compact expression that tells us the structure function of the phase in closed form for the case of a single turbulence layer.

3.4.2 Multiple Phase Screens

Now suppose that the structure parameter C_n^2 changes considerably along the propagation path. This may happen in astronomical viewing or over ground-to-ground imaging with different surfaces on the ground (road, grass, etc.). For this type of distribution, inhomogeneities may exist in the distribution. The approach by [24] that will result in the same solutions as more complex methods, leading us to introduce the concept of the **layered model for turbulence** to model the atmosphere, which requires us to introduce an assumption:

Assumption. Turbulent layers are statistically independent of one another.

We first define our resultant phase as

$$\begin{aligned}\phi(\boldsymbol{\xi}) &= k\int_0^L n_1(\boldsymbol{\xi},z)\,dz,\\ &= k\sum_{i=1}^{M-1}\int_{L_{i-1}}^{L_i} n_1(\boldsymbol{\xi},z)\,dz.\end{aligned} \tag{3.50}$$

As previously, we are interested in the structure function of this phase realization. As a result of our assumption of phase screen independence, we may write

$$\begin{aligned}\mathcal{D}_\phi(\boldsymbol{\xi},\boldsymbol{\xi}') &= k^2\mathbb{E}\left[\left(\sum_{i=1}^{M-1}\int_{L_{i-1}}^{L_i}\left[n_1(\boldsymbol{\xi},z)-n_1(\boldsymbol{\xi}',z)\right]dz\right)^2\right],\\ &\overset{(a)}{=} k^2\mathbb{E}\left[\sum_{i=1}^{M-1}\left(\int_{L_{i-1}}^{L_i}\left[n_1(\boldsymbol{\xi},z)-n_1(\boldsymbol{\xi}',z)\right]dz\right)^2\right],\\ &= k^2 2.91k^2|\boldsymbol{\xi}-\boldsymbol{\xi}'|^{5/3}\sum_{i=1}^{M-1}C_n^2[i]\Delta L_i,\end{aligned} \tag{3.51}$$

where $\Delta L_i = L_{i+1} - L_i$ and (a) utilizes the independence of the atmospheric slices, thus dropping the cross terms in the squared summation. We have simply used our previous result to describe a summation of independent turbulence layers – thus we now have the structure function for the phase which arises in the presence of inhomogeneities.

Two extensions of Equation (3.51) to continuous cases will be of importance to us which we will now present without proof. The first is the structure function of the phase for plane waves,

$$\mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}') = 2.91k^2|\boldsymbol{\xi} - \boldsymbol{\xi}'|^{5/3} \int_0^L C_n^2(z)dz. \tag{3.52}$$

A second extension of our result will be to spherical waves,

$$\mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}') = 2.91k^2|\boldsymbol{\xi} - \boldsymbol{\xi}'|^{5/3} \int_0^L C_n^2(z) \left(\frac{L-z}{L}\right)^{5/3} dz. \tag{3.53}$$

At this point I'd wish to remind the reader that $z = L$ occurs in the object plane. This convention absolutely *will* affect the form of (3.53), therefore when reading papers on turbulence it is important to know the convention adopted by the author.

3.4.3 Wave Structure Function

Earlier we saw Tatarskii's approach to deriving the phase and amplitude statistics, though we did not put as much work into simplifying them the way that we have here. I'd like to give a bit of insight into what happens when you use his results "the full way". Doing this leads to the **wave structure function** which combines both phase and amplitude distortions,

Definition 3.4.2 (**Wave Structure Function**). *The wave structure function is defined as [25]*

$$\mathcal{D}(\boldsymbol{\xi}, \boldsymbol{\xi}') = \mathcal{D}_l(\boldsymbol{\xi}, \boldsymbol{\xi}') + \mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}'), \tag{3.54}$$

*where $\mathcal{D}_l(\boldsymbol{\xi}, \boldsymbol{\xi}')$ and $\mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}')$ are the structure functions of the **log-amplitude** and phase, respectively.*

The log-amplitude captures the logarithm of the variation in the amplitude of the incident wave. A paper by Fried proves a useful description which is a bit more mathematically direct [25]

$$l(\boldsymbol{\xi}) = \ln\big(A(\boldsymbol{\xi})/\overline{A}\big) \tag{3.55}$$

where $A(\boldsymbol{\xi})$ is the amplitude and $\overline{A}$ is the root mean square (RMS) of A. Tatarskii does indeed discuss this term, but I cite Fried's paper here because of his additional comments can be useful in understanding it (see the paper for details).

In this thesis I am only concerned with the phase structure function and not the wave structure function. You may be wondering if that hurts us? Are we missing a wide range of cases because of this assumption? Fried [26] provides another paper in which he presents the two following approximations:

$$\mathcal{D}_{\phi}(\boldsymbol{\xi}, \boldsymbol{\xi}') \approx \mathcal{D}(\boldsymbol{\xi}, \boldsymbol{\xi}') \qquad D \gg (L\lambda)^{1/2}, \tag{3.56}$$

$$\mathcal{D}_{\phi}(\boldsymbol{\xi}, \boldsymbol{\xi}') \approx \frac{1}{2}\mathcal{D}(\boldsymbol{\xi}, \boldsymbol{\xi}') \qquad D \ll (L\lambda)^{1/2}. \tag{3.57}$$

The first case (3.56) is referred to as the near field whereas (3.57) is the far field. This can give us some insight as to how to interpret the different approximations. My argument is that we are interested in reconstruction – in the far field we will likely have a severely degraded image and be ill-posed for reconstruction. I would like to include difficutl cases, but not impossible ones. Therefore, the phase structure function is suitable (or at least useful as a starting place for now).

We do wish to talk about one final topic before moving on to applications. In all that has been discussed we have effectively used the Rytov approximation. The Rytov approximation is said to be valid for **weak fluctuations** (that is – the amplitude is not *too* distorted). This amplitude fluctuation is sometimes referred to as **scintillation**. For further details, we would refer the reader to Goodman [19] as a starting point, though Tatarskii discusses these directly in his manuscript [18] along with Ishimaru [27].

Is there an approach which captures **strong fluctuations**? Yes, and interestingly, Tatarskii had to do with *this* development as well roughly 5-10 years after his original manuscript. This approach utilizes the **path integral formulation** (alternatively known as a Feynman path integral) [28] from quantum mechanics to solve the inhomogeneous wave equation. These methods were applied to the case of turbulence in works from Tatarskii and collaborators [29–32] along with others such as Dashen [33]. To my knowledge, this framework represents the most general approach solving the inhomogeneous wave equation. We can capture high

order moments of the field using this approach and account for multiple scattering. Therefore, this method *can* describe the regime in which **strong fluctuations** exist. This approach, however, will not be a focus of this thesis.

3.5 Important Applications of the Model

The previous results are the foundations of the field. We now wish to produce a few simple measures which we can use a bit more simply which capture a large chunk of the equations previously written. One such measure will be the **Fried parameter** r_0. This parameter pops up in probably every paper that I have written (with one exception – I think) in the form of D/r_0. So what is this ratio all about?

3.5.1 Fried Parameter and Isoplanatic Angle

The atmospheric coherence diameter r_0 (or what I refer to as the **Fried parameter**) can be thought of the virtual aperture imposed by the atmosphere. However, the atmosphere is stochastic in nature. Therefore, it is an *average* virtual aperture.

Definition 3.5.1 (**Fried Parameter**). *Consider a turbulent medium with a structure constant of the index of refraction C_n^2, and propagation distance L. For a plane wave incident upon turbulence, the Fried parameter is defined as [34]*

$$r_0 = 0.185 \left[\frac{4\pi^2}{k^2 \int_0^L C_n^2(z) dz} \right]^{3/5}, \tag{3.58}$$

and for a spherical wave:

$$r_0 = 0.185 \left[\frac{4\pi^2}{k^2 \int_0^L \left(\frac{L-z}{L}\right) C_n^2(z) dz} \right]^{3/5} \tag{3.59}$$

Some books may write the planar wave of the Fried parameter as r_0 and leave the spherical wave to be $r_{0,sw}$ or something similar. Others will use slightly different notation or no notation at all! I will actually follow this case of no notation at all as it is truly the most common.

Therefore, when interpreting the equations you must keep in mind whether or not we are working with spherical or planar waves (of course, the fact they share notation often is for the reason that equations can usually be written in the same way – planar or spherical).

The Fried parameter captures the integrated effects of the atmosphere in a concise way. As mentioned previously, this ratio of aperture size D to r_0 often comes up. If $D/r_0 > 1$, then the atmosphere is the limiting factor. If instead $D/r_0 < 1$, then the camera is the limiting factor. We present some visualizations of this turbulence ratio in Figure 3.2.

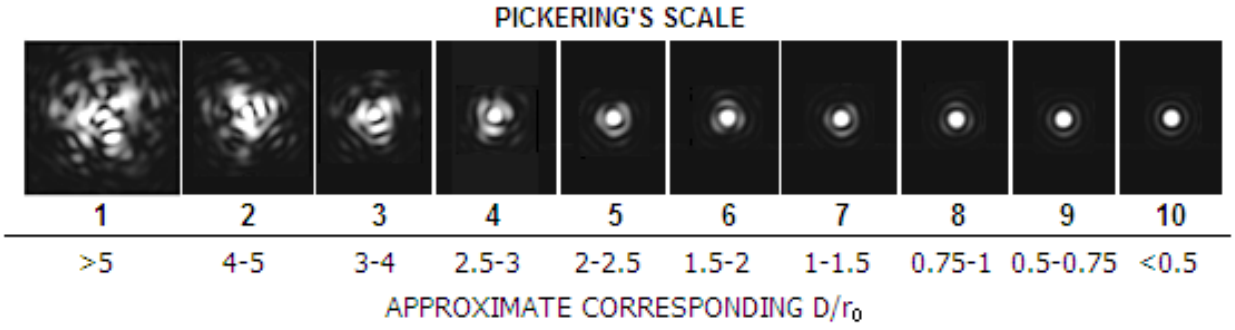


Figure 3.2. Visualizing the impact of different values of D/r_0. [Source]

The Fried parameter has another useful application. It is very carefully defined to make the structure function easier to write! The structure function of Equation (3.51) may be written with the Fried parameter as follows:

Definition 3.5.2 (**Kolmogorov Structure Function of the Phase**). *The structure function of the phase propagated through Kolmogorov turbulence is*

$$\mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}') = 6.88 \left(\frac{|\boldsymbol{\xi} - \boldsymbol{\xi}'|}{r_0} \right)^{5/3}, \tag{3.60}$$

where r_0 is the Fried parameter.

We would note that this equation is the same for both spherical and planar waves because of the varying definition of the Fried parameter.

In addition to capturing the observed strength of turbulence, we can also have some sense as to the similarity of effects. One way of quantifying these effects is the through the

isoplanatic angle. The definition comes from the adaptive optics community where it is often useful to calibrate a system using a nearby star. Therefore, how big of a range around our desired point can we look around to find an ideal guide star? The isoplanatic angle describes this through the following definition [24]:

Definition 3.5.3 (**Isoplanatic Angle**). *The isoplanatic angle is*

$$\theta_0 = 58.1 \times 10^{-3} \lambda^{6/5} \left[\frac{1}{\int_0^L z^{5/3} C_n^2(z) dz} \right]^{3/5}, \tag{3.61}$$

which denotes an angle by which adaptive optics will perform well.

Note that the definition is slightly vague. We can make it more precise, but it would invoke a bit too much detail than we are willing to say. I would suggest the interested reader to [35] for details on this.

3.5.2 Instantaneous OTF

If we recall the results of the previous Chapter, we now that a spatially invariant imaging system obeys the equation:

$$I_i(\mathbf{x}) = |h(\mathbf{x})|^2 \circledast I_g(\mathbf{x}). \tag{3.62}$$

In the Fourier domain with $\widetilde{I}_i(\mathbf{f}) = \mathfrak{Fourier}[I_i(\mathbf{x})]$, it follows that

$$\widetilde{I}_i(\mathbf{f}) = \mathcal{H}(\mathbf{f}) \widetilde{I}_g(\mathbf{f}). \tag{3.63}$$

Thus, the resolution and quality of the observed image are determined by $\mathcal{H}(\mathbf{f})$.

With turbulence present in the scene, there is now a *new* factor. First, let's use the concept of the complex pupil [1] to write the OTF of the entire system

$$\begin{aligned}\mathcal{H}(\mathbf{f}) &= H(\mathbf{f}) \circledast H^*(\mathbf{f}) \\ &= \left(P(\lambda z\mathbf{f})e^{-j\phi(\lambda z\mathbf{f})}\right) \circledast \left(P(\lambda z\mathbf{f})e^{j\phi(\lambda z\mathbf{f})}\right),\end{aligned}$$

We refer to this OTF as the **instantaneous OTF**. Note that this is random because of its dependence on the random function ϕ. Recall that the PSF is related to the OTF by

$$\underbrace{|h(\mathbf{x})|^2}_{\text{PSF}} = \mathfrak{Fourier}^{-1}\{\mathcal{H}(\mathbf{f})\}. \tag{3.64}$$

Since $\phi(\boldsymbol{\xi})$ is random, $|h(\mathbf{x})|^2$ is also random. Figure 3.3 shows a few snapshots of the PSFs generated from Kolmogorov statistics. These PSFs come from the distribution which will distort our images.

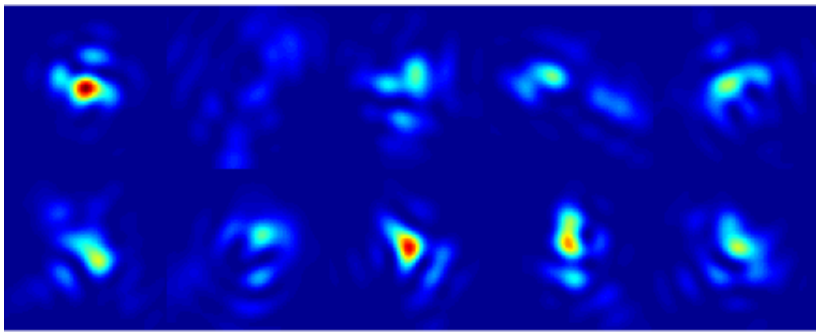


Figure 3.3. Instantaneous point spread functions generated from propagation through Kolmogorov phase statistics.

The phase function can be broke up into effectively two components:

- **Tilt**: The random pixel displacement by the plane of best fit of ϕ is said to be the tilt. This will cause objects to shift their position from their ideal location and cause geometric distortions.

- **Blur**: We refer to the blur effects as high-order effects in the phase and contain aberrations such as spherical, trefoil, and coma. These will make neighboring pixels mix and cause sharp edges to become blurry.

It is common to separate the phase into its two components which we summarize as follows:

$$\underbrace{\phi(\boldsymbol{\xi})}_{\text{overall distortion}} = \underbrace{\varphi(\boldsymbol{\xi})}_{\text{aberration}} + \underbrace{\boldsymbol{\alpha}^T\boldsymbol{\xi}}_{\text{tilt}}, \tag{3.65}$$

where $\boldsymbol{\alpha}$ is a vector defining the best-fitted first-order plane to the phase function. In what follows, we shall discuss the **long exposure (LE)** OTF and **short exposure (SE)** OTF. The LE OTF will keep the term $\boldsymbol{\alpha}^{\boldsymbol{\xi}}$ while the SE OTF will discard it. Thus, the LE OTF will include the tilt of the system while the SE OTF will focus only on the blur.

3.5.3 Long and short exposure OTFs

The long exposure function can be seen as the OTF that will arise from leaving the camera exposure open infinitely long. This motivates to take the expectation as follows:

$$\begin{aligned}
\mathbb{E}_\phi[\mathcal{H}(\mathbf{f})] &= \mathbb{E}_\phi\left[\left(W(\lambda z\mathbf{f})e^{-j\phi(\lambda z\mathbf{f})}\right) \circledast \left(W(\lambda z\mathbf{f})e^{j\phi(\lambda z\mathbf{f})}\right)\right] \\
&= \mathbb{E}_\phi\left[\int W(\boldsymbol{\xi})e^{-j\phi(\boldsymbol{\xi})} \cdot W(\boldsymbol{\xi}-\lambda z\mathbf{f})e^{j\phi(\boldsymbol{\xi}-\lambda z\mathbf{f})}d\boldsymbol{\xi}\right] \\
&= \int W(\boldsymbol{\xi})W(\boldsymbol{\xi}-\lambda z\mathbf{f})\mathbb{E}_\phi[e^{-j\phi(\boldsymbol{\xi})}e^{j\phi(\boldsymbol{\xi}-\lambda z\mathbf{f})}]d\boldsymbol{\xi},
\end{aligned} \tag{3.66}$$

where the subscript $\mathbb{E}_\phi[\cdot]$ emphasizes that the expectation is taken with respect to the random phase ϕ.

We can again use the trick of the moment generating function to produce the result

$$\mathbb{E}_\phi[e^{-j\phi(\boldsymbol{\xi})}e^{j\phi(\boldsymbol{\xi}-\lambda z\mathbf{f})}] = \exp\left\{-\frac{1}{2}6.88\left(\frac{\lambda z|\mathbf{f}|}{r_0}\right)\right\}, \tag{3.67}$$

where we have used the homogeneous property of the phase structure function. This means that we can split the total OTF into two components,

$$\mathbb{E}_\phi[\mathcal{H}(\mathbf{f})] = \underbrace{\int W(\boldsymbol{\xi})W(\boldsymbol{\xi} - \lambda z\mathbf{f})d\boldsymbol{\xi}}_{\mathcal{H}_{\text{diff}}(\mathbf{f})} \times \underbrace{\exp\left\{-\frac{1}{2}6.88\left(\frac{\lambda z|\mathbf{f}|}{r_0}\right)^{5/3}\right\}}_{\mathcal{H}_{\text{LE}}(\mathbf{f})},$$

which is a product of $\mathcal{H}_{\text{diff}}(\mathbf{f})$ and $\mathcal{H}_{\text{LE}}(\mathbf{f})$:

Theorem 3.5.1 (**Decomposition of Average OTF**). *The average OTF of a system degraded by turbulence can be written as*

$$\mathbb{E}_\phi[\mathcal{H}(\mathbf{f})] = \mathcal{H}_{diff}(\mathbf{f}) \times \mathcal{H}_{LE}(\mathbf{f}), \tag{3.68}$$

where $\mathcal{H}_{diff}(\mathbf{f})$ is the diffraction-limited OTF and $\mathcal{H}_{LE}(\mathbf{f})$ is the LE OTF.

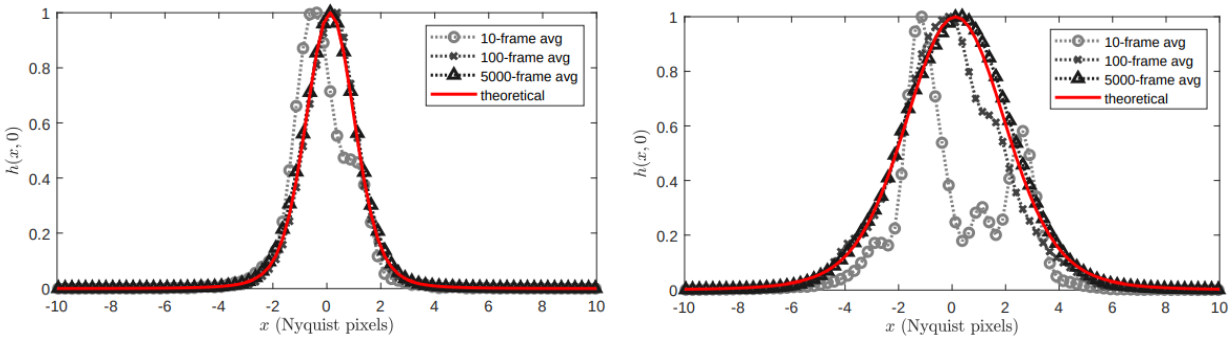
The atmospheric term $\mathcal{H}_{\text{LE}}(\mathbf{f})$ is known as the long exposure OTF, which is what we would observe if we turn on the shutter of the camera for a prolonged period of time.

Definition 3.5.4 (**Long Exposure OTF**). *The long exposure optical transfer function $\mathcal{H}_{LE}(\mathbf{f})$ is defined as*

$$\mathcal{H}_{LE}(\mathbf{f}) = \exp\left\{-3.44\left(\frac{\lambda z|\mathbf{f}|}{r_0}\right)^{5/3}\right\}, \tag{3.69}$$

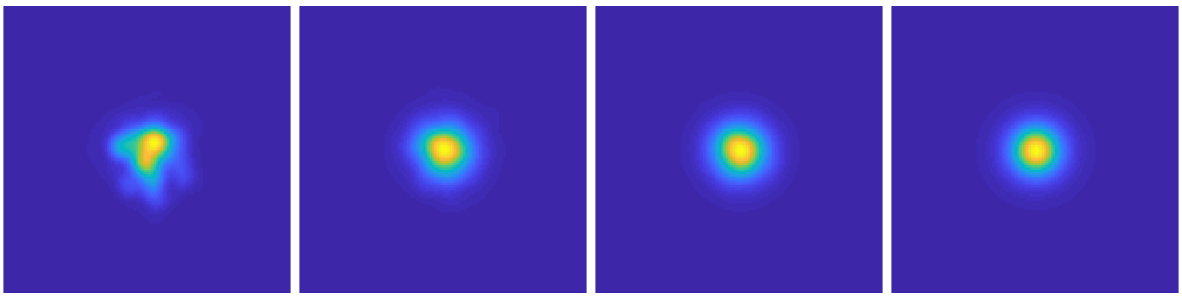
where z is the path length, and r_0 is the Fried parameter.

The LE OTF therefore captures the effects by just the turbulence and when multiplied by the diffraction OTF quantifies the average pass band of the system (in terms of frequency). We present Figure 3.4 in which we draw samples which converge to the expected long exposure PSFs (which are the inverse FFT of the LE OTF). Figure 3.5 shows a similar visual in a different way. Here we can see the LE PSFs as an image, which we note again converges to the expected value with more PSFs.



$C_n^2 = 2.5 \times 10^{-16}\mathrm{m}^{-2/3}$ $C_n^2 = 1 \times 10^{-15}\mathrm{m}^{-2/3}$

Figure 3.4. The theoretical long-exposure PSF compared with the empirical averaged instantaneous PSFs for low and high turbulence. In each figure we show groups of averages for 10, 100, and 5000 PSFs. The more PSFs, the closer it approaches the average.



(a) 50-avg (b) 500-avg (c) 5000-avg (d) Theoretical

Figure 3.5. Instantaneous PSFs for $C_n^2 = 1 \times 10^{-15}\mathrm{m}^{-2/3}$. (a) 50-frame average. (b) 500-frame average. (c) 5000-frame average. (d) Theoretical LE PSF.

We now wish to introduce the SE OTF which removes the tilt from the phase. This leads us to define

$$\varphi(\boldsymbol{\xi}) \stackrel{\text{def}}{=} \phi(\boldsymbol{\xi}) - \boldsymbol{\alpha}^T\boldsymbol{\xi}, \tag{3.70}$$

where $\boldsymbol{\alpha}^T\boldsymbol{\xi}$ is the best linear fit to $\phi(\boldsymbol{\xi})$. Through a very similar analysis, though now with φ in place of ϕ, we can find the expression of the SE OTF. Removing the tilt will allow us to focus on the blur of the system which is useful for adaptive optics systems which may have tilt-compensation systems.

We define t_φ in accordance with the previous approach

$$t_\varphi(\boldsymbol{\xi}) = \exp\{j\varphi(\boldsymbol{\xi})\} = \exp\left\{j(\phi(\boldsymbol{\xi}) - \boldsymbol{\alpha}^T\boldsymbol{\xi})\right\}. \tag{3.71}$$

The autocorrelation of the tilt-correct wavefront is

$$\begin{aligned}\Gamma_{t_\varphi}(\boldsymbol{\xi},\boldsymbol{\xi}') &= \mathbb{E}\left[t_\varphi(\boldsymbol{\xi})t_\varphi^*(\boldsymbol{\xi}')\right] \\ &= \mathbb{E}\left[\exp\left\{j(\phi(\boldsymbol{\xi}) - \phi(\boldsymbol{\xi}') - \boldsymbol{\alpha}^T(\boldsymbol{\xi}-\boldsymbol{\xi}'))\right\}\right].\end{aligned} \tag{3.72}$$

Using the same trick of the moment generating function of a zero-mean Gaussian, we can show that

$$\begin{aligned}\Gamma_{t_\varphi}(\boldsymbol{\xi},\boldsymbol{\xi}') = \exp\Bigg\{ &-\frac{1}{2}\mathbb{E}\left[(\phi(\boldsymbol{\xi}) - \phi(\boldsymbol{\xi}'))^2\right] \\ &+ \mathbb{E}\left[(\phi(\boldsymbol{\xi}) - \phi(\boldsymbol{\xi}'))\boldsymbol{\alpha}^T(\boldsymbol{\xi}-\boldsymbol{\xi}')\right] - \frac{1}{2}\mathbb{E}\left[(\boldsymbol{\alpha}^T(\boldsymbol{\xi}-\boldsymbol{\xi}'))^2\right]\Bigg\}.\end{aligned} \tag{3.73}$$

At this point, a great deal of approximations or assumptions must take place. I would suggest the reader to [35] for a textbook which covers these details or the papers by Fried [26] and Heidbreder [36] for more details. Through some simplifications which are beyond the intent of this thesis, we find that

$$\begin{aligned}\mathbb{E}_\varphi[\mathcal{H}(\mathbf{f})] = &\underbrace{\int W(\boldsymbol{\xi})W(\boldsymbol{\xi}-\lambda z\mathbf{f})d\boldsymbol{\xi}}_{\mathcal{H}_{\text{diff}}(\mathbf{f})} \\ &\times \underbrace{\exp\left\{-3.44\left(\frac{\lambda z|\mathbf{f}|}{r_0}\right)^{5/3}\left(1-\left(\frac{\lambda z|\mathbf{f}|}{D}\right)^{1/3}\right)\right\}}_{\mathcal{H}_{\text{SE}}(\mathbf{f})},\end{aligned}$$

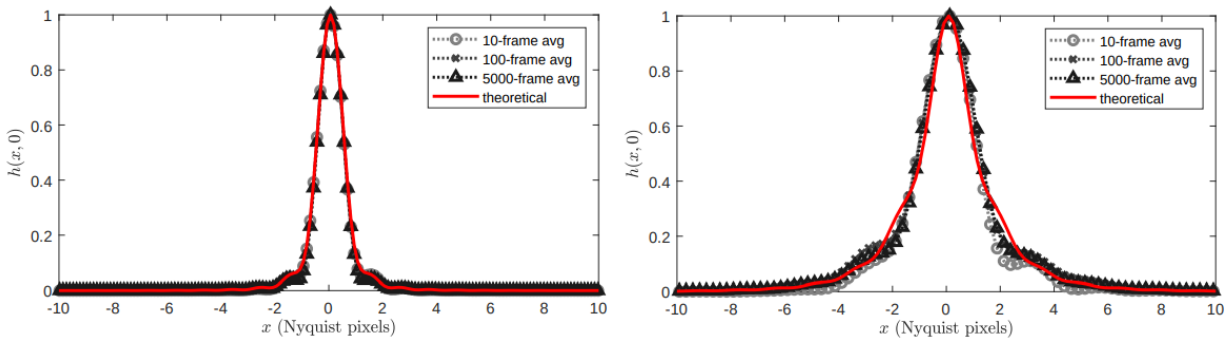
leading us to present the following definition:

Definition 3.5.5 (**Short Exposure OTF**). *The SE OTF $\mathcal{H}_{SE}(\mathbf{f})$ of a system degraded by turbulence is defined as*

$$\mathcal{H}_{SE}(\mathbf{f}) = \exp\left\{ -3.44 \left(\frac{\lambda d|\mathbf{f}|}{r_0}\right)^{5/3} \left(1 - \left(\frac{\lambda d|\mathbf{f}|}{D}\right)^{1/3}\right)\right\}, \tag{3.74}$$

where D is the aperture diameter, d is the focal length, and r_0 is the Fried parameter.

With this, we can show a similar figure as before now with the SE PSFs. Figure 3.6 presents the 10, 100, and 5000 averages of slices of PSFs which eventually converge to the SE PSF. We note that the turbulence strength here is the same as previously. The SE PSF is notably thinner than the LE PSF, this is because the tilting has been removed. The shifting adds a great deal of extra blur, therefore we can see that a system which compensates for such shifting will perform far better.



$C_n^2 = 2.5 \times 10^{-16}\mathrm{m}^{-2/3}$ $C_n^2 = 1 \times 10^{-15}\mathrm{m}^{-2/3}$

Figure 3.6. The theoretical short-exposure PSF compared with the empirical averaged instantaneous PSFs for $C_n^2 = 2.5 \times 10^{-16}\mathrm{m}^{-2/3}$ and $C_n^2 = 1 \times 10^{-15}\mathrm{m}^{-2/3}$. Shown in each sub-figure are the 10-frame average, 100-frame average and 5000-frame average PSFs, all in black. The red curve is the theoretically predicted result. The x-axis denotes the Nyquist pixel, where one pixel corresponds to $\lambda z/D$ with D being the aperture diameter.

3.6 Summary

In this Chapter we have discussed the model for atmospheric turbulence and its effects on waves. This included the more rigorous approach by Tatarskii and the simpler approach by

Roggemann and Welsh which I find particularly useful. We introduced the structure function and why we need it for turbulence modeling as well as some applications of the model.

Two key equations

With a brief survey of the major equations which dictate various processes in imaging through turbulence I would like to highlight two key equations which will be essential to this thesis. One is that the phase is the accumulation of the fluctuations in the atmosphere's index of refrction We first define our resultant phase as

$$\phi(\boldsymbol{\xi}) = k \int_0^L n_1(\boldsymbol{\xi}, z) dz \tag{3.75}$$

This equation will be our way of conceptualizing how the phase is affected by the atmosphere. The next is the structure function of the phase, where I will almost almost prefer to write in terms of the Fried parameter,

$$\mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}') = 6.88 \left(\frac{|\boldsymbol{\xi} - \boldsymbol{\xi}'|}{r_0} \right)^{5/3}. \tag{3.76}$$

These two equations will capture the majority of the turbulent statistics relevant to this thesis. However, in my opinion, without presenting the surrounding material the insight of these equations would be lost.

4. ZERNIKE-BASED SIMULATION OF IMAGING THROUGH TURBULENCE

In Chapters 2 and 3 the optical and atmospheric turbulence models were discussed. For simplicity, and not to cloud your vision by what others have done, I have neglected tying the two together in any real way. This Chapter, in part, will work towards unifying the two discussed concepts thus far. However, this Chapter has a bigger goal as well: to present the way which we think of and simulate imaging through atmospheric turbulence.

The main contributions of this dissertation is primarily contained within this Chapter in the form of our simulator which we refer to as **Zernike-based simulation**. It is fair to say that people have used *some* of the techniques we outline here but only in pieces and arguably no where near to the degree of speed and accuracy we are talking about here.

At the core of this method are theoretical developments, engineering tricks, and learning-based solutions which give us a differentiable, fast, and accurate atmospheric imaging simulator. We can simulate an incredibly wide variety of scenarios that capture the weak fluctuations discussed in the previous chapter. To me, this is the achievement. There will always be a debate as to "Can it be extended to strong fluctuations?". For now, the answer is no. However, we have personally observed a great deal of success from using this method for training reconstruction networks and have no plan on abandoning it because of its lack of strong fluctuations.

4.1 Preliminaries

I now wish to present a few ideas which will help us to combine the previously introduced concepts as well as give a preview as to the developments in this Chapter. This will be like priming you to think of atmospheric imaging in a specific way. To this end, I have intentionally left off how people simulate these effects in the optics community. Although I do not comment on them in this thesis, I will mention some great sources where you may learn about them.

4.1.1 The model of imaging through atmospheric turbulence

The imaging model that is relevant to this thesis is that the geometrically predicted image I_g is related to the observed image I_i through

$$I_i(\mathbf{x}) = (|h|^2 \circledast I_g)(\mathbf{x}), \tag{4.1}$$

where

$$|h(\mathbf{x})|^2 = |\mathfrak{Fourier}\{P(\boldsymbol{\xi})e^{j\phi(\boldsymbol{\xi})}\}|^2. \tag{4.2}$$

Now, given the last Chapter, you likely have a sense as to why the turbulence is inserted into ϕ. That being said, there is still a particular difficulty. The effects of the turbulent forward process is spatially varying. Therefore, we must inject this spatially varying nature into the model.

Because of the effort in the first Chapter, it is easy to extend this to the case of spatially varying phase functions as follows

$$I_i(\mathbf{x}) = (|h_{\mathbf{u}}|^2 \overset{\mathbf{u}}{\circledast} I_g)(\mathbf{x}) \tag{4.3}$$

with

$$|h_{\mathbf{u}}(\mathbf{x})|^2 = |\mathfrak{Fourier}\{P(\boldsymbol{\xi})e^{j\phi_{\mathbf{u}}(\boldsymbol{\xi})}\}|^2. \tag{4.4}$$

This gives us a few new things to think about when considering imaging through atmospheric turbulence. Namely, each PSF will come from (in general) a *different* random phase component. As you may expect, neighboring points will accumulate similar distortions, and therefore there will be a **spatial correlation** present in the images.

However, as an electrical engineer, when I look at the image formation equation through turbulence, I'd still like something to be a bit simpler. Specifically, I'm partial to a basis representation. Because of the effort of Chapter 1, I can safely say that a spatially varying convolution can be approximated as a sum of invariant ones (with weights injected into the process). This leads me to preview what is to be described in the rest of this Chapter.

4.1.2 A preview of what is to come

Through working on the simulator, we have eventually come to the conclusion that the spatially varying phase and PSFs should be represented by the following pair of equations:

$$\phi_{\mathbf{u}}(R\boldsymbol{\rho}) = \sum_i a_{\mathbf{u},i} Z_i(\boldsymbol{\rho}), \tag{4.5}$$

$$|h_{\mathbf{u}}(\mathbf{x})|^2 = \sum_j \beta_{\mathbf{u},j} \varphi_i(\mathbf{x}). \tag{4.6}$$

The consideration of the phase effects decomposed by the functions $Z_i(\boldsymbol{\rho})$ will introduce an interesting problem regarding $\mathbb{E}[a_{\mathbf{u},i} a_{\mathbf{u}',j}]$. Furthermore, although the spatially varying blur approximation is not necessarily novel, we will consider a mapping from $\mathbf{a}_{\mathbf{u}} = [a_{\mathbf{u},1}, a_{\mathbf{u},2}, \ldots, a_{\mathbf{u},M}]^T$ to $\boldsymbol{\beta}_{\mathbf{u}} = [\beta_{\mathbf{u},1}, \beta_{\mathbf{u},2}, \ldots, \beta_{\mathbf{u},N}]^T$, $f(\mathbf{a}_{\mathbf{u}}) = \boldsymbol{\beta}_{\mathbf{u}}$.

The main insights of our approach are solving these two problems: the phase coefficient correlations and the mapping from the phase coefficients to the PSF coefficients $f(\cdot)$. The rest will be tricks or speed-ups here and there that we employ to achieve the speed we are looking for.

4.1.3 Plan of development

The simulator has evolved over the course of my Ph.D, from nothing to where it is today. As one may expect, plenty of erasing, looping back, discarding, and so on have taken place with various approaches and concepts during its development. I will not describe *every* approach or thought that we have considered for the simulator, but rather a collection of them which make sense. I now outline the plan of this Chapter:

1. **The Zernike polynomials.** At the core of representing the atmospheric effects in $\phi(\boldsymbol{\xi})$ through a basis representation is the introduction of the **Zernike polynomials**. We will introduce these and discuss their application to atmospheric turbulence.

2. **Our first attempt at phase correlations.** First we will detail our first approach (i.e., from Chimitt and Chan [37]) towards describing the correlations of $\mathbb{E}[a_{\mathbf{u},i} a_{\mathbf{u}',j}]$. This

approach was useful, and was in the simulator from the beginning, only recently being replaced. This approach provides an approximate way of describing the correlations.

3. **The current phase correlations.** Returning to the goal of the first paper [37], we revisit the problem $\mathbb{E}[a_{\mathbf{u},i}a_{\mathbf{u}',j}]$ and find, what we consider to be, an *exact* solution [38]. Interestingly, the impact to the downstream steps in the simulator is negligible.

4. **The Zernike space.** The **Zernike space** is how we conceptualize modeling images observed through atmospheric turbulence. It is a concept which can be extended to any set of phase distortions.

5. **The mapping between basis coefficients and spatially varying convolution.** Finally, we consider the mapping between the two sets of coefficients $\mathbf{a}$ and $\boldsymbol{\beta}$ and performing the spatially varying convolution as a sum of invariant convolution with a weighted input image.

We then tie together the ideas and present an overview of the simulation approach. We additionally comment on validating the simulator.

4.1.4 Related work

I now wish to only briefly discuss the related work in simulating imaging through turbulence. The first and primary topic is **split-step simulation**, which is excellently described in the book by Schmidt [22]. I will then make a few comments on papers which do not fall within the boundaries of split-step but still are from the optics community.

Split-step propagation

Split-step is a process that virtually mimicks nature and our analysis of the previous Chapter by breaking up the propagation into two steps: **phase screens** and free space propagation. The phase screen is where the Kolmogorov turbulence statistics are injected. I would recommend the books by Schmidt [22] and Voelz [39] for learning exactly how one should simulate these things, which largely involves Fourier optics and **angular spectrum propagation**.

This idea of splitting the path up, as we did in theory and computationally for split-step, can also be done in real life. A great example of this is in a work by Pellizzari et al. [40] involving coherent imaging. Other ways of doing this include tunable phase modulators [41–43], heat sources [44, 45], or plastic screens [46]. Split-step itself can also be extended beyond turbulence, such as in works which consider sound propagation [47].

Sampling the phase screens properly is a big challenge for this method, with many methods attmpted to do so [48–55]. The prevailing approach appears to me **subharmonic generation**, which generates the phase screens in two major stages – capturing the high frequency information then the low frequency information.

I personally came to learn of split-step through the following papers: [35, 56, 57] and I would regard them as a good starting place along with [58], the most recent version of the approach by Hardie et al. However, it is undeniable that the book by Schimdt [22] is likely *the* place to learn split-step.

Beyond split-step

There are other methods which consider alternative ways of generating the effects, of which our approach is one such method. However, beyond our approach are the methods such as the brightness function method [59–61] which effectively blend ray tracing with simulating through phase screens. To me, it is very split-step in nature, but they derive some useful expressions for speeding up the evaluation of split-step which is notoriously slow.

In addition to this, a set of empirically-motivated methods by Repasi and Weiss [62, 63], Leonard et al. [64], or Potvin et al. [65] which use a blend of analytic and empirical properties (the empirical aspects partially based on the NATO RTG-40 dataset [66, 67]) to simulate PSFs and the subsequent images directly. The NATO dataset used various measurement devices, with one example being **scintillometers** to measure the effective turbulence level by providing an estimate for C_n^2. These simulation methodologies have been revisited more recently by Miller et al. [68, 69]. We would point the interested reader to Figure 1 of Miller et al. [68] for a clear and concise visual description of their processing pipeline. In essence,

these methods skip modeling in the phase domain and do image effects directly. Because of the careful collection of the NATO dataset, these methods can use it to great benefit.

4.2 The Zernike polynomials and atmospheric turbulence

The effects of the atmosphere as captured by $\phi(\boldsymbol{\xi})$ can be represented through a basis representation. Let's think of the requirements that such a basis would require. For one, orthogonality *over a unit disc* would be desirable – this is because we are concerned with optics, which typically has a circular (or, near circular) pupil. They should capture the atmospheric effects without needing 100's of polynomials – it should be an *efficient* basis. As is it turns out, someone by the name of Robert Noll has given us a great starting point with his study of modeling atmospheric turbulence with the Zernike polynomials [70].

4.2.1 The Zernike polynomials

The Zernike polynomials are named after Fritz Zernike, the Nobel laureate and inventor of phase-contrast microscopy. Interestingly, there are many different orderings or **indexing** schemes of the Zernike polynomials as well as somewhat varying definitions. In this thesis, we follow the conventions set by Noll [70].

The Zernike polynomials are a product of two functions: an angular (azimuthal) component and a radial component. We will denote its angular degree as m and its radial degree as n. Noll proposed an indexing scheme which maps the pair $(m, n) \to j$. With $\boldsymbol{\rho} = [\rho, \theta]$ defined over the unit circle, then

$$
\begin{aligned}
Z_{\text{even } j}(\boldsymbol{\rho}) &= \sqrt{2(n+1)} R_n^m(\rho) \cos(m\theta), && m \neq 0, \\
Z_{\text{odd } j}(\boldsymbol{\rho}) &= \sqrt{2(n+1)} R_n^m(\rho) \sin(m\theta), && m \neq 0, \\
Z_j(\boldsymbol{\rho}) &= \sqrt{n+1} R_n^m(\rho) \sin(m\theta), && m = 0,
\end{aligned}
$$

where $R_n^m(\rho)$ is

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s![(n+m)/2-s]![(n-m)/2-s]!} \rho^{n-2s}.$$

While the form of these equations indeed looks formidable, we will see that upon expanding them that they are rather mundane. The first Zernike polynomial Z_0 is known as the constant or piston term. We will almost always ignore this term as for incoherent imaging a DC shift in phase doesn't affect the PSF. Z_2 and Z_3 represent the x-**tilt** and y-**tilt**, respectively. Some higher order terms represent things such as astigmatism, coma, and so on. In Figure 4.1, we present a visual of the Zernike polynomials along with a table of the various forms they take.

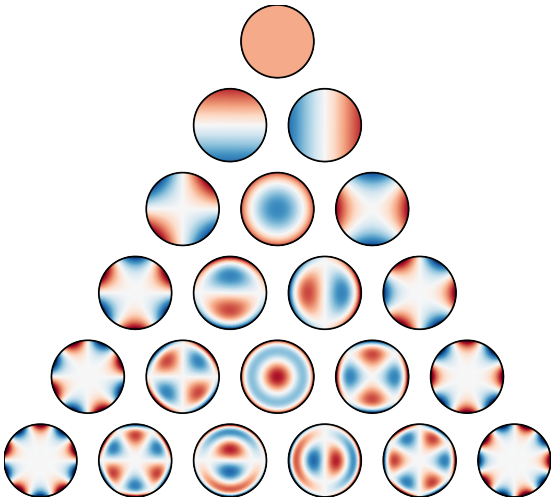


Table 1. Zernike polynomial up to and including the 4th radial order.

Order (n)	Freq. (m)	$Z_n^m(\rho, \theta)$	Meaning
0	0	1	Constant term, or Piston
1	-1	$2\rho\sin(\theta)$	x-tilt, Distortion
1	1	$2\rho\cos(\theta)$	y-tilt, Distortion
2	-2	$\sqrt{6}\rho^3\sin(2\theta)$	Astigmatism (±45°)
2	0	$\sqrt{3}(2\rho^3-1)$	Field curvature, Defocus
2	2	$\sqrt{6}\rho^3\cos(2\theta)$	Astigmatism (0° or 90°)
3	-3	$\sqrt{8}\rho^3\sin(3\theta)$	y-trefoil
3	-1	$\sqrt{8}(3\rho^3-2\rho)\sin(\theta)$	y-coma
3	1	$\sqrt{8}(3\rho^3-2\rho)\cos(\theta)$	x-coma
3	3	$\sqrt{8}\rho^3\cos(3\theta)$	x-trefoil
4	-4	$\sqrt{10}\rho^4\sin(4\theta)$	y-quadrafoil
4	-2	$\sqrt{10}(4\rho^4-3\rho^2)\sin(4\theta)$	y-secondary astigmatism
4	0	$\sqrt{5}(6\rho^4-6\rho^2+1)\sin(4\theta)$	Spherical, Defocus
4	2	$\sqrt{10}(4\rho^4-3\rho^2)\cos(4\theta)$	x-secondary astigmatism
4	4	$\sqrt{10}\rho^4\cos(4\theta)$	x-quadrafoil
...	...	...	...

Figure 4.1. The shape of the Zernike basis functions, and their corresponding equations. [Source].

4.2.2 Noll and the Zernike polynomials

The Zernike polynomials are defined over the unit circle, therefore, the inner product must also be defined on the unit circle. If we are given a phase function $\phi_{\mathbf{u}}(\boldsymbol{\rho})$, we can define the jth Zernike coefficient at location $\mathbf{u}$ as

$$\begin{aligned} a_{\mathbf{u},j} &= \langle \phi_{\mathbf{u}}(R\boldsymbol{\rho}), Z_j(\boldsymbol{\rho}) \rangle_P \\ &\stackrel{\text{def}}{=} \int_0^{2\pi} \int_0^1 P(\boldsymbol{\rho}) \phi_{\mathbf{u}}(R\boldsymbol{\rho}) Z_j(\boldsymbol{\rho}) d\rho d\theta, \end{aligned} \tag{4.7}$$

where $\langle \cdot, \cdot \rangle_P$ denotes the inner product using $P(\boldsymbol{\rho})$ as a weight. We would remind the reader that since Tatarskii argued that $\phi_{\mathbf{u}}$ is a zero-mean Gaussian [18], then the coefficients are *also* zero-mean Gaussians because they are obtained through linear projections.

Our accepted model of the phase from the previous Chapter was

$$\phi(\boldsymbol{\xi}) = k \int_0^L n_1(\boldsymbol{\xi}, z) dz, \tag{4.8}$$

in words, proportional to the fluctuations in the atmosphere. We further know that the structure function captured the two-point statistics of the phase function,

$$\mathcal{D}_\phi(|\boldsymbol{\xi} - \boldsymbol{\xi}'|) = 6.88 \left(\frac{|\boldsymbol{\xi} - \boldsymbol{\xi}'|}{r_0} \right)^{5/3}. \tag{4.9}$$

In addition to this, we have stated that $\phi(\boldsymbol{\xi})$ is a Gaussian random process.

This starting point was the same starting point considered by Robert Noll. Noll set out to describe the phase through a basis decomposition by the Zernike polynomials [70]

$$\phi(\boldsymbol{\xi}) = \sum_{j=1}^{N} a_j Z_j(\boldsymbol{\xi}). \tag{4.10}$$

Note that I have removed the dependence on source location $\mathbf{u}$. This is because Noll did *not* consider this possibility. Therefore, Noll's work is a great starting place, but we will need more. But first, let's appreciate the result which he derived.

Considering the vector of Zernike coefficients $\mathbf{a_u} = [a_{\mathbf{u},1}, a_{\mathbf{u},2}, \ldots, a_{\mathbf{u},M}]^T$ to be the same of the basis representation, the problem of Noll was to find the autocovariance $\mathbb{E}[\mathbf{a}\mathbf{a}^T]$. Through a very careful application of the theory in which the expression

$$\mathbb{E}[a_i\, a_j] = \frac{1}{\pi^2} \iint P(\boldsymbol{\rho}) P(\boldsymbol{\rho}') Z_i(\boldsymbol{\rho}) Z_j(\boldsymbol{\rho}') \times \mathbb{E}[\phi(R\boldsymbol{\rho})\phi(R\boldsymbol{\rho}')]\, d\boldsymbol{\rho} d\boldsymbol{\rho}'. \tag{4.11}$$

was considered, the answer was found to be

$$\begin{aligned}\mathbb{E}[a_i\ a_j] = 2.2698(-1)^{(n_i+n_j-2m)/2} &\sqrt{(n_i+1)(n_j+1)}\mathbb{I}_{m_i m_j} \\ &\times \frac{(D/r_0)^{5/3}}{\Gamma[(n_i+n_j+23/3)/2]} \\ &\times \frac{\Gamma[(n_i+n_j-5/3)/2]}{\Gamma[(n_j-n_i+17/3)/2]\Gamma[(n_i-n_j+17/3)/2]},\end{aligned} \tag{4.12}$$

if $i-j =$ even and $\mathbb{E}[a_i\ a_j] = 0$ if $i-j =$ odd. $\mathbb{I}_{m_i m_j}$ is the indicator function with $\mathbb{I}_{m_i m_j} = 1$ if $m_i = m_j$ and $\mathbb{I}_{m_i m_j} = 0$ if $m_i \neq m_j$. The normalized correlation matrix, which we will refer to as the **Noll matrix**, is shown in Figure 4.2, where we note most of the other entries are *sparse*.

Noll's result is excellent and is one of my personal favorite papers on atmospheric turbulence. However, we will need more than Noll has offered us. If we add back in our indexing of source position, Noll effectively found the expression for $\mathbb{E}[a_{\mathbf{u},i}a_{\mathbf{u},j}]$ whereas we will instead need $\mathbb{E}[a_{\mathbf{u},i}a_{\mathbf{u}',j}]$ – the position will *also* need to vary! The difference here is subtle but hopefully it can be appreciated by the following reasoning: Noll's result will only tell us how one point in an image behaves through a basis decomposition, it will not describe an image's distortions. Therefore, to simulate an image, we will need to extend to multiple points (in fact, we will need the correlation for every pair of pixels in an image).

4.3 A first attempt at Zernike correlations

In 2020, we attempted to first solve the problem of $\mathbb{E}[a_{\mathbf{u},i}a_{\mathbf{u}',j}]$ with an approximation [37, 71]. I originally had not intended to detail this in my thesis, however, I decided to instead

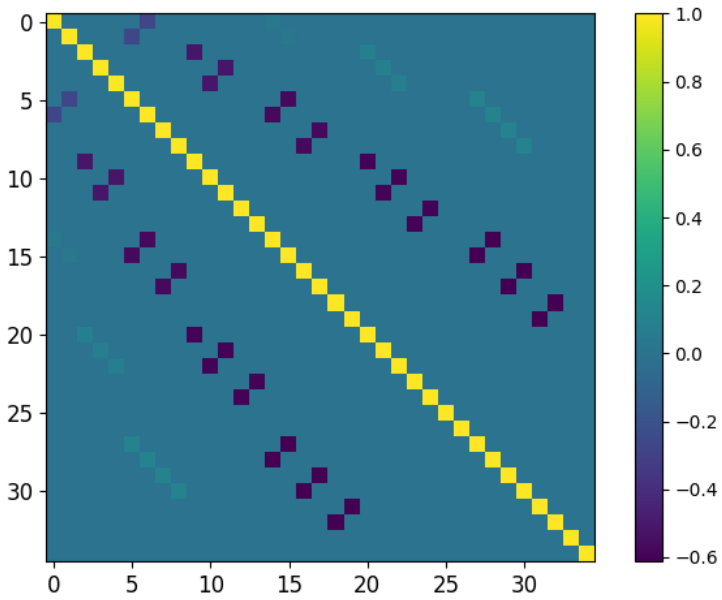


Figure 4.2. Inter-modal correlation of the Zernike coefficients $\mathbb{E}[a_i a_j]/\sqrt{\mathbb{E}[a_i^2]\mathbb{E}[a_j^2]}$ for a fixed spatial location $\mathbf{u}$.

put it in this thesis because of the interesting implication which arises from our later analysis (which generalized this one). As we will see, we were actually pretty right the first time! But later, we did it in a more elegant and general fashion.

4.3.1 Problem setup

To begin, let us recall the structure function of the phase

$$\mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}') = \mathbb{E}[(\phi(\boldsymbol{\xi}) - \phi(\boldsymbol{\xi}')]. \tag{4.13}$$

This tells us the structure function of *one* phase realization and how its points correlate in their distortions by the atmosphere. This can be written as

$$\mathcal{D}_\phi(\boldsymbol{\xi}, \boldsymbol{\xi}') = 6.88(|\boldsymbol{\xi} - \boldsymbol{\xi}'|/r_0)^{5/3} \tag{4.14}$$

where r_0 is the Fried parameter defined as

$$r_0 = 0.185 \left[\frac{4\pi^2}{k^2 \int_0^L dz C_n^2(z)} \right]^{3/5}. \tag{4.15}$$

We consider a very similar decomposition of the phase by the Zernike polynomials as Noll, i.e.,

$$\phi_{\mathbf{u}}(R\boldsymbol{\rho}) = \sum_i a_{\mathbf{u},i} Z_i(\boldsymbol{\rho}), \tag{4.16}$$

however, note that we have introduced a dependence of the basis coefficients on position $\mathbf{u}$. This led us to find an expression for the structure function which included geometry.

$$\mathcal{D}(R\boldsymbol{\rho} - R\boldsymbol{\rho}', \mathbf{u} - \mathbf{u}') = 2.91k^2 \int_0^L dz C_n^2(z) \left| R(\boldsymbol{\rho} - \boldsymbol{\rho}') + \left(\frac{z}{L - z} \right) (\mathbf{u} - \mathbf{u}') \right|^{5/3}. \tag{4.17}$$

We then make an approximation of this function with the assumption that C_n^2 is constant. Ignoring C_n^2 for the time being (which is aided by our assumption of it being a constant) we may write:

$$I = \int_0^L \left| (R\boldsymbol{\rho} - R\boldsymbol{\rho}') \left(1 - \frac{z}{L} \right) + z(\boldsymbol{\theta} - \boldsymbol{\theta}') \right|^{5/3} dz. \tag{4.18}$$

Let $u = \frac{z}{L}$, and define

$$f(u) = \left| (R\boldsymbol{\rho} - R\boldsymbol{\rho}') (1 - u) + Lu(\boldsymbol{\theta} - \boldsymbol{\theta}') \right|^{5/3}. \tag{4.19}$$

The first order Taylor approximation of $f(u)$ at $u = 1/2$ is

$$f(u) \cong f\left(\frac{1}{2} \right) + f'\left(\frac{1}{2} \right) \left(u - \frac{1}{2} \right). \tag{4.20}$$

Integrating $f(u)$ from 0 to 1, and recognizing that $\int_0^1 (u - 1/2) du = 0$, we have

$$I = L \int_0^1 f(u) du = L \left| \frac{(R\boldsymbol{\rho} - R\boldsymbol{\rho}')}{2} + \frac{L(\boldsymbol{\theta} - \boldsymbol{\theta}')}{2} \right|^{5/3}. \tag{4.21}$$

Substituting into Equation (4.17) yields Equation (4.22).

$$D_\phi(R\boldsymbol{\rho} - R\boldsymbol{\rho}', \boldsymbol{\theta} - \boldsymbol{\theta}') \cong 2.91k^2C_n^2L\left|\frac{(R\boldsymbol{\rho} - R\boldsymbol{\rho}')}{2} + \frac{L(\boldsymbol{\theta} - \boldsymbol{\theta}')}{2}\right|^{5/3}. \tag{4.22}$$

Note here that we have introduced $L\boldsymbol{\theta} = \mathbf{u}$.

4.3.2 Utilizing multi-aperture correlations

The previous result gives us a starting point to incorporate some new statistics. Previously, the structure function was "trapped" in an integration which we have approximated our way out of. This leads us to present a result by Takato et al. [72]

$$\mathbb{E}[a_{\mathbf{x},i}(\mathbf{0})a_{\mathbf{x},j}(\mathbf{s})] = \frac{1}{\pi^2}\iint d\boldsymbol{\rho}d\boldsymbol{\rho}'P(\rho)P(\rho')Z_i(\boldsymbol{\rho})Z_j(\boldsymbol{\rho}')\mathbb{E}[\phi_{\mathbf{x}}(R\boldsymbol{\rho})\phi_{\mathbf{x}}(R\boldsymbol{\rho}' + D\mathbf{s})], \tag{4.23}$$

which I will note comes from the problem of having two cameras looking at the same star through turbulence and asking of their correlation. In other words – a different but vaguely similar problem! Factoring out $2^{-5/3}$, we can write

$$D\mathbf{s} = L(\theta - \theta') \tag{4.24}$$

where D is the aperture diameter, the structure function in Equation (4.22) can be written as

$$D_\phi(R\boldsymbol{\rho} - R\boldsymbol{\rho}' + D\boldsymbol{\xi}) = \frac{2.91k^2C_n^2L}{2^{5/3}}\left|R\boldsymbol{\rho} - R\boldsymbol{\rho}' + D\boldsymbol{\xi}\right|^{5/3}. \tag{4.25}$$

Without loss of generality we may assume $\boldsymbol{\theta} = \mathbf{0}$. Thus, the correlation becomes

$$\mathbb{E}[a_j^* a_{j'}(D\boldsymbol{\xi})] = -\int\int W(\boldsymbol{\rho})W(\boldsymbol{\rho}')Z_j(\boldsymbol{\rho})Z_{j'}(\boldsymbol{\rho}') \times \mathcal{D}_\phi(R\boldsymbol{\rho} - R\boldsymbol{\rho}' + D\boldsymbol{\xi})\ d\boldsymbol{\rho}d\boldsymbol{\rho}', \tag{4.26}$$

which is the same expression as [72]. In other words, our approximation allows us to build a connection between our problem and the problem of Takato and Yamaguchi.

I will not present much more regarding this result other than the final result, which we found to be

$$\mathbb{E}[a_{\mathbf{u},i}a_{\mathbf{u}',j}] = \mathscr{A}_{i,j}\left(\frac{1}{2}\right)^{5/3} LC_n^2 f_{ij}\left(\frac{(\mathbf{u}-\mathbf{u}')}{D}, k_0\right). \tag{4.27}$$

This result I have re-written to accomodate the next discussion properly, but it is the same as found in [37] just with a different appears. The expression for f_{ij} comes from Takato et al., therefore, we have used their results for a different problem for our problem.

4.4 The Zernike space

This trick of using multi-aperture correlations to describe correlations in the Zernike coefficients was and is still used by our simulator. However, in 2023, we made a change to this correlation expression. This removed the Taylor approximation and allowed to be *exact* (to the degree that the Rytov theory can be considered exact!). Along the way, we also came to understand what we now refer to as the Zernike space. Therefore, this Section will be dedicated to how *we* understanding imaging through atmospheric turbulence and how we've decided to define the forward model.

4.4.1 Overview

The previous discussion highlighted that we had tried the problem before, but there were some issues. In fact, in that paper there were a few more that we could not solve. At this point, we have solved a majority of those either exactly or by approximation. This leads me to present the **Zernike space**, the main theoretical concept that this thesis has produced.

The Zernike space is the collection of $\{\mathbf{a}_\mathbf{u}\}$ for all points in an image. It is not even contained to describing atmospheric turbulence, but it can describe camera aberrations by position-dependent defocus or other aberrations such as chromatic and so on. If the aberrations that effect a camera are captured by the phase and can be decomposed appropriately by the Zernike polynomials, then the Zernike space is sufficient in describing the distortions. In our

case, because of the randomness of the coefficients, the Zernike space is a **random vector field**.

4.4.2 Derivation of the correlations

There are a great many similarities of our 2023 approach [73] to the 2020 approach [37]. However, we do not utilize a Taylor approximation *nor* do we assume a constant C_n^2. This, of course, opens up our simulator to a far wider variety of situations and only boosts the accuracy greater.

We begin with stating the problem at hand as

$$\mathbb{E}[a_{\mathbf{x},i}a_{\mathbf{x}',j}] = \frac{1}{\pi^2}\iint d\boldsymbol{\rho}d\boldsymbol{\rho}' P(\rho)P(\rho')Z_i(\boldsymbol{\rho})Z_j(\boldsymbol{\rho}')\mathbb{E}[\phi_{\mathbf{x}}(R\boldsymbol{\rho})\phi_{\mathbf{x}'}(R\boldsymbol{\rho}')]. \tag{4.28}$$

This differs from (4.23) by consideration of two point sources located at points $\mathbf{x}, \mathbf{x}'$. As before, this may be written using the phase structure function as

$$\mathbb{E}[a_{\mathbf{x},i}a_{\mathbf{x}',j}] = \frac{-1}{2\pi^2}\iint d\boldsymbol{\rho}d\boldsymbol{\rho}' P(\rho)P(\rho')Z_i(\boldsymbol{\rho})Z_j(\boldsymbol{\rho}')\mathcal{D}(R\boldsymbol{\rho} - R\boldsymbol{\rho}', \mathbf{x} - \mathbf{x}'). \tag{4.29}$$

The formulation of our problem is then most in accordance with Fried's approach, though notably we have changed from the case of the tilt vector to any arbitrary Zernike polynomial.

This section begins with presenting the main theoretical result of this work: the correlation of the Zernike coefficients for the case of a continuous turbulence profile. This will consist of the usage of [72]'s work to solve (4.28). We will then move to discretize the main continuous results, causing us to define a C_n^2-slice, which will draw some analogy to a phase screen from split-step. This second perspective will be the one taken for the sake of numerical evaluation. We finish with how the approximation of Chimitt and Chan [37] fits into this framework and discuss its limitations.

4.4.3 Continuous Case: Varying C_n^2

Before carrying out the main derivation, we must mention that our approach will be applied to the case of spherical waves. However, the results of Takato and Yamaguchi are

developed for planar waves. This creates a potential conflict, applying results from planar waves to spherical ones. To remedy this, we facilitate the general nature of their results which has no restriction on the C_n^2 profile. As shown in Takato et al. [72], the turbulence profile is a non-closed expression in their final result. We may therefore choose to write the turbulence profile to satisfy our requirements via

$$C_n^2(z) = \left(\frac{L-z}{L}\right)^{5/3} \tilde{C}_n^2(z)[u(z) - u(z-L)], \tag{4.30}$$

where $u(z-a)$ is the unit step function which is unity for $z > a$ and 0 otherwise and $\tilde{C}_n^2(z)$ is the original, unmodified turbulence profile in the planar case (of course, the two are the same). Substitution of this turbulence profile into Takato's main result [72] then satisfies our requirement for spherical waves. Our later comparisons will be done for spherical wave statistics, which lends credence to this approach.

With our approach to using Takato and Yamaguchi's results for spherical waves in mind, we first begin with rewriting (4.17) as

$$\mathcal{D}(R\boldsymbol{\rho} - R\boldsymbol{\rho}', \mathbf{x} - \mathbf{x}') = 2.91k^2 \int_0^L dz \left(\frac{L-z}{L}\right)^{5/3} C_n^2(z) \left| R(\boldsymbol{\rho} - \boldsymbol{\rho}') + \left(\frac{z}{L-z}\right)(\mathbf{x} - \mathbf{x}') \right|^{5/3}. \tag{4.31}$$

This allows us to write (4.29) as

$$\begin{aligned} \mathbb{E}[a_{\mathbf{x},i} a_{\mathbf{x}',j}] = \frac{-2.91k^2}{2\pi^2} \int dz \left(\frac{L-z}{L}\right)^{5/3} C_n^2(z) \iint d\boldsymbol{\rho} d\boldsymbol{\rho}' \\ \times P(\rho) P(\rho') Z_i(\boldsymbol{\rho}) Z_j(\boldsymbol{\rho}') \left| R(\boldsymbol{\rho} - \boldsymbol{\rho}') + \left(\frac{z}{L-z}\right)(\mathbf{x} - \mathbf{x}') \right|^{5/3}. \end{aligned} \tag{4.32}$$

If we seek to leverage the results of the two-aperture statistics, we must relate the magnitude term to some two-aperture separation in accordance with Takato and Yamaguchi. We therefore define

$$\mathbf{s}(z) = \left(\frac{z}{D(L-z)}\right)(\mathbf{x} - \mathbf{x}'), \tag{4.33}$$

to be a displacement that is changing with distance along the path of propagation. With this substitution, we can write

$$\mathbb{E}[a_{\mathbf{x},i}a_{\mathbf{x}',j}] = \frac{-2.91k^2}{2\pi^2}\int dz \left(\frac{L-z}{L}\right)^{5/3} C_n^2(z) \iint d\boldsymbol{\rho} d\boldsymbol{\rho}' \\ \times P(\rho)P(\rho')Z_i(\boldsymbol{\rho})Z_j(\boldsymbol{\rho}')\left|R(\boldsymbol{\rho}-\boldsymbol{\rho}')+D\mathbf{s}(z)\right|^{5/3}. \tag{4.34}$$

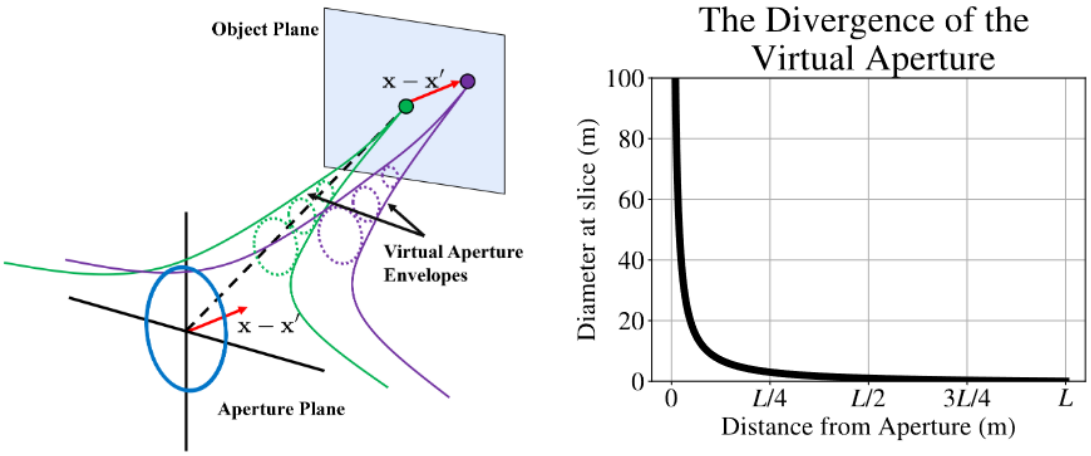
We then recognize the inner double integral to be of the same form as [72, 74]. Defining $\mathscr{A} = 0.00969k^2 2^{14/3}\pi^{2/3}R^{5/3}$, this allows us to simply leverage the results of Takato and Yamaguchi using a weighted integration of their solutions, resulting in

$$\mathbb{E}[a_{\mathbf{x},i}a_{\mathbf{x}',j}] = \mathscr{A}_{i,j}\int_0^L \left(\frac{L-z}{L}\right)^{5/3} C_n^2(z) f_{ij}\left(\mathbf{s}(z), k_0\right) dz, \tag{4.35}$$

where $\mathscr{A}_{i,j} = \mathscr{A}\sqrt{(n_i+1)(n_j+1)}$. This result, however, has an additional visual interpretation. Turning to (4.33), we can write this in terms of a "virtual" aperture which varies with distance, which we define to be

$$\hat{D}(z) = D\left(\frac{L-z}{z}\right). \tag{4.36}$$

We may then write the displacement as $\mathbf{s}(z) = (\mathbf{x}-\mathbf{x}')/\hat{D}(z)$. We provide a visualization of this virtual aperture in Figure 4.3 which illustrates two points in an object forming two cones with diverging radii. The overlap of the cross sections at each individual infinitesimal slice is the problem analyzed by Takato and Yamaguchi. However, our result differs by being a *sum* of these solutions; each infinitesimal slice contributes a correlation which is dictated by the results of [72]. Intuitively, the slice closest against the aperture will contribute global correlation. Physically this is due to the fact that every point source on an object will pass through this slice. Mathematically, the case of this final slice will result in $\hat{D}(z) \to \infty$, which



(a) Virtual aperture

(b) Diameter of virtual aperture

Figure 4.3. Visualization of the virtual aperture. In (a) we show in 3D space how the virtual aperture changes with position along the length of propagation. In (b) we show the increase in the virtual aperture diameter as a function of propagation distance.

will cause $\mathbf{s}(z) \to 0$ for all finite $\mathbf{x} - \mathbf{x}'$, implying perfect global correlation. To present the main result completely, we substitute (4.33) into (4.35), giving us

$$\mathbb{E}[a_{\mathbf{x},i} a_{\mathbf{x}',j}] = \mathscr{A}_{i,j} \int_0^L \left(\frac{L-z}{L}\right)^{5/3} C_n^2(z) f_{ij}\left(\left(\frac{z}{D(L-z)}\right)(\mathbf{x}-\mathbf{x}'), k_0\right) dz. \tag{4.37}$$

4.4.4 Discrete Case: C_n^2-slices

The expression for the correlation of two points in the object plane (4.37) is desirable for the sake of simulation. These results can be directly used in previous multi-aperture simulations [37, 75] with only minor modification. In addition to this, (4.37) is preferable to the results of Whiteley et al. [76] for numerical evaluation of the integral. This is due to the fact that their derivation used a particular Riemann summation rule as a component of their derivation. Equation (4.37) assumes no such rule. The result is an expression which can be adapted to suit the needs of the application without needing to re-derive for a separate integration rule. In particular, the way in which one performs a summation over $C_n^2(z)$ is the primary knob one may tune to utilize the results of (4.37).

Since one may choose the way in which the integral is represented as a Riemann sum, one possibility would be to average the turbulence profile along the interval of propagation. There is some analogy here to the phase screens used in split-step propagation, though they are not directly equivalent. The main difference is that within this framework, one does not have to actually generate phase screens, rather, just evaluate the statistical expression. To differentiate the two, we denote them as C_n^2-slices. We begin with defining a collection of C_n^2-slices to be

$$C_n^2(z;M) = \sum_{m=1}^{M} \delta\left(z - \frac{Lm}{M+1}\right) \int_{L(m-1)/M}^{Lm/M} C_n^2(v)dv, \tag{4.38}$$

where we are integrating along the path of propagation using v as a dummy variable. For some simplicity in notation, we define the locally collapsed mth C_n^2-slice to be

$$\overline{C_n^2}(z_m) = \int_{L(m-1)/M}^{Lm/M} C_n^2(v)dv. \tag{4.39}$$

Using the representation of C_n^2-slices in place of the turbulence profile in (4.37), we can arrive at

$$\mathbb{E}[a_{\mathbf{x},i}a_{\mathbf{x}',j};M] = \mathscr{A}_{i,j} \sum_{m=1}^{M} \left(\frac{M+1-m}{M+1}\right)^{5/3} \overline{C_n^2}(z_m) f_{ij}\left(\frac{m(\mathbf{x}-\mathbf{x}')}{D(M+1-m)}, k_0\right). \tag{4.40}$$

We use (4.40) as a basis for our numerical testing. Due to the generality of the expression provided in (4.37), one may use an alternative integration rule in order to decide each C_n^2 value as previously stated. Furthermore, instead of focusing on numerical integration, one may instead optimize the discrete C_n^2 values for objective functions which optimize quantities such as the Fried parameter, isoplanatic angle, and log amplitude variance as in Hardie et al. [56]. In this case, individual C_n^2 values which describe the phase screen parameters were optimized in a fashion similar to that by Schmidt [22]. Thus (4.40) can also be used to evaluate the impact of simulation parameters on the various Zernike correlations if one replaces $\overline{C_n^2}$ with other values.

4.4.5 Comparison to Earlier Multi-Aperture Methods

The fundamental work that enabled the first iteration of the multi-aperture simulation [37] imposed two main restrictions to achieve their results. The first is the assumption of a constant C_n^2 profile. However, the more limiting restriction is an approximation upon the structure function, which they use a first order Taylor series to simplify. The following consideration achieves the same results, shedding some light on the limitations of these previous results.

To arrive at these same results, we first choose to define a single C_n^2-slice along the entire path of propagation,

$$C_n^2(z;1) = \int_0^L \delta\left(z - \frac{L}{2}\right) C_n^2(v)dv. \tag{4.41}$$

With the additional assumption of a constant C_n^2 profile, then

$$C_n^2(z;1) = LC_n^2\delta\left(z - \frac{L}{2}\right). \tag{4.42}$$

The resulting substitution of this C_n^2 profile results in the same correlation function as in [37],

$$\mathbb{E}[a_{\mathbf{x},i}a_{\mathbf{x}',j};1] = \mathscr{A}_{i,j}\left(\frac{1}{2}\right)^{5/3} LC_n^2 f_{ij}\left(\frac{(\mathbf{x}-\mathbf{x}')}{D}, k_0\right). \tag{4.43}$$

This demonstrates the limiting assumption inherent to previous multi-aperture simulators more clearly. Specifically, these simulation methodologies assume a single C_n^2-slice at the halfway point of the propagation path. Beyond previously discussed limitations, even the case of constant C_n^2 profiles will experience moderate deviations given certain camera configurations, which has been observed [37]. To comment more directly on the usage of our results within theirs, the more general results in this work only requires the replacement of (4.43) with (4.40), keeping the rest of the simulation the same. This offers a considerable increase in accuracy with only an initial small loss in speed – after the generation of the theoretical spatial correlation, the generation method (and therefore the speed) is identical.

4.4.6 Comparison to Existing Zernike Correlations

The work most closely aligned with ours is that of Whiteley et al. [76]. Whiteley et al. provides an expression for spatial and temporal Zernike correlations, and similarly utilize Takato and Yamaguchi. The core difference is notably in approach taken and result. On the approach side, [76] uses an assumption on the independence of neighboring atmospheric slices. Ours additionally uses such a fundamental assumption, as this underlies the standard Markov approximation, however our inclusion of this independence is "built into" the result of Takato et al. [72]. Therefore, it is useful to make the comment that both works ultimately rely on this same layered atmospheric property – though Whiteley et al. [76] assumes this earlier in their derivation. The result is that if one wishes to change a Riemann sum approach (such as max/min vs. Simpson's rule) one must carry the derivation out again. With ours, one may start at the result of (4.37), and discretize as desired following Section 4.4.4 directly. Furthermore, the visual interpretation of the integration process via the virtual aperture along with the subjectively more convenient form of (4.37) are two added benefits of our approach.

It is important to note that Whiteley et al. [76] does offer a more general framework with regards to aperture motion and temporal correlations. This is done primarily through use of Taylor's frozen flow hypothesis. To illustrate how these concepts may be incorporated into our framework, the temporal effects can be modeled in a similar fashion via

$$\mathbf{s}(z) = \left(\frac{z}{D(L-z)}\right)(\mathbf{x}-\mathbf{x}') + \frac{\mathbf{v}(z)\tau}{D}, \tag{4.44}$$

where $\mathbf{v}(z)$ is the mean transverse wind velocity at position z along the path of integration. We note the division by D is a requirement to match the form of Takato and Yamaguchi's expression [72]. We may substitute (4.44) into (4.35), along with potential modifications of (4.44) as outlined in Sasiela's book [21].

4.4.7 The function f_{ij}

Here we detail the function f_{ij} which characterizes the correlations of the Zernike polynomials. The form of their equation is rather cumbersome and somewhat difficult to interpret for certain values of Noll indices. The purpose of this discussion is to simplify the resulting equation for the ease of interpretation and further study of these results. Following Takato and Yamaguchi [72], we first define the function

$$I_{a,b,c}(s,k_0) = \int dx \frac{J_a(sx)J_b(x)J_c(x)}{x(x^2+k_0)^2}, \tag{4.45}$$

with J_k as the kth order Bessel function of the first kind. With Noll indices $(n_i, m_i) \to i$, we then define

$$n^+ = n_i + n_j, \tag{4.46}$$

$$n^- = n_i - n_j, \tag{4.47}$$

$$m^+ = m_i + m_j, \tag{4.48}$$

$$m^- = m_i - m_j. \tag{4.49}$$

We also define an indicator function,

$$h(i,j) = \begin{cases} 1 & m_i \neq 0; m_j \neq 0; i+j \text{ even} \\ 2 & m_i \neq 0; m_j \neq 0; i+j \text{ odd} \\ 3 & m_i = 0; j \text{ even} \oplus m_j = 0; i \text{ even} \\ 4 & m_i = 0; j \text{ odd} \oplus m_j = 0; i \text{ odd} \\ 5 & m_i = 0; m_j = 0 \end{cases}, \tag{4.50}$$

with $\oplus$ denoting the XOR function. First, we will present a form in line with that of [72], though using some of this notation and the appropriate simplifications. For a displacement $\mathbf{s} = (s, \varphi)$ written in polar form, we can write the expression in [72] as in (4.51)

$$f_{ij}(\mathbf{s},k_0)=\begin{cases}\pm(-1)^{(n^+-m^+)/2}\cos(m^+\varphi)I_{m^+,n_i+1,n_j+1}(2s,2\pi Rk_0) & \\ \quad+(-1)^{(n^++2m_i+|m^-|)/2}\cos(m^-\varphi)I_{|m^-|,n_i+1,n_j+1}(2s,2\pi Rk_0) & h(i,j)=1\\ (-1)^{(n^+-m^+)/2}\sin(m^+\varphi)I_{m^+,n_i+1,n_j+1}(2s,2\pi Rk_0) & \\ \quad+(-1)^{(n^++2m_i+|m^-|)/2}\sin(m^-\varphi)I_{|m^-|,n_i+1,n_j+1}(2s,2\pi Rk_0) & h(i,j)=2\\ (-1)^{(n^+-m^+)/2}\sqrt{2}\cos(m^+\varphi)I_{m^+,n_i+1,n_j+1}(2s,2\pi Rk_0) & h(i,j)=3\\ (-1)^{(n^+-m^+)/2}\sqrt{2}\sin(m^+\varphi)I_{m^+,n_i+1,n_j+1}(2s,2\pi Rk_0) & h(i,j)=4\\ (-1)^{(n^+-m^+)/2}I_{m^+,n_i+1,n_j+1}(2s,2\pi Rk_0) & h(i,j)=5\end{cases}\ ,\tag{4.51}$$

with the $\pm$ corresponding to $+$ if both (i,j) are even, and $-$ if they are both odd.

However, our notation highlights further possible simplification. We can therefore write the function as

$$\begin{aligned}f_{ij}(\mathbf{s},k_0)=&(-1)^{(n^+-m^+)/2}\Theta^{(1)}(i,j)I_{m^+,n_i+1,n_j+1}(2s,2\pi Rk_0)\\&+(-1)^{(n^++2m_i+|m^-|)/2}\Theta^{(2)}(i,j)I_{|m^-|,n_i+1,n_j+1}(2s,2\pi Rk_0),\end{aligned}\tag{4.52}$$

with functions

$$\Theta^{(1)}(i,j)=\begin{cases}(-1)^j\cos(m^+\varphi) & h(i,j)=1\\ \sin(m^+\varphi) & h(i,j)=2\\ \sqrt{2}\cos(m^+\varphi) & h(i,j)=3\\ \sqrt{2}\sin(m^+\varphi) & h(i,j)=4\\ 1 & h(i,j)=5\end{cases}\tag{4.53}$$

and,

$$\Theta^{(2)}(i,j)=\begin{cases}\cos(m^-\varphi) & h(i,j)=1\\ \sin(m^-\varphi) & h(i,j)=2\\ 0 & h(i,j)=3\\ 0 & h(i,j)=4\\ 0 & h(i,j)=5\end{cases}\ ,\tag{4.54}$$

contributing the angular terms.

4.5 Sampling from the Zernike space

I would jokingly mention that in doing things related to turbulence you are never done. It may appear that with the statistics described we can easily pull random realizations from it and move on our way. Unfortunately, that is not the case. There is a small issue lingering in the correlations of the Zernike space that present us from using **Fast Fourier transform (FFT)**-based solutions to sampling. Therefore, we must come up with a trick in order to allow us to use the FFT methods for random sampling.

4.5.1 Main issue

When we look at the correlation expression Equation (4.37), we notice that it is *not* wide sense stationary. Why is this and what effect does it have on our steps forward? To answer why it is, we need to look no further than realizing that the function f_{ij} depends on (i, j) in a *pairwise* sense, and not its difference $i - j$. This can be understood that there is no notion of neighborhood in the index dimension – the ordering of the Zernike polynomials is arbitrary.

What impact does this have on generating samples from this distribution. Well, this means that the Wiener-Khinchin theorem *no longer applies!* This means that we cannot use FFT-based methods to sample from the Zernike space. The obvious alternative may be to take the large covariance tensor $\mathbf{A}$ and decompose it by the some method. The problem is that for a 256×256 image, this matrix will be massive; it will likely be on the order of many GB's. Therefore, we need an alternative.

The idea of drawing correlated Zernike coefficients according to a correlation matrix is not difficult if the number of grid points is small. Consider a multivariate Gaussian random variable with mean $\mathbb{E}[\mathbf{y}] = 0$ and the correlation matrix $\mathbf{\Sigma} = \mathbb{E}[\mathbf{y}\mathbf{y}^T]$. To draw a random vector $\mathbf{y}$ from this distribution, we can decompose the correlation matrix $\mathbf{\Sigma} = \mathbf{U}\mathbf{S}\mathbf{U}^T$ via the eigen-decomposition, and define $\mathbf{\Sigma}^{\frac{1}{2}} = \mathbf{U}\mathbf{S}^{\frac{1}{2}}\mathbf{U}^T$. Then, starting with a white noise vector $\mathbf{e} \sim \text{Gaussian}(0, \mathbf{I})$, the transformed vector $\mathbf{y} = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{e}$ will satisfy the desired property that $\mathbb{E}[\mathbf{y}] = 0$ and $\mathbb{E}[\mathbf{y}\mathbf{y}^T] = \mathbf{\Sigma}$.

When the size of the grid is small so that the dimension of the correlation matrix $\mathbf{\Sigma}$ is small, the matrix $\mathbf{\Sigma}^{\frac{1}{2}}$ can be generated using standard numerical techniques such as the Cholesky factorization. However, for a large grid of points, storing the matrix $\mathbf{\Sigma}$ and running the factorization would become infeasible. One exception is that if $\mathbf{y}$ is homogeneous (wide-sense stationary), then the correlation matrix $\mathbf{\Sigma}$ is *circulant* and so the eigen-decomposition is equivalent to the Fourier transform. In this case, generating the random vector $\mathbf{y}$ can be implemented via

$$\mathbf{y} = \mathbf{U}\mathbf{S}^{\frac{1}{2}}\mathbf{U}^T\mathbf{e} = \mathcal{F}^{-1}(\mathbf{S}^{\frac{1}{2}}\mathcal{F}(\mathbf{e})),$$

where $\mathcal{F}$ denotes the discrete-time Fourier transform, and the diagonal matrix $\mathbf{S}$ is the Fourier spectrum of one row (or column) of the correlation matrix $\mathbf{\Sigma}$.

The significance of the homogeneity is that it allows us to speed up the sampling process by performing all computations in the Fourier space. In addition, the memory bottleneck is resolved because we do need to construct the full correlation matrix $\mathbf{\Sigma}$ and run the Cholesky factorization. The question is: For the Zernike correlation matrix we are considering in this paper, does it have any kind of wide sense stationarity? If not, can we approximate it using something that has such a property?

4.5.2 Structure of the Correlation Matrix

The coefficients $a_{\mathbf{x},j}$ are zero-mean Gaussian, and have a correlation matrix (technically, a tensor stored in the matrix form) which we denote the $(\mathbf{x}, \mathbf{x}', i, j)$th element of the matrix $\mathbf{A}$ as

$$[\mathbf{A}]_{\mathbf{x},\mathbf{x}',i,j} = \mathbb{E}[a_{\mathbf{x},i}\ a_{\mathbf{x}',j}]. \tag{4.55}$$

This notation stresses that the matrix $\mathbf{A}$ is location dependent on the pair $\mathbf{x}$ and $\mathbf{x}'$. In the special case when the correlation matrix is spatially invarying, we write $[\mathbf{A}]_{\mathbf{x},\mathbf{x}',i,j}$ as $[\mathbf{A}]_{\mathbf{x}-\mathbf{x}',i,j}$. If we further assume that $\mathbf{x}' = \mathbf{x}$, the correlation matrix becomes $[\mathbf{A}]_{\mathbf{0},i,j} = \mathbb{E}[a_{\mathbf{x},i}\ a_{\mathbf{x},j}]$, which is equivalent to the covariance matrix given by Noll [70].

Let us now take a closer look at the structure of the Zernike correlation matrix $\mathbf{A}$. As defined in Equation (4.55), the $(\mathbf{x}, \mathbf{x}', i, j)$th entry of the correlation matrix at the location-

pair $(\mathbf{x}, \mathbf{x}')$ and the Zernike mode-pair (i, j) is $[\mathbf{A}]_{\mathbf{x},\mathbf{x}',i,j} = \mathbb{E}[a_{\mathbf{x},i}\ a_{\mathbf{x}',j}]$. Thus, $\mathbf{A}$ is a four-dimensional tensor, with two dimensions allocated to the pair of spatial coordinates (organized through a column-wise stack), and two dimensions allocated to the pair of Zernike modes. One way to visualize the Zernike space is to consider the illustration shown in Figure 4.4. For a fixed pixel location $\mathbf{x}$, there is a vector $\mathbf{a}_{\mathbf{x}} = [a_{\mathbf{x},1}, \ldots, a_{\mathbf{x},N}]$ where $n = 1, \ldots, N$ denotes the Zernike mode index. As we move to another pixel location $\mathbf{x}'$, the vector becomes $\mathbf{a}_{\mathbf{x}'} = [a_{\mathbf{x}',1}, \ldots, a_{\mathbf{x}',N}]$.

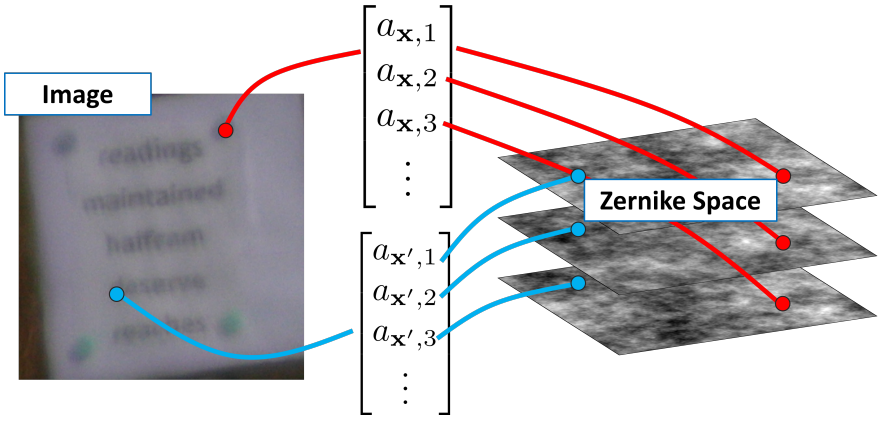


Figure 4.4. The Zernike Space is a representation describing the phase decomposition by its basis functions. For every point on an object, or analogously every pixel in an image, there is a Zernike vector that describes the distortion across the aperture. This motivates us to define the Zernike space as a tensor.

For a fixed Zernike mode pair (i, j), the correlation is limited to the spatial axis. This will give us the (i, j)th slice of the four-dimensional tensor

$$\mathbf{A}_{i,j} = \begin{bmatrix} [\mathbf{A}]_{\mathbf{x}_1,\mathbf{x}_1,i,j} & \cdots & [\mathbf{A}]_{\mathbf{x}_p,\mathbf{x}_1,i,j} \\ \vdots & \ddots & \vdots \\ [\mathbf{A}]_{\mathbf{x}_p,\mathbf{x}_1,i,j} & \cdots & [\mathbf{A}]_{\mathbf{x}_p,\mathbf{x}_p,i,j} \end{bmatrix},$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_p$ are the p coordinates in the grid. Assuming that the Zernike coefficients across the field of view is homogeneous [37], the $(\mathbf{x}, \mathbf{x}')$th element $[\mathbf{A}_{i,j}]_{\mathbf{x},\mathbf{x}'}$ will be the function of $\mathbf{x} - \mathbf{x}'$ instead of the absolute positions $(\mathbf{x}, \mathbf{x}')$. In this case, we can write $[\mathbf{A}_{i,j}]_{\mathbf{x},\mathbf{x}'}$ as

$[\mathbf{A}_{i,j}]_{\mathbf{s}}$ where $\mathbf{s} = (\mathbf{x} - \mathbf{x}')/D$ with D being the aperture diameter[1]. For a grid of points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$, the matrix $\mathbf{A}_{i,j}$ takes the form

$$\mathbf{A}_{i,j} = \begin{bmatrix} [\mathbf{A}]_{\mathbf{s}_0,i,j} & [\mathbf{A}]_{\mathbf{s}_1,i,j} & \cdots & [\mathbf{A}]_{\mathbf{s}_p,i,j} \\ [\mathbf{A}]_{\mathbf{s}_1,i,j} & [\mathbf{A}]_{\mathbf{s}_0,i,j} & \cdots & [\mathbf{A}]_{\mathbf{s}_{p-1},i,j} \\ \vdots & \vdots & \ddots & \vdots \\ [\mathbf{A}]_{\mathbf{s}_p,i,j} & [\mathbf{A}]_{\mathbf{s}_{p-1},i,j} & \cdots & [\mathbf{A}]_{\mathbf{s}_0,i,j} \end{bmatrix}.$$

Because of homogeneity, $\mathbf{A}_{i,j}$ can be decomposed via the Fourier transform. This "slice" of the tensor is representative of a single Zernike coefficient field.

For a fixed location pair $(\mathbf{x}, \mathbf{x}')$, we can obtain another slice of the correlation matrix

$$\mathbf{A}_{\mathbf{x},\mathbf{x}'} = \mathbf{A}_{\mathbf{s}} = \begin{bmatrix} [\mathbf{A}]_{\mathbf{s},1,1} & \cdots & [\mathbf{A}]_{\mathbf{s},1,N} \\ \vdots & \ddots & \vdots \\ [\mathbf{A}]_{\mathbf{s},N,1} & \cdots & [\mathbf{A}]_{\mathbf{s},N,N} \end{bmatrix}.$$

The special case where $\mathbf{s} = \mathbf{0}$, which gives matrix $\mathbf{A}_{\mathbf{0}}$ is exactly the Noll matrix [70] that specifies the correlation between the Zernike modes at a single pixel location $\mathbf{x}$. In general, the matrix $\mathbf{A}_{\mathbf{s}}$ does not have the circulant structure and cannot be decomposed via the Fourier transform.

The prior notation allows us to easily identify the homogeneous and the non-homogeneous components of the tensor. The entire Zernike correlation tensor $\mathbf{A}$ can be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \cdots & \mathbf{A}_{1,N} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{N,1} & \cdots & \mathbf{A}_{N,N} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{\mathbf{s}_0} & \cdots & \mathbf{A}_{\mathbf{s}_p} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{\mathbf{s}_p} & \cdots & \mathbf{A}_{\mathbf{s}_0} \end{bmatrix},$$

[1]↑The necessity to normalize the coordinate using D comes from the fact that the geometry of the optical system has a certain impact to the Zernike coefficients. In particular, two aperture diameters D_1 and D_2 will lead to two different Zernike space because the imaging systems are viewing through different effective slices as a function of aperture size. Therefore, we define a standardized unitless vector corresponding to correlation length in accordance with [37] to be $\mathbf{s} = \frac{L(\boldsymbol{\theta}-\boldsymbol{\theta}')}{D} = \frac{\mathbf{x}-\mathbf{x}'}{D}$, where $\boldsymbol{\theta}, \boldsymbol{\theta}'$ are two vectors pointing from the center of the imaging system to points $\mathbf{x}, \mathbf{x}'$, respectively.

The entire tensor structure is non-homogeneous. Thus, it is not possible to draw samples via the FFT method, which highlights a fundamental limitation of [37] and [75].

4.5.3 Generation of the Random Zernike Fields

Due to the non-homogeneity of $\mathbf{A}$, we propose an approximate tensor $\widetilde{\mathbf{A}}$ which we claim captures a majority of the statistical behavior while having the property of homogeneity. Reserving the accuracy of this approximation for section 4.5.4, we propose the following method for the generation of the Zernike fields:

1. Generate $i = \{1, 2, \ldots, N\}$ unit-variance, spatially correlated random fields according to $\mathbf{A}_{i,i}$. Note this generation uses only *auto*covariance functions. At this stage, our N fields are independent, thus utilizing the FFT-based method based on their homogeneity.

2. Perform a point-wise mixing of the random fields according to the Noll matrix $\mathbf{A_0}$. This mixing is done pixel-wise (across the coefficient index dimension) per pixel.

This generation process will give us random fields which are in accordance with the autocovariance functions, however, for the cross-covariance terms there will exist some deviation. This is most simply presented by the form of the covariance structure of the tensor,

$$\widetilde{\mathbf{A}} = \mathbf{L} \begin{bmatrix} \mathbf{A}_{1,1} & 0 & \ldots & 0 \\ 0 & \mathbf{A}_{2,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}_{N,N} \end{bmatrix} \mathbf{L}^T, \tag{4.56}$$

where $\mathbf{L}\mathbf{L}^T = \mathbf{A_0}$. We note the resultant matrix is no longer diagonal. Therefore, the off-diagonal entries of $\widetilde{\mathbf{A}}$ differ from $\mathbf{A}$, which will be the focus of our numerical analysis.

We provide a visualization of the resulting covariance matrix in Figure 4.5 using a simplistic example in the case of an 8×8 image with 3 coefficient fields. Initially, white noise is generated for an $8 \times 8 \times 3$ random volume, which is then spatially correlated for each slice according to its autocovariance function. The resulting covariance matrix has a considerable amount of zero-entries corresponding to the other 8×8 random fields. After this, the fields

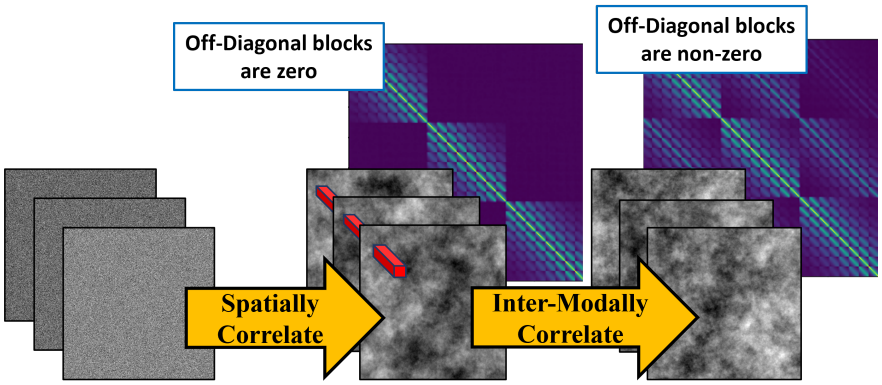


Figure 4.5. We give a visual representation of the covariance structure as it changes through the generation process. First, white noise has a covariance structure applied to it, resulting in spatially correlated, yet independent random fields. Next, along the index axis, a mixing matrix is applied via Cholesky decomposition. Finally, the resulting fields are correlated both spatially and along the index dimension and have the covariance structure shown.

are mixed according to the 3×3 covariance matrix along the final index axis, after which the covariance matrix becomes much denser. This same principle is extended to the case of an $W \times H$ image with N coefficient fields.

4.5.4 The Impact of Cross-Correlation Functions

The core innovation in this work is the ability to utilize property of homogeneity for individual fields $a_{\mathbf{s},i}$ (or equivalently, $\widetilde{a}_{\mathbf{s},i}$). This allows us to quickly draw samples using FFT-based generation followed by a mixing matrix. To justify this approximation, we begin with the following considerations:

1. The correlation functions $\mathbf{A}_{i,j}$ for $i \neq j$ sharply decay and approximately vanish for $s > 4$;

2. The off-diagonal terms for the matrix $\mathbf{A}_{\mathbf{x},\mathbf{x}}$ (corresponding to indices $i \neq j$) contain a small proportion of the energy.

Together, these two properties suggest that the overall energy lost by the removal of the cross-correlation functions will be negligible. Intuitively, the reason for this lies within the

sharp decay of the correlation functions (1), which are already significantly smaller than their counterparts (2).

With this observation, we turn to the numerical analysis performed to more concretely justify this approximation. We wish to measure the energy that will be contained within our approximation and compare it with the energy contained if the entire cross-correlation functions were to be retained. To do so, we integrate outward in the Zernike space and quantify the deviation of $\widetilde{\mathbf{A}}$ from $\mathbf{A}$ as a function of $\mathbf{s}$. While the correlation functions $\mathbf{A}_{i,j}$ have angular dependencies, we integrate over the angular components for ease of presenting our comparison. We note that along any particular direction individually, the results do not vary considerably. Mathematically, we write the energy contained with all functions considered via the Frobenius norm as

$$E(s) = \int_0^s \int_0^{2\pi} d\mathbf{s}' \operatorname{tr}\left\{\mathbf{A}_{\mathbf{s}'}^T \mathbf{A}_{\mathbf{s}'}\right\}, \tag{4.57}$$

where we've integrated spatially outwards up to s, as we intend show the accuracy of the approximation as the size of our field in the Zernike space grows. As s becomes large with respect to the spread of the function, $E(s)$ will converge. The energy within our approximation can be written similarly as

$$\widetilde{E}(s) = \int_0^s \int_0^{2\pi} d\mathbf{s}' \operatorname{tr}\left\{\widetilde{\mathbf{A}}_{\mathbf{s}'}^T \widetilde{\mathbf{A}}_{\mathbf{s}'}\right\}. \tag{4.58}$$

In addition to comparing the energies separately, we may also consider measuring the difference between them. To this end, we propose to measure

$$E^{(-)}(s) = \int_0^s \int_0^{2\pi} d\mathbf{s}' \left|\operatorname{tr}\left\{\mathbf{A}_{\mathbf{s}'}^T \mathbf{A}_{\mathbf{s}'} - \widetilde{\mathbf{A}}_{\mathbf{s}'}^T \widetilde{\mathbf{A}}_{\mathbf{s}'}\right\}\right|, \tag{4.59}$$

where we are only interested in the magnitude of the residual, therefore measuring its absolute value.

These three functions will help to provide insight into the accuracy of our approximation. Ideally, we want $E(s)$ and $\widetilde{E}(s)$ to match. However, this alone is not enough to claim accuracy, as this is only a way of measuring the total joint behavior. We additionally then propose to

use $E^{(-)}(s)$ to measure the total index-wise residuals. An optimal result for $E^{(-)}(s)$ would be for it to vanish at every point. We present the results of this numerical analysis in Figure 4.6 which we note is cumulative, therefore reaching a steady state means no additional errors will be incurred. We also choose to not include the first three Zernike coefficients (piston and x,y-tilts) for this plot, as these terms dominate the plot significantly in magnitude, though the conclusion of the analysis is unchanged. With this, we present the following observations:

1. At $s = 0$ there is no loss in energy/statistical accuracy (e.g. for a single point we are perfectly accurate);

2. A majority of the errors are for small $\mathbf{s}$, which is expected as the correlation functions have not yet vanished at this point;

3. For separations $s > 4$, there is no additional loss in energy and remains at steady state. The energy of the difference $E^{(-)}(s)$ is two orders of magnitude smaller than the total energy, suggesting its overall accuracy.

These three observations suggest that our approximation retains a majority of the field's statistical behavior, while allowing for considerable speed-up.

4.5.5 Spatio-Temporally Correlated Fields

The extension of the proposed ideas to spatio-temporally correlated fields is possible through the adoption of Taylor's frozen flow hypothesis [24]. Under this hypothesis, a displacement in time $\mathbf{t}$ may be written as a spatial displacement $\mathbf{y}$ through the relationship $v\mathbf{t} = \mathbf{y}$, with v as the mean transverse velocity of the turbulent medium. Taylor's frozen flow hypothesis allows us to extend the results by considering $\mathbf{z} = \mathbf{x} + \mathbf{y}$. This will give us a spatio-temporal tensor $\mathbf{A}_{\mathbf{z},\mathbf{z}'}$ of analogous structure to the previously described $\mathbf{A}_{\mathbf{x},\mathbf{x}'}$.

A downside of the frozen flow hypothesis is that we need to pre-compute and store multiple three-dimensional random fields. The workaround solution is to enforce the temporal correlation via an auto-regressive process. Therefore, the random seed used to generate the Zernike fields are updated according to $\mathbf{e}_t = \gamma\mathbf{e}_{t-1} + \sqrt{1-\gamma^2}\mathbf{n}$, with $\mathbf{n}$ as a zero-mean, unit-variance Gaussian vector and $\gamma \in [0, 1]$. While the theoretical justification of such a

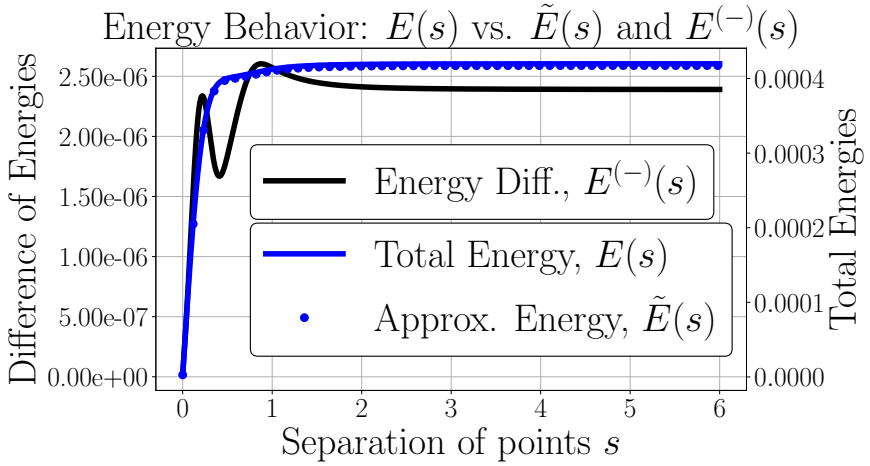


Figure 4.6. The results from our numerical evaluation of the three proposed energy functions. The axis for $E(s)$ and $\widetilde{E}(s)$ are on the right, while the axis for residual difference energy function $E^{(-)}$ is on the left. We note the apparent match between the two individual energy functions. More importantly, there is a comparably small magnitude for the difference function, suggesting a reasonable match.

model is limited, the speed of simulation is marginally impacted. We observe the visual effects under the auto-regressive approximation to be suitable in practice.

4.6 Phase-to-Space transform

While the Zernike space represents the main theoretical development of our simulation technique, the **Phase-to-space (P2S) transform** represents the key practical novelty of our simulator.

4.6.1 Motivation

Suppose we have a sample of the Zernike space generated on our computer. Whether or not it utilizes the two-step method previously described, we will still have to form the PSF in order to apply them to an image. If we do this in the standard fashion, we must compute PSF formation equation for *each pixel in the image*:

$$|h_{\mathbf{u}}(\mathbf{x})|^2 = |\mathfrak{Fourier}\{P(\boldsymbol{\xi})e^{j\phi_{\mathbf{u}}(\boldsymbol{\xi})}\}|^2. \tag{4.60}$$

This means taking a 2D FFT per pixel. This is certainly doable, but it will be slow.

In addition to this, there is another problem: we will have to take a spatially varying convolution! Referring back to our analysis of Chapter 2 regarding spatially varying convolution, and considering that we are indeed talking about image formation, we can recall that spatially varying convolution takes the form

$$I_i(\mathbf{x}) = \sum_i \int \beta_{\mathbf{u},i}\varphi_i(\mathbf{x}-\mathbf{u})I_g(\mathbf{u})d\mathbf{u}, \tag{4.61}$$

$$= \sum_i (\varphi_i \circledast \beta_{\mathbf{u},i}I_g)(\mathbf{x}). \tag{4.62}$$

However, this assumes we *know* what the φ should be, and furthermore than the decomposition

$$|h_{\mathbf{u}}|^2 = \sum_i \beta_{\mathbf{u},i}\varphi_i(\mathbf{x}-\mathbf{u}) \tag{4.63}$$

is even feasible. It may save us some time to first compute the PSFs then decompose them by this equation, if we can find $\{\varphi_i\}_i$, however, we'd like to do something that will get two birds with one stone, so to speak. Therefore, our goal is the following: don't take the 2D transform and utilize our (scattering) spatially varying approximation.

4.6.2 The phase-to-space transform

By our goal, hopefully we have made it clear that we are attempting to go from the Zernike space representation (in the phase domain) to apply its effects to the image (in the spatial domain). Therefore, we are wishing to convert the phase effects to the spatial effects, hence the name **phase-to-space (P2S) transform**.

Recall that we have decomposed the phase functions by the Zernike coefficients

$$\phi_{\mathbf{u}}(\boldsymbol{\xi}) = \sum_{j=1}^{N} a_{\mathbf{u},j} Z_j(\boldsymbol{\xi}). \tag{4.64}$$

In the previous discussion, we also introduced the basis decomposition

$$|h_{\mathbf{u}}|^2 = \sum_i \beta_{\mathbf{u},i} \varphi_i(\mathbf{x} - \mathbf{u}). \tag{4.65}$$

Our goal is then as follows: we wish to find the mapping $f(\cdot)$ such that

$$\boldsymbol{\beta}_{\mathbf{u}} = f(\mathbf{a}_{\mathbf{u}}). \tag{4.66}$$

In words, we would like to have a pixel-wise transformation of each Zernike vector to the PSF basis coefficients.

Why does this seem like a reasonable goal. Well, if we can produce $f(\cdot)$ which is computationally inexpensive, then we may be able to do all of these computations *in parallel* in one go. In fact, we manage to achieve this with our chosen transformation. Then, we are already in a position to apply the converted coefficients to the image through the scattering convolution approximation. This translates to a huge speed up in our process.

For a couple months, it was attempted to find an analytical transformation which satisfies the P2S transformation. After a while of failing, we eventually settled on another approach. However, I find it interesting to mention that the attempt of analytically finding the transformation was doomed from the start. Someone already had the transformation we were looking for, and in my understanding, it would be too computationally costly to work for our case. Interestingly, the theory was developed by Zernike himself and a student of his by the name of Nijboer, thus creating Nijboer-Zernike theory. This theory, and its extensions, and well detailed in the thesis: blah.

The solution we eventually landed on may be surprising: a *learned* mapping $f(\cdot)$. In simple terms, we chose a shallow neural network! Why does this make sense in our case? Well, we can generate a database of $\mathbf{a}$ and $\boldsymbol{\beta}$ pairs by actually doing the PSF formation equation and corresponding decomposition, and training a network to sit in place of the transformation for us. As it turns out, this network works shockingly well. It learns quickly, is lightweight, and can handle images of size 1024×1024 in one pass due to its size and parallelizable nature.

4.6.3 Training the P2S

Training the P2S consists of a few steps, though all of which are relatively straightforward to do in Python. Firstly, we specify the network architecture along with conceptually the goal of the P2S transform in Figure 4.7. How can we train this network?

We need to train a transform that does a pixel-wise transformation from Zernike coefficients to PSF basis coefficients. Note that the fact that it is pixel-wise is a simplification: we only need to consider one Zernike vector input and a PSF basis vector output. Therefore, pulling samples from the Noll matrix directly *is* an option. Thus, if there was any skepticism about the two-stage mixing approximation, this result will not rely on this method as it is done per pixel.

To generate the samples from the Noll matrix $\mathbf{A}_0$, we can simply decompose

$$\mathbf{A}_0 = \mathbf{L}\mathbf{L}^T. \tag{4.67}$$

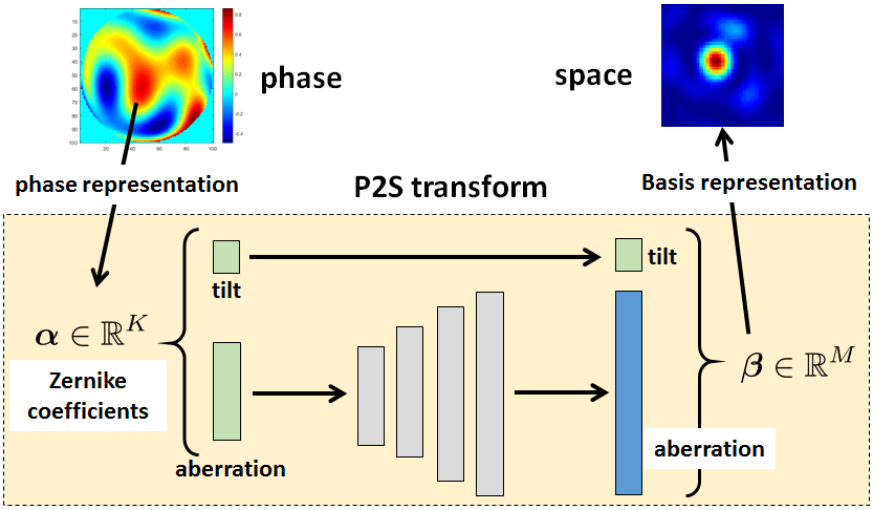


Figure 4.7. An overview of the phase-to-space transform

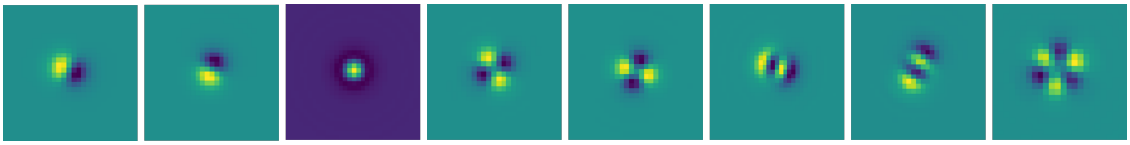


Figure 4.8. The first 8 P2S basis functions found by a PCA.

Therefore we can draw Zernike samples through

$$\mathbf{a} = \mathbf{L}\mathbf{X} \tag{4.68}$$

and $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$. From this we can perform the PSF formation equation

$$|h(\mathbf{x})|^2 = |\mathfrak{Fourier}\{P(\boldsymbol{\xi})e^{j\sum_i a_i Z_i(\boldsymbol{\xi})}\}|^2. \tag{4.69}$$

From here, we can form a dataset of PSF and Zernike coefficients pairs $(|h(\mathbf{x})|^2, \mathbf{a})$.

Now, we must consider how to find the corresponding PSF basis function. To do this, we employ **principal component analysis (PCA)** to find the corresponding $\{\varphi_i\}_i$. From this, we can then decompose each PSF in our database by the numerically computed principal components. Therefore, we can reduce the size of our dataset (in memory) and produce its final form of Zernike and PSF basis coefficients pairs $(\boldsymbol{\beta}, \mathbf{a})$. We train the network as shown in Figure 4.7 to find the mapping

$$\mathbf{a} \underbrace{\longrightarrow}_{\text{P2S}} \boldsymbol{\beta}, \tag{4.70}$$

where we are training based on mean square error (MSE) between $f(\mathbf{a})$ and $\boldsymbol{\beta}$.

A few extra considerations go into generating the P2S basis functions φ_i. When generating the dataset, we are careful to choose a sufficient enough resolution so that we can capture high frequency details. This additionally has helped us in training the transform to be allowed to resize – when done at a low enough resolution we observed aliasing upon resizing. We visualize a few of the φ_i terms in Figure 4.8. We note their clear structure and vague similarity to the Zernike polynomials.

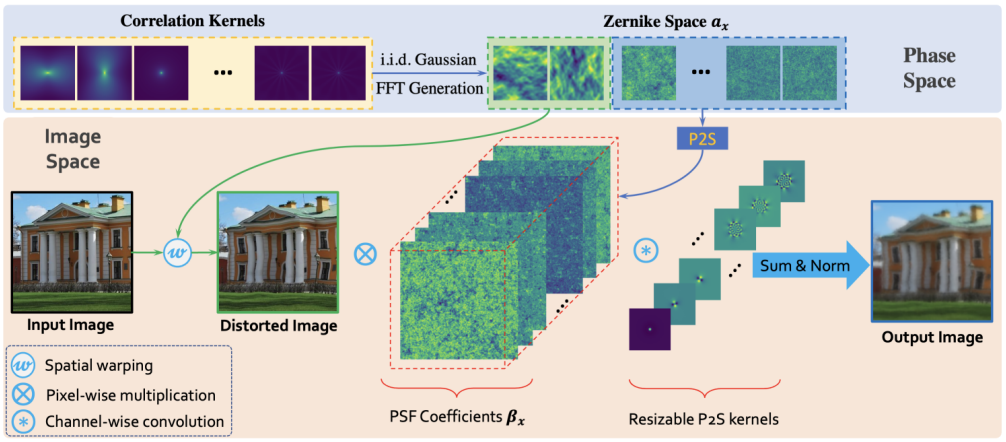


Figure 4.9. An overview of our Zernike-based simulator.

4.7 Zernike-based simulation

A pictoral overview of Zernike-based simulation is presented in Figure 4.9. Let's step through each component.

To begin, the user will input a set of parameters ranging from aperture size, propagation distance, the size of the image in meters, and so on. These parameters dictate the sampling of the Zernike space. We have pre-evaluated the integrals which describes the correlation of the Zernike coefficients, and thus we pull from these pre-evaluated distributions. This step produces $\mathbb{E}[a_{\mathbf{u},i}a_{\mathbf{u}',i}]$, where we would remind the reader that the correlation here is homogeneous.

After this step, the two-stage mixing process will occur. First we generate independent Zernike fields which are then mixed by the Noll matrix. The forms our entire realization of the (approximate) Zernike space. Each time the simulation runs (assumed the parameters are fixed), this is where the *randomness* will enter. We optionally can insert temporal correlation into the generation at this point.

The Zernike space realization is then split into two groups: tilt and blur. As described by [77], tilt must come first. The image J is tilted to form $\mathcal{T}(J)$. Although the tilt has a rather simplistic relationship to its spatial counterpart, the blurs do not. We then employ the P2S network to map the higher order Zernike terms to their spatial coefficient equivalents. After

this, we follow our spatially varying convolution discussion and use the scattering formulation. The result is $\mathcal{B}(\mathcal{T}(J))$ which is the final output image.

4.7.1 Validating the simulator

In analyzing the simulation approach, adherence to the desired statistical behavior is a key metric by which we judge the quality of our simulation approach. The validation here is divided into two categories: (1) aperture statistics and (2) spatial statistics. For statistics on the aperture, the key comparison is our generated statistics plotted against the theoretical structure function. With respect to spatial statistics, our comparison is with the known tilt statistics; there is a limitation on the known behavior of the spatial statistics with respect to the blur without the approximation in [37], so no direct comparison is possible.

Aperture Statistics

For the evaluation of aperture statistics, the most important function to match is that of the structure function given by. By matching this function, as we show in Figure 4.10, we are able to match any statistical value that can be written in terms of the structure function, such as the Fried parameter. We show a good match to the structure function at varying levels of turbulent distortions.

We also perform another experiment on the temporal averages of the distortions, the short exposure (SE) and long exposure (LE) optical transfer functions (OTFs). The OTF in general is defined as

$$H(\boldsymbol{\xi}) = \left(P(\boldsymbol{\xi})e^{-j\phi(\boldsymbol{\xi})}\right) \circledast \left(P(-\boldsymbol{\xi})e^{j\phi(-\boldsymbol{\xi})}\right), \tag{4.71}$$

which is the autocorrelation operation of the overall pupil function and phase distortion. The LE OTF is then given by a temporal average over realizations of individual OTFs. Mathematically, the LE OTF is given as

$$H_{\mathrm{LE}}(\boldsymbol{\xi}) = \mathbb{E}\left[\left(P(\boldsymbol{\xi})e^{-j\phi(\boldsymbol{\xi})}\right) \circledast \left(P(-\boldsymbol{\xi})e^{j\phi(-\boldsymbol{\xi})}\right)\right]. \tag{4.72}$$

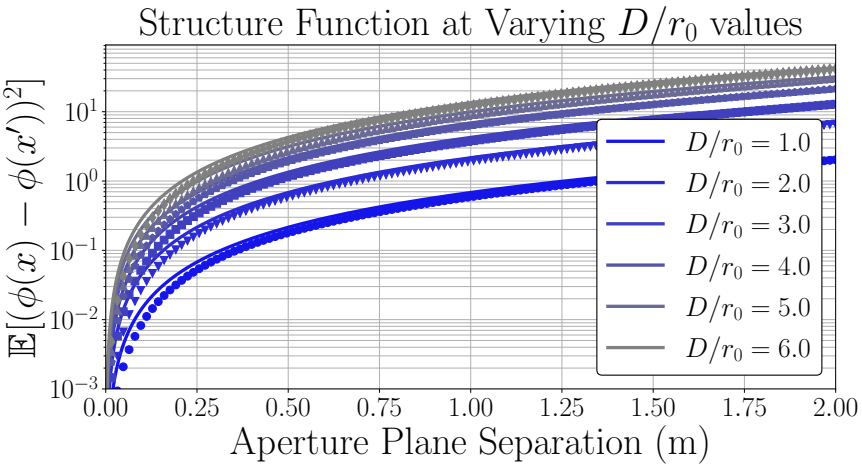


Figure 4.10. A comparison between the theoretical structure function (solid curves) at different distortion levels vs. our generated statistics (dotted lines). We observe a reasonable match across multiple levels of distortions.

The SE OTF is similarly defined using the temporal average, however, this measure uses a "tilt-corrected" phase function φ given by

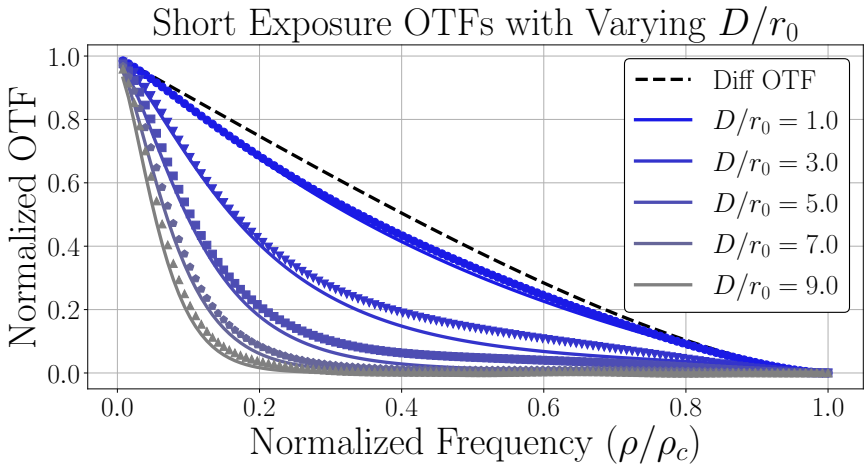
$$\varphi(\boldsymbol{\xi}) = \phi(\boldsymbol{\xi}) - \boldsymbol{\alpha}^T \boldsymbol{\xi}, \tag{4.73}$$

where $\boldsymbol{\alpha}^T \mathbf{f}$ is the plane of best fit, effectively removing the tilt. The remaining phase distortion φ will then only describe the high-order distortions. This differs from the LE OTF, which includes both blur and tilt/shifting. The SE is then described as

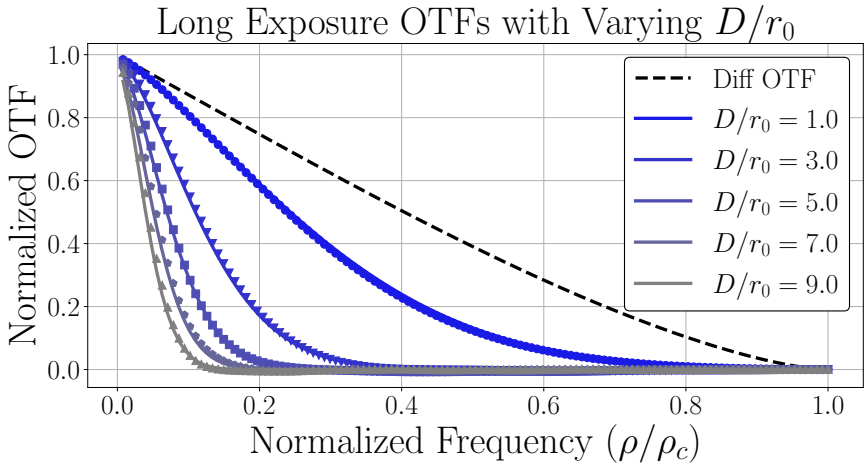
$$H_{\mathrm{SE}}(\boldsymbol{\xi}) = \mathbb{E}\left[\left(P(\boldsymbol{\xi})e^{-j\varphi(\boldsymbol{\xi})}\right) \circledast \left(P(-\boldsymbol{\xi})e^{j\varphi(-\boldsymbol{\xi})}\right)\right]. \tag{4.74}$$

The LE and SE OTFs have analytic expressions [24], which we compare with the results of our simulation in Figure 4.11 which also includes the diffraction OTF. We observe a wide variety of D/r_0 levels that we perform with sufficient accuracy. Some slight deviation may be improved by subharmonic methods such as those described by Schmidt [22].

In addition to what we would call aperture statistics, we can also discuss spatial correlations.



(a) Short-Exposure OTF



(b) Long-Exposure OTF

Figure 4.11. Comparing the long and short exposure OTFs at different distortion levels with the theoretical curves (solid lines) and our statistics (dotted lines).

To begin our comparisons, we first look at the comparison between previously reported angle-of-arrival results [78, 79] with our expression (4.40). We note that the tilt expression in our framework is proportional to the sum of the Zernike tilt coefficient variances. Therefore, we may write the result for angle-of-arrival using our results by summation of the terms corresponding to the tilt Zernike terms within (4.40). We contrast the analytical form of the correlation here to those given in [78, 79]. We observe this to be a nearly identical match between the expressions provided in these two works. These expressions are developed for the case of spherical waves, matching our development.

To compare the two, we can evaluate the angle-of-arrival integral directly using the Python library scipy's integration class. Specifically, we use the triple integration method 'tplquad', writing the angle-of-arrival expression with lambda functions. Therefore, no elements of our C_n^2-slice concept are a component of the angle-of-arrival integral evaluation. For generating the curves predicted by our analysis, we may instead evaluate (4.40). To save on a bit of time, we find the values for the C_n^2-slices by taking $10M$ points and averaging over groups of 10 (thus an approximation of (4.38)). For example, if we require 10 C_n^2-slices, these values are estimated from 100 samples of the C_n^2 profile via local averaging.

The results of the constant C_n^2 profile are presented in Figure 4.12, while the path-varying turbulence profile results can be seen in Figure 4.13. Here we normalize by the isoplanatic angle on the x-axis, given as

$$\theta_0 = 58.1 \times 10^{-3} \lambda^{6/5} \left[\int_0^L z^{5/3} C_n^2(z) dz \right]^{-3/5}. \tag{4.75}$$

We observe a convincing match for all parameters within our tests. That is, the angle-of-arrival integral appears to be similar to the results predicted by this analysis. We further note that all evaluations of (4.40) were performed with the number of C_n^2-slices kept constant at 200. We have noticed some minor improvement with an increase of slices (i.e. our curves match the angle-of-arrival curves more closely), but we find 200 to be sufficient for the purposes of this comparison. We again note that the angle-of-arrival results from Basu et al. (now Bose-Pillai) [78] come from a separate analysis and are evaluated with methods separate from ours.

4.7.2 The Problem with Small Numbers of Phase Screens

When analyzing or simulating propagation through turbulence using the discrete phase screen approach [24], the number of phase screens can be chosen to match various requirements of the application. For analysis which utilizes phase screens, the number of phase screens is typically left to be unspecified so long as assumption of phase screen independence holds. In simulation, 10 phase screens are used in Hardie et al. [56] which we deem to be relatively

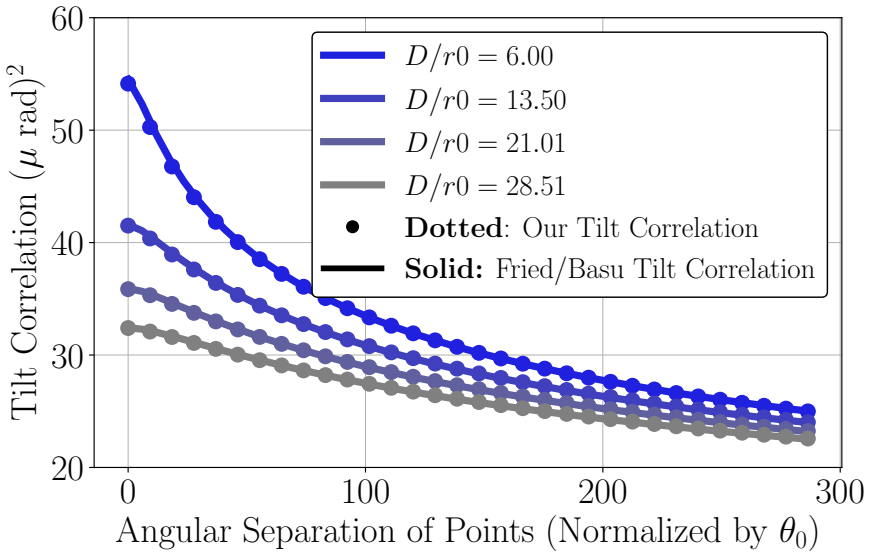


Figure 4.12. Theoretical comparison of our expression for tilt correlation compared to Basu et al. (now Bose-Pillai) [78] for varying aperture diameters. These results are for a constant C_n^2 profile of $C_n^2 = 2 \times 10^{-15}$ m$^{-2/3}$.

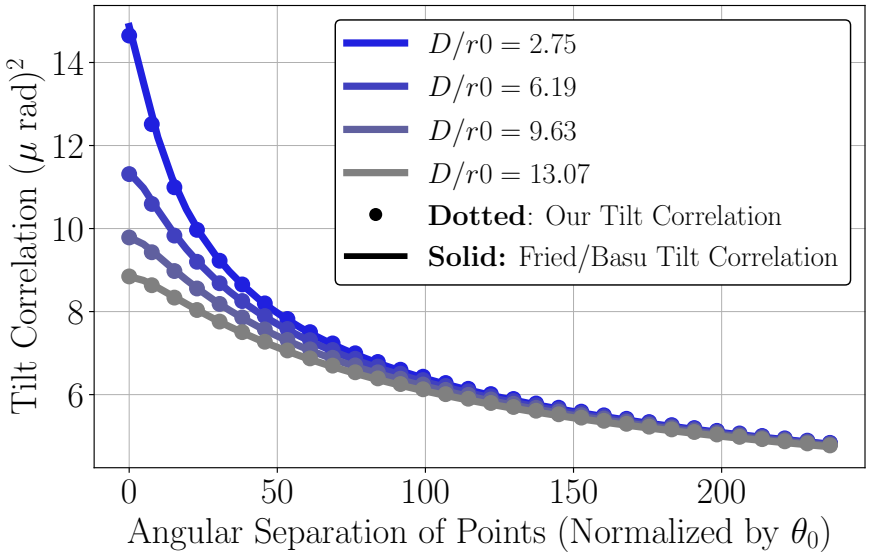


Figure 4.13. Theoretical comparison of the expression for our tilt correlation compared to [78]. The results shown is for a path-varying turbulence profile that is high at the aperture $C_n^2(z) = 2(z/L) \times 10^{-15}$ m$^{-2/3}$ and with varying aperture sizes.

standard, though one may use more/less if the situation dictates. This leads to a difference in analysis vs. simulation: one uses a large number of phase screens whereas in simulation a small number is used.

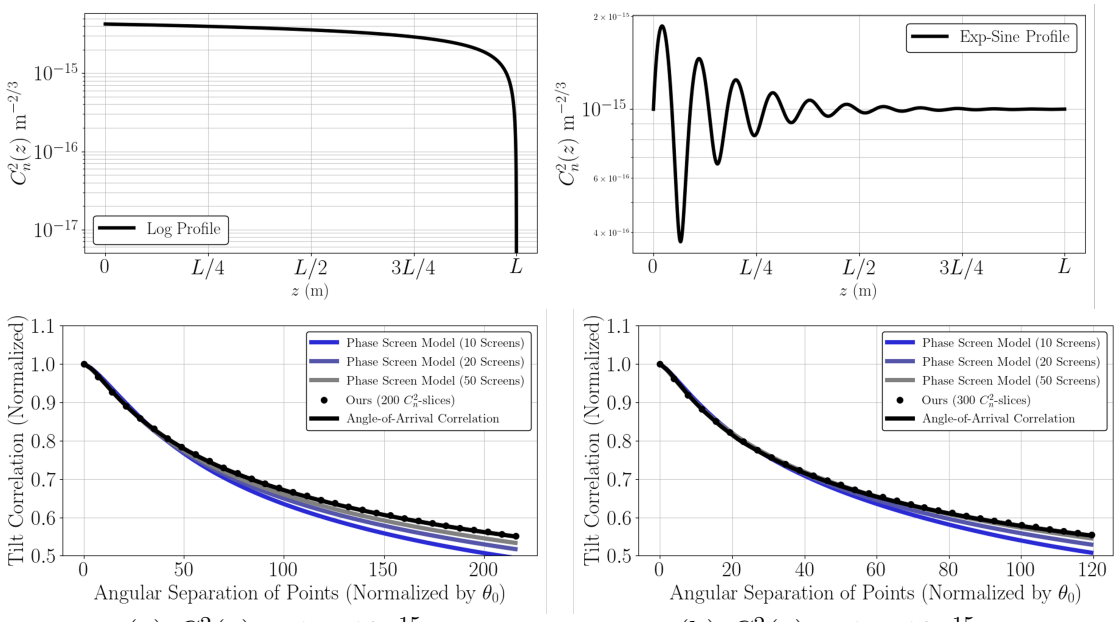
To quantify this difference in analysis and simulation, one aspect we may study is the tilt correlation. We may ask: Does a simulation with a small amount of phase screens match the analytic prediction with a larger amount of independent phase screens? To answer this question without reliance on empiricism, we first note that a properly performed simulation should perform identically to its analytic counterpart. Therefore, we may analyze simulation by a representative analytical model.

The model we choose for our purposes is similar to the C_n^2-slice model, though the values of the Riemann sum terms are chosen to optimize the objective function provided in Hardie et al. [56]. We note that this was done to choose phase screen parameters to closely match isoplanatic angle, Fried parameter, and log amplitude variance in a least-squares fashion. Thus, the model we choose is

$$C_n^2(z) = \sum_{m=1}^{M} \delta\left(z - \frac{Lm}{M+1}\right) \tilde{C}_{n,m}^2, \tag{4.76}$$

where $\tilde{C}_{n,m}^2$ is the optimally chosen C_n^2 value of the mth phase screen (optimized according to Hardie et al. [56]). We will refer to (4.76) as the phase screen model.

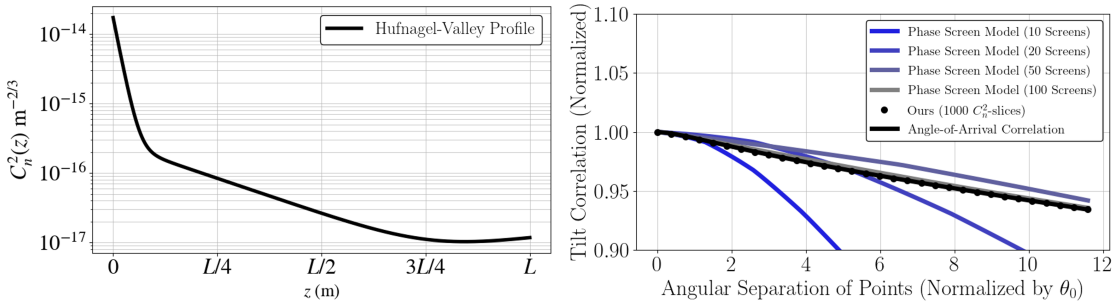
Studying the impact of M on the accuracy of the tilt correlation may be done through using the phase screen model from (4.76) with (4.37) to compute the phase screen correlations shown in Figure 4.14 and Figure 4.15. In these cases, we provide both the angle-of-arrival correlations[78] (via direct scipy integration) for reference along with the case of 200 C_n^2-slices using (4.40). In this case, we observe that if we are interested in a narrow field of propagation (corresponding to a small value on the x-axis of Figure 4.14 and Figure 4.15) then we may use a small number of phase screens to model the situation. However, for situations with more anisoplanatism a larger number of phase screens may be required depending on the desired accuracy.



(a) $C_n^2(z) = 1 \times 10^{-15}$ $\times \log((z - L)/100 + 1)$ m$^{-2/3}$

(b) $C_n^2(z) = 1 \times 10^{-15}$ $\times (e^{-z/1000} \sin(z/100) + 1)$ m$^{-2/3}$

Figure 4.14. [Top] For two different C_n^2 profiles (shown in the log domain) we can plot the [Bottom] tilt correlation for each using (i) angle-of-arrival correlations, (ii) C_n^2-slice correlations, (iii) the phase screen model correlations. We note that the phase screen model's C_n^2 values are optimized as described in Hardie et al.[56]. We can see that with increasing the number of phase screens, the phase screen model approaches the angle-of-arrival and high C_n^2-slice curves.



(a) Hufnagel-Valley Profile

(b) Empirical Correlations

Figure 4.15. [Left] For the Hufnagel-Valley C_n^2 profiles (shown in the log domain) we can plot the [Right] tilt correlation for each using (i) angle-of-arrival correlations, (ii) C_n^2-slice correlations, (iii) the phase screen model correlations. We note that the phase screen model's C_n^2 values are optimized as described in Hardie et al.[56]. We can see that with increasing the number of phase screens, the phase screen model approaches the angle-of-arrival and high C_n^2-slice curves.

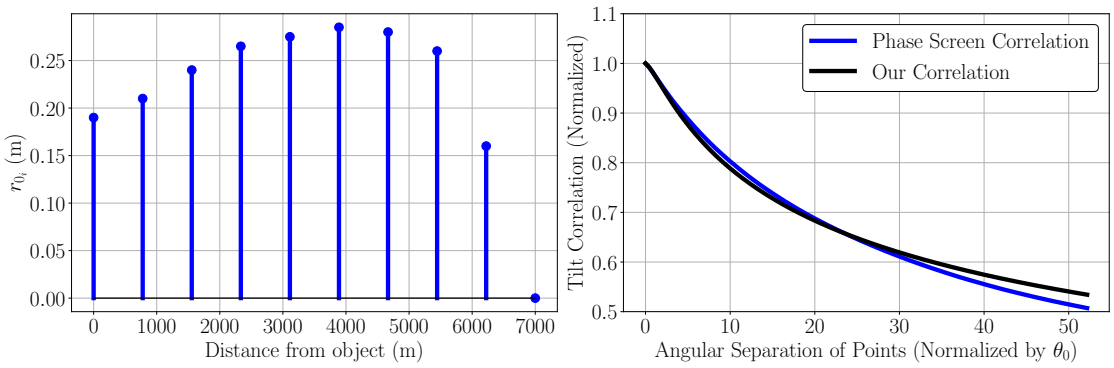


Figure 4.16. [Left] A set of phase screen r_0 values which model the C_n^2 profile, $C_n^2 = 0.25 \times 10^{-15}$ m$^{-2/3}$ as given by Hardie et al.[56]. [Right] The deviation in evaluation of our correlation integral with the phase screen model matches previously reported results[56].

In addition to these more complicated cases, we also use simpler profiles which match those of Hardie et al. [56] as shown in Figure 4.16 and Figure 4.17. We find there to be a match with their reported empirical results when using their reported values for $\tilde{C}_{n,m}^2$ within the phase screen model (4.76). We show these examples for two reasons: (i) we use this as supporting evidence that (4.76) correctly models a properly performed split-step simulation; (ii) that errors in tilt correlation are not a fault of split-step, but rather a function of either the optimization chosen to select the C_n^2 parameters or number of phase screens. With respect to (ii), we note this particularly as stated in Hardie et al. [56] that the mismatch in their reported tilt correlation was to be investigated in future work. Along similar lines as previously, the number of phase screens may be increased in order to match the tilt correlation function more closely.

4.7.3 Speed and Resolution Comparisons

For our runtime comparison, we consider the generation contained in Figure **??**. That is, we want to compare the time to generate only the phase distortions, not the application of the point spread functions. We choose not to include this as this can be replaced in either approach with either the analytic formula for PSF formation or the P2S network. We feel this is the most fair comparison of an atmospheric turbulent simulation tool, as the core goal

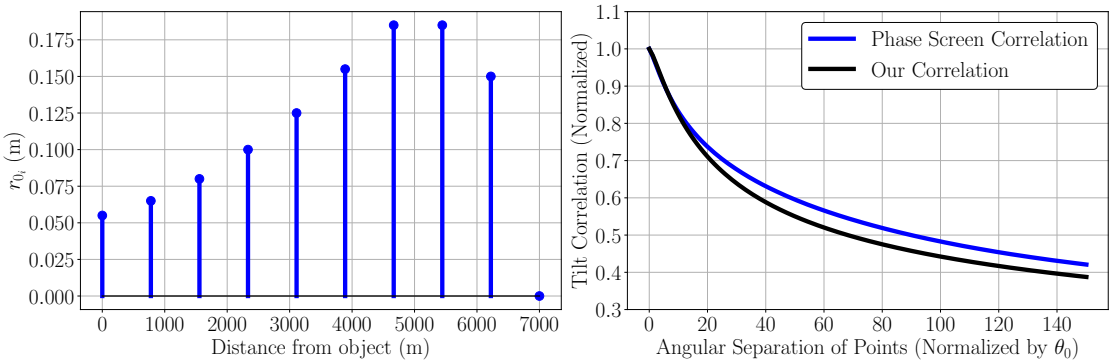
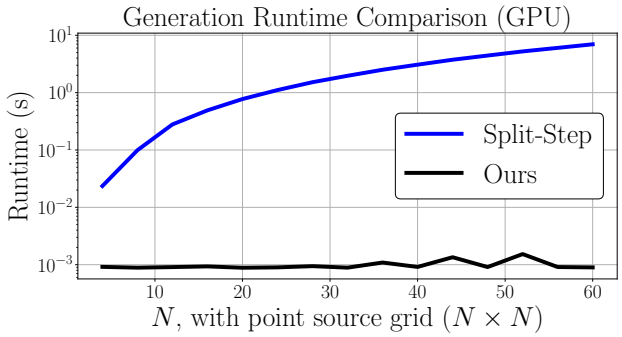
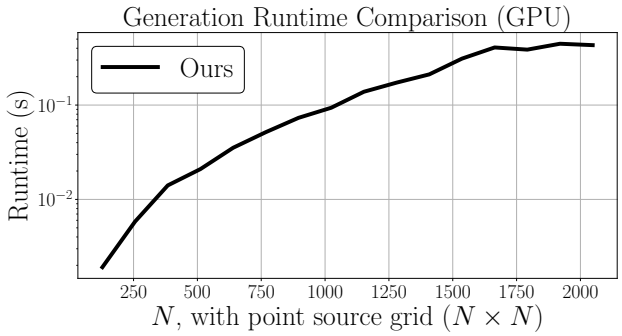


Figure 4.17. [Left] A set of phase screen r_0 values which model the C_n^2 profile, $C_n^2(z) = 2(1 - z/L)$ as given by Hardie et al.[56]. [Right] The deviation in evaluation of our correlation integral with the phase screen model matches previously reported results[56].



(a) Low resolution comparison



(b) High resolution comparison

Figure 4.18. A comparison of the runtimes of split-step [56] and ours. (a) Low resolution. (b) High resolution. Note that split-step cannot handle any high resolution.

is to produce the phase distortions. Therefore, Figure 4.18 (a) reflects the time to generate the turbulent phase distortions at varying resolutions between our approach and split step. We then show additional resolution for our simulation in Figure 4.18 (b) to which split-step is not scalable. Notably, our simulator can generate an image at an image resolution of 512×512 in the same time as approximately a 10×10 image with split-step, which would then typically be upscaled $4\times$, resulting in a 40×40 image.

4.7.4 GUI for Real-Time Simulation

As a result of our simulation being computationally efficient, a real-time generation of turbulent images can be developed. We show a real-time demonstration using a camera and GPU in which we can display the distorted video stream, pixel displacements, and an 8×8 grid of sub-sampled PSFs in Figure 4.19. Given some interfacing with the camera and

calculation of statistics in real-time, this is not performed at the same speed as reported in Figure 4.18. However, we can achieve approximately 8 frames per second (FPS) for a 512×512 image on a nVidia GeForce GTX 1080 Ti GPU with the GUI and additional statistical information displayed. To our knowledge, this is the first simulator in the literature that has the capability of performing a simulation with this level of accuracy at this speed, making this the first real-time demo of turbulent imaging simulation. Additionally, we can modify the imaging geometry and turbulence strength in real-time, for which we have tunable knobs in our GUI.

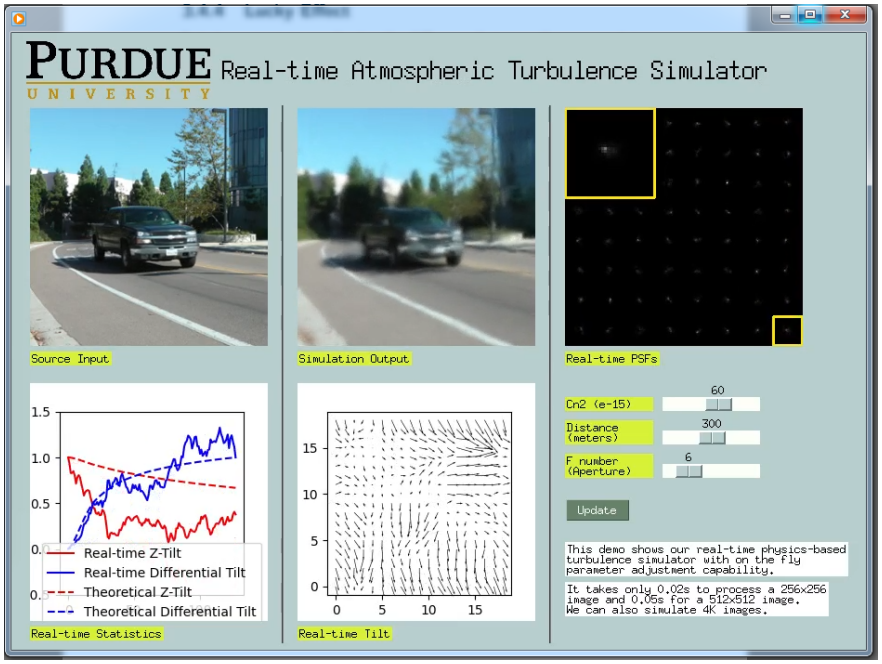


Figure 4.19. Graphics User Interface (GUI). Our GUI can be connected to a standard camera and generate dense-field turbulence effects at 7 fps for a 512×512 image.

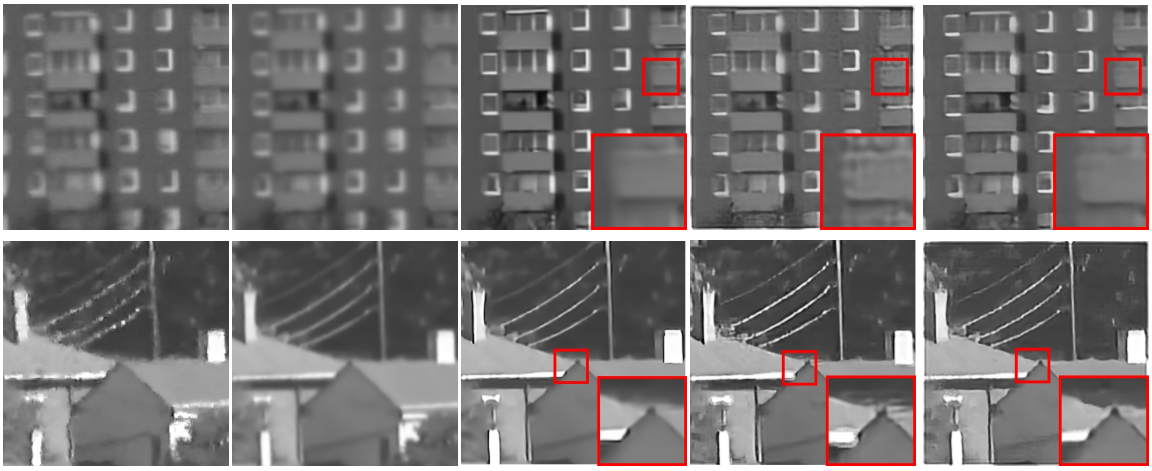
4.8 Impact of a good simulator and summary

This Chapter is the bulk of the novel ideas produces by this thesis. The concept of the Zernike space, deriving their correlations, drawing random realizations and applying them to an image through spatially varying convolution and the P2S are the main focus of this

Chapter. We have also spent a while validating the simulator through various methodologies such as aperture and spatial statistics. Now I wish to make a few parting comments.

4.8.1 Impact on restoration

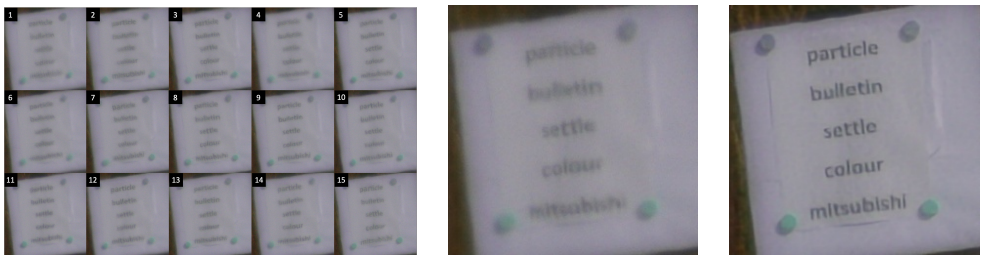
The impact of a good simulator on restoration is undeniable. Over the past few years, we have observed that when using data which matches the real observations by turbulence, we often recieve a better reconstruction. The first demonstration of this for turbulence was by Mao et al. [75] in which a vanilla U-Net was trained on data from a weaker atmospheric turbulence simulator and ours (at that time – version 2!). The result was that the U-Net performed better from our data. We show this in Figure 4.20.



(a) Input (b) Average (c) [80] (d) [81]+U-Net (e) Ours+U-Net

Figure 4.20. Image reconstruction using **real** data (so ground truth is not available). For (d) and (e), we train a U-Net using data synthesized by [81] and our simulator, respectively. Notice the artifacts in (d).

This has further been done using version 3 of our simulator by [82] through application of a GAN method [83] in which two forms were made: one which was trained on an alternative methods data generation and one which was trained on ours. The results of this are presented in Figure 4.21. We do have some newer results utilizing our newest version of the simulator, however, it is still in development.



(a) Input (b) Trained with [83]'s data (c) Trained with [84] data

Figure 4.21. The impact of training data generated by [83] and our method [84]. We can see a clear improvement in quality of reconstruction with the same method as described in [83].

4.8.2 Summary

This Chapter is the bulk of the novel ideas produces by this thesis. The concept of the Zernike space, deriving their correlations, drawing random realizations and applying them to an image through spatially varying convolution and the P2S are the main focus of this Chapter. We have also spent a while validating the simulator through various methodologies such as aperture and spatial statistics.

The reason for the ideas contained in this Chapter do not stem from the standard physics-like motivations in simulating, but instead simulating for restoration. A great deal of time has been devoted to forming reconstruction models, I personally believe a similar amount must be dedicated to forward models, to which I would point to this simulator and other emerging studies which seem to suggest the importance of such a simulation capability.

I would end this Chapter with mentioned that just because a simulator is available does not mean that applying any chosen parameter to it means a good simulation. You must take time in understanding how the parameters are chosen. Doing so may greatly impact the quality of simulation and the generalization to real data. Therefore, one must appreciate the sophistication which goes into simulation to apply it properly

5. CONCLUSION

In this thesis I have presented Zernike-based simulation which speeds up atmospheric turbulence simulation by up to a factor of 1000× as well as maintaining a high degree of accuracy and offering differentiability to be integrated into the training process of a reconstruction network. We have additionally provided an abridged background of imaging through turbulence theory as well as a background on optics for the purposes of making the ideas which Zernike-based simulation rests on to be suitably clear. We have demonstrated that with increased accuracy in training data, we can expect to see better reconstruction results. This also suggests that with our simulator there is a closer match to the statistics of atmospheric degradations than a less sophisticated simulator.

5.1 Future work

Zernike-based simulation was not born overnight, and developed over the course of a few publications [37, 38, 73, 75, 84]. This progression has involved a lot of revisiting and changing to previous ideas which we believed to be the best at the time, only later to find out that we could improve upon it by some trick.

What stands out to me as the current limiting factor(s)? There are two that I would primarily name:

1. **Speed.** As surprising as it may be, the speed is a limiting factor at this point in the simulator. If we wish to use in in-the-loop for reconstruction, then we will need to have its generation time be nearly as fast or ever *faster* than the forward pass of a deep network – the simulator is just not there yet. There are some more engineering type tricks we are investigating to help speed up the simualtor.

2. **Incorporation of a better camera model.** So much of this thesis has rested upon the thin lens model. However, what if we want to introduce a true camera model into the mix? This matters for imaging over a long distance as the camera is typically specialized to see at such a great distance to the point of its characteristics being unique to each camera. Integration of these effects, as we've seen with accurate turbulence

simulation, may help to resolve some of the non-turbulent effects that still linger in the images.

There are also a few other topics which may be of interest to those wishing to extend this result. One is possibly glaringly obvious – the complete discarding of the amplitude effects! Incorporating these effects, to me, will either come in a complete re-do of the Zernike-based theory *or* a very cheap approximation which works for most scenarios. I believe that the latter will serve us well for a time, but the former will likely come by adopting the Feynman path integral approach to modeling wave propagation through the atmosphere.

REFERENCES

[1] J. W. Goodman, *Introduction to Fourier Optics*, Third. Roberts and Company, 2005.

[2] M. Born and E. Wolf, *Principles of Optics; Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Seventh. Cambridge University Press, 1999.

[3] E. Hecht, *Optics*, Fifth. Pearson Education, Inc., 2015.

[4] D. Griffiths, *Introduction to Electrodynamics*. Pearson Education, 2014.

[5] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge books online). Cambridge University Press, 2003, ISBN: 9780521540513.

[6] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000. DOI: 10.1109/34.888718.

[7] D. P. Feder, "Automatic lens design methods∗," *J. Opt. Soc. Am.*, vol. 47, no. 10, pp. 902–912, Oct. 1957. DOI: 10.1364/JOSA.47.000902. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josa-47-10-902.

[8] D. P. Feder, "Calculation of an optical merit function and its derivatives with respect to the system parameters∗," *J. Opt. Soc. Am.*, vol. 47, no. 10, pp. 913–925, Oct. 1957. DOI: 10.1364/JOSA.47.000913. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josa-47-10-913.

[9] D. P. Feder, "Differentiation of ray-tracing equations with respect to construction parameters of rotationally symmetric optics," *J. Opt. Soc. Am.*, vol. 58, no. 11, pp. 1494–1505, Nov. 1968. DOI: 10.1364/JOSA.58.001494. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josa-58-11-1494.

[10] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*. Cambridge University Press, 1995, ISBN: 9781139643023. [Online]. Available: https://books.google.com/books?id=dVYhAwAAQBAJ.

[11] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals & Systems*. USA: Prentice-Hall, Inc., 1996, ISBN: 0138147574.

[12] C. Bouman, *Foundations of Computational Imaging: A Model-Based Approach* (Other Titles in Applied Mathematics). Society for Industrial and Applied Mathematics, 2022, ISBN: 9781611977134. [Online]. Available: https://books.google.com/books?id=TiJ6EAAAQBAJ.

[13] S. Chan, *Introduction to Probability for Data Science*. Michigan Publishing, 2021, ISBN: 9781607857464. [Online]. Available: https://books.google.com/books?id=GR2jzgEACAAJ.

[14] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (Electrical Engineering Series). McGraw-Hill, 1991, ISBN: 9780070484771. [Online]. Available: https://books.google.com/books?id=4IwQAQAAIAAJ.

[15] A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers," *Doklady Akademii Nauk SSSR*, vol. 30, pp. 301–305, 1941.

[16] A. N. Kolmogorov, "Dissipation of energy in the locally isotropic turbulence," *Doklady Akademii Nauk SSSR*, vol. 32, pp. 16–18, 1941.

[17] U. Frisch, *Turbulence: The Legacy of A. N. Kolmogorov*. Cambridge University Press, 1995. DOI: 10.1017/CBO9781139170666.

[18] V. I. Tatarskii, *Wave Propagation in a Turbulent Medium*. New York: Dover Publications, 1961.

[19] J. W. Goodman, *Statistical Optics*, Second. John Wiley and Sons Inc., 2015.

[20] O. Korotkova, *Random Light Beams*. CRC Press, 2017.

[21] R. Sasiela, *Electromagnetic Wave Propagation in Turbulence: Evaluation and Application of Mellin Transforms* (Press Monographs). SPIE, 2007, ISBN: 9780819467287. [Online]. Available: https://books.google.com/books?id=tBxRAAAAMAAJ.

[22] J. D. Schmidt, *Numerical simulation of optical wave propagation: With examples in MATLAB*. SPIE Press, 2010, pp. 1–197.

[23] J. W. Goodman, *Speckle Phenomena in Optics: Theory and Applications*. Roberts & Company, 2007, ISBN: 9780974707792.

[24] M. C. Roggemann and B. M. Welsh, *Imaging through Atmospheric Turbulence*. Taylor & Francis, 1996.

[25] D. Fried, "Optical heterodyne detection of an atmospherically distorted signal wave front," *Proceedings of the IEEE*, vol. 55, no. 1, pp. 57–77, 1967. DOI: 10.1109/PROC.1967.5377.

[26] D. L. Fried, "Optical resolution through a randomly inhomogeneous medium for very long and very short exposures," *Journal of the Optical Society of America*, vol. 56, no. 10, pp. 1372–1379, 1966.

[27] A. Ishimaru, *Wave propagation and scattering in random media*. Academic press New York, 1978, vol. 2.

[28] R. P. Feynman and A. R. Hibbs, *Quantum mechanics and path integrals* (International series in pure and applied physics). McGraw-Hill, 1965. [Online]. Available: https://cds.cern.ch/record/100771.

[29] V. U. Zavorotnyi, V. I. Kliatskin, and V. I. Tatarskii, "Strong fluctuations of electromagnetic-wave intensity in randomly inhomogeneous media," *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, vol. 73, pp. 481–497, Aug. 1977.

[30] V. Tatarskii and V. Zavorotnyi, "Strong fluctuations in light propagation in a randomly inhomogeneous medium," *Progress in Optics*, vol. 18, E. Wolf, Ed., pp. 204–256, 1980, ISSN: 0079-6638. DOI: https://doi.org/10.1016/S0079-6638(08)70214-2. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0079663808702142.

[31] V. Tatarskii, A. Ishimaru, and V. Zavorotny, *Wave Propagation in Random Media (Scintillation): Proceedings of the Conference Held 3-7 August, 1992 at the University of Washington, USA.* (SPIE press monographs). Taylor & Francis, 1993, ISBN: 9780819410627. [Online]. Available: https://books.google.com/books?id=4f5AAQAAIAAJ.

[32] M. Charnotskii, J. Gozani, V. Tatarskii, and V. Zavorotny, "Wave propagation theories in random media based on the path-integral approach," *Progress in Optics*, vol. 32, E. Wolf, Ed., pp. 203–266, 1993, ISSN: 0079-6638. DOI: https://doi.org/10.1016/S0079-6638(08)70164-1. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0079663808701641.

[33] R. Dashen, "Path integrals for waves in random media," *Journal of Mathematical Physics*, vol. 20, no. 5, pp. 894–920, May 1979. DOI: 10.1063/1.524138.

[34] D. L. Fried, "Statistics of a geometric representation of wavefront distortion," *Journal of the Optical Society of America*, vol. 55, no. 11, pp. 1427–1435, Nov. 1965.

[35] M. C. Roggemann, B. M. Welsh, D. Montera, and T. A. Rhoadarmer, "Method for simulating atmospheric turbulence phase effects for multiple time slices and anisoplanatic conditions," *Applied Optics*, vol. 34, no. 20, pp. 4037–4051, Jul. 1995. DOI: 10.1364/AO.34.004037. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-34-20-4037.

[36] G. Heidbreder, "Image degradation with random wavefront tilt compensation," *IEEE Transactions on Antennas and Propagation*, vol. 15, no. 1, pp. 90–98, 1967. DOI: 10.1109/TAP.1967.1138846.

[37] N. Chimitt and S. H. Chan, "Simulating anisoplanatic turbulence by sampling intermodal and spatially correlated Zernike coefficients," *Optical Engineering*, vol. 59, no. 8, p. 083 101, 2020.

[38] N. Chimitt, X. Zhang, Y. Chi, and S. H. Chan, *Scattering and gathering for spatially varying blurs*, Available online: https://arxiv.org/abs/2303.05687., 2023.

[39] D. Voelz, *Computational Fourier Optics: A MATLAB Tutorial*. SPIE Press, 2011. [Online]. Available: https://books.google.com/books?id=fd9ZnQAACAAJ.

[40] C. J. Pellizzari, M. T. Banet, M. F. Spencer, and C. A. Bouman, "Demonstration of single-shot digital holography using a bayesian framework," *Journal of the Optical Society of America A*, vol. 35, no. 1, pp. 103–107, Jan. 2018. DOI: 10.1364/JOSAA.35.000103. [Online]. Available: http://opg.optica.org/josaa/abstract.cfm?URI=josaa-35-1-103.

[41] F. Wang, I. Toselli, and O. Korotkova, "Two spatial light modulator system for laboratory simulation of random beam propagation in random media," *Applied Optics*, vol. 55, no. 5, pp. 1112–1117, Feb. 2016. DOI: 10.1364/AO.55.001112. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-55-5-1112.

[42] B. Rodenburg *et al.*, "Simulating thick atmospheric turbulence in the lab with application to orbital angular momentum communication," *New Journal of Physics*, vol. 16, no. 3, p. 033 020, Mar. 2014. DOI: 10.1088/1367-2630/16/3/033020. [Online]. Available: https://doi.org/10.1088/1367-2630/16/3/033020.

[43] J. D. Phillips, M. E. Goda, and J. Schmidt, "Atmospheric turbulence simulation using liquid crystal spatial light modulators," in *Advanced Wavefront Control: Methods, Devices, and Applications III*, International Society for Optics and Photonics, vol. 5894, SPIE, 2005, p. 589 406. DOI: 10.1117/12.620407. [Online]. Available: https://doi.org/10.1117/12.620407.

[44] M. A. Vorontsov and G. W. Carhart, "Anisoplanatic imaging through turbulent media: Image recovery by local information fusion from a set of short-exposure images," *Journal of the Optical Society of America A*, vol. 18, no. 6, pp. 1312–1324, Jun. 2001. DOI: 10.1364/JOSAA.18.001312. [Online]. Available: http://opg.optica.org/josaa/abstract.cfm?URI=josaa-18-6-1312.

[45] Y. Tian, S. G. Narasimhan, and A. J. Vannevel, "Depth from optical turbulence," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 246–253. DOI: 10.1109/CVPR.2012.6247682.

[46] C. J. Pellizzari, M. F. Spencer, and C. A. Bouman, "Imaging through distributed-volume aberrations using single-shot digital holography," *Journal of the Optical Society of America A*, vol. 36, no. 2, A20–A33, Feb. 2019. DOI: 10.1364/JOSAA.36.000A20. [Online]. Available: http://opg.optica.org/josaa/abstract.cfm?URI=josaa-36-2-A20.

[47] C. Macaskill and T. E. Ewart, "Computer Simulation of Two-dimensional Random Wave Propagation," *IMA Journal of Applied Mathematics*, vol. 33, no. 1, pp. 1–15, Jul. 1984, ISSN: 0272-4960. DOI: 10.1093/imamat/33.1.1. eprint: https://academic.oup.com/imamat/article-pdf/33/1/1/2262445/33-1-1.pdf. [Online]. Available: https://doi.org/10.1093/imamat/33.1.1.

[48] J. P. Bos, M. C. Roggemann, and V. S. R. Gudimetla, "Anisotropic non-kolmogorov turbulence phase screens with variable orientation," *Applied Optics*, vol. 54, no. 8, pp. 2039–2045, Mar. 2015. DOI: 10.1364/AO.54.002039. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-54-8-2039.

[49] R. Frehlich, "Simulation of laser propagation in a turbulent atmosphere," *Applied Optics*, vol. 39, no. 3, pp. 393–397, Jan. 2000. DOI: 10.1364/AO.39.000393. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-39-3-393.

[50] W. P. Burckel and R. N. Gray, "Turbulence phase screens based on polar-logarithmic spectral sampling," *Applied Optics*, vol. 52, no. 19, pp. 4672–4680, Jul. 2013. DOI: 10.1364/AO.52.004672. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-52-19-4672.

[51] B. J. Herman and L. A. Strugala, "Method for inclusion of low-frequency contributions in numerical representation of atmospheric turbulence," in *Propagation of High-Energy Laser Beams Through the Earth's Atmosphere*, Proc. SPIE 1221, 1990. DOI: 10.1117/12.18342. [Online]. Available: https://doi.org/10.1117/12.18342.

[52] B. M. Welsh, "Fourier-series-based atmospheric phase screen generator for simulating anisoplanatic geometries and temporal evolution," in *Propagation and Imaging through the Atmosphere*, Proc. SPIE 3125, 1997. DOI: 10.1117/12.279029. [Online]. Available: https://doi.org/10.1117/12.279029.

[53] D. L. Fried and T. Clark, "Extruding kolmogorov-type phase screen ribbons," *Journal of the Optical Society of America A*, vol. 25, no. 2, pp. 463–468, Feb. 2008. DOI: 10.1364/JOSAA.25.000463. [Online]. Available: http://opg.optica.org/josaa/abstract.cfm?URI=josaa-25-2-463.

[54] R. G. Lane, A. Glindemann, and J. C. Dainty, "Simulation of a Kolmogorov phase screen," *Waves in Random Media*, vol. 2, no. 3, pp. 209–224, 1992. DOI: 10.1088/0959-7174/2/3/003.

[55] E. Anzuola and S. Gladysz, "Modeling dynamic atmospheric turbulence using temporal spectra and KarhunenLoève decomposition," *Optical Engineering*, vol. 56, no. 7, p. 071 508, 2017. DOI: 10.1117/1.OE.56.7.071508. [Online]. Available: https://doi.org/10.1117/1.OE.56.7.071508.

[56] R. C. Hardie, J. D. Power, D. A. LeMaster, D. R. Droege, S. Gladysz, and S. Bose-Pillai, "Simulation of anisoplanatic imaging through optical turbulence using numerical wave propagation with new validation analysis," *Optical Engineering*, vol. 56, no. 7, p. 071 502, 2017.

[57] J. P. Bos and M. C. Roggemann, "Technique for simulating anisoplanatic image formation over long horizontal paths," *Optical Engineering*, vol. 51, no. 10, p. 101 704, 2012.

[58] R. C. Hardie, M. A. Rucci, S. R. Bose-Pillai, R. V. Hook, and B. K. Karch, "Modeling and simulation of multispectral imaging through anisoplanatic atmospheric optical turbulence," *Optical Engineering*, vol. 61, no. 9, p. 093 102, 2022. DOI: 10.1117/1.OE.61.9.093102. [Online]. Available: https://doi.org/10.1117/1.OE.61.9.093102.

[59] M. A. Vorontsov and V. Kolosov, "Target-in-the-loop beam control: Basic considerations for analysis and wave-front sensing," *Journal of the Optical Society of America A*, vol. 22, no. 1, pp. 126–141, Jan. 2005. DOI: 10.1364/JOSAA.22.000126. [Online]. Available: http://opg.optica.org/josaa/abstract.cfm?URI=josaa-22-1-126.

[60] S. L. Lachinova, M. A. Vorontsov, V. V. Dudorov, V. V. Kolosov, and M. T. Valley, "Anisoplanatic imaging through atmospheric turbulence: brightness function approach," in *Atmospheric Optics: Models, Measurements, and Target-in-the-Loop Propagation*, Proc. SPIE 6708, 2007. DOI: 10.1117/12.738796. [Online]. Available: https://doi.org/10.1117/12.738796.

[61] S. L. Lachinova, M. A. Vorontsov, G. A. Filimonov, D. A. LeMaster, and M. E. Trippel, "Comparative analysis of numerical simulation techniques for incoherent imaging of extended objects through atmospheric turbulence," *Optical Engineering*, vol. 56, no. 7, 2017. DOI: 10.1117/1.OE.56.7.071509. [Online]. Available: https://doi.org/10.1117/1.OE.56.7.071509.

[62] E. Repasi and R. Weiss, "Analysis of image distortions by atmospheric turbulence and computer simulation of turbulence effects," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XIX*, Proc. SPIE 6941, 2008. DOI: 10.1117/12.775600. [Online]. Available: https://doi.org/10.1117/12.775600.

[63] E. Repasi and R. Weiss, "Computer simulation of image degradations by atmospheric turbulence for horizontal views," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXII*, Proc. SPIE 8014, 2011. DOI: 10.1117/12.883805. [Online]. Available: https://doi.org/10.1117/12.883805.

[64] K. R. Leonard, J. Howe, and D. E. Oxford, "Simulation of atmospheric turbulence effects and mitigation algorithms on stand-off automatic facial recognition," in *Optics and Photonics for Counterterrorism, Crime Fighting, and Defence VIII*, Proc. SPIE 8546, Oct. 2012.

[65] G. Potvin, J. L. Forand, and D. Dion, "A simple physical model for simulating turbulent imaging," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXII*, G. C. Holst and K. A. Krapels, Eds., International Society for Optics and Photonics, vol. 8014, SPIE, 2011, 80140Y. DOI: 10.1117/12.884989. [Online]. Available: https://doi.org/10.1117/12.884989.

[66] D. H. Tofsted, D. Quintis, S. G. O'Brien, J. Yarbrough, M. D. Bustillos, and G. M. T. Vaucher, "Test report on the november 2005 nato rtg-40 active imager land field trials," Army Research Lab, Tech. Rep., 2006.

[67] D. Tofsted, S. O'Brien, J. Yarbrough, D. Quintis, and M. Bustillos, "Characterization of atmospheric turbulence during the NATO RTG-40 land field trials," in *Atmospheric Propagation IV*, Proc. SPIE 6551, 2007. DOI: 10.1117/12.720696. [Online]. Available: https://doi.org/10.1117/12.720696.

[68] K. J. Miller, B. Preece, T. W. D. Bosq, and K. R. Leonard, "A data-constrained algorithm for the emulation of long-range turbulence-degraded video," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXX*, G. C. Holst and K. A. Krapels, Eds., International Society for Optics and Photonics, vol. 11001, SPIE, 2019, 110010J. DOI: 10.1117/12.2519069. [Online]. Available: https://doi.org/10.1117/12.2519069.

[69] K. J. Miller and T. D. Bosq, "A machine learning approach to improving quality of atmospheric turbulence simulation," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXXII*, Proc. SPIE 11740, 2021, 117400N. DOI: 10.1117/12.2587749. [Online]. Available: https://doi.org/10.1117/12.2587749.

[70] R. J. Noll, "Zernike polynomials and atmospheric turbulence," *Journal of the Optical Society of America*, vol. 66, no. 3, pp. 207–211, Mar. 1976.

[71] N. Chimitt and S. H. Chan, "Simulating anisoplanatic turbulence by sampling correlated zernike coefficients," in *2020 IEEE International Conference on Computational Photography (ICCP)*, 2020, pp. 1–12. DOI: 10.1109/ICCP48838.2020.9105270.

[72] N. Takato and I. Yamaguchi, "Spatial correlation of Zernike phase-expansion coefficients for atmospheric turbulence with finite outer scale," *Journal of the Optical Society of America A*, vol. 12, no. 5, pp. 958–963, May 1995.

[73] N. Chimitt and S. Chan, "Anisoplanatic optical turbulence simulation for near-continuous Cn2 profiles without wave propagation," *Optical Engineering*, vol. 62, no. 7, p. 078 103, 2023. DOI: 10.1117/1.OE.62.7.078103.

[74] G. A. Chanan, "Calculation of wave-front tilt correlations associated with atmospheric turbulence," *Journal of the Optical Society of America A*, vol. 9, no. 2, pp. 298–301, Feb. 1992.

[75] Z. Mao, N. Chimitt, and S. H. Chan, "Accelerating atmospheric turbulence simulation via learned phase-to-space transform," in *IEEE/CVF International Conference on Computer Vision*, Oct. 2021, pp. 14 759–14 768.

[76] M. R. Whiteley, M. C. Roggemann, and B. M. Welsh, "Temporal properties of the zernike expansion coefficients of turbulence-induced phase aberrations for aperture and source motion," *Journal of the Optical Society of America A*, vol. 15, no. 4, pp. 993–1005, Apr. 1998. DOI: 10.1364/JOSAA.15.000993. [Online]. Available: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-15-4-993.

[77] S. H. Chan, "Tilt-then-blur or blur-then-tilt? clarifying the atmospheric turbulence model," *IEEE Signal Processing Letters*, vol. 29, pp. 1833–1837, 2022. DOI: 10.1109/LSP.2022.3200551.

[78] S. Basu, J. E. McCrae, and S. T. Fiorino, "Estimation of the path averaged atmospheric refractive index structure constant from time lapse imagery," in *Laser Radar Technology and Applications XX; and Atmospheric Propagation XII*, Proc. SPIE 9465, May 2015, 94650T.

[79] D. L. Fried, "Varieties Of Isoplanatism," in *Imaging Through the Atmosphere*, Proc. SPIE 0075, 1976. DOI: 10.1117/12.954733. [Online]. Available: https://doi.org/10.1117/12.954733.

[80] Z. Mao, N. Chimitt, and S. H. Chan, "Image reconstruction of static and dynamic scenes through anisoplanatic turbulence," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1415–1428, Oct. 2020.

[81] C. P. Lau and L. M. Lui, "Subsampled turbulence removal network," *Mathematics, Computation and Geometry of Data*, vol. 1, no. 1, pp. 1–33, 2021.

[82] X. Zhang, Z. Mao, N. Chimitt, and S. H. Chan, *Imaging through the atmosphere using turbulence mitigation transformer*, Available online: https://arxiv.org/abs/2207.06465.

[83] D. Jin *et al.*, "Neutralizing the impact of atmospheric turbulence on complex scene imaging via deep learning," *Nature Machine Intelligence*, vol. 3, pp. 876–884, 2021.

[84] N. Chimitt, X. Zhang, Z. Mao, and S. H. Chan, "Real-time dense field phase-to-space simulation of imaging through atmospheric turbulence," *IEEE Transactions on Computational Imaging*, 2022.

INDEX