WAVEFRONT CURVATURE SENSING FROM IMAGE PROJECTIONS

DISSERTATION

Jonathan C. Buffington, Major, USAF

AFIT/DS/ENG/07-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright Patterson Air Force Base, Ohio

Distribution Unlimited

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT/DS/ENG/07-01

WAVEFRONT CURVATURE SENSING FROM IMAGE PROJECTIONS

DISSERTATION

Presented to the Faculty
Graduate School of Engineering and Management
Air Force Institute of Technology
Air University
Air Education and Training Command
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

Jonathan C. Buffington, B.S., M.S.
Major, USAF

September, 2006

Distribution Unlimited

AFIT/DS/ENG/07-01

WAVEFRONT CURVATURE SENSING FROM IMAGE PROJECTIONS

Jonathan C. Buffington, B.S., M.S.
Major, USAF

Approved:

LtCol Matthew E. Goda, PhD (Chairman)	Date
Dr. Aihua W. Wood (Dean's Representative)	Date
Dr. Stephen C. Cain (Member)	Date
Dr. William P. Baker (Member)	Date

Accepted:

M. U. Thomas Dean, Graduate School of Engineering and Management	Date

AFIT/DS/ENG/07-01

Abstract

This research outlines the development and simulation of a signal processing approach to real time wavefront curvature sensing in adaptive optics. The signal processing approach combines vectorized Charge Coupled Device (CCD) read out with a wavefront modal estimation technique. The wavefront sensing algorithm analyzes vector projections of image intensity data to provide an estimate of the wavefront phase as a combination of several low order Zernike polynomial modes. This wavefront sensor design expands on an existing idea for vector based tilt sensing by providing the ability to compensate for additional modes. Under the proposed wavefront sensing approach, the physical wavefront sensor would be replaced by a pair of imaging devices capable of generating vector projections of the image data. Using image projections versus two-dimensional image data allows for faster CCD read out and decreased read noise.

The primary research contribution is to create an effective method for estimating low order wavefront modes from image vector information. This dissertation provides simulation results and Cramér-Rao performance bounds for two wavefront sensor designs. The first sensor provides estimates of tilt and defocus: Zernike polynomials 2 through 4. The second sensor estimates Zernike polynomials 2 through 10. Sensors are simulated in guide star applications under the influence of von Kármán atmospheric phase aberrations and CCD noise models. Secondary research contributions include identifying key algorithm performance parameters, and parameter sensitivity as well as an investigation of strategies for improving extensible phase screen generation.

A simulated performance comparison is conducted between the Z_{2-4} and the Z_{2-10} sensors, and a centroiding tilt sensor and a projection based maximum likelihood tilt sensor. Simulation trials using a subaperture diameter of 0.07m stepped through values of r_0 from 0.04 to 0.14m and average photon counts of 100 to 1000. The Z_{2-4} sensor provides superior performance over both tilt sensors in all trials conducted. The Z_{2-10} sensor outperforms both tilt sensors when the average photon count is greater than 200 photons, and performance on par with both tilt sensors when the average photon count is between 100 and 200 photons.

Acknowledgements

First and foremost, I wish to express my sincerest gratitude toward my wife for her patience with academic life and for her outspoken belief in my abilities. Your confidence became my confidence and cause for perseverance and for that I owe the largest portion of my success to you.

I would like to express my appreciation to LtCol Goda for his tireless efforts as my research advisor. Thank you for asking the difficult questions and keeping me on the right track to complete this research. Thank you Dr. Cain for sparking my interest in estimation theory and wavefront sensing and for leading me down the path of projection based wavefront estimation. Your insightful advice and uplifting commentary over the last three years has made all the difference. Thank you Dr. Baker for your mathematical insight and for convincing me to push the envelop. It has been my honor to have you as a part of my Ph.D. committee. To each of my committee members: I hope to have the opportunity to work with you again in the future.

Finally, I would like to express my gratitude to my parents, family members, and friends who remembered my plight in their prayers. Without you all, this endeavor surely would have ended fruitless.

Jonathan C. Buffington

Table of Contents

List of Figures

List of Tables

WAVEFRONT CURVATURE SENSING FROM IMAGE PROJECTIONS

1. *Introduction*

This work includes the derivation and simulated performance of a fast, efficient algorithm for real time wavefront curvature sensing. Real time wavefront sensing falls into two categories: interferometric measurement of phase or phase slope, and estimation of the phase from image intensity characteristics. The proposed wavefront sensing method falls into the latter category. Phase estimators may be further distinguished by the number and the order of the aberrations, or modes, they estimate. The category "tilt sensors," for instance, is often used in reference to linear mode estimators. Estimators that provide phase information beyond the linear tilt modes may be referred to as "curvature sensors." Linear modes are primarily comprised of low frequency phase characteristics. Linear mode estimators are most accurate over small regions of the wavefront; and consequently, tilt estimators must use highly parallel systems with many small subapertures to provide a global wavefront map. Curvature sensors estimate higher frequency modes and are generally effective over larger regions of the wavefront. This research will outline a fast, effective tilt sensing technique [1] and extend the technique to include higher order parameter estimation. The following sections will define common methods of wavefront sensing and identify the motivation behind wavefront sensing devices.

1.1 The Random Atmosphere

The degree to which two point sources will be resolved by an imaging device in free space will be limited by diffraction effects directly tied to the size of the aperture, D, and the wavelength, λ [2]. This is because the width of a point source in the image plane is essentially the width of the central spot of a circular diffraction pattern, commonly referred to as a Rayleigh distance:

$$\text{Rayleigh distance} = 1.22\frac{\lambda f}{D}, \tag{1.1}$$

where λ is the wavelength of the source and f is the geometric focal length. The size of a Rayleigh distance is inversely proportional to the aperture diameter indicating that large apertures will yield better resolving power. This effect is shown in Figure 1.1 demonstrating overlapping diffraction patterns from a circular aperture.

Figure 1.1 Simulated Airy's disks for two different aperture diameters. The aperture used for the images on the right side is double the diameter of the aperture used to form the images on the left side.

Optical imaging systems designed to resolve objects through the earth's atmosphere must contend with the degrading effects of its continuously fluctuating index of refraction. This condition is commonly referred to as atmospheric turbulence. Newton, though not convinced of the wave theory of light, was aware of diffraction effects and the benefits of a large aperture on resolution. He was also aware of the added limitations of imaging through atmospheric turbulence [3].

> "Long Telescopes may cause Objects to appear brighter and larger than the short ones can do, but they cannot be so formed as to take away the confusion of the Rays which arises from the Tremors of the Atmosphere [4]."

When speaking of the "confusion of the Rays," Newton was describing the effects of the atmosphere's fluctuating index of refraction. Due to wind and temperature gradients, the atmosphere churns and tumbles as it flows over the Earth. The turbulence contains continuously evolving temperature and pressure variations. Since temperature and pressure relate directly to the index of refraction, the index of refraction varies continuously as well [5]. Imagine that a column of atmosphere is divided into many discrete segments each with a different index of refraction. Snell's law in ray optics predicts that the path of a single ray will bend as it transitions through each of these segments. Hence, propagation through atmospheric turbulence gives rise to random variations in the optical path. Unlike in free space, a ray's path through turbulence will not propagate in a straight line, but will wander slightly as suggested by the top diagram in Figure 1.2. Combine this random wander in the ray path with the diffraction pattern for a point source and the result is a spot image that wanders around in the image plane as the atmosphere evolves. The imaging system will apply some integrated exposure to these random spot movements and, consequently, the image becomes a broadened diffraction pattern. The amount of broadening is related to the turbulence in the optical path. For ground to space seeing conditions, these atmospheric effects become the dominant contributor to resolving power when the aperture size is larger than a few centimeters. Technology continues to offer inventive ways to counter these atmospheric effects. Today's most powerful terrestrial telescopes "sense" the conditions of the atmosphere and react to improve seeing conditions. The sensing capability relies on the concept of an optical wavefront which contains a measure of the atmospheric effects. This dissertation will review the concepts necessary for a basic understanding of these atmospheric phenomenon. The background will provide the foundation for developing an improved method for wavefront sensing.

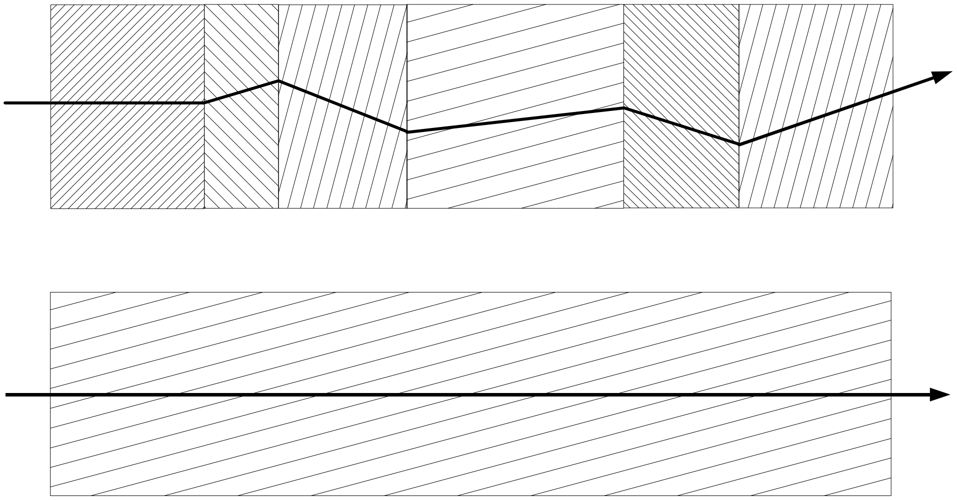


Figure 1.2 Top: Ray propagating through a layered medium with multiple indexes of refraction. Bottom: Ray propagating through a medium with a single index of refraction.

1.2 The Wavefront

Since this work is concerned with wavefront sensing, it is necessary to develop the concept of a wavefront or phasefront. For the purpose of this work, the wavefront is defined as the difference between some reference field, predicted by free space propagation, and the actual field in the aperture of an imaging system. For a point source object, this reference field has a simple geometric formulation. Consider optical energy emanating from a point source. The wave propagation is equal in all directions. The sphere of radius R with center located at the point source represents a surface of constant amplitude and phase. Now, imagine that the point source is far away and R becomes very large. If R is very large then the small portion of the spherical wavefront interfacing with the imaging system is planar to close approximation. In many circumstances, light from a distant source is accurately modeled as a plane wave (planar wavefront) over the optical system aperture. In some instances it is more practical to discuss the effects of an optical system after attempting to focus the planar wavefront. In these cases, it may be more appropriate to use a spherical reference wavefront. For instance, consider the effects of an imperfect optical system on a planar wavefront. The difference between the focused wavefront from a perfect spherical wave reveals the imperfections in the system. This situation is demonstrated in Figure 1.3. Whatever the form of the reference wavefront, wavefront sensors are designed to measure the difference between the incoming wavefront that reference. The turbulent atmosphere will acts as an aberrating thick lens warping the wavefront as it propagates. Figure 1.5

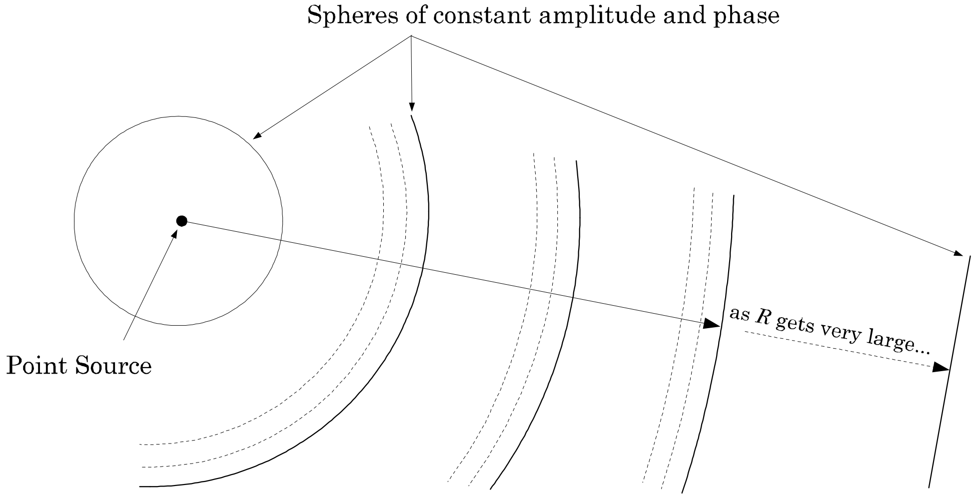


Figure 1.3 Graphical depiction of a sperical wavefront emanating from a point source.

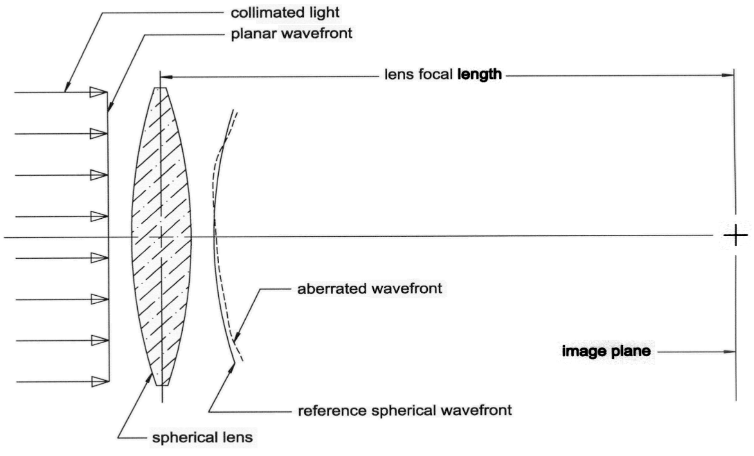


Figure 1.4 Diagram shows a reference planar wavefront passing through an imperfect sperical lens. Aberrations in the lens system can be described by comparing the outgoing aberrated wavefront to a reference spherical wavefront.

demonstrates the effects of a large volume of atmosphere on a plane wave. A sensor capable of measuring the amount of distortion created by the atmosphere would enable an optical system to sense atmospheric effects and, given a reactionary capability, somehow compensate for these effects.

1.3 Adaptive Optics

An adaptive optics system employs a wavefront sensor in a feedback path. The wavefront sensor provides an error measure to some system of actively controlled optics. The controllable optics are then capable of compensating for the wavefront error. Adaptive optics systems may be used to improve performance of imaging systems or laser propagation systems. Although the purpose of the two types of systems is dramatically different, the feedback and control mechanisms used to increase performance are remarkably similar. The technology for such systems has been evolving since conception in the 1950s [6]. These systems are typically constructed using a telescope, an active or passive beacon, a wavefront sensor, a deformable mirror and control electronics. Figure 1.6 shows a block diagram of an adaptive optics system [3]. The beacon is used to provide the reference wavefront discussed in the previous section. In celestial imaging, the beacon may be formed from a neighboring bright star, called a natural guide star. When no such guide star exists, the object of

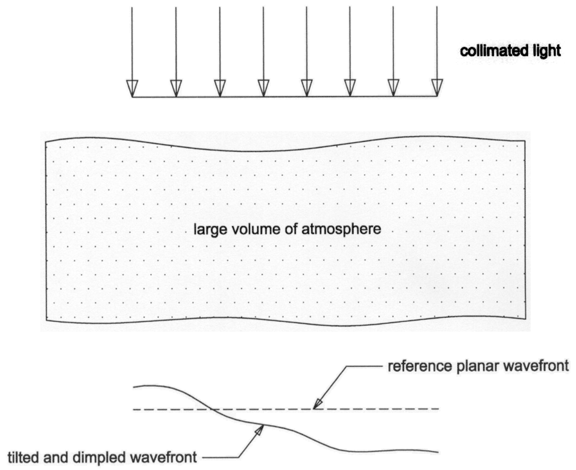


Figure 1.5 Diagram demostrates how the atmosphere will tilt and dimple an incoming planar wavefront.

interest itself may be used. Dim objects and extended objects present further issues, and in such cases an artificial guide star formed by laser reflection from the upper atmosphere may be used. The wavefront sensor provides a measurement of the atmospheric distortion at the input aperture in the form of a direct wavefront measurement or a wavefront slope measurement. A brief discussion of various types of wavefront sensors follows in Section 1.4 and a discussion in greater detail is included in Chapter 4. The deformable mirror consists of a mechanically actuated device capable of forming the conjugate phase measured by the wavefront sensor. The conjugate phase may be divided into a tilt component and higher order components, in which case the system may include a gimbaled mirror designated for global tilt correction and a deformable mirror used to conjugate higher order effects. The control electronics are responsible for mapping the conjugate wavefront from the wavefront sensor measurement to the actuator commands for a deformable mirror.

Two major challenges to be overcome when designing an adaptive optics system include: obtaining adequate levels of light for wavefront sensor performance, and maintaining the bandwidth necessary for active atmospheric compensation. Light from the beacon must be routed to all necessary wavefront sensing devices. Ensuring adequate signal to noise ratio is present in all optical detectors is critical to performance. If the beacon light shares

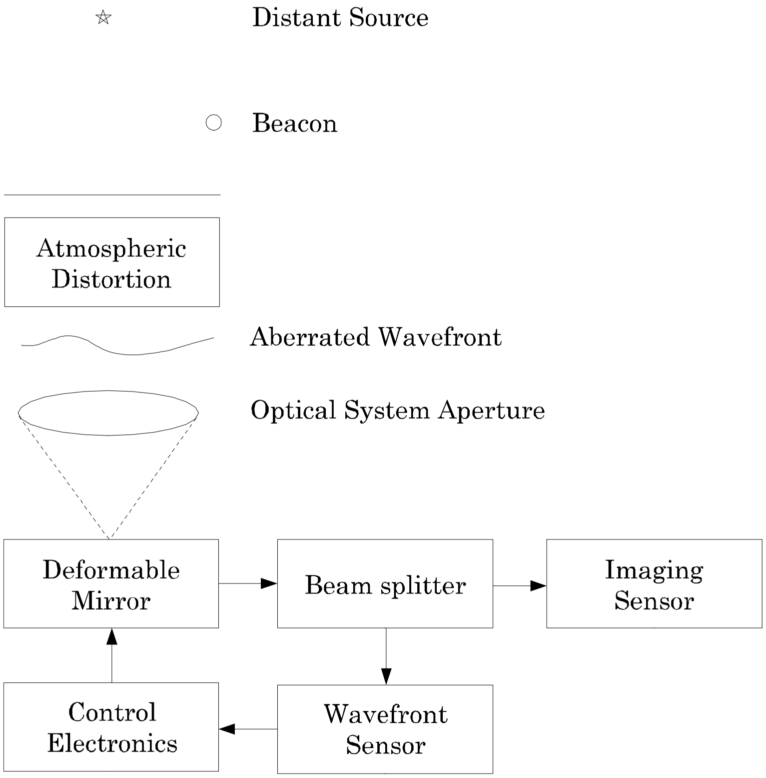


Figure 1.6 Block diagram of an Adaptive Optics system [3].

the same path with the object light then conserving light for the imaging system becomes a trade-off with providing light to the wavefront sensor. Real time correction for atmospheric effects requires that the control electronics and deformable mirror make corrections on the order of several hundred Hz or greater. These bandwidths can be very demanding specifications for the wavefront sensor and the control electronics. Wavefront sensors are diverse in size, power and maintenance requirements. Choosing a wavefront sensor often drives the design of the remainder of an adaptive optics system. Providing a new option for wavefront sensing is the focus of this research.

1.4 Wavefront Sensors

In order to describe the wavefront sensor in further detail, it is beneficial to first transform the figurative concept of a wavefront into a tractable mathematical model. Consider that the wavefront sensor must somehow estimate the complex electromagnetic field at the optical system entrance pupil. The generalized pupil function, denoted $\mathcal{P}$, provides a basic mathematical model for the optical field at the system pupil:

$$\begin{aligned}
\mathcal{P}(x,y;R_P) &= A_P(x,y)W_P(x,y;R_P)\exp(\mathrm{j}P_\psi(x,y)), && (1.2)\\
\text{where } A_P(x,y) &\equiv \text{pupil amplitude function}, && (1.3)\\
W_P(x,y;R_P) &\equiv \text{pupil windowing function}, && (1.4)\\
\text{and } P_\psi(x,y) &\equiv \text{pupil phase function}. && (1.5)
\end{aligned}$$

A_P represents the amplitude of the field, W_P is a unit amplitude windowing function used to mask out a circular aperture with radius R_P, and P_ψ represents the phase of the field. The atmosphere effects both the amplitude and phase of the field as it propagates. Under many conditions, however, the phase distortions create far more pronounced effects in the resulting image. Furthermore, although amplitude effects may be present, the dynamics of those effects often occur on a spatial scale greater than the size of a wavefront sensor subaperture, especially for "weak" turbulence. A_P is relatively constant for such cases. This is a pleasant characteristic of nature since amplitude effects are far more difficult to compensate. For these reasons, the wavefront sensor is designed to detect the difference between the wavefront phase and some reference phase function, P_ψ. Field amplitude is typically ignored. Because the wavefront phase is the quantity of interest, the terms wavefront and phasefront are often used interchangeably in the literature. The phase function P_ψ, being an error measure, is also commonly referred to as the atmospheric aberration function or simply the phase aberration function. Unless specified otherwise, the term wavefront in this document refers to the phase function P_ψ which is assumed to represent the difference in wavefront phase from some desired reference phase.

The wavefront sensor must estimate P_ψ to some level of precision. At optical frequencies, only intensity can be measured directly, not the field amplitude and phase. The wavefront sensor must then map from intensity measurements to field measurements. Some sensors measure P_ψ through interferometry. To do so, a portion of the incoming light is used to create a reference wavefront which is then interfered with the original wavefront. Interference fringes in the intensity reveal relative phase differences between the reference and the aberrated wavefront. The self-referencing Point Diffraction Interferometer (PDI) is an example of this type of wavefront sensor. Figure 1.7 provides a block overview of the

PDI [6]. Some wavefront sensors measure the slope of P_ψ by interfering the wavefront with

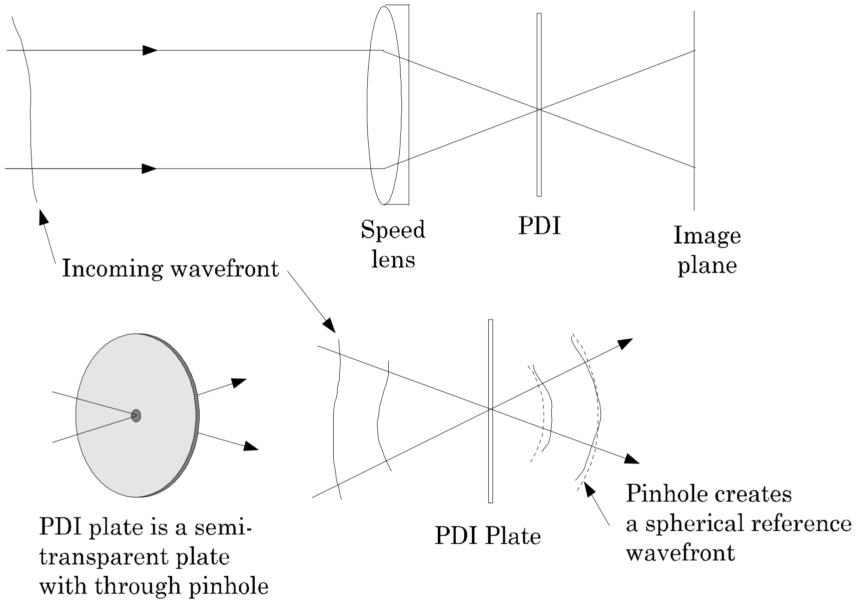


Figure 1.7 Diagram of the self-referencing Point Diffraction Interferometer.

a spatially shifted version of itself. The most common example of this interferometer is the Lateral Shearing Interferometer (LSI). Figure 1.8 provides a block overview of the LSI [3].

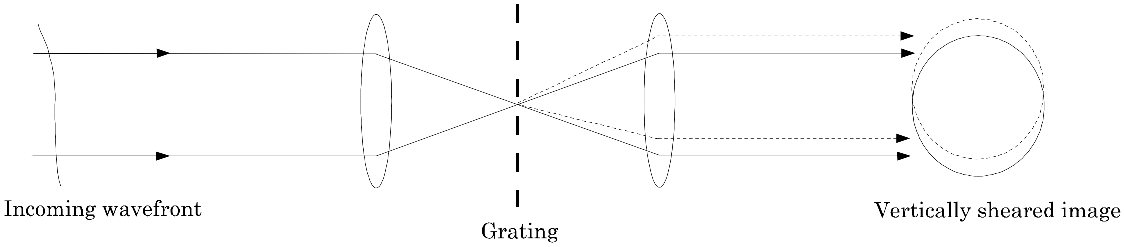


Figure 1.8 Diagram of the Lateral Shear Interferometer.

Measuring the wavefront phase without using interference techniques can seem a bit daunting. The concept of modal estimation offers a way to simplify the problem. Like any other function, the wavefront phase has some frequency domain representation. Transforming portions of the frequency content into the spatial domain produces a set of two-dimensional functions. These functions could be referred to as basis functions or modes.

Common Name	**Zernike Polynomial**
Piston	1
x-tilt	$2r\cos\theta$
y-tilt	$2r\sin\theta$
defocus	$\sqrt{3}(2r^2-1)$
astigmatism-xy	$\sqrt{6}r^2\sin 2\theta$
astigmatism	$\sqrt{6}r^2\cos 2\theta$

Table 1.1 The first six Zernike polynomials

The wavefront phase function may then be approximated by a sum of the ordered basis functions:

$$P_\psi(x,y) \approx \sum_{i=1}^{N} a_i f_i(x,y). \tag{1.6}$$

In the case of the modal estimator, the functions $f_i(x,y)$ are a two-dimensional polynomial basis set, and the coefficients a_i are weights applied to each polynomial. As $N \to \infty$, the approximation becomes exact. How does this simplify the problem of phase estimation? The modal estimator approximates the phase function as a combination of only a small number of polynomials. In the case of the tilt estimator, only 2 polynomials are used. Choosing the class of polynomials to use can be crucial. One such set of polynomials is the set of Seidel polynomials. Seidel polynomials are mentioned here because they appear quite often in the literature. Seidel polynomials are used to describe lens specifications for fabrication. A more convenient set of polynomials for measuring the aberrations in an optical system are the Zernike polynomials. The first six Zernike polynomials are listed in Table 1.1 (in polar coordinates). Notice that the first Zernike is simply a phase delay applied to the entire aperture. When comparing phase from multiple subapertures, relative piston measurements can be very helpful, but truly an engineering challenge due to the precision required. The wavefront sensors discussed in this work will provide phase information from a single subaperture and therefore piston is neglected in the measured aberration function. Define P_ϕ to be the piston removed phase. The piston removed phase can be approximated as a sum of scaled Zernike polynomials beginning with x-tilt:

$$P_\phi(r,\theta) \approx \sum_{i=2}^{N} a_i Z_i\left(\frac{r}{R_P},\theta\right). \tag{1.7}$$

Due to the nature of the atmospheric induced phase aberration, the average Zernike coefficients become successively smaller as the order of the Zernike increases [7]. In fact, the average variance of the tilt coefficients will be nearly 20 times greater than the defocus and astigmatism coefficients. Under the right conditions, wavefront sensors can compensate for up to 86% of the piston removed wavefront phase error by correcting for x and y-tilt only [3]. Figure 1.9 demonstrates how tilt coefficients can be derived from shifted intensity patterns. The off-center location of an Airy pattern can be translated into wavefront tilt by the simple equations:

$$\theta_x = \tan^{-1}\left(\frac{\Delta x}{f}\right) \approx \frac{\Delta x}{f}, \tag{1.8}$$

$$\theta_y \approx \frac{\Delta y}{f}. \tag{1.9}$$

Many of these tilt sensors can be combined together to form a wavefront sensor. Within the wavefront sensor, the primary aperture is divided into a grid of smaller subapertures each contributing a local tilt measurement. The combination of multiple subaperture tilt measurements compensates for the lack of relative piston information. Using a surface fitting algorithm, the grid of tilt or slope samples is used to reconstruct the wavefront phase. The resulting wavefront is an estimate of the actual wavefront from linear phase measurements. This type of wavefront sensor is commonly referred to as a Hartmann type

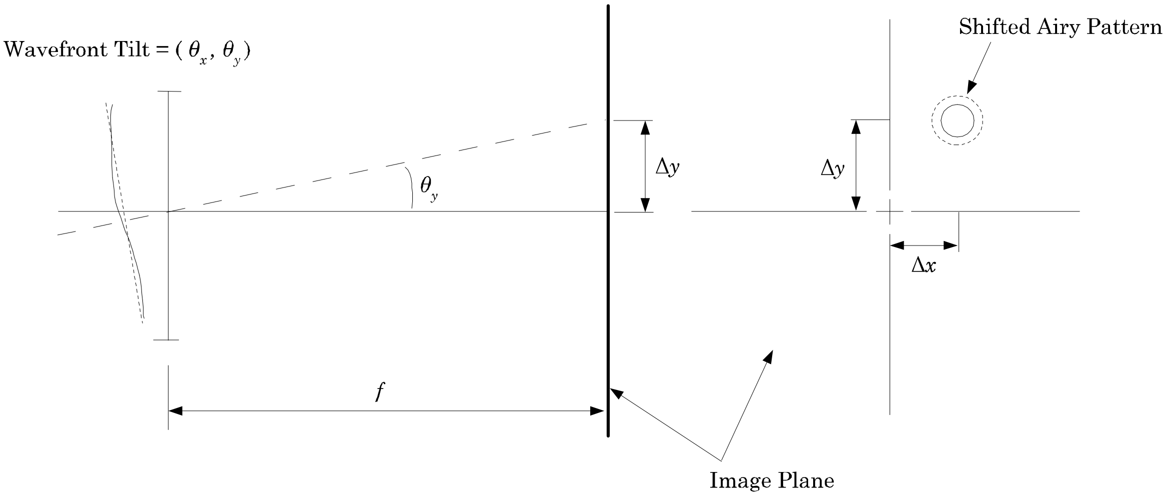


Figure 1.9 Demonstration of how wavefront tilt can be estimated from the off-center shift of an Airy pattern.

wavefront sensor. A diagram of how local wavefront tilt estimates can be used to reconstruct the wavefront is shown in Figure 1.10. Modifications to the Hartmann wavefront sensor and more sophisticated versions of interferometric wavefront sensors are discussed in Chapter 4.

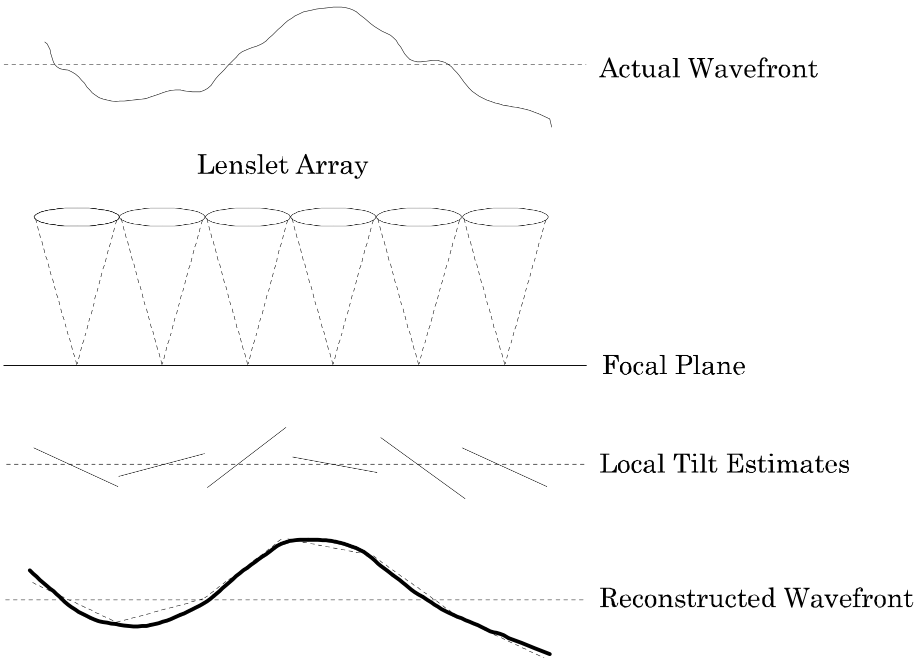


Figure 1.10 The Hartmann type wavefront sensor uses an array of subapertures each contributing a local tilt meaurement. The local tilt measurements are extrapolated to reconstruct the wavefront. [3].

1.5 Research Contributions

The research contributions contained in this dissertation are motivated by the need for higher order modal estimation in real time adaptive optics. The first contribution is a wavefront curvature sensor that provides estimates of Zernike polynomials Z_2 through Z_4. The Z_{2-4} sensor estimates x-tilt, y-tilt and defocus from image projections. The image projection reduces read out time and CCD read noise. Combining the time savings associated with image projection read out and an innovative algorithm design, the Z_{2-4} sensor operates in real time. The second contribution is a curvature sensor capable of estimating Zernike polynomials Z_2 through Z_{10}. The Z_{2-10} sensor uses additional image projections in order to estimate curvature terms Z_5 through Z_{10}. The third contribution involves performance bounding for both curvature sensors. The Cramér Rao lower bound for estimator variance is used to bound the performance of each sensor and to provide insight into design variable selection. The lower bound on performance also serves to validate sensor simulation. Each sensor is simulated using von Kármán phase aberrations and CCD noise modeling. The simulation provides a means to compare performance to existing wavefront sensor designs. The simulation is also used to provide an analysis of sensor sensitivity to errors in environment variable estimates. The last contribution is a unique implementation within the atmospheric simulation. The von Kármán phase screens are generated using a log-polar sampled phase screen generator. Phase screen generators are commonly used in atmospheric turbulence simulation. The log-polar phase screen generation technique offers improved isotropy and increased accuracy over existing phase screen generation techniques.

1.6 Organization

This dissertation is divided into 10 chapters including this introduction. This chapter is meant to provide some insight into adaptive optics, the need for wavefront sensing and a few introductory concepts required to understand the major design challenges involved. Chapter 2 discusses several background concepts necessary to understand the derivation of the tilt and curvature estimators. The background concepts include an introduction to parameter estimation, and atmospheric turbulence modeling. Fourier optics concepts such as the optical transfer function are also discussed. Chapter 3 introduces a discrete model for the optical system and the detected image. The noise model for the sensor detector is detailed as a random process which leads to a probabilistic mapping from image intensity to some set of wavefront modes. Chapter 3 concludes with a description of the image projection operator notation and a derivation of the parameter estimator used in each sensor. Chapter 4 serves as a literature review of related research. It contains a description of the types of curvature sensing devices currently available. The literature review concludes with an outline of the projection based tilt estimator. Chapter 5 describes in detail the extension of the vector based tilt estimator required in order to estimate the defocus parameter. Chapter 6 provides a method for bounding the performance of the projection based estimator. Using the performance bound as a metric for determining ideal design configurations is demonstrated. The performance bound is computed for both the Z_{2-4} and the Z_{2-10} sensors under a typical range of operating parameters. Chapter 7 outlines an existing method for generating random realizations of atmospheric phase. The phase screen generator is an essential part of the wavefront sensor simulation. The polar sampled phase screen generation technique is described in detail. Chapter 7 concludes with a performance comparison between the polar phase screen generator and an existing phase screen generation technique. Chapter 8 outlines the techniques for simulating the projection based Z_{2-4} curvature sensor and provides a summary of the sensor simulation results. The Z_{2-4} curvature sensor is compared to its lower bound and a simulation of existing tilt sensor designs. A sensitivity analysis is also performed in order to demonstrate the robustness of the sensor to erroneous environmental variable estimates. Chapter 9 provides an overview of the Z_{2-10} sensor design and concludes with simulated performance, a comparison to existing tilt sensor designs, and a sensitivity analysis.

2. Background

The projection based wavefront curvature sensors presented in this dissertation are essentially parameter estimators. In order to facilitate a better understanding of the wavefront sensor designs, I will begin by providing background material in this chapter. This material is essential to highlight the set of fundamental principles, any assumptions that I have applied, and the mathematical motivation behind the design and simulation of the wavefront curvature sensors. The background begins with a review of the Bayesian estimator and performance bounds. A discussion of Kolmogorov's turbulence model and its application to atmospheric dynamics follows. From there, the atmospheric dynamics are parameterized. This provides a set of atmospheric phase characteristics to be estimated along with their statistics: the crucial link between the random nature of the atmosphere and some finite set of parameters. Finally, the sensor's intensity measurements are linked to the field phase characteristics (the parameter set) via a linear optics model. The concept of an optical transfer function (OTF) will be the final ingredient that offers a method for mapping from the sensor's observation space to a small set of atmospheric parameters.

2.1 Parameter Estimation

The following parameter estimation background follows the treatment from Van Trees [8]. Consider an experiment where some observed quantity, $\mathbf{R}$, is the outcome when the environment is influenced by some parameter, A. Merely making an observation may not reveal the exact parameter or set of parameters which led to the observed environment. However, given an observation and some knowledge about the experiment, one may guess at the parameters. Prior knowledge about the experiment typically consists of a probabilistic mapping, $p_{\mathbf{r}|a}(\mathbf{R}|A)$, from the parameter space to the observation space. Parameter estimation will replace "guessing" or, more formally, forming a probabilistic map from observation space, $\mathbf{R}$, to a parameter estimate, $\hat{A}$. The map from the observation to the estimate is called an estimation rule, $\hat{a}\left(\mathbf{R}\right)$. The diagram in Figure 2.1 describes the estimation model. Note several variable naming conventions: lower case letters denote random variables, upper case letters indicate instances of random variables or nonrandom quantities, bold letters indicate vector quantities, and a carat indicates the estimate of a quantity. Table 2.1, defines several likelihood expressions that will be used in this section.

Expression	**Description**
$p_a(A)$	probability density for A
$p_{\mathbf{r},a}(\mathbf{R}, A)$	joint density for A and $\mathbf{R}$
$p_{\mathbf{r}\|a}(\mathbf{R}\|A)$	conditional density for $\mathbf{R}$ given A
$p_{a\|\mathbf{r}}(A\|\mathbf{R}) = \frac{p_{\mathbf{r}\|a}(\mathbf{R}\|A)p_a(A)}{p_{\mathbf{r}}(\mathbf{R})}$	Bayes' Rule, a useful identity
a posteriori density = $\frac{\text{conditional likelihood} \times \text{a priori density}}{\text{marginal density}}$	defining the terms in Bayes' Rule

Table 2.1 Useful definitions from estimation theory.

The estimation rule should result in a parameter estimate that minimizes risk, $\mathcal{R}$. Risk is defined to be the expected value of a predefined cost function, C:

$$\begin{aligned} \mathcal{R} &\equiv E\{C[a, \hat{a}(\mathbf{R})]\}, && (2.1) \\ \mathcal{R} &= \int_{-\infty}^{\infty} dA \int_{-\infty}^{\infty} d\mathbf{R} C[a, \hat{a}(\mathbf{R})] p_{a,\mathbf{r}}(A, \mathbf{R}), && (2.2) \\ \mathcal{R} &= \int_{-\infty}^{\infty} d\mathbf{R} p_{\mathbf{r}}(\mathbf{R}) \int_{-\infty}^{\infty} dA C[a, \hat{a}(\mathbf{R})] p_{a|\mathbf{r}}(A|\mathbf{R}). && (2.3) \end{aligned}$$

Since cost is subjective, the cost function selected may vary. The purpose of the cost function is to assign some penalty to error in the estimate:

$$A_\epsilon = A - \hat{A}, \qquad (2.4)$$

$$\begin{aligned} \text{where } A &\equiv \text{realization of the random parameter, } a, \\ \text{and } \hat{A} &\equiv \text{estimate of } A. \end{aligned}$$

A few common cost functions are shown in Figure 2.2. From left to right the example plots

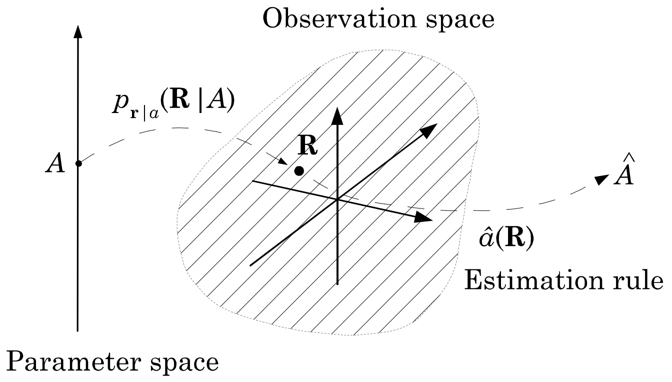


Figure 2.1 The estimation model [8].

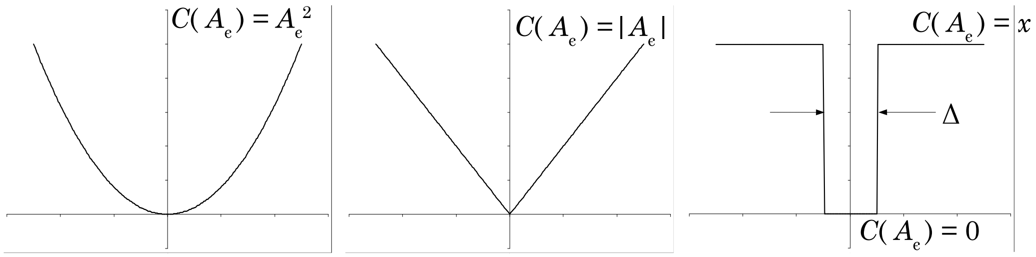


Figure 2.2 Example cost functions.

demonstrate quadratic, linear, and uniform cost functions.

Assume a uniform cost function where the cost of error is unity outside some region, Δ, and the cost of error is zero within the region Δ. Given uniform cost, the risk becomes:

$$\mathcal{R}_{unf} = \int_{-\infty}^{\infty} d\mathbf{R} p_{\mathbf{r}}(\mathbf{R}) \left[\int_{-\infty}^{\hat{a}_{unf}(\mathbf{R})-\frac{\Delta}{2}} dA p_{a|\mathbf{r}}(A|\mathbf{R}) + \int_{\hat{a}_{unf}(\mathbf{R})+\frac{\Delta}{2}}^{\infty} dA p_{a|\mathbf{r}}(A|\mathbf{R}) \right], \quad (2.5)$$

$$\mathcal{R}_{unf} = \int_{-\infty}^{\infty} d\mathbf{R} p_{\mathbf{r}}(\mathbf{R}) \left[1 - \int_{\hat{a}_{unf}(\mathbf{R})-\frac{\Delta}{2}}^{\hat{a}_{unf}(\mathbf{R})+\frac{\Delta}{2}} dA p_{a|\mathbf{r}}(A|\mathbf{R}) \right]. \quad (2.6)$$

Minimizing risk, in this case, means choosing the estimation rule, $\hat{a}_{unf}(\mathbf{R}) = \hat{A}$, such that the inner integral in (2.6) is maximized. Now consider the limiting case where the region Δ in the cost function approaches some arbitrarily small nonzero value. In the limit, the inner integral is maximized when $\hat{a}_{unf}(\mathbf{R})$ equals the parameter that maximizes the a posteriori density, $p_{a|\mathbf{r}}(A|\mathbf{R})$:

$$\begin{matrix}\min\\ A\end{matrix}\left\{\begin{matrix}\lim\\ \Delta\rightarrow 0\end{matrix}\mathcal{R}_{unf}\right\}\Bigg|_{A=\hat{a}_{unf}}. \tag{2.7}$$

The notation, $\begin{matrix}\min\\ A\end{matrix} f\left(A\right)\Big|_{A=\hat{a}}$, is used as a compact form for the expression,

$$\hat{a} = A : f(A) = \begin{matrix}\min\\ A\end{matrix} f\left(A\right), \tag{2.8}$$

which means that $\hat{a}$ is set equal to the value of the input variable A such that the function f is minimized over A. Substituting (2.6) into (2.7) and simplifying:

$$\int_{-\infty}^{\infty} d\mathbf{R}p_{\mathbf{r}}(\mathbf{R})\left[1-\begin{matrix}\max\\ A\end{matrix}\int_{\hat{a}_{unf}(\mathbf{R})^-}^{\hat{a}_{unf}(\mathbf{R})^+} dA p_{a|\mathbf{r}}(A|\mathbf{R})\right]\Bigg|_{A=\hat{a}_{unf}}, \tag{2.9}$$

$$\int_{-\infty}^{\infty} d\mathbf{R}p_{\mathbf{r}}(\mathbf{R})\left[1-\begin{matrix}\max\\ A\end{matrix}\left\{p_{a|\mathbf{r}}(A|\mathbf{R})\right\}\int_{\hat{a}_{unf}(\mathbf{R})^-}^{\hat{a}_{unf}(\mathbf{R})^+} dA\right]\Bigg|_{A=\hat{a}_{unf}}, \tag{2.10}$$

$$\begin{matrix}\max\\ A\end{matrix}\left\{p_{a|\mathbf{r}}(A|\mathbf{R})\right\}\Big|_{A=\hat{a}_{unf}\equiv\hat{a}_{map}}. \tag{2.11}$$

This estimation rule, denoted $\hat{a}_{map}(\mathbf{R})$, is commonly known as the maximum a posteriori (MAP) estimator. If the a posteriori density is continuous and has first partial derivatives then the MAP estimator can be found by solving for the function maximum in the usual manner. Furthermore, since the a posteriori density is necessarily monotonic, its maximum and the maximum of its natural logarithm will both occur at the same value of A. This is advantageous because applying Bayes' Rule (see Table 2.1) and taking the natural logarithm allows for convenient simplification of the MAP estimator expression. Begin by solving for the critical point (the maximum value in this case) of a function in the typical manner. Take the first derivative and set the result equal to zero:

$$\begin{matrix}\max\\ A\end{matrix}\left\{p_{a|\mathbf{r}}(A|\mathbf{R})\right\}\Big|_{A=\hat{a}_{map}}, \tag{2.12}$$

$$\frac{\partial}{\partial A}p_{a|\mathbf{r}}(A|\mathbf{R})\Big|_{A=\hat{a}_{map}} = 0. \tag{2.13}$$

Since the variable A at which the a posteriori density is maximized is also the point at which the logarithm of the a posteriori density is maximized, substitute in the logarithm of the a posteriori density:

$$\left.\frac{\partial}{\partial A}\ln\{p_{a|\mathbf{r}}(A|\mathbf{R})\}\right|_{A=\hat{a}_{map}} = 0. \tag{2.14}$$

Now apply Bayes' Rule and evaluate the logarithm:

$$\begin{aligned}\left.\frac{\partial}{\partial A}\ln\left\{\frac{p_{\mathbf{r}|a}(\mathbf{R}|A)p_a(A)}{p_{\mathbf{r}}(\mathbf{R})}\right\}\right|_{A=\hat{a}_{map}} &= 0, &(2.15)\\ \left.\frac{\partial}{\partial A}\ln p_{\mathbf{r}|a}(\mathbf{R}|A)+\frac{\partial}{\partial A}\ln p_a(A)-\frac{\partial}{\partial A}\ln p_{\mathbf{r}}(\mathbf{R})\right|_{A=\hat{a}_{map}} &= 0. &(2.16)\end{aligned}$$

Note that the derivative of the marginal density with respect to the parameter is zero. Removing the dependence on the marginal density gives:

$$\left.\frac{\partial}{\partial A}\ln p_{\mathbf{r}|a}(\mathbf{R}|A)+\frac{\partial}{\partial A}\ln p_a(A)\right|_{A=\hat{a}_{map}} = 0.$$

Once again, taking the first derivative, setting the result equal to zero and solving for the variable A is equivalent to maximizing the sum of logarithms of the conditional and the a priori densities. Rewriting the differential expression above as a maximization yields:

$$\max_{A}\left.\left\{\ln p_{\mathbf{r}|a}(\mathbf{R}|A)+\ln p_a(A)\right\}\right|_{A=\hat{a}_{map}}. \tag{2.17}$$

From this result, it is easy to see that there are two probabilistic mappings that are required to form the MAP estimator. The first term is the conditional probability of the observation given some set of parameters, $p_{\mathbf{r}|a}(\mathbf{R}|A)$, and the second term is the a priori probability distribution of the parameter space, $p_a(A)$. Unfortunately, many cases arise where the parameter a priori probability is unknown. In these cases, it is common to define some range for the parameter and then assume a uniform probability distribution within the range. If the a priori density is constant then its partial with respect to the parameter A is zero and the expression for $\hat{a}$ becomes simpler still:

$$\begin{aligned}\left.\frac{\partial}{\partial A}\ln p_{\mathbf{r}|a}(\mathbf{R}|A)\right|_{A=\hat{a}_{ml}} &= 0, &(2.18)\\ \max_{A}\left.\left\{\ln p_{\mathbf{r}|a}(\mathbf{R}|A)\right\}\right|_{A=\hat{a}_{ml}}. & &(2.19)\end{aligned}$$

The estimator in this case is often referred to as the Maximum Likelihood (ML) estimator, denoted $\hat{a}_{ml}$.

Now suppose that it is necessary to measure the level of performance of the estimator. A common method for evaluating estimator performance, called the Monte Carlo method, involves simulating or conducting many experiments and evaluating the variance of the estimator over a large sample of observations. This method offers an estimate of the estimator variance. However, the estimated variance is simply a number. It may also be useful to know how the sample variance compares to a theoretical lower bound. The Cramér-Rao lower bound (CRLB) provides a benchmark for the lowest achievable estimator mean squared error. Van Trees provides derivations of the CRLB for both the single and multiple parameter cases [8]. The CRLB on mean squared error for any *unbiased* estimator is presented here in two forms:

$$\begin{aligned} E\left\{(\hat{a}(\mathbf{R})-a)^2\right\} &\geq \frac{1}{E\left\{\left[\frac{\partial}{\partial A}\ln p_{\mathbf{r},a}(\mathbf{R},A)\right]^2\right\}}; && (2.20)\\ &\geq -\frac{1}{E\left\{\frac{\partial^2}{\partial A^2}\ln p_{\mathbf{r},a}(\mathbf{R},A)\right\}}, && (2.21)\end{aligned}$$

where the expectation is taken over both a and $\mathbf{r}$. The term unbiased indicates that the mean or expected value of the estimator equals the true parameter: $E\{\hat{a}(\mathbf{R})\} = A$. If the parameter is nonrandom or if the parameter is given an assumed uniform pdf, then the CRLB simplifies:

$$\begin{aligned} E\left\{(\hat{a}(\mathbf{R})-A)^2\right\} &\geq \frac{1}{E\left\{\left[\frac{\partial}{\partial A}\ln p_{\mathbf{r}|a}(\mathbf{R}|A)\right]^2\right\}}; && (2.22)\\ &\geq -\frac{1}{E\left\{\frac{\partial^2}{\partial A^2}\ln p_{\mathbf{r}|a}(\mathbf{R}|A)\right\}}. && (2.23)\end{aligned}$$

When the variance of an estimator is equal to the CRLB, then the estimator is *efficient.*

If an estimator is biased, then the CRLB above does not apply. The Cramér-Rao inequality for biased estimators is sometimes referred to as the lower bound on mean squared error:

$$E\{\hat{a}(\mathbf{R})\} = A + B(A), \tag{2.24}$$

$$E\left\{(\hat{a}(\mathbf{R}) - a)^2\right\} \geq \frac{\left[1 + \frac{d}{dA}B(A)\right]^2}{E\left\{\left[\frac{\partial}{\partial A}\ln p_{\mathbf{r},a}(\mathbf{R},A)\right]^2\right\}}. \tag{2.25}$$

A CRLB exists for multiple parameter cases. Assume K parameters, the CRLB has the following form (once again for unbiased estimates):

$$E\left\{(\hat{a}_i(\mathbf{R}) - a_i)^2\right\} \geq \mathbf{J}^{ii}, \tag{2.26}$$

where $\mathbf{J}^{ii}$ is the iith element in the $K \times K$ square matrix, $\mathbf{J}_T^{-1}$. $\mathbf{J}_T$ is defined as follows:

$$\mathbf{J}_T = \mathbf{J}_D + \mathbf{J}_P, \tag{2.27}$$

$$J_{D_{ij}} \equiv -E\left\{\frac{\partial^2}{\partial A_i \partial A_j}\ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})\right\}, \tag{2.28}$$

$$J_{P_{ij}} \equiv -E\left\{\frac{\partial^2}{\partial A_i \partial A_j}\ln p_{\mathbf{a}}(\mathbf{A})\right\}. \tag{2.29}$$

Thus the MAP and ML estimators offer methods for minimizing the risk associated with approximating parameters from experimental observations. The caveat is that something must be known about the environment. In either case, a probabilistic map of the parameter space, given some observation $p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})$, must be known. The MAP estimator requires an a priori probability for the estimated parameter(s) as well. This begins with generating an accurate, yet tractable, model for the experiment. In the case of the wavefront sensor problem presented here, it is necessary to develop models for the atmospheric turbulence and detector noise. I will begin with the turbulence model.

2.2 Turbulence Modeling

From the description of atmospheric turbulence provided in the introduction, the random nature of the index of refraction leads to optical system performance far worse than the limits imposed by diffraction effects. This section provides a review of the most common model for atmospheric fluctuations in index of refraction and the assumptions inherent in the model. Once the model for index of refraction is established, it is transformed into a more useful phase model. The importance of a phase model as opposed to an index of refraction model should be evident from the brief discussion in Section 1.2 concerning the optical wavefront. Recall that the wavefront model presents the aberration function as a relative phase difference between the wavefront in the system aperture and some reference wavefront. A wavefront sensor must detect and compensate for this atmospheric phase distortion. In order to design such a device, a keen understanding of the nature of the phase distortion is required along with a tractable model for use in simulation and testing of the sensor design. The most popular place to begin deriving such an atmospheric model is from the research contributions of A. N. Kolmogorov.

In the 1920s and 30s, Andrei Nikolaevich Kolmogorov made significant contributions to mathematics in the area of probability theory and function spaces. These accomplishments led to an applied mathematical treatment concerning the turbulent motion of fluids [9]. Kolmogorov hypothesized a 2/3 power law for the mean square difference in velocity between two points (often referred to as a *structure function*) in an isotropic, homogeneous medium. The terms isotropic and homogeneous refer to the spatial statistics of the fluid. Homogeneous means that the statistical moments are only a function of the displacement vector between the two points of interest and not the location of either point. The term isotropic further restricts the spatial statistics to depend only on the magnitude of the displacement vector without regard for the displacement direction. Kolmogorov's velocity structure function was of the form [5]:

$$D_v(\mathbf{R}_1, \mathbf{R}) = E\left\{[v(\mathbf{R}_1 + \mathbf{R}) - v(\mathbf{R}_1)]^2\right\}, \tag{2.30}$$

$$D_v(R) = C_v^2 R^{2/3}. \tag{2.31}$$

Where C_v^2 is the velocity structure function constant. This structure function applies to a region in the fluid called the inertial range. The inertial range is confined to a separation of points less than the outer scale and greater than the inner scale. The outer scale is the separation distance beyond which the turbulent motion is no longer considered isotropic. For the purpose of this research, the atmosphere is the fluid of interest. Near the Earth's surface, the atmospheric outer scale is generally considered equal to the height above the ground. The outer scale at higher altitudes is often estimated in the 10's of meters. The inner scale is the separation distance where the turbulence gives way to molecular friction. Reasonable values for atmospheric inner scale are on the order of a few millimeters to 15 cm.

Kolmogorov's power law provides a statistical model for relative particle velocity. It is necessary to extend this statistic to the atmospheric index of refraction, n. The first part of this extension lies in finding an expression for index of refraction that relates it to particle velocity, for which Kolmogorov's statistic applies. The second critical step involves a contribution by Corrsin concerning the concept of a conservative passive additive [10]. The index of refraction of air depends on density which is largely a function of temperature, pressure and humidity. The approximate expression for index of refraction at optical wavelengths, excluding humidity effects, is given by Andrews [11]:

$$\begin{aligned} n &= 1 + 7.76 \times 10^{-7}(1 + 7.52 \times 10^{-3}\lambda^{-2})\frac{P}{T}, && (2.32) \\ n &\approx 1 + 7.9 \times 10^{-7}\frac{P}{T}. && (2.33) \end{aligned}$$

The approximated n includes an assumed wavelength in the optical band: $\lambda = 0.5 \times 10^{-6}$m. Now examine the differential:

$$\begin{aligned} \delta n &= 7.9 \times 10^{-7}\frac{P}{T}\left(\frac{\delta P}{P} - \frac{\delta T}{T}\right), && (2.34) \\ \delta n &\approx -7.9 \times 10^{-7}\frac{P}{T^2}\delta T. && (2.35) \end{aligned}$$

This last approximation results from the fact that, at optical frequencies, temperature effects dominate the fluctuations in n and therefore pressure effects can be ignored [5]. Corrsin explains that quantities can be categorized as conservative if they are not dependent on a position in space. He further notes that a passive quantity bears the same atmospheric statistics regardless of position. Given that conservative passive additives do not effect the turbulence statistics, they obey the same 2/3 power established for velocity fluctuations. Temperature is not a conservative quantity in general, because it is dependent on altitude. Consider potential temperature, however, or temperature about a specific altitude. Potential temperature is a conservative quantity. Define potential temperature, Φ_T, as follows

[10]:

$$\begin{aligned}
\Phi_T &= T - 9.8\frac{^\circ\mathrm{C}}{\mathrm{km}}, && (2.36)\\
\delta\Phi_T &= \delta T, && (2.37)\\
\delta n &\approx -7.9\times 10^{-7}\frac{P}{T^2}\delta\Phi_T. && (2.38)
\end{aligned}$$

Now it is obvious that index of refraction fluctuations bear a direct relationship to potential temperature fluctuations making potential temperature a passive quantity. Since Φ_T is a conservative passive additive, its structure function obeys the same 2/3 power law as velocity:

$$D_{\Phi_T}(R) = C_{\Phi_T}^2 R^{\frac{2}{3}}. \tag{2.39}$$

It follows then that n follows a 2/3 power law as well, thus the desired statistic is given:

$$D_n(R) = C_n^2 R^{\frac{2}{3}}. \tag{2.40}$$

It is necessary to transform this spatial statistic into a spectral representation. An expression for the power spectral density is necessary in order to describe the process spectrally. Transforming the structure function into a power spectrum is made possible via the Fourier-Stieltjes integral [11]:

$$\begin{aligned}
x(t) &= \int_{-\infty}^{\infty} e^{\mathrm{j}\omega t} d\nu(\omega), && (2.41)\\
\nu(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{j}\omega t} x(t)dt, && (2.42)\\
x(t) &\equiv \text{spatial or temporal correlation}, && (2.43)\\
d\nu(\omega) &\equiv \text{infinitesimal spectral band}. && (2.44)
\end{aligned}$$

Solving the integral will require working with the correlation, B_n, rather than the structure function, D_n. Recall the following relationship between the structure function and the correlation for a homogeneous random process [11]:

$$B_n(0) - B_n(R) = \frac{1}{2}D_n(R). \tag{2.45}$$

It will be helpful to extend the property of statistical homogeneity into the temporal domain by assuming that the atmospheric statistics do not vary with time. When the temporal moments of a random process do not vary with time, the process is considered stationary. In order to assume stationarity, a temporal quality, it is necessary to assume ergodicity. Assuming that the atmospheric statistics are ergodic is to assume that taking many random samples of the atmosphere in different locations will yield the same statistics as sampling the same location over many time instances. In other words, the temporal statistics are the same as the spatial statistics, in a mean square sense. Finally, this derivation will require a transform pair between the spatial correlation function and the spectral density. To this purpose, $B_n(R)$ must be band limited in order to ensure that the inverse transform exists. Substituting B_n into the Fourier-Stieltjes integral transform and simplifying will require some mathematical rigor. Both Andrews [11] and Strohbehn [5] provide a more detailed version of the derivation. The following summarizes the treatment from Strohbehn.

Begin by defining a zero mean index of refraction random variable, n_1:

$$\begin{aligned} n(\mathbf{R}) &= E\left\{n(\mathbf{R})\right\} + n_1(\mathbf{R}), && (2.46) \\ E\left\{n_1(\mathbf{R})\right\} &= 0. && (2.47) \end{aligned}$$

Apply the Fourier-Stieltjes integral to the zero mean random variable, n_1:

$$n_1(\mathbf{R}) = \int_{-\infty}^{\infty} dN(\mathbf{K}) e^{j\mathbf{K}\cdot\mathbf{R}}, \tag{2.48}$$

where $\mathbf{K} = (K_x, K_y, K_z)$ is the three-dimensional spatial wave number and $dN(\mathbf{K})$ is some small spectral harmonic of the zero mean index of refraction. I am only interested in the zero mean random process n_1, not n. However, for notational simplicity, I would like to retain the variable n and dispense with the subscripted variable n_1. For this reason, the reader may assume that all subsequent references to "the index of refraction" and the variable n are indeed referring to the zero mean random process. Writing the correlation for the index of refraction:

$$\begin{aligned}
E\left\{n(\mathbf{R}_1)n^*(\mathbf{R}_2)\right\} &= B_n(\mathbf{R}_1,\mathbf{R}_2), && (2.49)\\
B_n(\mathbf{R}_1,\mathbf{R}_2) &= E\left\{\begin{array}{c}\displaystyle\int dN(\mathbf{K}_1)e^{\mathrm{j}\mathbf{K}_1\cdot\mathbf{R}_1}\times\\ \displaystyle\left[\int dN(\mathbf{K}_2)e^{\mathrm{j}\mathbf{K}_2\cdot\mathbf{R}_2}\right]^*\end{array}\right\}, && (2.50)\\
&= E\left\{\iint dN(\mathbf{K}_1)dN^*(\mathbf{K}_2)e^{\mathrm{j}(\mathbf{K}_1\cdot\mathbf{R}_1-\mathbf{K}_2\cdot\mathbf{R}_2)}\right\}, && (2.51)\\
&= \iint E\left\{dN(\mathbf{K}_1)dN^*(\mathbf{K}_2)\right\}e^{\mathrm{j}(\mathbf{K}_1\cdot\mathbf{R}_1-\mathbf{K}_2\cdot\mathbf{R}_2)}. && (2.52)
\end{aligned}$$

Making the substitutions $\mathbf{R}_2 = \mathbf{R}_1 + \mathbf{R}$**:**

$$B_n(\mathbf{R}_1,\mathbf{R}_1{+}\mathbf{R}) = \iint E\left\{dN(\mathbf{K}_1)dN^*(\mathbf{K}_2)\right\}e^{\mathrm{j}(\mathbf{K}_1\cdot\mathbf{R}_1-\mathbf{K}_2\cdot(\mathbf{R}_1+\mathbf{R}))}. \qquad (2.53)$$

According to the assumptions of stationarity and ergodicity of the process: $B_n(\mathbf{R}_1{+}\mathbf{R},\mathbf{R}_1) = B_n(\mathbf{R})$. The only form for the right hand side for which the correlation will be independent of $\mathbf{R}_1$ is to force the index spectrum to be delta correlated on $\mathbf{K}$:

$$E\left\{dN(\mathbf{K}_1)dN^*(\mathbf{K}_2)\right\} = \delta(\mathbf{K}_1-\mathbf{K}_2)\Phi_n(\mathbf{K}_2)d^3\mathbf{K}_1d^3\mathbf{K}_2. \qquad (2.54)$$

Substituting (2.54) into (2.52) and evaluating the $\mathbf{K}_2$ integral:

$$\begin{aligned}
B_n(\mathbf{R}) &= \int d^3\mathbf{K}_1e^{\mathrm{j}\mathbf{K}_1\cdot\mathbf{R}}\int\delta(\mathbf{K}_1-\mathbf{K}_2)\Phi_n(\mathbf{K}_2)d^3\mathbf{K}_2e^{-\mathrm{j}\mathbf{K}_2\cdot\mathbf{R}}, && (2.55)\\
B_n(\mathbf{R}) &= \int d^3\mathbf{K}_1e^{\mathrm{j}\mathbf{K}_1\cdot\mathbf{R}}\Phi_n(\mathbf{K}_1). && (2.56)
\end{aligned}$$

Thus, the following three-dimensional Fourier pair:

$$\begin{aligned}
B_n(\mathbf{R}) &= \int_{-\infty}^{\infty} e^{\mathrm{j}\mathbf{K}\cdot\mathbf{R}}\Phi_n(\mathbf{K})d^3\mathbf{K}, && (2.57)\\
\Phi_n(\mathbf{K}) &= \frac{1}{8\pi^3}\int_{-\infty}^{\infty} e^{-\mathrm{j}\mathbf{K}\cdot\mathbf{R}}B_n(\mathbf{R})d^3\mathbf{R}. && (2.58)
\end{aligned}$$

Recalling that the process is also isotropic, it is possible to further simplify the expression by converting to spherical coordinates:

$$
\begin{aligned}
B_n(R) &= \frac{4\pi}{R}\int_0^{\infty} d\kappa \sin(\kappa R)\Phi_n(\kappa)\kappa, && (2.59)\\
\Phi_n(\kappa) &= \frac{1}{2\pi^2\kappa}\int_0^{\infty} dRR\sin(\kappa R)B_n(R), && (2.60)\\
\text{where } \mathbf{K} &= (\kappa,\theta,\phi), && (2.61)\\
\text{and } d^3\mathbf{K} &= \kappa^2\sin\theta d\theta d\phi d\kappa. && (2.62)
\end{aligned}
$$

Combining the isotropic form for $B_n(R)$ in (2.59) and (2.45), the structure function can be expressed in terms of the spectral density:

$$
D_n(R) = 8\pi\int_0^{\infty} d\kappa\kappa^2\Phi_n(\kappa)\left[1-\frac{\sin(\kappa R)}{\kappa R}\right]. \tag{2.63}
$$

Taking the inverse Fourier transform of (2.63) Strohbehn provides the very important result [12]:

$$
\begin{aligned}
\Phi_n(\kappa) &= \frac{1}{4\pi^2\kappa^2}\int_0^{\infty}\frac{\sin(\kappa R)}{\kappa R}\frac{d}{dR}\left[R^2\frac{d}{dR}D_n(R)\right]dR, && (2.64)\\
\Phi_n(\kappa) &= \frac{5}{18\pi^2}C_n^2\kappa^{-3}\int_{l_0}^{L_0}\sin(\kappa R)R^{-1/3}dR, && (2.65)\\
\Phi_n(\kappa) &= 0.033C_n^2\kappa^{-11/3}, && (2.66)\\
&\quad \text{when } \frac{1}{L_0} \ll \kappa \ll \frac{1}{l_0}. && (2.67)
\end{aligned}
$$

These results provide tractable models for index of refraction in both the spatial, (2.40), and spatial frequency, (2.66), domains. At this point, it would also be beneficial to transform the results for index of refraction into a phase structure function and phase spectrum respectively. The transformation from index of refraction to phase models follows in the next section. Table 2.2 highlights the assumptions that were required to arrive at (2.40) and (2.66).

	Assumption	**Explanation**
1.	homogeneity	Atmospheric difference statistics depend only on the displacement vector and not its location in space.
2.	isotropy	Difference statistics are dependent only on the magnitude of the displacement vector.
3.	n solely dependent on temperature	Neglect effects of humidity and pressure on n.
4.	temperature is conservative	Large distance temperature effects are lumped into an atmospheric profile constant C_n^2.
5.	ergodicity	Turbulence evolution time scale is long when compared to wind flow across the aperture. Spatial and temporal statistics are the same.
6.	band limited turbulence	Distant particle velocities become uncorrelated.

Table 2.2 Assumptions required to derive the index of refraction spectrum from Kolmogorov velocity structure function.

The phase spectrum will serve two purposes. The first useful property of the phase spectrum is that it is necessary for deriving the expected power in each Zernike mode. A detailed discussion of Zernike mode variance is included in Section 2.3. The power in each Zernike mode is also a statistical variance which is essential for approximating $p_a(A)$ and the effective operation of the parameter based wavefront sensor. The second property is that the phase spectrum provides a fast, efficient method for simulating atmospheric turbulence. Assumption number 6 in Table 2.2 suggests making one more modification to the Kolmogorov spectrum before using it to create a phase spectrum. The Kolmogorov spectrum should include parameters for the inner and outer scale. Doing so effectively band limits the spectrum by removing the singularity in (2.66) at $\kappa = 0$ and adding exponential roll-off at high frequencies. Making these changes will force the spectrum to be absolutely integrable while still matching the Kolmogorov model within the inertial range. The following refractive index power spectrum is referred to as the von Kármán spectrum:

$$\begin{aligned} \Phi_n(\kappa) &= 0.033 C_n^2 \frac{\exp\left(\frac{-\kappa^2}{\kappa_m^2}\right)}{\left(\kappa^2 + \kappa_0^2\right)^{11/6}}, && (2.68) \\ \text{where } 0 &\leq \kappa \leq \infty, && (2.69) \\ \kappa_m &= 5.92/l_0, && (2.70) \\ \text{and } \kappa_0 &= 1/L_0. && (2.71) \end{aligned}$$

Transforming the index of refraction spectrum into a phase spectrum begins by applying thin screen theory [13]. In the context of thin screen theory, turbulence effects are condensed into a thin screen such that only the phase (not the amplitude) of the propagating field is modulated by the screen. For instance, the field at the entrance and exit of a thin screen may be represented as $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively:

$$\begin{aligned} \mathcal{P}_1(x,y) &= \exp(jP_\psi(x,y)), && (2.72) \\ \mathcal{P}_2(x,y) &= \exp(j\left[P_\psi(x,y) + P_\phi(x,y)\right]). && (2.73) \end{aligned}$$

Note that the phase screen has unit amplitude and a phase function denoted $P_\phi(x,y)$. Within this theory, a column of atmosphere could be simulated by multiple thin *phase screens* each separated by free space as shown in Figure 2.3. The field propagating between

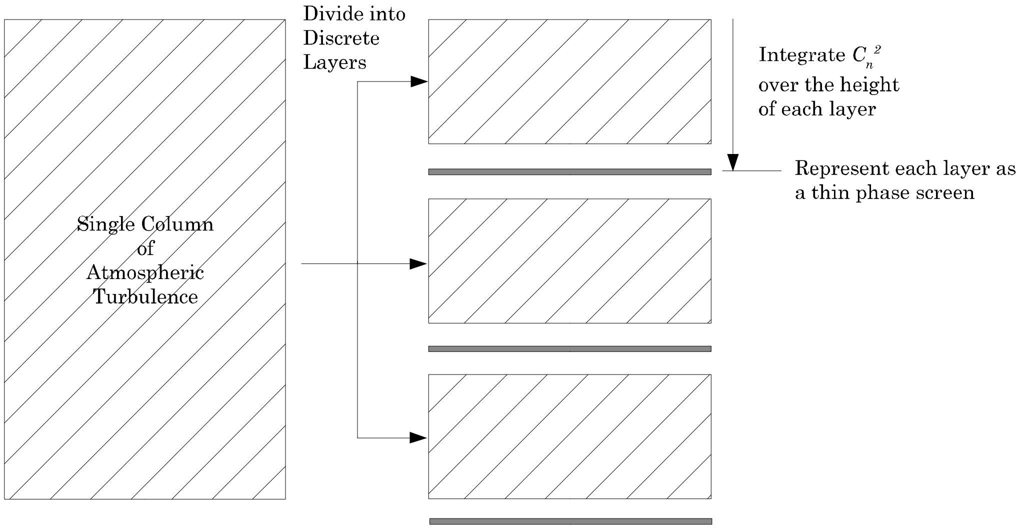


Figure 2.3 Atmospheric turbulence divided into discrete layers and modeled as a series of thin phase screens.

these thin screens experiences fluctuation in both phase *and* amplitude. Within the model described by Figure 2.3, the thin screen phase spectrum may be constructed from a piecewise integral of the index of refraction spectrum multiplied by the wave number [14]:

$$\Phi_{P_\phi}(\kappa_x, \kappa_y) = 2\pi k^2 \int_{z_0}^{z_0+\Delta z} \Phi_n(\kappa_x, \kappa_y, \kappa_z = 0; \xi) d\xi. \tag{2.74}$$

Substituting the von Kármán spectrum for Φ_n provides the thin phase screen phase spectrum:

$$\Phi_{P_\phi}(\kappa_r) = 2\pi k^2 (0.033) (\kappa_r^2 + \kappa_0^2)^{-11/6} \exp\left(\frac{-\kappa_r^2}{\kappa_m^2}\right) \int_{z_0}^{z_0+\Delta z} C_n^2(\xi) d\xi, \tag{2.75}$$

$$\text{where } \kappa_r = \sqrt{\kappa_x^2 + \kappa_y^2}. \tag{2.76}$$

It is convenient to simplify the expression for Φ_{P_ϕ} by introducing a constant called the coherence diameter, r_0. r_0 is also commonly referred to as the Fried parameter [15]. The

Fried parameter accounts for the integrated C_n^2 and the wavelength of interest:

$$r_0 = 0.185 \left(\frac{\lambda^2}{\int\limits_{z_0}^{z_0+\Delta z} C_n^2(\xi) d\xi} \right)^{\frac{3}{5}} . \tag{2.77}$$

Where the coefficient 0.185 is an approximation:

$$\left(\frac{5\Gamma\left(\frac{2}{3}\right)}{4\pi^{\frac{5}{2}}\Gamma\left(\frac{1}{6}\right)} \right)^{\frac{3}{5}} \left(\frac{24}{5} \Gamma\left(\frac{6}{5}\right) \right)^{\frac{1}{2}} \approx 0.185.$$

The Fried parameter offers insight as well as a compact notation. As a general rule, r_0 describes the strength of optical turbulence: as r_0 increases, the strength of the turbulence decreases. It represents the spatial dimension for an optical system aperture beyond which the resolving power advantages typically associated with increasing aperture diameter give way to turbulence effects. In other words, increasing the size of the aperture beyond r_0, while increasing the amount of light entering the system, does not improve resolution. Substituting r_0 into (2.75), provides the following expression for the thin screen phase spectrum:

$$\Phi_{P_\phi}(\kappa_r) = 0.4898 r_0^{-5/3} (\kappa_r^2 + \kappa_0^2)^{-11/6} \exp\left(\frac{-\kappa_r^2}{\kappa_m^2} \right) . \tag{2.78}$$

Substituting the thin screen phase spectrum into the Fourier-Stieltjes integral yields the corresponding thin phase structure function. Begin with a convenient form of the transform integral in (2.63) provided by Tatarskii [16]. This form of the integral assumes plane wave propagation and local isotropy. Also, it is convenient to remove the subscript r on κ for notational simplicity. Remember that κ, in this case, is only varying in an infinitesimally thin slice of the atmosphere perpendicular to the path of propagation:

$$\begin{aligned}
D_{P_\phi}(R_x, R_y, 0) &= 2\iint\limits_{-\infty}^{\infty}[1-\cos(\kappa_x R_x + \kappa_y R_y)]\Phi_{P_\phi}(\kappa_x, \kappa_y, 0)d\kappa_x d\kappa_y, && (2.79)\\
D_{P_\phi}(R) &= 4\pi\int\limits_0^{\infty} d\kappa\kappa\,[1-J_0(\kappa R)]\,0.4898r_0^{-5/3}\times \\
&\qquad (\kappa^2+\kappa_0^2)^{-11/6}\exp\left(\frac{-\kappa^2}{\kappa_m^2}\right), && (2.80)\\
&= 0.4898r_0^{-5/3}4\pi\begin{bmatrix} \int\limits_0^{\infty} d\kappa\frac{\kappa}{(\kappa^2+\kappa_0^2)^{11/6}}\exp\left(\frac{-\kappa^2}{\kappa_m^2}\right) - \\ \int\limits_0^{\infty} d\kappa\frac{\kappa}{(\kappa^2+\kappa_0^2)^{11/6}}J_0(\kappa R)\exp\left(\frac{-\kappa^2}{\kappa_m^2}\right)\end{bmatrix}. && (2.81)
\end{aligned}$$

Note that the symbol J_0 represents the Bessel function of the first kind. Examining (2.81), the first integral can be computed in closed form using a table from Andrews [11]:

$$\begin{aligned}
\int\limits_0^{\infty} d\kappa\kappa^{2\mu}\frac{\exp\left\{\frac{-\kappa^2}{\kappa_m^2}\right\}}{(\kappa^2+\kappa_0^2)^{11/6}} &= \frac{1}{2}\kappa_0^{2\mu-\frac{8}{3}}\Gamma(\mu+\frac{1}{2})U(\mu+\frac{1}{2};\mu-\frac{1}{3};\frac{\kappa_0^2}{\kappa_m^2}), && (2.82)\\
&\cong \frac{1}{2}\kappa_0^{2\mu-\frac{8}{3}}\Gamma(\mu+\frac{1}{2})\times \\
&\qquad \left[\frac{\Gamma(\frac{4}{3}-\mu)}{\Gamma(\frac{11}{6})} + \frac{\Gamma(\mu-\frac{4}{3})}{\Gamma(\mu+\frac{1}{2})}\left(\frac{\kappa_0^2}{\kappa_m^2}\right)^{\frac{4}{3}-\mu}\right]. && (2.83)
\end{aligned}$$

Where the function symbol, $U(a;c;z)$ represents a confluent hypergeometric function of the second kind. Substituting $\mu = \frac{1}{2}$ yields the following result:

$$\int\limits_0^{\infty} d\kappa\kappa^{2\mu}\frac{\exp\left\{\frac{-\kappa^2}{\kappa_m^2}\right\}}{(\kappa^2+\kappa_0^2)^{11/6}} \cong \frac{1}{2}\kappa_0^{-\frac{5}{3}}\left[\frac{\Gamma(\frac{5}{6})}{\Gamma(\frac{11}{6})} + \Gamma(-\frac{5}{6})\left(\frac{\kappa_0^2}{\kappa_m^2}\right)^{\frac{5}{6}}\right]. \tag{2.84}$$

Approximating the effects of the third hypergeometric term, $\frac{\kappa_0^2}{\kappa_m^2} = \frac{l_0^2}{35.05L_0^2}$, to be nearly zero yields the following result:

$$\int\limits_0^{\infty} d\kappa\kappa\frac{\exp\left\{\frac{-\kappa^2}{\kappa_m^2}\right\}}{(\kappa^2+\kappa_0^2)^{11/6}} = \frac{1}{2}\kappa_0^{-\frac{5}{3}}U(1;\frac{1}{6};\simeq 0) \cong \frac{1}{2}\kappa_0^{-\frac{5}{3}}\frac{\Gamma(\frac{5}{6})}{\Gamma(\frac{11}{6})} \cong \frac{3}{5}\kappa_0^{-\frac{5}{3}}. \tag{2.85}$$

The second integral in (2.81) presents a slight problem since it has no analytical solution. In order to simplify the integrand to a form that will provide an analytical result, the inner scale term must be removed. The are a few items to consider when determining whether such an assumption is viable. The primary purpose of this derivation is for computer phase screen simulation. For the purposes of computer modeling, the inner scale effects will be negligible as long as the smallest spatial sampling in the phase screen is on the order of l_0 or larger. In this case, the roll-off in the structure function due to κ_m may go unnoticed. A plot comparison of the analytical structure function derived below and a numerical evaluation of the structure function in (2.81) is presented in Figure 2.4. Removing the exponential term, the remaining integral can be solved using a table from Gradshteyn [17]:

$$\int_0^\infty d\kappa \frac{J_\nu(b\kappa)\kappa^{\nu+1}}{(\kappa^2+a^2)^{\mu+1}} = \frac{a^{\nu-\mu}b^\mu}{2^\mu\Gamma(\mu+1)} K_{\nu-\mu}(ab). \tag{2.86}$$

Substituting $\mu = \frac{5}{6}$, $\nu = 0$, $b = R$, and $a = \kappa_0$ yields the following result:

$$\int_0^\infty d\kappa \frac{\kappa}{(\kappa_r^2+\kappa_0^2)^{11/6}} J_0(\kappa R) = \frac{\kappa_0^{-\frac{5}{6}} R^{\frac{5}{6}}}{2^{\frac{5}{6}}\Gamma(\frac{11}{6})} K_{-\frac{5}{6}}(\kappa_0 R) = \frac{\left(\frac{R}{2\kappa_0}\right)^{\frac{5}{6}}}{\Gamma(\frac{11}{6})} K_{\frac{5}{6}}(\kappa_0 R). \tag{2.87}$$

Note that K_x represents a modified Bessel function of the second kind. Combining the results in (2.85) and (2.87) produces the familiar closed form expression for $D_{P_\phi}(R)$:

$$D_{P_\phi}(R) = 0.4898 r_0^{-5/3} 4\pi \left[\frac{3}{5}\kappa_0^{-\frac{5}{3}} - \frac{\left(\frac{R}{2\kappa_0}\right)^{\frac{5}{6}}}{\Gamma(\frac{11}{6})} K_{\frac{5}{6}}(\kappa_0 R) \right]. \tag{2.88}$$

These results provide both spatial (2.88) and spatial frequency (2.78) statistical models for the atmospheric turbulence phase. These two very important expressions were carefully derived from first principles. The phase spectrum and phase structure function will prove vital when creating a computer-based atmospheric simulation with which to test the wavefront sensor.

2.3 Defining the Parameter Space

Given a statistical model for the wavefront phase, the next step is to relate the model to the Zernike polynomial expansion introduced earlier. In doing so, it is possible to

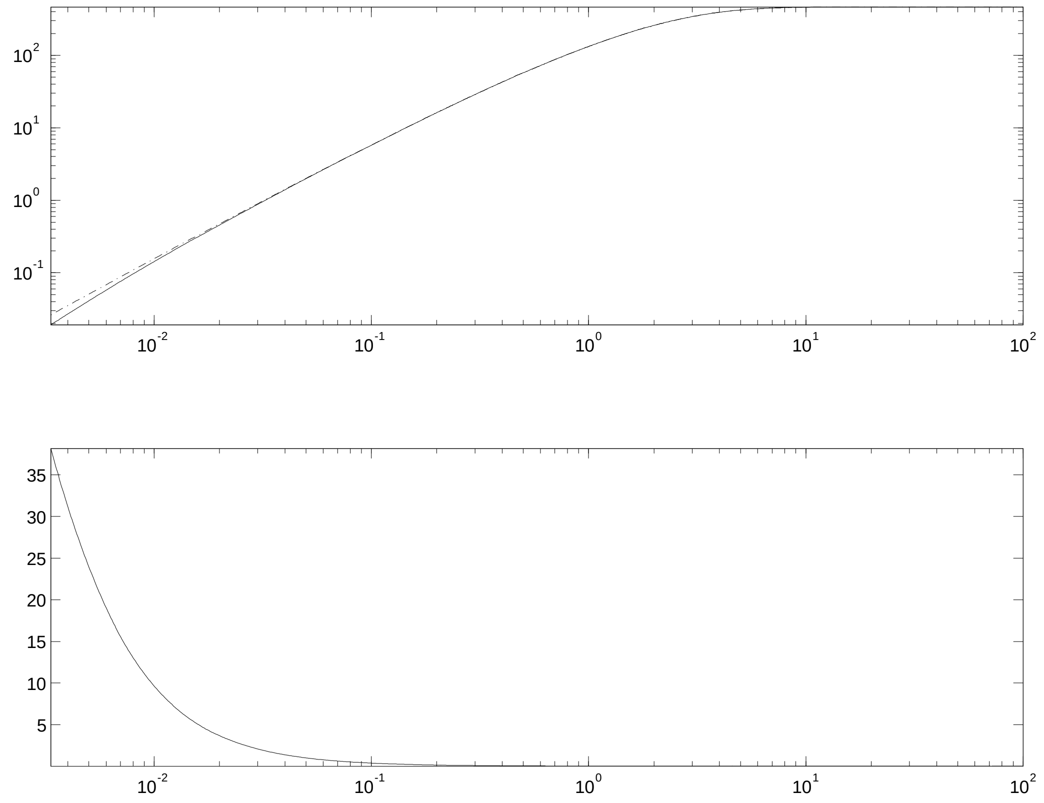


Figure 2.4 Top: the numeric integral structure function [*solid line*] in (2.81) and the analytical form [*dashed line*] in (2.88), [$r_0 = 0.088$m, $l_0 = 0.01$m, $L_0 = 10$m]. Bottom: percent difference between the numeric integral and analytic structure functions.

combine the ideas from the parameter estimation section with the modal statistics from the atmospheric model. Recall that deriving a probabilistic mapping from the image intensity to the set of estimated lower order Zernike modes, is the essence of the parameter estimator. Chapter 1 introduced the fact that the phase aberration function can be approximated by a linear combination of Zernike polynomials:

$$P_{\phi}(\mathbf{r}) \approx W_P(\mathbf{r}; R_P) \sum_{i=2}^{N} a_i Z_i\left(\mathbf{r}; R_P\right). \tag{2.89}$$

$$\begin{aligned} \text{where } R_P &\equiv \text{radius of the aperture}, \\ \mathbf{r} &= (r,\theta), \\ 0 &\leq r < \infty, \\ 0 &\leq \theta < 2\pi, \\ \text{and } Z_i(\mathbf{r};R_P) &\equiv Z_i\left(\frac{r}{R_P},\theta\right). \end{aligned}$$

The function W_P, is a pupil windowing function:

$$W_P(\mathbf{r};R_P) = \begin{cases} 1, & r \leq R_P \\ 0, & r > R_P \end{cases}. \tag{2.90}$$

Note that the two-dimensional coordinates (r,θ) have been condensed into a single vector representation, **r,** whenever convenient. The pupil or aperture radius, R_P, may be included in Zernike functions and windowing functions to give the notation more generality. Given some measurement of the field phase, the only unknowns are the a_i's. Assuming that it is possible to develop a method for transforming intensity, which can be measured directly, into field phase, then the estimator must guess at the values for the a_i's. Within the atmospheric statistics there must be an average value and a variance for each of the estimated coefficients. Given an average value and a range of say $\pm x$ standard deviations the estimator could define a most likely starting range for each coefficient much like the assumption required for the ML estimator. Further, the MAP estimator could be fashioned from an average value and an accurate variance for each parameter. The set of expected values and variances of each parameter will be referred to as the parameter space. This section will discuss a few basics concerning the Zernike basis set and the theory which links the Zernike basis to a particular atmospheric phase spectrum.

The Zernike background begins with a demonstration of how to construct each of the Zernike modes, calculate Zernike coefficients from a given phasefront measurement, and how to calculate the mean and variance of each Zernike mode. This treatment follows Roggemann [3] with the exception of some minor notation changes. The coefficients, a_i, can be found by projecting each Zernike onto the wavefront phase:

$$a_i = \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i(\boldsymbol{\rho};1) P_\psi(\rho R_P, \theta), \tag{2.91}$$

$$\boldsymbol{\rho} = \left(\frac{r}{R_P}, \theta\right), \tag{2.92}$$

$$W_Z(\boldsymbol{\rho}) = \begin{cases} \frac{1}{\pi}, & |\boldsymbol{\rho}| \le 1 \\ 0, & |\boldsymbol{\rho}| > 1 \end{cases}. \tag{2.93}$$

Note the use of the scaled coordinates, $\boldsymbol{\rho}$, is required because the Zernike polynomials are only valid on the unit circle. The weighted windowing function, $W_Z(\boldsymbol{\rho};1)$, provides the limits of integration and the appropriate scaling such that the Zernikes are orthonormal on the unit circle. Recall that Zernike polynomials are defined as an orthonormal basis strictly over the unit circle:

$$\int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i(\boldsymbol{\rho};1) = 0, \tag{2.94}$$

for $i \ge 2$, and

$$\int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i(\boldsymbol{\rho};1) Z_{i'}(\boldsymbol{\rho};1) = \delta_{ii'}, \tag{2.95}$$

$$\delta_{ii'} = \begin{cases} 0, & i \ne i' \\ 1, & i = i' \end{cases}, \tag{2.96}$$

and hence, the need for the scaled windowing function and radial coordinate scaling. Recall that the wavefront sensor will estimate modes beginning with Z_2. For this reason, the phase function of interest is the piston removed, or zero mean, phase aberration, P_ϕ. Consequently, each coefficient, $i \ge 2$, has zero mean:

$$\begin{aligned} E\{a_i\} &= E\left\{\int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i(\boldsymbol{\rho}) P_\phi(\rho R_P, \theta)\right\}, && (2.97) \\ &= \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i(\boldsymbol{\rho}) E\{P_\phi(\rho R_P, \theta)\}, && (2.98) \\ &= 0. && (2.99) \end{aligned}$$

The radius argument in W_Z and Z_i is removed to compact the notation for the case: $R_P = 1$. The coefficient variance is another important statistical moment:

$$\begin{aligned} E\{a_i a_{i'}\} &= E\left\{ \begin{array}{c} \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i(\boldsymbol{\rho}) P_\phi(\rho R_P, \theta) \times \\ \int d\boldsymbol{\rho}' W_Z(\boldsymbol{\rho}') Z_{i'}(\boldsymbol{\rho}') P_\phi(\rho' R_P, \theta') \end{array} \right\}, && (2.100) \\ &= \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) \int d\boldsymbol{\rho}' W_Z(\boldsymbol{\rho}') Z_i(\boldsymbol{\rho}) Z_{i'}(\boldsymbol{\rho}') E\{P_\phi(\rho R_P, \theta) P_\phi(\rho' R_P, \theta')\}, && (2.101) \\ &= \begin{cases} \mathrm{Var}\{a_i\}, & i = i' \\ \mathrm{Cov}\{a_i a_{i'}\}, & i \neq i' \end{cases}. && (2.102) \end{aligned}$$

The amount of power or phase error associated with any coefficient is related to the coefficient variance. Given a plane wave reference, the aperture mean square error will be shown to be (see Section 6.1):

$$\left\langle P_{\phi_e}^2 \right\rangle = \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) E\left\{P_\phi^2(\rho R_P, \theta)\right\} = \sum_{i=2}^{\infty} E\left\{a_i^2\right\}. \tag{2.103}$$

The mean square phase aberration is commonly used as a measure of the atmospheric distortion present in an optics system. The mean square measure of distortion equates to the sum of the variances in each Zernike coefficient. This relationship makes the modal expansion extremely useful in identifying the amount of phase error expected of the atmospheric model and how that error is distributed among the modes.

Section 1.4 in the Introduction, identified the first 6 Zernike polynomials. The set of rules below provide a means to derive any Zernike polynomial [7]:

$$\left. \begin{array}{r} Z_{i\ \mathrm{even}}(r,\theta) = \sqrt{2(n+1)} R_n^m(r) \cos(m\theta) \\ Z_{i\ \mathrm{odd}}(r,\theta) = \sqrt{2(n+1)} R_n^m(r) \sin(m\theta) \end{array} \right\} m \neq 0, \tag{2.104}$$

$$Z_i(r,\theta) = \sqrt{n+1} R_n^0, \qquad m = 0, \tag{2.105}$$

$$R_n^m(r) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s! \left(\frac{n+m}{2} - s\right)! \left(\frac{n-m}{2} - s\right)!} r^{n-2s}. \tag{2.106}$$

There is more than one ordering scheme for Zernike polynomials. To minimize confusion, I will quickly outline two common ordering schemes. Ordering Zernike polynomials requires at least two indices due to the two degrees of freedom r (radial) and θ (azimuthal). Most

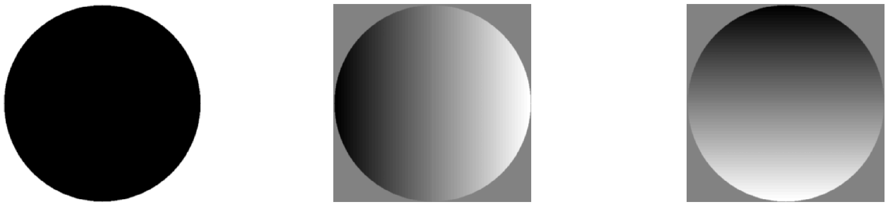
```
function [$n$, $m$] = index2nm($i$)
x = 1;
xsum = 1;
while xsum < $i$
    x = x + 1;
    xsum = xsum + x;
end
$n$ = x - 1;
$m$ = ($i$ - xsum + x);
if $n/2$ == round($n/2$)
    if $m/2 \neq$ round($m/2$)
        $m$ = $m$ - 1;
    end
else
    if $m/2$ == round($m/2$)
        $m$ = $m$ - 1;
    end
end
return
```

Table 2.3 Example algorithm designed to generate n and m from i using Noll's Zernike ordering scheme.

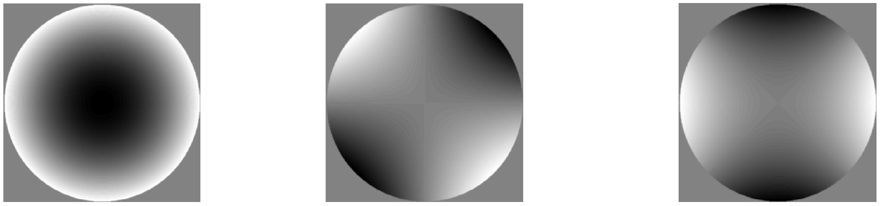
ordering schemes use a single index, i, to sort the polynomials and two other indices, (n, m), to identify radial order and azimuthal order. Two common ordering methods are provided by Malacara [18] and Noll [7]. Each ordering scheme has its respective benefits. For instance, Malacara's ordering offers a simple relationship between the primary index i and the dual indices n and m:

$$\begin{aligned} n &= \text{next integer greater than } \frac{-3+\sqrt{1+8i}}{2}, & (2.107) \\ m &= i - \frac{n(n+1)+2}{2}. & (2.108) \end{aligned}$$

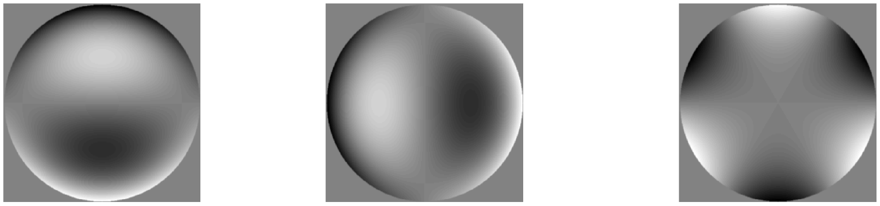
This paper will adopt Noll's ordering scheme. Noll's ordering places an odd or even dependency between the overall index i and the symmetry of each polynomial. This is extremely beneficial to the derivation to follow in that it allows for a very compact solution to the coefficient variance. Unfortunately, Noll's scheme does not share the benefit of having a simple relationship between the primary and dual indices. Instead an algorithm similar to the one shown in Table 2.3 is required to generate n and m from i:



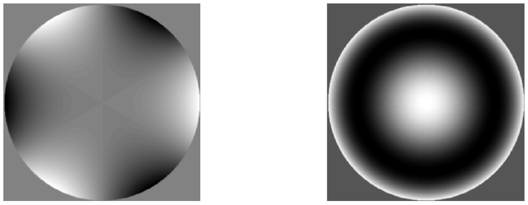
Piston: $(1, 0, 0)$ Tilt-x: $(2, 1, 1)$ Tilt-y: $(3, 1, 1)$



Defocus: $(4, 2, 0)$ Astigmatism-xy: $(5, 2, 2)$ Astigmatism: $(6, 2, 2)$



Coma: $(7, 3, 1)$ Coma: $(8, 3, 1)$ Trefoil: $(9, 3, 3)$



Trefoil: $(10, 3, 3)$ Spherical: $(11, 4, 0)$

Table 2.4 The first 11 Zernike polynomials and their corresponding i, n, and m Noll ordering.

Given the expressions for the Zernike polynomials above and Noll's ordering scheme, it is possible to unambiguously describe each polynomial and single out its effects on the wavefront. Table 2.4 contains example phase plots and demonstrates the indexing for the first eleven Zernikes. Each Zernike image in Table 2.4 is labeled according to the convention, Zernike name: (i, n, m).

This section began with the purpose of deriving a relationship between the statistical turbulence model and the modal decomposition coefficients. Demonstrating how the atmospheric statistics derived in the previous section relate to the Zernike coefficients requires a frequency domain representation of the Zernike modes. Let $Q_i(\kappa, \psi)$ be the frequency domain representation of $Z_i(r,\theta)$. Born provides the following transform pair [19]:

$$\begin{aligned} W_P(\mathbf{r};1)Z_i(\mathbf{r}) &= \int d^2\mathbf{K}Q_i(\kappa,\psi)\exp\{-\mathrm{j}\mathbf{K}\cdot\mathbf{r}\}, && (2.109)\\ \text{where } \mathbf{K} &= (K_x,K_y)\,, && (2.110)\\ \kappa &= \sqrt{K_x^2+K_y^2}, && (2.111)\\ \text{and } \psi &= \operatorname{atan2}(K_y,K_x)\,, && (2.112)\end{aligned}$$

where the atan2 function is a call to the Matlab® function arctan which allows arguments in the range $[0,2\pi)$. The frequency domain Zernike functions, Q_i, are given by:

$$\left.\begin{aligned} Q_{\text{even }i}(\kappa,\psi) &= \\ Q_{\text{odd }i}(\kappa,\psi) &= \end{aligned}\right\} \sqrt{n+1}\frac{2J_{n+1}(\kappa)}{\kappa}\left\{\begin{aligned} &(-1)^{(n-m)/2}\mathrm{j}^m\sqrt{2}\cos m\psi \\ &(-1)^{(n-m)/2}\mathrm{j}^m\sqrt{2}\sin m\psi \end{aligned}\right. \quad \text{, for } m\neq 0$$

$$Q_i(\kappa,\psi) = \sqrt{n+1}\frac{2J_{n+1}(\kappa)}{\kappa}(-1)^{n/2}, \text{ for } m=0. \qquad (2.113)$$

Note that J_v represents the Bessel function of the first kind with order v. With this property in hand, return to the expression for Zernike coefficient variance (2.101):

$$E\left\{a_i^* a_{i'}\right\} = \int d\boldsymbol{\rho}W_Z(\boldsymbol{\rho})\int d\boldsymbol{\rho}'W_Z(\boldsymbol{\rho}')Z_i(\boldsymbol{\rho})\,Z_{i'}(\boldsymbol{\rho}')\,E\{P_\phi(\rho R_P,\theta)\,P_\phi(\rho' R_P,\theta')\}. \quad (2.114)$$

Making use of (2.113) and (2.78) it is possible to rewrite this expression in the frequency domain:

$$E\left\{a_i^* a_{i'}\right\} = \int d\psi\int d\psi'\int\frac{1}{R_P}\frac{\kappa}{R_P}d\kappa\int\frac{1}{R_P}\frac{\kappa'}{R_P}d\kappa' Q_i^*(\kappa,\psi)\Phi_{P_\phi}(\frac{\kappa}{R_P},\psi,\frac{\kappa'}{R_P},\psi')Q_{i'}(\kappa',\psi'). \quad (2.115)$$

Enforcing the condition that Φ_{P_ϕ} is delta correlated, the expression for $E\left\{a_i^* a_{i'}\right\}$ can be reduced to two integrals:

$$\begin{aligned} E\left\{a_i^* a_{i'}\right\} &= \int d\psi\int d\psi'\int\frac{1}{R_P}\frac{\kappa}{R_P}d\kappa\int\frac{1}{R_P}\frac{\kappa'}{R_P}d\kappa' \times \\ &\quad Q_i^*(\kappa,\psi)\Phi_{P_\phi}(\frac{\kappa}{R_P},\psi,\frac{\kappa'}{R_P},\psi')Q_{i'}(\kappa',\psi')\delta(\kappa-\kappa')\delta(\psi-\psi'), && (2.116)\\ E\left\{a_i^* a_{i'}\right\} &= \int d\psi\int\frac{1}{R_P}\frac{\kappa}{R_P}d\kappa Q_i^*(\kappa,\psi)\Phi_{P_\phi}(\frac{\kappa}{R_P},\psi)Q_{i'}(\kappa,\psi). && (2.117)\end{aligned}$$

Now recall that Φ_{P_ϕ} is symmetrical in ψ. Removing the dependence of Φ_{P_ϕ} on ψ and reordering the integrals yields:

$$E\left\{a_i^* a_{i'}\right\} = \int \frac{1}{R_P}\frac{\kappa}{R_P} d\kappa \Phi_{P_\phi}(\frac{\kappa}{R_P}) \int d\psi Q_i^*(\kappa,\psi) Q_{i'}(\kappa,\psi). \tag{2.118}$$

Substituting $\frac{\kappa}{R_P}$ for κ_r in (2.78), Φ_{P_ϕ} is given by:

$$\Phi_{P_\phi}\left(\frac{\kappa}{R_P}\right) = 0.4898 r_0^{-5/3}\left(\frac{\kappa^2}{R_P^2}+\kappa_0^2\right)^{-11/6}\exp\left(\frac{-\kappa^2}{R_P^2\kappa_m^2}\right). \tag{2.119}$$

This expression needs a few points of discussion. The ratio, $\frac{R_P}{r_0}$, should be factored separately within the equation as it will allow introducing the ratio $\frac{D_P}{r_0}$. Recalling the brief introduction to r_0 should clarify the importance of the ratio. Also, for mathematical tractability, a separate expression for Φ_{P_ϕ} neglecting the effects of inner and outer scale can be derived. The cost of this assumption is that the results derived from this form of the expression are only valid in the inertial range. High frequency modes with periods smaller than the inner scale will suffer variance overage errors due to improper inner scale roll off. Likewise, low frequency modes with periods longer than the outer scale will have high variance estimates. With these cautions in mind, I offer the following two forms of Φ_{P_ϕ}: the first with both inner and outer scale,

$$\Phi_{P_\phi}^{\mathrm{vK}}\left(\frac{\kappa}{R_P},\frac{\kappa'}{R_P},\psi,\psi'\right) = 0.4898\left(\frac{R_P}{r_0}\right)^{5/3} R_P^{-5/3}\left(\frac{\kappa^2}{R_P^2}+\kappa_0^2\right)^{-11/6}\exp\left(\frac{-\kappa^2}{R_P^2\kappa_m^2}\right), \tag{2.120}$$

$$= 0.4898\left(\frac{R_P}{r_0}\right)^{5/3} R_P^2\left(\kappa^2+R_P^2\kappa_0^2\right)^{-11/6}\exp\left(\frac{-\kappa^2}{R_P^2\kappa_m^2}\right), \tag{2.121}$$

and the second (Kolmogorov turbulence) with inner and outer scale terms removed,

$$\Phi_{P_\phi}^{\mathrm{K}}\left(\frac{\kappa}{R_P},\frac{\kappa'}{R_P}\psi,\psi'\right) = 0.4898\left(\frac{R_P}{r_0}\right)^{5/3} R_P^2\kappa^{-11/3}. \tag{2.122}$$

The spectrum containing both inner scale and outer scale compensation will be reserved for numerical results to compare with the analytical solution. Proceeding to form an analytical result to the correlation integral, substitute the simplified phase spectrum into (2.118) and make use of the frequency domain Zernike expression in (2.113):

$$\begin{aligned} E\left\{a_i^* a_{i'}\right\} &= \int \frac{1}{R_P^2}\kappa d\kappa 0.4898\left(\frac{R_P}{r_0}\right)^{5/3} R_P^2 \kappa^{-11/3}\int d\psi Q_i^*(\kappa,\psi)Q_{i'}(\kappa,\psi), && (2.123)\\ &= 0.4898\left(\frac{R_P}{r_0}\right)^{5/3}\int \kappa^{-8/3}d\kappa\int Q_i^*(\kappa,\psi)Q_{i'}(\kappa,\psi)d\psi. && (2.124)\end{aligned}$$

Substituting (2.113) for Q yields:

$$\begin{aligned} E\left\{a_i^* a_{i'}\right\} &= 2\pi\cdot 0.4898\left(\frac{R_P}{r_0}\right)^{5/3}\sqrt{(n+1)(n'+1)}(-1)^{(n+n'-2m)/2}\delta_{mm'}\times \\ &\quad \int \kappa^{-8/3}\frac{2J_{n+1}(\kappa)}{\kappa}\frac{2J_{n'+1}(\kappa)}{\kappa}d\kappa, && (2.125)\\ &= 0.4898\cdot 2^{4/3}\pi\left(\frac{2R_P}{r_0}\right)^{5/3}\sqrt{(n+1)(n'+1)}(-1)^{(n+n'-2m)/2}\delta_{mm'}\times \\ &\quad \int d\kappa\kappa^{-14/3}J_{n+1}(\kappa)J_{n'+1}(\kappa), \qquad i-i' \text{ even} && (2.126)\\ &= 0, \qquad i-i' \text{ odd}. && (2.127)\end{aligned}$$

It is important to highlight a few simplifications required for the form in (2.126). The extra factor of 2 in $\left(\frac{2R_P}{r_0}\right)^{5/3}$ was added for convenience. The expression now relates Zernike coefficient covariance to the ratio of the pupil *diameter*, D_P, to r_0. The ratio $\frac{D_P}{r_0}$ is dominates the Zernike variance expression within the inertial range. For this reason, the ratio will be used as an indicator of turbulence strength throughout the remainder of this document. Also, several simplifications reducing the integral over ψ are made possible by the symmetry in the Zernike modes combined with Noll's ordering scheme. First, if $i-i'$ is odd then the integral over ψ:

$$\int_0^{2\pi}\cos(m\psi)\sin(m'\psi)d\psi, \qquad (2.128)$$

will integrate to zero, given $m, m' \in \mathbb{Z}$. By similar reasoning, only those covariances where $m = m'$ are nonzero, because:

$$\text{if } m, z \in \mathbb{Z}, \qquad (2.129)$$

$$\text{then } \int_0^{2\pi}\cos(m\psi)\cos((m+z)\psi)d\psi = 0. \qquad (2.130)$$

Hence, the $\delta_{mm'}$ term. The factor of 2π leading the expression in (2.125) is due to nonzero cases for the integral over ψ. Each nonzero case for the integral over ψ results in a factor of 2π:

$$\int_0^{2\pi} 2\cos^2(m\psi)d\psi = 2\pi, \tag{2.131}$$

$$\int_0^{2\pi} 2\sin^2(m\psi)d\psi = 2\pi, \tag{2.132}$$

$$\int_0^{2\pi} d\psi = 2\pi. \tag{2.133}$$

The term, $(-1)^{(n+n'-2m)/2}$, results from the following cases:

$$\left((-1)^{(n-m)/2}\mathrm{j}^m\right)^* (-1)^{(n'-m)/2}\mathrm{j}^m = (-1)^{(n-m)/2}(-1)^{(n'-m)/2}, \quad \text{for } m \text{ even,} \tag{2.134}$$

and
$$\left((-1)^{(n-m)/2}\mathrm{j}^m\right)^* (-1)^{(n'-m)/2}\mathrm{j}^m = -\mathrm{j}(\mathrm{j})(-1)^{(n-m)/2}(-1)^{(n'-m)/2}, \quad \text{for } m \text{ odd.} \tag{2.135}$$

The remaining integral can be solved via table [17]. The form is as follows:

$$\int_0^\infty J_\nu(\alpha t)J_\mu(\alpha t)t^{-\lambda}dt = \frac{\alpha^{\lambda-1}\Gamma(\lambda)\Gamma\left(\frac{\nu+\mu-\lambda+1}{2}\right)}{2^\lambda\Gamma\left(\frac{-\nu+\mu+\lambda+1}{2}\right)\Gamma\left(\frac{\nu+\mu+\lambda+1}{2}\right)\Gamma\left(\frac{\nu-\mu+\lambda+1}{2}\right)} \tag{2.136}$$

Letting $\nu = n+1$, $\mu = n'+1$, $\lambda = \frac{14}{3}$, and $\alpha = 1$, the final simplified solution to the covariance is:

$$E\left\{a_i^* a_{i'}\right\}\left(\frac{D_P}{r_0}\right)^{-5/3} = 0.4898 \cdot 2^{4/3}\pi\sqrt{(n+1)(n'+1)}(-1)^{(n+n'-2m)/2}\delta_{mm'} \times \frac{\Gamma(\frac{14}{3})\Gamma\left(\frac{n+n'-\frac{5}{3}}{2}\right)}{2^{14/3}\Gamma\left(\frac{n'-n+\frac{17}{3}}{2}\right)\Gamma\left(\frac{n+n'+\frac{23}{3}}{2}\right)\Gamma\left(\frac{n-n'+\frac{17}{3}}{2}\right)}. \tag{2.137}$$

This result is convenient for generating entries in the covariance matrix for the Zernike coefficients. Table 2.5 provides the scaled covariance results, $E\left\{a_i^* a_{i'}\right\}\left(\frac{D_P}{r_0}\right)^{-5/3}$, for the

	2	3	4	5	6	7	8	9	10	11
2	0.449	0	0	0	0	0	-0.0141	0	0	0
3	0	0.449	0	0	0	-0.0141	0	0	0	0
4	0	0	0.0232	0	0	0	0	0	0	0
5	0	0	0	0.0232	0	0	0	0	0	0
6	0	0	0	0	0.0232	0	0	0	0	0
7	0	-0.0141	0	0	0	0.00619	0	0	0	0
8	-0.0141	0	0	0	0	0	0.00619	0	0	0
9	0	0	0	0	0	0	0	0.00619	0	0
10	0	0	0	0	0	0	0	0	0.00619	0
11	0	0	0	0	0	0	0	0	0	0.00245

Table 2.5 Normalized Zernike coefficient covariance: $L_0 = \infty$, $l_0 = 0$.

first 11 Zernike coefficients neglecting piston. This table reflects the results from Noll for the case of infinite outer scale and inner scale equal to 0. There is no analytical result to the expression for $E\{a_i^* a_{i'}\}$ that includes the effects of von Kármán inner and outer scale, but it is easy enough to approximate the integration numerically. Forming the covariance integral with the von Kármán turbulence spectrum yields:

$$\begin{aligned} E\{a_i^* a_{i'}\} &= 0.4898 \cdot 2^{4/3}\pi \left(\frac{2R_P}{r_0}\right)^{5/3} \sqrt{(n+1)(n'+1)}(-1)^{(n+n'-2m)/2}\delta_{mm'} \times \\ &\quad \int d\kappa \frac{J_{n+1}(\kappa)J_{n'+1}(\kappa)}{\kappa(\kappa^2 + R_P^2\kappa_0^2)^{11/6}} \exp\left(\frac{-\kappa^2}{R_P^2\kappa_m^2}\right), \qquad i - i' \text{ even} \qquad (2.138) \\ &= 0, \qquad i - i' \text{ odd}. \qquad (2.139) \end{aligned}$$

Tables 2.6 and 2.7 demonstrate the results on lower order Zernike variance due to finite outer scale and inner scale greater than zero. Table 2.6 shows the effects of varying outer scale on the first 11 Zernike modes when inner scale is 0. Table 2.7 shows the effects of varying of inner scale when outer scale is infinite. These results show that the analytical covariance result serves as a guideline only and the numerical covariance should be used as the ideal reference in situations where the outer scale and inner scale are known or where they may be estimated.

This section provided a modal decomposition for the wavefront using the Zernike polynomial basis set. Expressions (2.137) and (2.138) demonstrate that, given the Fried parameter, the aperture diameter, and the inertial range, one can construct a model for the variance of the Zernike modes. The variance of each Zernike mode is directly related

	L_0 [m]							
	10^0	10^1	10^2	10^3	10^4	10^5	10^6	10^7
$Z_{2,3}$	0.171	0.317	0.388	0.421	0.436	0.443	0.446	0.448
Z_{4-6}	0.0220	0.0232	0.0232	0.0232	0.0232	0.0232	0.0232	0.0232
Z_{7-10}	0.00610	0.00619	0.00619	0.00619	0.00619	0.00619	0.00619	0.00619
Z_{11}	0.00244	0.00245	0.00245	0.00245	0.00245	0.00245	0.00245	0.00245
$l_0 = 0$, $r_0 = 0.088$m, $D_P = 0.088$m								

Table 2.6 Zernike variance versus L_0.

	l_0 [m]				
	$r_0/50$	$r_0/10$	$r_0/5$	$r_0/3$	$r_0/2$
$Z_{2,3}$	0.449	0.448	0.447	0.444	0.438
Z_{4-6}	0.0232	0.0230	0.0224	0.0211	0.0190
Z_{7-10}	0.00619	0.00607	0.00573	0.00504	0.00399
Z_{11}	0.00245	0.00237	0.00215	0.00173	0.00117
$L_0 = \infty$, $r_0 = 0.088$m, $D_P = 0.088$m					

Table 2.7 Zernike variance versus l_0.

to the power spectral density of the random atmospheric phase distortion. Clearly this statistical model aids in the construction of the wavefront sensor described earlier. The parameter space for each coefficient theoretically maps to the entire real line because each one is modeled as a zero mean Gaussian random variable. Given the variance however, the sensor algorithm may choose to truncate the parameter space to a smaller portion of the real line containing the bulk of the probability mass. This sets the range of possible estimates for a maximum likelihood estimation process. Further, the Gaussian prior density may be used to form a maximum a posteriori estimator. With the parameter variance calculations in hand, all that remains is to offer some method for discerning the incoming wavefront from intensity measurements. This leads to the final background topic: a brief review of the linear systems optics model and the concept of an optical transfer function.

2.4 The Optical Transfer Function (OTF)

Simulating optical wave propagation and the wavefront sensor environment using the mathematical convenience of linear systems theory will require applying a few assumptions to electromagnetic wave theory. It would be quite beneficial, for instance, to be able to apply the superposition and convolution properties that apply in the realm of linear systems theory to the generic optical system. Maxwell's equations describe the physical properties of electromagnetic waves. The set of assumptions for wave propagation using linear systems begins with those assumptions necessary to derive a scalar wave equation. The scalar wave equation is then combined with diffraction theory to create an integral formulation of the optical field at some aperture location. Simplifying the integral formulation leads to a duality between the field at one aperture location and its Fourier domain representation at some originating aperture. Thus, under reasonable assumptions, an optical system image plane can be considered a convolution of the geometric image and the diffraction pattern created by the aperture. The series of important results to follow are by no means a thorough treatment of Fourier optics, but should provide enough highlights to reinforce the concepts that will be necessary to simulate optical propagation through the atmosphere and the interaction with the wavefront sensor. The following derivations are summary of the results presented in the popular works of Goodman [20] and Born [19].

Beginning with Maxwell's equations, assume that the medium is linear, isotropic, homogeneous, nondispersive and nonmagnetic. Linearity in the medium may be explained by describing the medium as a system with complex fields as inputs and outputs. The property of linearity applies to the system, f, if the following superposition holds for all functions u_1 and u_2 and all complex constants a and b:

$$f\left\{au_1\left(P\right)+bu_2\left(P\right)\right\}=af\left\{u_1\left(P\right)\right\}+bf\left\{u_2\left(P\right)\right\}. \tag{2.140}$$

Under this assumption, the resulting field propagating from the sum of two scaled source fields is equivalent to summing the scaled results of the two source fields propagated independently through the medium. Isotropic indicates that the propagation is independent of direction of polarization of the field. Homogeneity indicates that the permittivity, ϵ, is constant. The term nondispersive means that the permittivity, μ, is not a function of wavelength. Lastly, the medium is assumed to be nonmagnetic meaning that the medium has vacuum permeability, $\mu=\mu_0$. Under these assumptions, the solution to Maxwell's equations reduces to the scalar wave equation:

$$\nabla^2 u(P,t)-\frac{n^2}{c^2}\frac{\partial^2 u(P,t)}{\partial t^2}=0, \tag{2.141}$$

where $n = \sqrt{\frac{\epsilon}{\epsilon_0}}$ is the index of refraction, $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$is the vacuum speed of light, P indicates spatial location, u is any component of the vector fields: $\mathcal{E}$, or $\mathcal{H}$, and t is time. Representing the field u as the real part of a complex phasor U gives:

$$\begin{aligned} U(P) &= A(P)\exp\{-\mathrm{j}\phi(P)\}; && (2.142) \\ u(P,t) &= \mathrm{Re}\{U(P)\}, && (2.143) \\ &= A(P)\cos\{2\pi\nu t + \phi(P)\}. && (2.144) \end{aligned}$$

Forcing the field to satisfy the scalar wave equation produces the familiar time independent Helmholtz equation as follows:

$$\begin{aligned} \nabla^2 u(P,t) - \frac{n^2}{c^2} A(P) \frac{\partial^2}{\partial t^2}\left[\cos\{2\pi\nu t + \phi(P)\}\right] &= 0, && (2.145) \\ \nabla^2 u(P,t) + (2\pi)^2 \nu^2 \frac{n^2}{c^2} A(P)\cos\{2\pi\nu t + \phi(P)\} &= 0, && (2.146) \\ (\nabla^2 + k^2)u &= 0. && (2.147) \end{aligned}$$

The integral theorem of Helmholtz and Kirchoff can be developed (see Goodman chapter 3) from the divergence theorem of Gauss [21]:

$$\iiint\limits_V \left(U\nabla^2 G - G\nabla^2 U\right) dv = \iint\limits_S \left(\left(\hat{n}\cdot\nabla U\right)G - U\left(\hat{n}\cdot\nabla G\right)\right) ds, \tag{2.148}$$

and the Helmholtz equation, (2.147), which provides a relationship between the field at a point and the closed surface around the point [20]:

$$U(P_0) = \frac{1}{4\pi}\iint\limits_S \left(\left(\hat{n}\cdot\nabla U\right)G - U\left(\hat{n}\cdot\nabla G\right)\right) ds. \tag{2.149}$$

Where the surface S is some surface surrounding the point P_0, and the expression $\hat{n}\cdot\nabla(\cdot)$ is equivalent to the derivative taken normal to S. This crucial relation is the key to wave optical simulation. It provides an integral formulation for the field at some boundary or aperture at a distance from a known source. The choice of Green's function, G, is critical. Rayleigh and Sommerfeld are attributed with the formulation in Figure 2.5 which suggests the Half Space Green's function of the form:

$$G(P_1) = \frac{\exp(\mathrm{j}kr_{01})}{r_{01}} - \frac{\exp(\mathrm{j}k\tilde{r}_{01})}{\tilde{r}_{01}}. \tag{2.150}$$

This choice of G represents the linear combination of fields from two sources at P and $\tilde{P}$

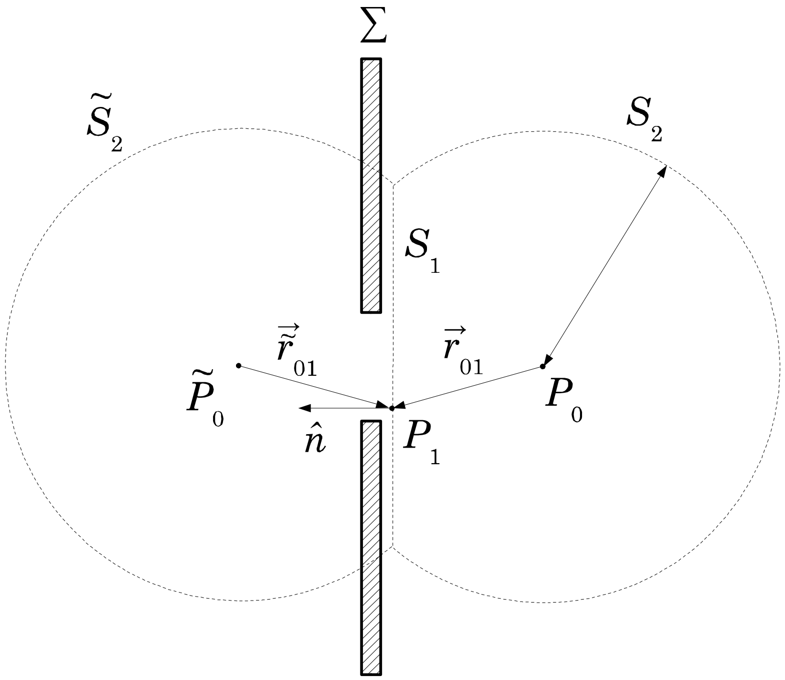


Figure 2.5 Rayleigh-Sommerfeld formulation of diffraction by a plane screen [20].

oscillating 180 out of phase. Notice the change in integration limits in Figure 2.5. According to the diagram, the integral must now be evaluated over S_1, S_2 and $\tilde{S}_2$. Sommerfeld simplified the limits of integration by assuming that the field U vanishes at least as fast as a diverging spherical wave. This assumption, known as the Sommerfeld radiation condition, reduces the integral over all of the dashed surface: $S = \bigcup \left\{S_1, S_2, \tilde{S}_2\right\}$, to an integral over the plane of the aperture, S_1:

$$\lim_{R\to\infty} \left(R\frac{\partial U}{\partial n} - jkU \right) = 0, \tag{2.151}$$

$$U(P_0) = \frac{1}{4\pi} \iint\limits_{S_1} \left((\hat{n} \cdot \nabla U) \, G - U \, (\hat{n} \cdot \nabla G) \right) ds. \tag{2.152}$$

Substituting G from (2.150) into the integral theorem of Helmholtz and Kirchoff, (2.149), and assuming that the Sommerfeld radiation condition holds, yields the following integral over the plane of the screen:

$$U(P_0) = \frac{1}{\mathrm{j}\lambda} \iint\limits_{S_1} U(P_1) \frac{\exp(\mathrm{j}kr_{01})}{r_{01}} \cos(\hat{n}, \mathbf{r}_{01}) ds, \tag{2.153}$$

called the Rayleigh-Sommerfeld diffraction formula. Note that the vector $\hat{n}$ is the direction normal to the aperture and the cosine expression with two vector arguments is shorthand for the cosine of the angle between the two argument vectors. Using the Rayleigh-Sommerfeld diffraction formula for optical propagation, the simple thin lens imaging system may be conveniently modeled as two propagations: one from the object to the aperture plane and one from the aperture plane to the image plane. This model will become the basis for discussions to come, and as such, warrants a defined coordinate system. Figure 2.6 portrays the basic three plane imaging system model and respective Cartesian coordinate system labeling. Assuming a spherical wavefront and converting to Cartesian coordinates the propagation integral in (2.153) may be rewritten as:

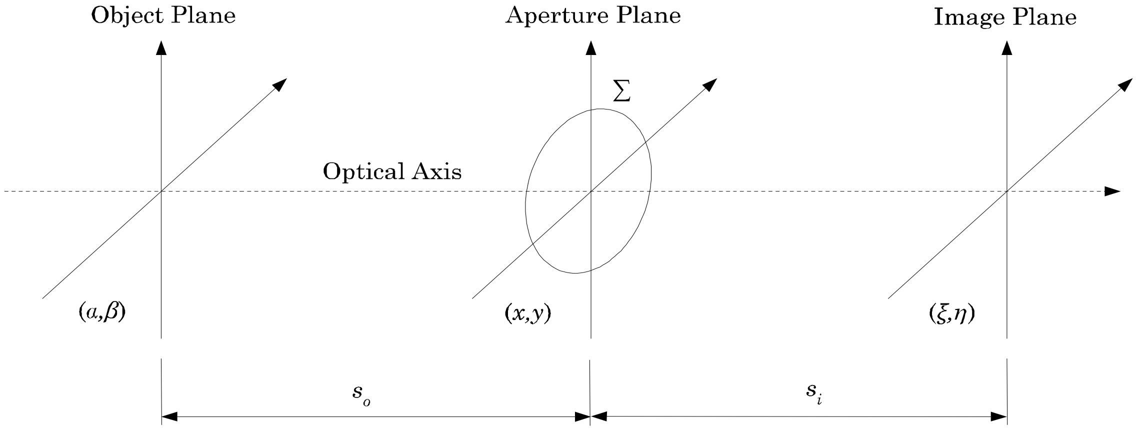


Figure 2.6 Three plane Cartesian coordinate system.

$$U(\xi, \eta) = \frac{s_i}{\mathrm{j}\lambda} \iint\limits_{\Sigma} U(x, y) \frac{\exp(\mathrm{j}kr_{01})}{r_{01}^2} dxdy, \tag{2.154}$$

$$\text{where } r_{01} = \sqrt{s_i^2 + (x - \xi)^2 + (y - \eta)^2}, \tag{2.155}$$

and the cosine of the normal angle has been replaced by the small angle approximation:

$$\cos(\hat{n}, \mathbf{r}_{01}) = \frac{s_i}{r_{01}}. \tag{2.156}$$

The linear systems approach will require an approximation for the square root in the r_{01} terms. The radial distance in the exponential phase term, $\exp(\mathrm{j}kr_{01})$, and in the denominator, $\frac{1}{r_{01}^2}$, may each be replaced by a truncated binomial series expansion. The binomial expansion for r_{01} is shown here:

$$\begin{aligned} r_{01} &= s_i\sqrt{1+\frac{(x-\xi)^2}{s_i^2}+\frac{(y-\eta)^2}{s_i^2}}, && (2.157) \\ \sqrt{1+b} &= 1+\frac{1}{2}b-\frac{1}{8}b^2+..., && (2.158) \\ r_{01} &= s_i\left(\begin{array}{c} 1+\frac{1}{2}\left(\frac{(x-\xi)^2}{s_i^2}+\frac{(y-\eta)^2}{s_i^2}\right)- \\ \frac{1}{8}\left(\frac{(x-\xi)^2}{s_i^2}+\frac{(y-\eta)^2}{s_i^2}\right)^2+... \end{array}\right). && (2.159) \end{aligned}$$

Retaining the first term yields the approximation:

$$r_{01} \approx s_i. \tag{2.160}$$

Substituting this approximation for r_{01} into the denominator term gives:

$$\frac{1}{r_{01}^2} \approx \frac{1}{s_i^2}. \tag{2.161}$$

Retaining the first two terms yields the approximation:

$$r_{01} \approx s_i + \frac{1}{2s_i}\left((x-\xi)^2+(y-\eta)^2\right). \tag{2.162}$$

Substituting this approximation for r_{01} into the phase term gives:

$$\begin{aligned} \exp(\mathrm{j}kr_{01}) &\approx \exp\left(\mathrm{j}k\left[s_i+\frac{1}{2s_i}\left((x-\xi)^2+(y-\eta)^2\right)\right]\right), && (2.163) \\ &= \exp\left(\mathrm{j}ks_i\right)\exp\left(\frac{\mathrm{j}k}{2s_i}(x-\xi)^2\right)\exp\left(\frac{\mathrm{j}k}{2s_i}(y-\eta)^2\right). && (2.164) \end{aligned}$$

Making both substitutions in (2.154) and simplifying produces the following two familiar forms of the Fresnel diffraction integral:

$$\begin{aligned} U(\xi,\eta) &= \frac{\exp(\mathrm{j}ks_i)}{\mathrm{j}\lambda s_i}\exp\left\{\mathrm{j}\frac{k}{2s_i}(\xi^2+\eta^2)\right\}\times \\ &\quad \iint\limits_{-\infty}^{\infty} U(x,y)\exp\left\{\mathrm{j}\frac{k}{2s_i}(x^2+y^2)\right\}\exp\left\{-\mathrm{j}\frac{2\pi}{\lambda s_i}(\xi x+\eta y)\right\}dxdy, \quad (2.165) \\ &= \frac{\exp(\mathrm{j}ks_i)}{\mathrm{j}\lambda s_i}\iint\limits_{-\infty}^{\infty} U(x,y)\exp\left\{\mathrm{j}\frac{k}{2s_i}\left[(\xi-x)^2+(\eta-y)^2\right]\right\}dxdy. \quad (2.166) \end{aligned}$$

Under certain conditions, the quadratic phase term in the Fresnel integral can be removed. If a converging spherical lens is in place, for instance, then the quadratic phase is exactly canceled by the focusing properties of the lens. Also, in cases where the aperture does not contain a converging lens but the propagation is applied over a sufficiently long distance, the effects of the quadratic phase over some small aperture becomes negligible. The propagation distance at which the quadratic phase term in the Fresnel integral becomes negligible is often referred to as the far-field condition:

$$s_i > \frac{2D_P^2}{\lambda}. \quad (2.167)$$

Assuming that far-field conditions hold, the integral in (2.166) simplifies to the Fraunhofer diffraction integral:

$$U(\xi,\eta) = \frac{\exp(\mathrm{j}ks_i)\exp\left\{\mathrm{j}\frac{k}{2s_i}(\xi^2+\eta^2)\right\}}{\mathrm{j}\lambda s_i}\iint\limits_{-\infty}^{\infty} U(x,y)\exp\left\{-\mathrm{j}\frac{2\pi}{\lambda s_i}(\xi x+\eta y)\right\}dxdy. \quad (2.168)$$

Under these circumstances, the solution for the image field, $U(\xi,\eta)$, becomes a scaled Fourier transform. Using this theory, a single spherical lens optical system can be modeled as a linear filter, where the impulse response for the system is the Fraunhofer diffraction integral of the system pupil. In order to recognize this property and the additional assumptions required, consider first, the assumption mentioned previously, that the medium is linear and therefore the following superposition holds for monochromatic light:

$$\begin{aligned}
U_i(\xi,\eta) &= \iint\limits_{-\infty}^{\infty} h(\xi-\alpha,\eta-\beta)U_o(\alpha,\beta)d\alpha d\beta, && (2.169)\\
\text{where } U_i &\equiv \text{field in the image plane,} && (2.170)\\
U_o &\equiv \text{field in the object plane,} && (2.171)\\
\text{and } h &\equiv \text{optical system impulse response.} && (2.172)
\end{aligned}$$

A side note is required before proceeding with the derivation. The wavefront sensor is designed to operate using polychromatic incoherent light. However, the case is made by Goodman that the solution for the monochromatic field can be transformed into a similar approach for polychromatic incoherent light by modeling such a system as the average of contributions from many incoherent monochromatic sources [20]. Therefore the case will be made following this derivation that an incoherent system is linear, not in the field, but rather in intensity and, as such, a similar convolution integral may be introduced for the case of incoherent light. Returning now to the monochromatic case, propagation through a single lens system can be represented as a convolution of the object field with some impulse response, h. In order to derive the impulse response, consider the response of the system in Figure 2.7 to a point source. The paraxial representation of a spherical wave at the

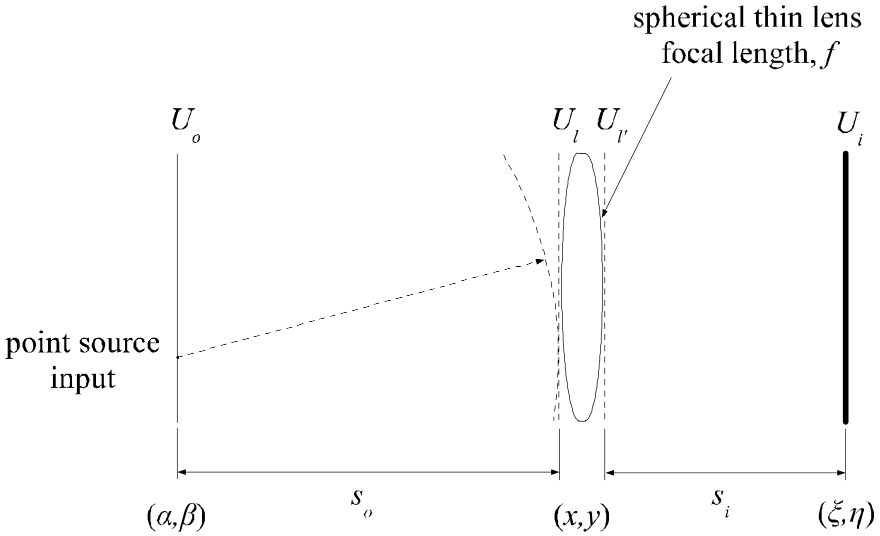


Figure 2.7 Model of a simple thin lens imaging system illuminated by a point source.

aperture plane location (x, y) emanating from the object plane at point (α, β) is given by:

$$U_l(x,y) = \frac{1}{\mathrm{j}\lambda s_o} \exp\left\{\mathrm{j}\frac{k}{2s_o}\left[(x-\alpha)^2 + (y-\beta)^2\right]\right\}. \tag{2.173}$$

The lens applies a quadratic phase and an aperture window function, W_P:

$$W_P(x,y;R_P) \exp\left\{-\mathrm{j}\frac{k}{2f}\left(x^2+y^2\right)\right\}. \tag{2.174}$$

Combining (2.173) and (2.174) in the second form of the Fresnel integral in (2.166), the impulse response is given by (neglect constant phase terms):

$$\begin{aligned} h(\xi,\eta;\alpha,\beta) &= \iint\limits_{-\infty}^{\infty} \frac{W_P(x,y;R_P)}{\lambda s_i \lambda s_o} \exp\left\{\mathrm{j}\frac{k}{2s_o}\left[(x-\alpha)^2 + (y-\beta)^2\right]\right\} \times \\ &\quad \exp\left\{-\mathrm{j}\frac{k}{2f}\left(x^2+y^2\right)\right\} \exp\left\{\mathrm{j}\frac{k}{2s_i}\left[(\xi-x)^2 + (\eta-y)^2\right]\right\} dxdy. \end{aligned} \tag{2.175}$$

Expanding and analyzing similar terms within the three phase components yields:

$$\begin{aligned} \exp\left\{\mathrm{j}\frac{k}{2s_o}\left[(x-\alpha)^2 + (y-\beta)^2\right]\right\} &= \exp\left\{\mathrm{j}\frac{k}{2s_o}\left(x^2+y^2\right)\right\} \exp\left\{-\mathrm{j}k\left(\frac{x\alpha}{s_o} + \frac{y\beta}{s_o}\right)\right\} \times \\ &\quad \exp\left\{\mathrm{j}\frac{k}{2s_o}\left(\alpha^2+\beta^2\right)\right\}, \end{aligned} \tag{2.176}$$

$$\exp\left\{-\mathrm{j}\frac{k}{2f}\left(x^2+y^2\right)\right\} = \exp\left\{-\mathrm{j}\frac{k}{2s_o}\left(x^2+y^2\right)\right\} \exp\left\{-\mathrm{j}\frac{k}{2s_i}\left(x^2+y^2\right)\right\}, \tag{2.177}$$

$$\begin{aligned} \exp\left\{\mathrm{j}\frac{k}{2s_i}\left[(\xi-x)^2 + (\eta-y)^2\right]\right\} &= \exp\left\{\mathrm{j}\frac{k}{2s_i}\left(\xi^2+\eta^2\right)\right\} \exp\left\{-\mathrm{j}k\left(\frac{x\xi}{s_i} + \frac{y\eta}{s_i}\right)\right\} \times \\ &\quad \exp\left\{\mathrm{j}\frac{k}{2s_i}\left(x^2+y^2\right)\right\}. \end{aligned} \tag{2.178}$$

If the image plane is placed such that: $\frac{1}{f} = \frac{1}{s_o} + \frac{1}{s_i}$, then the quadratic phase in x and y is cancelled by the quadratic phase contribution of the lens. Furthermore, assuming that the quadratic phase in α and β is nearly zero over the region of the image plane effected by the point source, allows that quadratic to be removed as well. After making these simplifications, the remaining integral is given:

$$\begin{aligned} h(\xi,\eta;\alpha,\beta) &= \frac{\exp\left\{\mathrm{j}\frac{k}{2s_i}\left(\xi^2+\eta^2\right)\right\}}{\lambda^2 s_i s_o} \times \\ &\iint\limits_{-\infty}^{\infty} W_P(x,y;R_P)\exp\left\{-\mathrm{j}k\left[\left(\frac{\xi}{s_i}+\frac{\alpha}{s_o}\right)x+\left(\frac{\eta}{s_i}+\frac{\beta}{s_o}\right)y\right]\right\}dxdy. \end{aligned} \tag{2.179}$$

Now define the transverse magnification to be: $M = -\frac{s_i}{s_o}$, and make the coordinate changes: $\tilde{\alpha} = M\alpha$, $\tilde{\beta} = M\beta$, $\tilde{x} = \frac{x}{\lambda s_i}$, $\tilde{y} = \frac{y}{\lambda s_i}$, and $\tilde{h} = \frac{1}{|M|}h$:

$$\begin{aligned} h\left(\xi,\eta;\tilde{\alpha},\tilde{\beta}\right) &= \frac{\exp\left\{\mathrm{j}\frac{k}{2s_i}\left(\xi^2+\eta^2\right)\right\}}{\lambda^2 s_i s_o} \times \\ &\iint\limits_{-\infty}^{\infty} \lambda s_i \lambda s_i d\tilde{x}d\tilde{y} W_P(\lambda s_i\tilde{x},\lambda s_i\tilde{y};R_P) \times \\ &\exp\left\{-\mathrm{j}\frac{2\pi}{\lambda}\left[\begin{array}{c}\left(\frac{\xi}{s_i}+\left(-\frac{s_o}{s_i}\right)\frac{\tilde{\alpha}}{s_o}\right)\lambda s_i\tilde{x}+ \\ \left(\frac{\eta}{s_i}+\left(-\frac{s_o}{s_i}\right)\frac{\tilde{\beta}}{s_o}\right)\lambda s_i\tilde{y}\end{array}\right]\right\}, \end{aligned} \tag{2.180}$$

$$\begin{aligned} \tilde{h}\left(\xi,\eta;\tilde{\alpha},\tilde{\beta}\right) &= \exp\left\{\mathrm{j}\frac{k}{2s_i}\left(\xi^2+\eta^2\right)\right\}\iint\limits_{-\infty}^{\infty} W_P(\lambda s_i\tilde{x},\lambda s_i\tilde{y};R_P) \times \\ &\exp\left\{-\mathrm{j}2\pi\left[\left(\xi-\tilde{\alpha}\right)\tilde{x}+\left(\eta-\tilde{\beta}\right)\tilde{y}\right]\right\}d\tilde{x}d\tilde{y}. \end{aligned} \tag{2.181}$$

Making the appropriate variable change from α and β to $\tilde{\alpha}$ and $\tilde{\beta}$ in (2.169) h becomes $\tilde{h}$ and the field in the image plane of a converging spherical lens system can be represented as:

$$U_i(\xi,\eta) = \iint\limits_{-\infty}^{\infty} \tilde{h}(\xi-\tilde{\alpha},\eta-\tilde{\beta})U_g(\tilde{\alpha},\tilde{\beta})d\tilde{\alpha}d\tilde{\beta}, \tag{2.182}$$

$$\text{where } \tilde{h}(\xi,\eta) = \exp\left\{\mathrm{j}\frac{k}{2s_i}\left(\xi^2+\eta^2\right)\right\}\iint\limits_{-\infty}^{\infty} W_P(\lambda s_i\tilde{x},\lambda s_i\tilde{y};R_P)\exp\left\{-\mathrm{j}2\pi\left(\xi\tilde{x}+\eta\tilde{y}\right)\right\}d\tilde{x}d\tilde{y}, \tag{2.183}$$

$$M \equiv \text{transverse magnification}, \tag{2.184}$$

$$\text{and } U_g(\tilde{\alpha},\tilde{\beta}) = \frac{1}{|M|}U_o\left(\frac{\tilde{\alpha}}{M},\frac{\tilde{\beta}}{M}\right). \tag{2.185}$$

Thus, for the monochromatic case, the image field is a convolution of the image field predicted by geometric optics, $U_g(\tilde{\alpha}, \tilde{\beta})$, and the amplitude impulse response, $\tilde{h}(\xi, \eta)$. Where the amplitude impulse response of the optical system, given in (2.183) above, is the Fraunhofer diffraction integral applied to the pupil window, $W_P(x, y)$.

If the input source is polychromatic incoherent light, modeled as the average of many contributions from incoherent monochromatic sources then it follows that the incoherent imaging system is linear in *intensity* [20]. Under this condition, the field quantities are replaced by field intensities and the amplitude impulse response becomes an intensity impulse response, $\left|\tilde{h}(\xi, \eta)\right|^2$, which is the magnitude squared Fraunhofer diffraction pattern. Therefore, for incoherent light, the spatial convolution integral is given by [20]:

$$\begin{aligned} I_i(\xi, \eta) &= \kappa \iint_{-\infty}^{\infty} \left|\tilde{h}(\xi - \tilde{\alpha}, \eta - \tilde{\beta})\right|^2 I_g(\tilde{\alpha}, \tilde{\beta}) d\tilde{\alpha} d\tilde{\beta}, && (2.186) \\ \kappa &\equiv \text{real scaling constant.} && (2.187) \end{aligned}$$

The dual of this expression in the spatial frequency domain is:

$$\mathcal{G}_i(f_X, f_Y) = \mathcal{H}(f_X, f_Y)\mathcal{G}_g(f_X, f_Y). \tag{2.188}$$

Where $\mathcal{G}_g$ and $\mathcal{G}_i$ are the normalized frequency domain transforms of the geometric and diffraction image intensities, and $\mathcal{H}$ is the transform of the impulse response, commonly referred to as the Optical Transfer Function (OTF):

$$\mathcal{G}_g(f_X, f_Y) = \frac{\displaystyle\iint\limits_{-\infty}^{\infty} I_g(\xi, \eta) \exp\{-\mathrm{j}2\pi(f_X\xi + f_Y\eta)\}\, d\xi d\eta}{\displaystyle\iint\limits_{-\infty}^{\infty} I_g(\xi, \eta)\, d\xi d\eta}, \tag{2.189}$$

$$\mathcal{G}_i(f_X, f_Y) = \frac{\displaystyle\iint\limits_{-\infty}^{\infty} I_i(\xi, \eta) \exp\{-\mathrm{j}2\pi(f_X\xi + f_Y\eta)\}\, d\xi d\eta}{\displaystyle\iint\limits_{-\infty}^{\infty} I_i(\xi, \eta)\, d\xi d\eta}, \tag{2.190}$$

$$\mathcal{H}(f_X, f_Y) = \frac{\displaystyle\iint\limits_{-\infty}^{\infty} |h(\xi, \eta)|^2 \exp\{-\mathrm{j}2\pi(f_X\xi + f_Y\eta)\}\, dudv}{\displaystyle\iint\limits_{-\infty}^{\infty} |h(\xi, \eta)|^2\, d\xi d\eta}. \tag{2.191}$$

As a consequence of the linear systems assumptions, the OTF provides a relationship between the pupil phase and the image intensity. For incoherent imaging, the OTF is the normalized autocorrelation of the pupil function:

$$\mathcal{H}(f_X, f_Y) = \frac{\displaystyle\iint\limits_{-\infty}^{\infty} \mathcal{P}(x + \tfrac{\lambda s_i f_X}{2}, y + \tfrac{\lambda s_i f_Y}{2})\mathcal{P}(x - \tfrac{\lambda s_i f_X}{2}, y - \tfrac{\lambda s_i f_Y}{2})dxdy}{\displaystyle\iint\limits_{-\infty}^{\infty} \mathcal{P}(x, y)\mathcal{P}^*(x, y)dxdy}. \tag{2.192}$$

Combining (2.192) and the modal wavefront representation in (2.89), gives a direct method for calculating the effects of any combination of Zernike modes on the OTF. Consider the example of a diffraction limited imaging system with a plane wave input. Use the circular windowing function in (2.90) and the Zernike expansion for phase to represent the pupil expression. In this special case, the OTF becomes:

$$\mathcal{P}(x, y, R_P) = W_P(x, y; R_P) \exp\left\{ \mathrm{j} \sum_{i=2}^{N} a_i Z_i\left(\frac{x}{R_P}, \frac{y}{R_P}\right) \right\}, \tag{2.193}$$

$$\mathcal{H}(f_X, f_Y) = \iint_{-\infty}^{\infty} \mathcal{P}\left(x + \frac{\lambda s_i f_X}{2}, y + \frac{\lambda s_i f_Y}{2}\right) \mathcal{P}\left(x - \frac{\lambda s_i f_X}{2}, y - \frac{\lambda s_i f_Y}{2}\right) dx dy. \tag{2.194}$$

In this example, the normalized transform of the geometric image, $\mathcal{G}_i(f_X, f_Y)$, is unity. This implies that the OTF and the image are direct Fourier transforms:

$$\mathcal{H}(f_X, f_Y) = \mathcal{G}_i(f_X, f_Y). \tag{2.195}$$

Figure 2.8 provides a visual comparison of OTFs for this example. The OTF with no aberrations is compared to the OTF under the influence of the first 6 Zernikes independently. The unique effects created by each Zernike mode provide a means for distinguishing the presence of one mode over another. Here the OTF, the Fourier transform of the image,

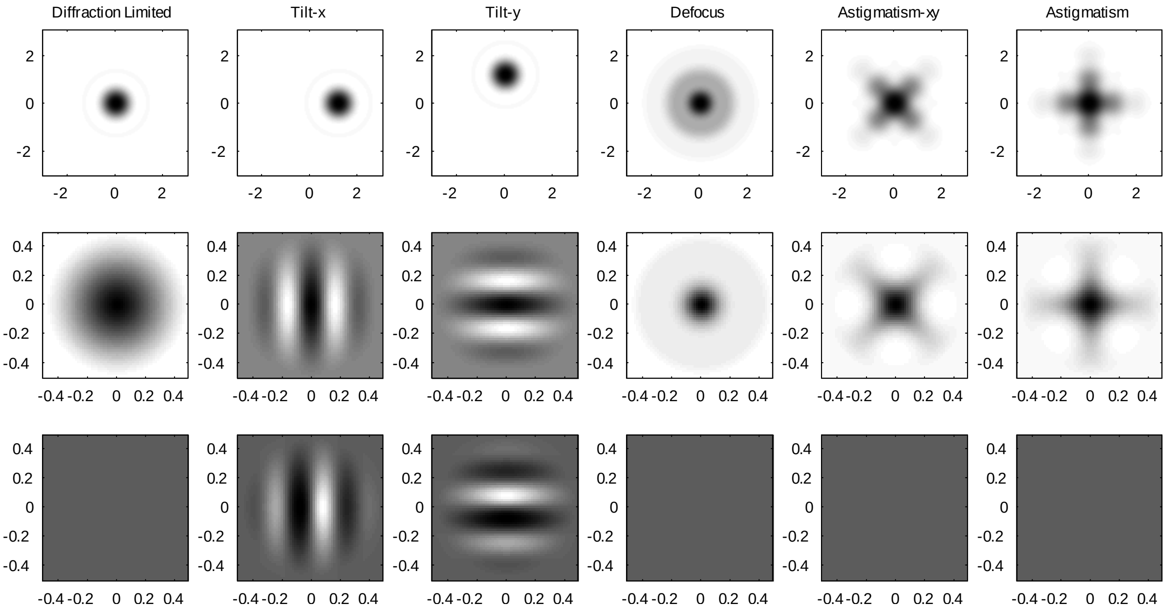


Figure 2.8 Top: Simulated point spread functions for a diffraction limited optical system and systems under independent influence from Zernikes 2-6. Middle: the real part of the Optical Transfer Functions (OTFs). Bottom: the imaginary part of the OTF.

contains the information needed to estimate the amount of each Zernike parameter present in the optical system. Granted, this is a very simplified case, but it demonstrates the relationship between pupil phase, which cannot be measured, and image intensity, which can be measured directly. Consequently, this theory reveals the possibility of estimating pupil phase from image intensity measurements.

2.5 Summary

The chapter began with a discussion on parameter estimation which concluded that parameter estimation problems require a probabilistic mapping from the parameter space to the observation space, $p_{\mathbf{r}|a}(\mathbf{R}|A)$. Additionally, parameter statistical properties such as mean and variance are helpful, while a probability distribution for the parameters, $p_a(A)$, is highly desired. In the case of the wavefront curvature estimator, the observation space, $\mathbf{R}$, is the image intensity. The wavefront phase was parameterized using the well known set of Zernike polynomial coefficients.

I carefully reviewed the origin of Kolmogorov's turbulence model and how it is related to atmospheric phase fluctuations in the optical wavelength range. The Kolmogorov model was then modified outside of the inertial range to produce the well known von Kármán turbulence model. The von Kármán statistic was used to derive both phase power spectral density and structure functions. From the turbulence model, I was able to deduce that the Zernike coefficients can be modeled by zero mean Gaussian distributions. Additionally, from Noll's efforts, the variance in each Zernike coefficient is related to atmospheric seeing conditions and the diameter of the optical system. Under these assumptions, the mean and variance of each coefficient fully describes its distribution, $p_a(A)$.

Finally, a linear model for optical wave propagation was discussed. Goodman's background on the derivation of the linear model revealed the assumptions necessary to view the optical system and free space propagation as a linear system. The linear model led to the optical transfer function and its relationship to the pupil phase. The OTF provides a means for linking the observed image intensity to the optical field in the aperture, the final ingredient needed to begin constructing a wavefront sensor.

3. The Discrete Model

The optical field and the optical system's influence on that field are naturally continuous electromagnetic wave phenomenon. For this reason, the derivations and discussions have been based on continuous variables and functions. However, there are two key reasons for converting the set of continuous mathematical constructs into a discrete parallel. First, simulation using a PC requires that all the continuous models be converted to some discrete approximation. Second, and perhaps most importantly, the digital imaging system is a naturally discrete system. Understanding the discrete nature of this type of system and its interface to the environment allows for better mathematical representation than simply sampling or approximating some analog equivalent.

This chapter begins by establishing discrete versions of the continuous reference frames and propagation integrals provided in Chapter 2. After establishing a discrete version of the linear systems techniques from Chapter 2, the discussion focuses on modeling the Charged Coupled Device array in the image plane. Modeling each detected pixel as a stochastic process will provide the final ingredient for the estimator: a probabilistic map from the detected image intensity to specific modes in the wavefront phase. Once the probabilistic map is in place, the concept of an image projection is introduced. A mathematical construct called the image projection operator is used to describe the process of extracting image projections from a set of CCD arrays. The image projection is then combined with the concepts of a discrete linear system and the detected image probability map to form the wavefront curvature estimator. The estimator derived here provides the mathematical foundation for the wavefront curvature sensor. All that will remain will be to establish a fast and efficient algorithm for evaluating the estimator expressions.

3.1 The Discrete Reference Frame

In order to discuss the discrete image space, it is best to begin by defining a complete set of discrete variables and definitions around the simple imaging system described in Section 2.4. Many of the variables and reference frames introduced here will be referred to throughout the remainder of the dissertation. Figure 3.1 provides an overview of the object, aperture and image plane axes labeling convention. Each plane is divided into grids of sample points or pixels. The pixel grids are built around the requirements that

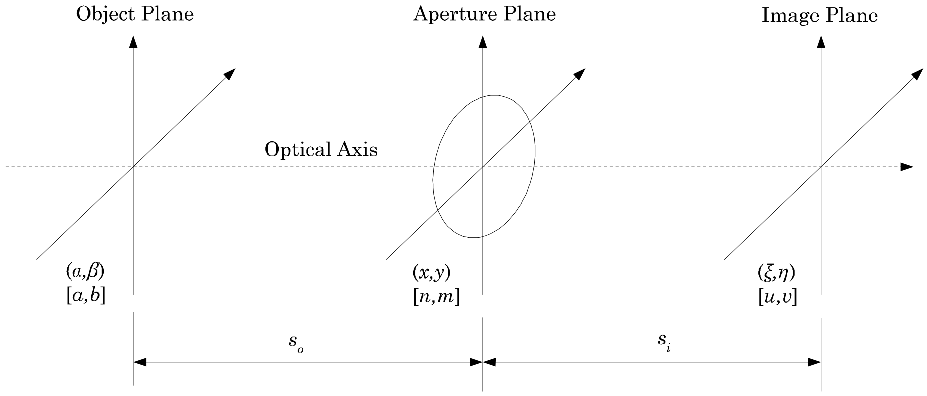


Figure 3.1 Axes labeling convention.

grids are equispaced Cartesian meshes, apertures are circular and an even number of sample points spans each side of the square grid covering the aperture. Actual dimensions of a grid pixel are superficial to functions that manipulate arrays of discrete data. In such cases, the associated index variables, $[n, m]$ or $[u, v]$ for instance, will be used versus discrete increments of (x, y) and (ξ, η) to provide a more general description. When discrete increments of continuous variables are required, the distinction from continuous variables will be made by replacing parentheses with brackets. For instance, the function $f(x)$ is assumed to exist for all x while the same function denoted $f[x]$ is meant to indicate the values of f over some sampled set of x values. Begin with the notation for the continuous aperture field $\mathcal{P}$ with arbitrary amplitude, A_P, and phase, P_ψ:

$$\mathcal{P}\left(x, y; \bar{\mathbf{a}}, R_P\right) = A_P\left(x, y\right) W_P\left(x, y; R_P\right) \exp\left\{\mathrm{j} P_\psi\left(x, y; \bar{\mathbf{a}}\right)\right\}. \tag{3.1}$$

Bold index and coordinate variables may be used to compact notation where possible. For instance, the bold variable $\mathbf{x}$ is the compact representation for the coordinate pair (x, y):

$$\mathcal{P}\left(\mathbf{x}; \bar{\mathbf{a}}, R_P\right) = A_P\left(\mathbf{x}\right) W_P\left(\mathbf{x}; R_P\right) \exp\left\{\mathrm{j} P_\psi\left(\mathbf{x}; \bar{\mathbf{a}}\right)\right\}. \tag{3.2}$$

Recall that the pupil phase can be represented by a series of weighted Zernike polynomials:

$$\begin{aligned} P_\psi\left(\mathbf{x}; \bar{\mathbf{a}}\right) &= \sum_{i=1}^{\infty} a_i Z_i\left(\mathbf{x}, R_P\right), && (3.3) \\ \text{where } \bar{\mathbf{a}} &\equiv \text{the infinite set of Zernike coefficients.} && (3.4) \end{aligned}$$

Assuming that the pupil is circular, the window function, $W_P(x, y; R_P)$, has amplitude 1 over the circle with radius R_P and zero outside the circle. The discrete coordinates $[n, m]$ can be derived from the continuous coordinates via the relationship: $(x, y) = (n\Delta x, m\Delta y)$. Thus, in the discrete coordinate frame, the aperture field with atmospheric phase aberrations is given by:

$$\mathcal{P}[n, m; \bar{\mathbf{a}}, R_P] = A_P[n, m] W_P[n, m; R_P] \exp\{\mathrm{j} P_\psi[n, m; \bar{\mathbf{a}}]\}. \tag{3.5}$$

Once again, more compactly:

$$\mathcal{P}[\mathbf{n}; \bar{\mathbf{a}}, R_P] = A_P[\mathbf{n}] W_P[\mathbf{n}; R_P] \exp\{\mathrm{j} P_\psi[\mathbf{n}; \bar{\mathbf{a}}]\}. \tag{3.6}$$

In simulation, the plane grids must be comprised of equispaced Cartesian samples. The circular aperture weighting function is modeled as accurately as possible. The diagram in Figure 3.2 demonstrates a 16 × 16 aperture weighting function W_P. Notice that pixels along the edge of the aperture mask are weighted proportional to the amount of the aperture included within the area of the pixel.

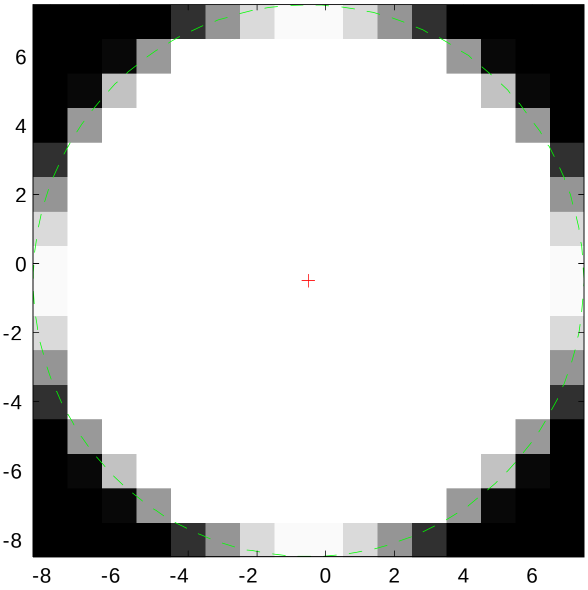


Figure 3.2 Example aperture mask for a 16 × 16 pixel aperture grid.

The Background chapter closed with a derivation of the linear systems model for propagation. The linear systems model depicted the intensity impulse response of a simple thin lens as the magnitude squared of the Fraunhofer diffraction integral applied to the aperture window function:

$$\begin{aligned}\left|\tilde{h}(\xi,\eta)\right|^2 &= \left|\begin{array}{c}\exp\left\{\mathrm{j}\frac{k}{2s_i}\left(\xi^2+\eta^2\right)\right\}\displaystyle\iint\limits_{-\infty}^{\infty} W_P(\lambda s_i\tilde{x},\lambda s_i\tilde{y};R_P)\times \\ \exp\left\{-\mathrm{j}2\pi\left(\xi\tilde{x}+\eta\tilde{y}\right)\right\}d\tilde{x}d\tilde{y}\end{array}\right|^2, &(3.7)\\ &= \left|\iint\limits_{-\infty}^{\infty} W_P(\lambda s_i\tilde{x},\lambda s_i\tilde{y};R_P)\exp\left\{-\mathrm{j}2\pi\left(\xi\tilde{x}+\eta\tilde{y}\right)\right\}d\tilde{x}d\tilde{y}\right|^2. &(3.8)\end{aligned}$$

The magnitude of the leading complex exponential is unity. The remaining integral is a scaled Fourier transform which I will denote by $\mathcal{F}_s\left\{\cdot\right\}$. Using this integral transform, the intensity impulse response for the complex pupil expression can be formed:

$$\left|\tilde{h}\left(\xi,\eta;\bar{\mathbf{a}},R_P\right)\right|^2 = \left|\mathcal{F}_s\left\{\mathcal{P}\left(\mathbf{x};\bar{\mathbf{a}},R_P\right)\right\}\right|^2. \tag{3.9}$$

The intensity in the image plane is then the convolution of the image intensity predicted by geometric optics and the intensity impulse response:

$$I_i(\xi,\eta;\bar{\mathbf{a}},R_P) = \kappa\iint\limits_{-\infty}^{\infty}\left|\tilde{h}(\xi-\tilde{\alpha},\eta-\tilde{\beta};\bar{\mathbf{a}},R_P)\right|^2 I_g(\tilde{\alpha},\tilde{\beta})d\tilde{\alpha}d\tilde{\beta}. \tag{3.10}$$

If I_g is a point source then the intensity impulse response is the image intensity:

$$\begin{aligned}I(\xi,\eta;\bar{\mathbf{a}},R_P) &= \iint\limits_{-\infty}^{\infty}\left|\tilde{h}(\xi-\tilde{\alpha},\eta-\tilde{\beta};\bar{\mathbf{a}},R_P)\right|^2\delta(\tilde{\alpha},\tilde{\beta})d\tilde{\alpha}d\tilde{\beta}, &(3.11)\\ I(\xi,\eta;\bar{\mathbf{a}},R_P) &= \left|\tilde{h}\left(\xi,\eta;\bar{\mathbf{a}},R_P\right)\right|^2, &(3.12)\end{aligned}$$

where image intensity scaling has been neglected. Here I have introduced the new variable I which is the intensity impulse response for a pupil with atmospheric phase contributions. I is often referred to as a point spread function or PSF, however, since the wavefront sensor will be simulated using a point source, the variable I will henceforth serve as the expected image variable.

Using the discrete reference frames and variables, it is straightforward to convert the continuous propagation integral to its discrete counterpart. The Fourier integrals and their respective discrete forms are presented here. Recall the two-dimensional Fourier transform pair:

$$\begin{aligned}\mathcal{F}\{D\} &= \mathcal{D}(f_X, f_Y) = \iint\limits_{-\infty}^{\infty} D(x,y)\exp\{-\mathrm{j}2\pi(f_X x + f_Y y)\}dxdy, && (3.13)\\ \mathcal{F}^{-1}\{\mathcal{D}\} &= D(x,y) = \frac{1}{(2\pi)^2}\iint\limits_{-\infty}^{\infty} \mathcal{D}(f_X, f_Y)\exp\{\mathrm{j}2\pi(f_X x + f_Y y)\}df_X df_Y. && (3.14)\end{aligned}$$

The discrete counterparts of these operations are given here as the discrete Fourier series, $\mathcal{DFS}\{\cdot\}$, and the discrete Fourier transform, $\mathcal{DFT}\{\cdot\}$:

$$\begin{aligned}\mathcal{DFS}\{D\} = \mathcal{D}[f_X, f_Y] = \quad & \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} D[n\Delta x, m\Delta y] \times \\ & \exp\left\{-\mathrm{j}2\pi(n\Delta x f_X + m\Delta y f_Y)\right\}\Delta x\Delta y, && (3.15)\\ \mathcal{DFT}\{D\} = \mathcal{D}[u,v] = \quad & \sum_{n=-N/2}^{N/2-1}\sum_{m=-N/2}^{N/2-1} D[n,m]\exp\left\{-\mathrm{j}\frac{2\pi}{N}(nu+mv)\right\}, && (3.16)\\ \mathcal{DFT}^{-1}\{\mathcal{D}\} = D[n,m] = \quad & \frac{1}{N^2}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\mathcal{D}[u,v]\exp\left\{\mathrm{j}\frac{2\pi}{N}(nu+mv)\right\}. && (3.17)\end{aligned}$$

Sampling the $(\tilde{x}, \tilde{y})$ coordinate frame in (3.7) produces the discrete Fourier series form for $\tilde{h}$:

$$\tilde{h}\left[\xi,\eta\right] = \sum_{n=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} W_P(\lambda s_i n\Delta\tilde{x}, \lambda s_i m\Delta\tilde{y}; R_P)\exp\left\{-\mathrm{j}2\pi\left(\xi n\Delta\tilde{x} + \eta m\Delta\tilde{y}\right)\right\}\Delta\tilde{x}\Delta\tilde{y} \tag{3.18}$$

The discrete Fourier series representation, evaluated over an appropriate set of $[\xi, \eta]$ locations, can be posed as a discrete Fourier transform. There is only the minor difference of axis scaling between the scaled Fourier transform, $\mathcal{F}_s\left\{\cdot\right\}$, and the Fourier transform. This difference will be accounted for in a careful relationship between aperture sampling and image sampling conventions. The relationship between sample spacing in the aperture plane and sampling in the image plane is governed by the Nyquist sampling theorem. The choice of Δx, and Δy will be based on the dynamics of the field in the aperture. Applying Nyquist to Δx, Δy and the size of the aperture the maximum sample spacing in the image plane is given by:

$$\begin{aligned} \Delta\xi &= \frac{\lambda s_i}{2N\Delta x}, \Delta\eta = \frac{\lambda s_i}{2N\Delta y}, && (3.19) \\ N\Delta x &= D_P \equiv \text{aperture diameter}, && (3.20) \end{aligned}$$

Recall the relationship: $\Delta\tilde{x} = \frac{\Delta x}{\lambda s_i}$, $\Delta\tilde{y} = \frac{\Delta y}{\lambda s_i}$, from the Background chapter. Making the appropriate substitutions for sampling dimensions in the aperture and image planes yields:

$$\begin{aligned} \tilde{h}\left[u\Delta\xi, v\Delta\eta\right] &= \sum_{-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{-\frac{N}{2}}^{\frac{N}{2}-1} W_P\left[\lambda s_i n\frac{\Delta x}{\lambda s_i}, \lambda s_i m\frac{\Delta y}{\lambda s_i}; R_P\right] \times \\ & \exp\left\{-\mathrm{j}2\pi\left(u\Delta\xi n\frac{\Delta x}{\lambda s_i} + v\Delta\eta m\frac{\Delta y}{\lambda s_i}\right)\right\}\frac{\Delta x}{\lambda s_i}\frac{\Delta y}{\lambda s_i}, && (3.21) \\ \tilde{h}\left[u\Delta\xi, v\Delta\eta\right] &= \sum_{-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{-\frac{N}{2}}^{\frac{N}{2}-1} W_P\left[\lambda s_i n\frac{\Delta x}{\lambda s_i}, \lambda s_i m\frac{\Delta y}{\lambda s_i}; R_P\right] \times \\ & \exp\left\{-\mathrm{j}2\pi\left(u\frac{\lambda s_i}{2N\Delta x}n\frac{\Delta x}{\lambda s_i} + v\frac{\lambda s_i}{2N\Delta x}m\frac{\Delta y}{\lambda s_i}\right)\right\}\frac{\Delta x}{\lambda s_i}\frac{\Delta y}{\lambda s_i}, && (3.22) \\ \tilde{h}\left[u\Delta\xi, v\Delta\eta\right] &= \frac{\Delta x\Delta y}{\left(\lambda s_i\right)^2}\sum_{-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{-\frac{N}{2}}^{\frac{N}{2}-1} W_P\left[n\Delta x, m\Delta y; R_P\right]\exp\left\{-\mathrm{j}\frac{2\pi}{2N}\left(un+vm\right)\right\}, \\ & && (3.23) \\ \tilde{h}\left[u\Delta\xi, v\Delta\eta\right] &= \frac{\Delta x\Delta y}{\left(\lambda s_i\right)^2}\mathcal{DFT}\{W_P\left[n\Delta x, m\Delta y; R_P\right]\}. && (3.24) \end{aligned}$$

Thus, the discrete transformation of the aperture field into the image plane intensity is given by:

$$I\left[u\Delta\xi, v\Delta\eta; \bar{\mathbf{a}}\right] = \left|\frac{\Delta x\Delta y}{\left(\lambda s_i\right)^2}\mathcal{DFT}\{\mathcal{P}\left[n\Delta x, m\Delta y; \bar{\mathbf{a}}\right]\}\right|^2, \tag{3.25}$$

where the dependence of the pupil function on R_P remains, but the variable has been dropped to compact the notation. The bold image plane variable $\mathbf{I}$ will be substituted for the continuous function I to indicate that, while I is defined over the space of all real numbers, the discrete variable $\mathbf{I}$ is defined on a finite set of pixel locations $\mathcal{S}$:

$$\mathbf{I}\left[u, v; \bar{\mathbf{a}}\right] = I\left[u\Delta\xi, v\Delta\eta; \bar{\mathbf{a}}\right]. \tag{3.26}$$

$\mathcal{S}$ is the set of all pixel locations available in a CCD image. The bold $\mathbf{I}$ may also appear with or without index variables. When $\mathbf{I}$ is presented with index variables the expression is meant to indicate a single location in the image set. When $\mathbf{I}$ is presented without index variables, the expression refers to the entire image set:

$$\mathbf{I}\left[\bar{\mathbf{a}}\right] = \left\{\mathbf{I}[u, v; \bar{\mathbf{a}}] : u, v \in \mathcal{S}\right\}. \tag{3.27}$$

Finally, the discrete version of the OTF is given by the normalized discrete Fourier transform of $\mathbf{I}$ where all atmospheric parameters are zero:

$$\mathcal{H} = \frac{\mathcal{DFT}\left\{\mathbf{I}\right\}}{\sum\limits_{\mathbf{u}} \mathbf{I}[\mathbf{u}]}.$$

To reference a specific location in the OTF frequency domain, I will reuse the aperture plane index variable **n.**

The discrete reference frames and linear systems operations presented here can be used to simulate wave optics phenomenon. This simulation is intended to provide some demonstration of wavefront sensor performance. The fidelity of the simulation will suffer if the noise characteristics in the CCD are not included. Physically, noise occurs during the photon conversion process. Thus, to model CCD noise, some stochastic process must be included in the model to distinguish the detected image from the ideal image $\mathbf{I}$.

3.2 The Detected Image

The interface between the environment and the wavefront sensor occurs in the charge coupled device (CCD). If the CCD is sensitive enough and the exposure time is short, then the detector may be modeled as a photon counting device. In this case, intensity detected in the image plane is a count of discrete events as each incoming photon interacts with the CCD. Therefore, the detected image is spatially discrete, as it is formed from an array of image sample points, and discrete in the level of intensity measured at each sample point. The largest noise contribution is due to random photon arrival, often referred to as shot noise. Under the influence of shot noise only, the CCD image can be considered a grid of photon bins where the detected image is a count of Poisson distributed events. Let $\mathbf{d}[u,v]$ be a random variable representing the photon count with added shot noise in a single CCD pixel. Let $\mathbf{D}[u,v]$ be a realization of the random variable $\mathbf{d}[u,v]$. Note the use of the common convention that lower case variables indicate random variables and upper case variables indicate realizations of those random variables. The value of any single detected pixel is given by:

$$\mathbf{d}[u,v] = \text{Poisson}\left\{\mathbf{I}[u,v;\bar{\mathbf{a}}]\right\}. \tag{3.28}$$

Thus, each pixel in the detected image, $\mathbf{d}[u,v]$, is a Poisson random variable with its parameter being the associated value from the image predicted by the discrete linear system model with input parameter set $\bar{\mathbf{a}}$. Using this probability model for each pixel in $\mathbf{d}$, the conditional density for a detected pixel given an aperture field constructed from the set of Zernike coefficients $\bar{\mathbf{A}}$ is:

$$\begin{aligned} p_{\mathbf{d}[u,v]|\bar{\mathbf{a}}}(\mathbf{D}[u,v]|\bar{\mathbf{A}}) &= \frac{\mathbf{I}[u,v;\bar{\mathbf{A}}]^{\mathbf{D}[u,v]}\exp\left\{-\mathbf{I}[u,v;\bar{\mathbf{A}}]\right\}}{\mathbf{D}[u,v]!}, && (3.29)\\ \text{where } \mathbf{D}[u,v] &\equiv \text{photo detection events in image pixel } [u,v], && (3.30)\\ \text{and } \mathbf{I}[u,v;\bar{\mathbf{A}}] &\equiv \text{Poisson rate function is the noiseless image.} && (3.31)\end{aligned}$$

There may be multiple image planes associated with a subaperture. Suppose that the optical path is split such that there are $N_{\mathbf{D}}$ imaging planes. Each image and its respective set of pixel locations will be indexed with the subscript i:

$$\begin{aligned} \mathbf{D}_i &= \left\{\mathbf{D}_i[\mathbf{u}] : \mathbf{u} \in \mathcal{S}_i\right\}, && (3.32)\\ \text{where } \mathcal{S}_i &\equiv i\text{th sample space of pixels}, && (3.33)\\ \text{and } \mathbf{u} &= [u,v]. && (3.34)\end{aligned}$$

The combined set of detected images and expected images are denoted $\mathbf{D}_{\cup}$ and $\mathbf{I}_{\cup}$ respectively:

$$\mathbf{D}_{\cup} = \bigcup_{i=1}^{N_{\mathbf{D}}} \mathbf{D}_i. \tag{3.35}$$

The estimator will require a joint density for all pixels in all image planes. It is possible to extend the pdf in (3.29) above to a joint pdf including pixels from all image arrays by assuming that each pixel is independently and identically distributed (i.i.d.) Poisson. For independent random variables, the joint density is the product of the marginal densities. Take the product of every pixel in all images and the conditional density becomes:

$$p_{\mathbf{d}_{\cup}|\bar{\mathbf{a}}}(\mathbf{D}_{\cup}|\bar{\mathbf{A}}) = \prod_{i=1}^{N_{\mathbf{D}}} \prod_{\mathbf{u}\in\mathcal{S}_i} \frac{\mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}]^{\mathbf{D}_i[\mathbf{u}]}\exp(-\mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}])}{\mathbf{D}_i[\mathbf{u}]!}. \tag{3.36}$$

The maximum likelihood estimator $\hat{\mathbf{a}}_{ml}$ is formed by assuming the pdf for the parameter set is uniform and maximizing the log-likelihood expression over the range of the parameter set $\bar{\mathbf{A}}$ as in (2.19):

$$\max_{\bar{\mathbf{A}}} \left\{\ln\left\{p_{\mathbf{d}|\bar{\mathbf{a}}}(\mathbf{D}|\bar{\mathbf{A}})\right\}\right\}\Big|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{ml}}, \tag{3.37}$$

$$\max_{\bar{\mathbf{A}}} \left\{\sum_{i=1}^{N_{\mathbf{D}}} \sum_{\mathbf{u}\in\mathcal{S}_i} \mathbf{D}_i[\mathbf{u}]\ln\left\{\mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}]\right\} - \mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}] + \ln\left\{\mathbf{D}_i[\mathbf{u}]!\right\}\right\}\Bigg|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{ml}}, \tag{3.38}$$

$$\max_{\bar{\mathbf{A}}} \left\{\sum_{i=1}^{N_{\mathbf{D}}} \sum_{\mathbf{u}\in\mathcal{S}_i} \mathbf{D}_i[\mathbf{u}]\ln\left\{\mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}]\right\} - \mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}]\right\}\Bigg|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{ml}}, \tag{3.39}$$

where the detected image $\mathbf{D}$ represents the observation vector, $\mathbf{R}$ in (2.19). Also, the parameter A in (2.19) has been replaced with the entire set of Zernike coefficients, $\bar{\mathbf{a}}$. If the pdf for the parameter set is not uniform then the log-likelihood includes the additional term, $\ln\left\{p_{\bar{\mathbf{a}}}(\bar{\mathbf{A}})\right\}$. Recall that the phase at each point within a single phase screen is the sum of phase contributions along some optical path. Furthermore, each Zernike coefficient is computed by a projection sum using the resulting phase screen points. Given that the Zernike coefficients are formed from the sum of many zero mean random variables, it is reasonable to assume that the Central Limit Theorem applies and thus each Zernike coefficient is zero mean Gaussian with a variance that depends on the atmospheric turbulence model. Therefore, let the joint pdf for the parameter set be a multivariate Gaussian:

$$p_{\bar{\mathbf{a}}}(\bar{\mathbf{A}}) = \frac{1}{(2\pi)^{\frac{n}{2}}\left(\det\left(\mathbf{\Lambda}_{\bar{\mathbf{a}}}\right)\right)^{\frac{1}{2}}}\exp\left\{-\frac{\bar{\mathbf{A}}\mathbf{\Lambda}_{\bar{\mathbf{a}}}^{-\mathbf{1}}\bar{\mathbf{A}}^{t}}{2}\right\}, \tag{3.40}$$

where the matrix, $\mathbf{\Lambda}_{\bar{\mathbf{a}}}$, is the covariance matrix for the jointly Gaussian parameter set $\bar{\mathbf{a}}$. An example of the matrix, $\mathbf{\Lambda}_{\bar{\mathbf{a}}}$, for parameters a_2 through a_{11} was generated in Table 2.5, Section 2.3. Substituting $\ln\left\{p_{\bar{\mathbf{a}}}(\bar{\mathbf{A}})\right\}$ and $\ln\left\{p_{\mathbf{d}|\bar{\mathbf{a}}}(\mathbf{D}|\bar{\mathbf{A}})\right\}$ into (2.17), the maximum a posteriori estimator $\hat{\mathbf{a}}_{map}$ is given by:

$$\max_{\bar{\mathbf{A}}} \left\{\ln\left\{p_{\mathbf{d}|\bar{\mathbf{a}}}(\mathbf{D}|\bar{\mathbf{A}})\right\} + \ln\left\{p_{\bar{\mathbf{a}}}(\bar{\mathbf{A}})\right\}\right\}\Big|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{map}}, \tag{3.41}$$

$$\max_{\bar{\mathbf{A}}} \left\{ \begin{array}{c} \left[\sum_{i=1}^{N_{\mathbf{D}}} \sum_{\mathbf{u}\in\mathcal{S}_i} \mathbf{D}_i[\mathbf{u}] \ln\left\{\mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}]\right\} - \mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}]\right] + \\ \ln\left\{\frac{1}{(2\pi)^{\frac{n}{2}}\left(\det(\mathbf{\Lambda}_{\bar{\mathbf{a}}})\right)^{\frac{1}{2}}}\right\} - \frac{\bar{\mathbf{A}}\mathbf{\Lambda}_{\bar{\mathbf{a}}}^{-1}\bar{\mathbf{A}}^t}{2} \end{array} \right\}\Bigg|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{map}}, \tag{3.42}$$

$$\max_{\bar{\mathbf{A}}} \left\{ \left[\sum_{i=1}^{N_{\mathbf{D}}} \sum_{\mathbf{u}\in\mathcal{S}_i} \mathbf{D}_i[\mathbf{u}] \ln\left\{\mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}]\right\} - \mathbf{I}_i[\mathbf{u};\bar{\mathbf{A}}]\right] - \frac{\bar{\mathbf{A}}\mathbf{\Lambda}_{\bar{\mathbf{a}}}^{-1}\bar{\mathbf{A}}^t}{2} \right\}\Bigg|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{map}}. \tag{3.43}$$

Using the linear system model and the CCD noise model, ML and MAP estimators have been derived. These estimators make use of the final ingredient required for a parameter estimating wavefront sensor, the conditional probability mapping, $p_{\mathbf{r}|\bar{\mathbf{a}}}(\mathbf{R}|\bar{\mathbf{A}})$ introduced in Section 2.1. While these estimators will yield good performance, they may not be capable of being implemented in a real time wavefront sensor due to CCD read out time and mathematical complexity. The following sections discuss modifications to the estimators that offer trade-offs between complexity and performance.

3.3 The Image Projection

The wavefront sensor design must map from the detected image intensity to the aperture phase. This implies that the first step in the wavefront correction process will be to detect and store the intensity in the image plane. Though the two-dimensional image contains a wealth of data, the processing time required to read out and evaluate that number of data points can exceed the maximum bandwidth for a real time adaptive optics system. The time spent reading out the charge from each pixel location in the CCD can become the largest portion of the time required for processing the image signal. Additionally, read out noise is proportional to the number of pixels read from the CCD. The read out noise model will be introduced in the next section. Compressing the number of data points without losing vital information for wavefront sensing can become a time saving step if not a necessity when trying to achieve real time adaptive optics system bandwidths. One technique for reducing the number of data points is to project the image plane into a vector format prior to reading the CCD. The projection operation can be performed quickly on the chip using simple shifting and summing of data. The simplest image projections are the one-dimensional projections along either the u or v axis:

$$\begin{aligned} \mathbf{D}^{v}[u] &= \sum_{m=0}^{N-1} \mathbf{D}[u,m], && (3.44) \\ \mathbf{D}^{u}[v] &= \sum_{n=0}^{N-1} \mathbf{D}[n,v], && (3.45) \end{aligned}$$

which are the sums of the two-dimensional image in each direction.

Image Projections and the OTF. The true benefit of creating the pair of "vectorized" images can be realized by first recalling the discrete Fourier transform (3.16). The two-dimensional DFT is a separable summation. It is possible to perform the DFT for the zeroth order frequency in each direction independently. These operations reveal a special benefit of the vectorized image:

$$\begin{aligned} \mathcal{D}[n,0] &= \sum_{u=0}^{N-1} \left[\sum_{v=0}^{N-1} \mathbf{D}[u,v] \right] \exp\left\{ \mathrm{j}\frac{2\pi}{N}\left(nu + m\left(0\right)\right) \right\}, && (3.46) \\ &= \sum_{u=0}^{N-1} \left[\sum_{v=0}^{N-1} \mathbf{D}[u,v] \right] \exp\left\{ \mathrm{j}\frac{2\pi}{N} nu \right\}, && (3.47) \\ &= \sum_{u=0}^{N-1} \mathbf{D}^{v}[u] \exp\left\{ \mathrm{j}\frac{2\pi}{N} nu \right\}; && (3.48) \\ \mathcal{D}[0,m] &= \sum_{v=0}^{N-1} \mathbf{D}^{u}[v] \exp\left\{ \mathrm{j}\frac{2\pi}{N} mv \right\}. && (3.49) \end{aligned}$$

This shows that performing a one-dimensional Fourier transform on the vectorized image produces the zeroth order frequency vector from the two-dimensional transform. To realize the importance of this, recall the OTF examples in Figure 2.8. The OTF, in this case, was a simple Fourier transform of the image. Vectorizing the image on the CCD, then performing a one-dimensional Fourier transform on the result produces a slice of the OTF much like the cross-section plots provided in Figure 3.3. Given the uniqueness in the OTF

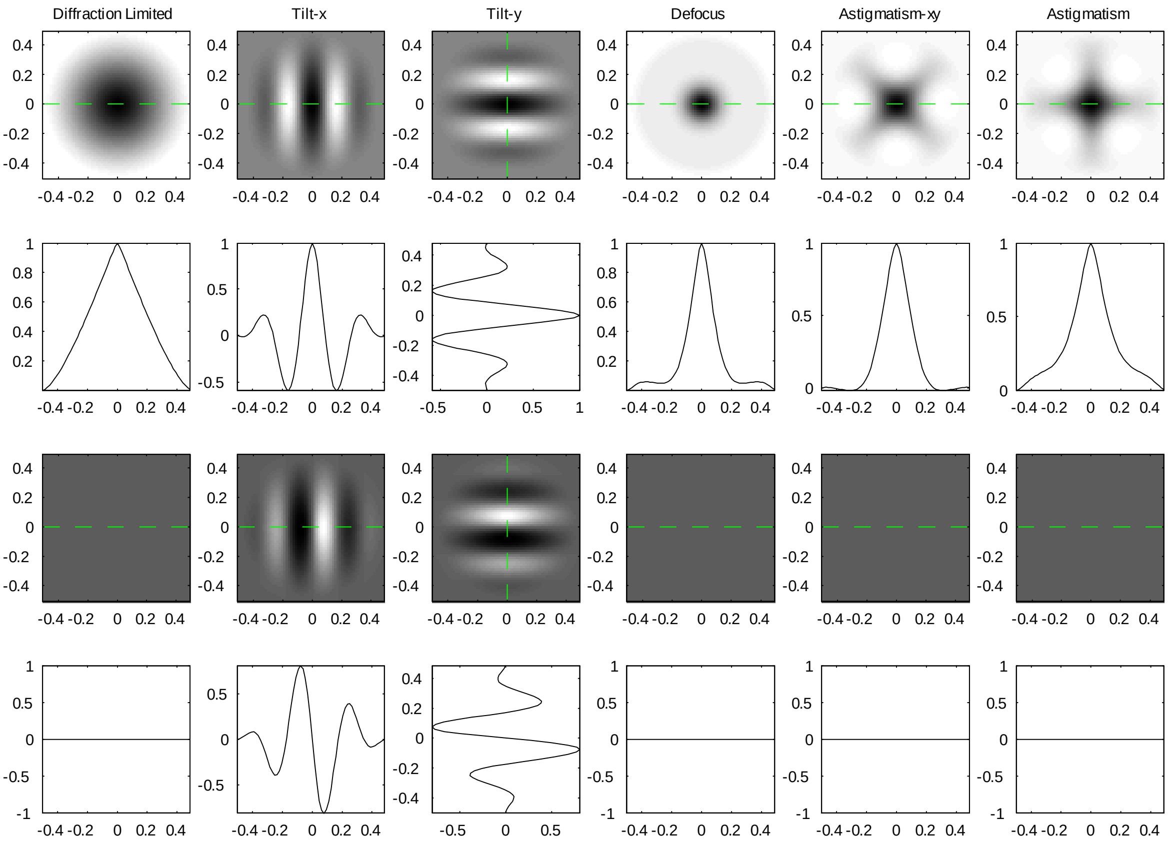


Figure 3.3 Simulated OTFs for a diffraction limited optical system (column 1) and systems under independent influence from Zernikes 2-6 (columns 2-6 respectively). Rows 1 and 3: real and imaginary parts of the OTFs respectively. Rows 2 and 4: projections corresponding to dashed section lines in rows 1 and 3.

cross-section under the influence of some limited set of Zernike modes, it may be possible to distinguish, and therefore estimate, the effects from each of the Zernikes. Thus, the vectorized image promises a much faster read out time from the CCD along with some amount of information useful for estimating pupil phase.

A General Image Projection Operator. A more general description of the projection operation is required before including it in the estimator models. The notation for key quantities and reference frames within the discretized optical system model will be used to define the general form of the image projection operation. The image projection operator accepts an array or set of arrays for input and returns a vector. Additionally, at least two images are formed in separate CCD arrays. Given multiple arrays, each CCD may be rotated such that its $\hat{v}_i$ unit vector forms a unique angle, θ_i, with the original image plane $\hat{v}$ direction. The CCD image arrays are given by:

$$\mathbf{D}_{i,\theta_i}\left[u_i, v_i\right] \equiv \text{pixel from CCD image array } i, \tag{3.50}$$

$$\theta_i \equiv \frac{180}{\pi}\cos^{-1}\left(\hat{v} \cdot \hat{v}_i\right) \text{ [degrees].} \tag{3.51}$$

The image projection operator $\mathbf{v}(\cdot)$ accepts the set of image arrays: $\{\mathbf{D}_{i,\theta_i} : i \in 1, 2, ..., N_{\mathbf{D}}\}$ for input. First, each input array is windowed by discarding all pixels except the $N_W \times N_W$ pixel region centered around the optical axis. Since the windowed pixels are the only pixels of interest from the larger set $\mathcal{S}_i$, continuing to use the old index values $[u_i, v_i]$ becomes unnecessary. It is more convenient to renumber the windowed array using a 1 to N_W row, column numbering system. The resulting set of windowed pixels are summed along the v_i direction according to a set of starting and ending row number pairs contained in the set $\mathbf{s}$:

$$\mathbf{s} = \left\{\left(1, \mathbf{s}_2\right), ..., \left(\mathbf{s}_{N_{\mathbf{s}}-1}, \mathbf{s}_{N_{\mathbf{s}}}\right), N_W\right\}. \tag{3.52}$$

The last entry in the set $\mathbf{s}$ does not provide a summation pair, rather, it identifies the window length in pixels, N_W. For notational simplicity, if the N_W entry in $\mathbf{s}$ is omitted, it is assumed that the final ordered pair, $(\mathbf{s}_{N_{\mathbf{s}}-1}, \mathbf{s}_{N_{\mathbf{s}}})$, ends on the window length index, N_W (i.e. $\mathbf{s}_{N_{\mathbf{s}}} = N_W$). The following are examples of $\mathbf{s}$ for a 6×6 window: $\mathbf{s} = \{(1, 6)\}$ identifies the projection operation which sums along the entire windowed region; $\mathbf{s} = \{(1, 3), 6\}$ indicates that the projection only includes the first 3 rows, but the window size is 6×6, and $\mathbf{s} = \{(1, 3), (4, 6)\}$ describes the projection operation which sums the upper half and lower half (rows 1 to 3 and 4 to 6) of the window into two separate vector projections. The set $\mathbf{s}$ will be included as a preceding subscript on $\mathbf{v}(\cdot)$ whenever the projection operation requires clarification. The resulting vector output is a concatenation of N_{v} projection sums where N_{v} is the number of ordered pairs in $\mathbf{s}$. The variable $l \in \{1, 2, ..., N_{\mathrm{v}} N_W\}$ indexes the location in the resulting vector. Given this convention, the general form of the projection operator output with individual vector location index is either denoted:

$$_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{1,\theta_1}, \mathbf{D}_{2,\theta_2}, ..., \mathbf{D}_{N_{\mathbf{D}},\theta_{N_{\mathbf{D}}}}\right), \tag{3.53}$$

or by the more compact notation:

$$
{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_\cup\right), \tag{3.54}
$$

$$
\text{where} \qquad \mathbf{D}_\cup = \bigcup_{i=1}^{N_\mathbf{D}} \mathbf{D}_{i,\theta_i}. \tag{3.55}
$$

As an example, consider the operation which employs whole plane projections from two CCDs at 0° and 90° rotations using a 6 × 6 window. The lth location in the projection operation output is denoted:

$$
{}_{\{(1,6)\}}\mathbf{v}_l\left(\mathbf{D}_{1,0}, \mathbf{D}_{2,90}\right). \tag{3.56}
$$

Given the general form for the projection operator it is possible to develop a pair of projection based estimator expressions. The joint density for an image projection is given by:

$$
p_{\mathbf{sv}\,(\mathbf{d}_\cup)|\bar{\mathbf{a}}}\left({}_{\mathbf{s}}\mathbf{v}\left(\mathbf{D}_\cup\right)|\bar{\mathbf{A}}\right) = \prod_{l=1}^{N_{\mathrm{v}}N_W} \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)^{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_\cup\right)} \exp\left({}_{\mathbf{s}}\mathbf{v}_l\left(-\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)\right)}{{}_{\mathbf{s}}\mathbf{v}_l\left(\bigcup_{i=1}^{N_\mathbf{D}}\left\{\left(\mathbf{D}_{i,\theta_i}\left[\mathbf{u}\right]\right)! : \mathbf{u} \in \mathcal{S}_i\right\}\right)}. \tag{3.57}
$$

The notation $\left(\mathbf{D}_{i,\theta_i}\left[\mathbf{u}\right]\right)!$ indicates that the factorial is applied at each pixel location $\mathbf{u} \in \mathcal{S}_i$. Substituting (3.57) into (2.19), the form for the maximum likelihood estimator is given by:

$$
\max_{\bar{\mathbf{A}}} \left\{ \begin{array}{c} \displaystyle\sum_{l=1}^{N_{\mathrm{v}}N_W} {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_\cup\right) \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)\right\} - {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right) + \\ \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\bigcup_{i=1}^{N_\mathbf{D}}\left\{\mathbf{D}_{i,\theta_i}\left[\mathbf{u}\right]! : \mathbf{u} \in \mathcal{S}_i\right\}\right)\right\} \end{array} \right\}\Bigg|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{ml}}, \tag{3.58}
$$

$$
\max_{\bar{\mathbf{A}}} \left\{ \sum_{l=1}^{N_{\mathrm{v}}N_W} {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_\cup\right) \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)\right\} - {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right) \right\}\Bigg|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{ml}}. \tag{3.59}
$$

Similarly, the form for the maximum a posteriori estimator is given:

$$
\max_{\bar{\mathbf{A}}} \left\{ \left[\sum_{l=1}^{N_{\mathrm{v}}N_W} {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_\cup\right) \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)\right\} - {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right) \right] - \frac{\bar{\mathbf{A}}\mathbf{\Lambda}_{\bar{\mathbf{a}}}^{-\mathbf{1}}\bar{\mathbf{A}}^t}{2} \right\}\Bigg|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{map}}. \tag{3.60}
$$

3.4 Other Detected Image Models

While the Poisson pdf model for an image pixel is sufficient, it does not account for all forms of CCD noise. In addition to shot or Poisson noise, read noise is always present. While shot noise is due to the random arrival of photons and is the basis of the Poisson model, read noise is generated by the buffers and amplifiers used to digitize the photon count voltage. Poisson noise is signal dependent because the Poisson rate parameter is the noiseless image value. Read noise is independent of the signal. Instead, read noise depends on the number of pixels read from the CCD.

Accounting for Read Noise in the Image pdf. Let $\mathbf{n}_{ro}[\mathbf{u}]$ represent the read noise in a CCD pixel. Let the read noise, $\mathbf{n}_{ro}[\mathbf{u}]$, be a zero mean Gaussian process with variance σ_{ro}^2. Considering the effects of read noise and shot noise to be independent, the pixel with combined shot noise and read noise effects, $\mathbf{d}_{ro}[\mathbf{u}]$, is modeled as the sum of independent Poisson and Gaussian random variables. The pdf of the sum $z = x + y$ is a convolution of pdfs of x and y when x and y are independent [22]:

$$\begin{aligned}
p_{\mathbf{d}[\mathbf{u}]|\bar{\mathbf{a}}}(\mathbf{D}\left[\mathbf{u}\right]|\bar{\mathbf{A}}) &= \exp\left\{-\mathbf{I}\left[\mathbf{u};\bar{\mathbf{A}}\right]\right\}\frac{\mathbf{I}\left[\mathbf{u};\bar{\mathbf{A}}\right]^{\mathbf{D}[\mathbf{u}]}}{\mathbf{D}\left[\mathbf{u}\right]!}, && (3.61)\\
p_{\mathbf{n}_{ro}[\mathbf{u}]}(\mathbf{N}_{ro}\left[\mathbf{u}\right]) &= \frac{1}{\sqrt{2\pi}\sigma_{ro}}\exp\left\{-\frac{\mathbf{N}_{ro}^2\left[\mathbf{u}\right]}{2\sigma_{ro}^2}\right\}, && (3.62)\\
\mathbf{d}_{ro}\left[\mathbf{u}\right] &= \mathbf{d}\left[\mathbf{u}\right] + \mathbf{n}_{ro}\left[\mathbf{u}\right], && (3.63)\\
p_{\mathbf{d}_{ro}[\mathbf{u}]|\bar{\mathbf{a}}}(\mathbf{D}_{ro}\left[\mathbf{u}\right]|\bar{\mathbf{A}}) &= \frac{1}{\sqrt{2\pi}\sigma_{ro}}\sum_{\mathbf{D}[\mathbf{u}]=0}^{\infty}\exp\left\{-\frac{\left(\mathbf{D}_{ro}\left[\mathbf{u}\right]-\mathbf{D}\left[\mathbf{u}\right]\right)^2}{2\sigma_{ro}^2}\right\}\times\\
&\quad \exp\left\{-\mathbf{I}\left[\mathbf{u};\bar{\mathbf{A}}\right]\right\}\frac{\mathbf{I}\left[\mathbf{u};\bar{\mathbf{A}}\right]^{\mathbf{D}[\mathbf{u}]}}{\mathbf{D}\left[\mathbf{u}\right]!}. && (3.64)
\end{aligned}$$

This expression must be modified slightly to account for the discrete nature of the CCD electronics. The A/D conversion process will apply a rounding function to the continuous set of values allowed by the Gaussian pdf. This will have the effect of forcing $\mathbf{d}_{ro}\left[\mathbf{u}\right]$ to take on only nonnegative integer values. The A/D conversion can be included in the statistical model as an integration around each integer value for the Poisson random variable:

$$p_{\mathbf{d}_{ro}[\mathbf{u}]|\bar{\mathbf{a}}}(\mathbf{D}_{ro}\left[\mathbf{u}\right]|\bar{\mathbf{A}}) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_{ro}} \sum\limits_{\mathbf{D}[\mathbf{u}]=0}^{\infty} \exp\left\{-\mathbf{I}\left[\mathbf{u};\bar{\mathbf{A}}\right]\right\} \frac{\mathbf{I}\left[\mathbf{u};\bar{\mathbf{A}}\right]^{\mathbf{D}[\mathbf{u}]}}{\mathbf{D}[\mathbf{u}]!} \times \\ \int\limits_{-\infty}^{0.5} \exp\left\{-\frac{(\boldsymbol{\xi}-\mathbf{D}[\mathbf{u}])^2}{2\sigma_{ro}^2}\right\} d\xi, \qquad \text{for } \mathbf{D}_{ro}\left[\mathbf{u}\right]=0, \\ \frac{1}{\sqrt{2\pi}\sigma_{ro}} \sum\limits_{\mathbf{D}[\mathbf{u}]=0}^{\infty} \exp\left\{-\mathbf{I}\left[\mathbf{u};\bar{\mathbf{A}}\right]\right\} \frac{\mathbf{I}\left[\mathbf{u};\bar{\mathbf{A}}\right]^{\mathbf{D}[\mathbf{u}]}}{\mathbf{D}[\mathbf{u}]!} \times \\ \int\limits_{\mathbf{D}_{ro}[\mathbf{u}]-0.5}^{\mathbf{D}_{ro}[\mathbf{u}]+0.5} \exp\left\{-\frac{(\boldsymbol{\xi}-\mathbf{D}[\mathbf{u}])^2}{2\sigma_{ro}^2}\right\} d\xi, \text{ for } \mathbf{D}_{ro}\left[\mathbf{u}\right] \in \{1,2,3,...\}. \end{cases} \tag{3.65}$$

Unfortunately, this distribution does not simplify and, although it can be evaluated numerically, it is far too computationally expensive for use in the sensor model. However, the numerical evaluation of $\mathbf{d}_{ro}\left[\mathbf{u}\right]$ begins to look like a biased Poisson statistic. This characteristic allows for a far simpler mathematical expression that can be used in the real time sensor model.

Approximating the Read Noise pdf. The combination of Poisson shot noise and Gaussian read noise can be approximated by a biased Poisson statistic [23]:

$$\mathbf{d}\left[\mathbf{u}\right] = \text{Poisson}\left\{\mathbf{I}\left[\mathbf{u};\bar{\mathbf{a}}\right]+\sigma_{ro}^2\right\}-\sigma_{ro}^2, \tag{3.66}$$

$$\mathbf{d}\left[\mathbf{u}\right]+\sigma_{ro}^2 = \text{Poisson}\left\{\mathbf{I}\left[\mathbf{u};\bar{\mathbf{a}}\right]+\sigma_{ro}^2\right\}. \tag{3.67}$$

Using this approximate distribution, it is possible to develop a new expression for the MAP estimator. Begin with the property: the sum of independent Poisson random variables is distributed Poisson with a mean equal to the sum of the individual means. Using (3.67), the joint density for all pixels in the image projection conditioned on the set of Zernike coefficients becomes:

$$\begin{aligned} p_{{}_{\mathbf{s}}\mathbf{v}_l(\mathbf{d}_\cup)+\sigma_{ro}^2|\bar{\mathbf{a}}}\left({}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_\cup\right)+\sigma_{ro}^2|\bar{\mathbf{A}}\right) &= \exp\left\{-\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)+\sigma_{ro}^2\right]\right\}\times \\ &\quad \frac{\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)+\sigma_{ro}^2\right]^{\mathbf{v}_i(\mathbf{D}_\cup)+\sigma_{ro}^2}}{{}_{\mathbf{s}}\mathbf{v}_l\left(\bigcup_{i=1}^{N_{\mathbf{D}}}\left\{\mathbf{D}_{i,\theta_i}\left[\mathbf{u}\right]! : \mathbf{u}\in\mathcal{S}_i\right\}\right)+\sigma_{ro}^2}. \end{aligned} \tag{3.68}$$

The joint density for locations in the image projection vector is simply the product of the marginals:

$$\begin{aligned} p_{{}_{\mathbf{s}}\mathbf{v}(\mathbf{d}_\cup)+\boldsymbol{\sigma}_{ro}^2|\bar{\mathbf{a}}}\left({}_{\mathbf{s}}\mathbf{v}\left(\mathbf{D}_\cup\right)+\boldsymbol{\sigma}_{ro}^2|\bar{\mathbf{A}}\right) &= \prod_{l=1}^{N_{\mathrm{v}}N_W}\exp\left\{-\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)+\sigma_{ro}^2\right]\right\}\times \\ &\quad \frac{\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)+\sigma_{ro}^2\right]^{{}_{\mathbf{s}}\mathbf{v}_l(\mathbf{D}_\cup)+\sigma_{ro}^2}}{{}_{\mathbf{s}}\mathbf{v}_l\left(\bigcup_{i=1}^{N_{\mathbf{D}}}\left\{\mathbf{D}_{i,\theta_i}\left[\mathbf{u}\right]! : \mathbf{u}\in\mathcal{S}_i\right\}\right)+\sigma_{ro}^2}. \end{aligned} \tag{3.69}$$

The bold read out variance $\boldsymbol{\sigma}_{ro}^2$ indicates the product: $\mathbf{1}\sigma_{ro}^2$, where $\mathbf{1}$ represents a vector of ones of length $N_{\mathrm{v}}N_W$. Taking the natural log of the conditional density gives:

$$\begin{aligned} \ln\left\{p_{{}_{\mathbf{s}}\mathbf{v}(\mathbf{d}_\cup)+\boldsymbol{\sigma}_{ro}^2|\bar{\mathbf{a}}}\left({}_{\mathbf{s}}\mathbf{v}\left(\mathbf{D}_\cup\right)+\boldsymbol{\sigma}_{ro}^2|\bar{\mathbf{A}}\right)\right\} &= \sum_{l=1}^{N_{\mathrm{v}}N_W}\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_\cup\right)+\sigma_{ro}^2\right]\times \\ &\quad \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)+\sigma_{ro}^2\right\}- \\ &\quad {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_\cup\left[\bar{\mathbf{A}}\right]\right)-\sigma_{ro}^2- \\ &\quad \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\bigcup_{i=1}^{N_{\mathbf{D}}}\left\{\mathbf{D}_{i,\theta_i}\left[\mathbf{u}\right]! : \mathbf{u}\in\mathcal{S}_i\right\}\right)+\sigma_{ro}^2\right\}. \end{aligned} \tag{3.70}$$

Recall the log of the jointly Gaussian prior density:

$$\ln\left\{p_{\bar{\mathbf{a}}}(\bar{\mathbf{A}})\right\}=\ln\left\{\frac{1}{(2\pi)^{\frac{n}{2}}\left(\det\left(\boldsymbol{\Lambda}_{\bar{\mathbf{a}}}\right)\right)^{\frac{1}{2}}}\right\}-\frac{\bar{\mathbf{A}}\boldsymbol{\Lambda}_{\bar{\mathbf{a}}}^{-\mathbf{1}}\bar{\mathbf{A}}^t}{2}. \tag{3.71}$$

Combining these results yields the MAP estimator:

$$\max_{\bar{\mathbf{A}}} \left\{ \left[\begin{array}{c} \sum_{l=1}^{N_v N_W} \left[{}_{\mathbf{s}}\mathbf{v}_l \left(\mathbf{D}_{\cup} \right) + \sigma_{ro}^2 \right] \times \\ \ln \left\{ {}_{\mathbf{s}}\mathbf{v}_l \left(\mathbf{I}_{\cup} \left[\bar{\mathbf{A}} \right] \right) + \sigma_{ro}^2 \right\} - \\ {}_{\mathbf{s}}\mathbf{v}_l \left(\mathbf{I}_{\cup} \left[\bar{\mathbf{A}} \right] \right) \end{array} \right] - \frac{\bar{\mathbf{A}} \mathbf{\Lambda}_{\bar{\mathbf{a}}}^{-\mathbf{1}} \bar{\mathbf{A}}^t}{2} \right\} \Bigg|_{\bar{\mathbf{A}} = \hat{\mathbf{a}}_{map}}, \qquad (3.72)$$

where, once again, all terms with no dependence on $\bar{\mathbf{A}}$ are discarded from the estimator expression.

3.5 Summary

The chapter began by developing discrete reference frames and a discrete linear systems model from their continuous counterparts introduced in the background chapter. Then a model for CCD noise was provided. This led to the idea that the detected image is not simply the output of the linear system, but rather that the system output combined with some random noise. From there, the joint density for a group of pixels in several image planes conditioned on the set of atmospheric phase coefficients was formed. The conditional density was used to create the first attempts at ML and MAP estimators.

The CCD image contains too many pixels for read out in real time. To limit the number of pixels read out of the CCD, the image projection was used. The image projection has the benefits of shorter read out time and decreased read noise. The ML and MAP estimators were adapted to include the generic image projection operator. An approximated read noise pdf was provided. Combining the read noise pdf with the image projection concept gave the final form for the ML and MAP estimators. The chapters to follow will detail an algorithm for a curvature sensor implementation using these estimators along with sensor performance bounding and simulated performance of the curvature sensor.

4. A Survey of Wavefront Sensing Techniques

This chapter provides a brief overview of several wavefront sensing methods. The discussion begins with interferometric techniques, followed by a review of a method known as phase retrieval or phase diversity. I introduce the role of parameter estimation in wavefront sensing with a description of the Hartmann sensor. Finally, I have included a more detailed discussion of two parameter estimation methods which use image projections. The wavefront sensor designs presented in this dissertation are modified versions of Cain's projection based estimator [1].

4.1 Wavefront Sensing through Interferometry

Interferometric techniques for measuring wavefront phase involve interpreting the intensity patterns arising from the interaction of two or more coherent fields. Two fields arriving at a point in space and time are perfectly coherent if they are very narrow band (composed primarily of a single wavelength) and their phase variation is identical. Broadening the spectrum and varying the relative phase reduces the level of coherence. When coherent fields are combined, periodic fringe patterns are visible in the resulting intensity. The classic examples involve the Michelson interferometer and Young's double slit experiment shown side by side in Figure 4.1. These two examples put to practical use the effects of temporal and spatial coherence respectively. The level of both temporal coherence and spatial coherence in a source can determine the degree of calibration and precision necessary to successfully conduct an interference experiment. Temporal coherence is inversely proportional to the bandwidth of the source. In interferometry, temporal coherence is quantified by the length of time delay one can impose on a portion of a beam and still create measurable interference fringes when that delayed portion is recombined with the original beam. The amount of spatial coherence is measured by spatially shifting a portion of a field and recombining it with the original field. The spatial coherence is quantified by the spatial separation beyond which the two portions cease to create measurable interference patterns. Each style of interferometric wavefront sensor gathers information about the wavefront phase by sampling the intensity and applying software algorithms to detect known characteristics from the interference patterns.

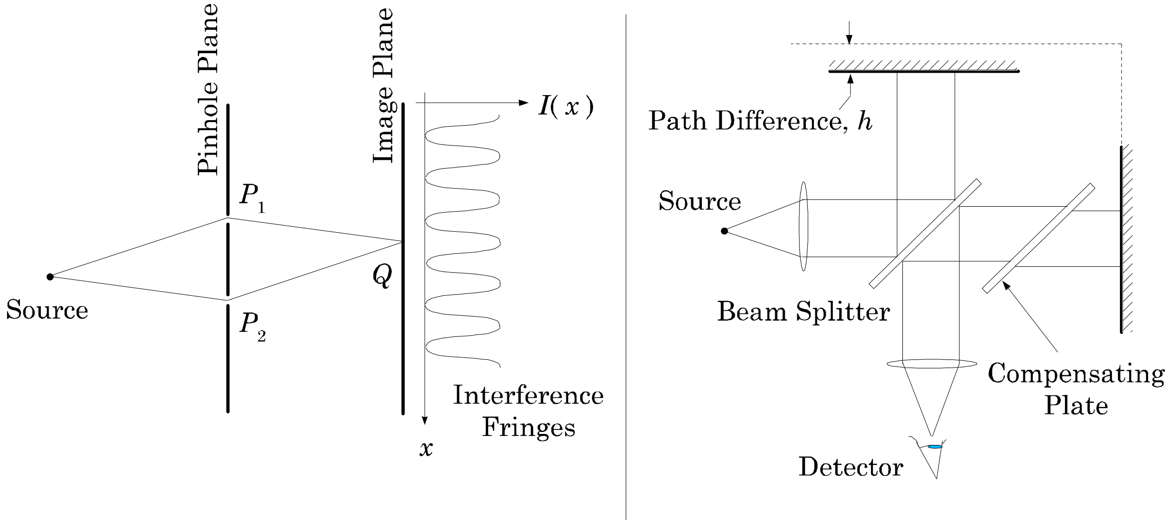


Figure 4.1 At left: geometric interpretation of Young's double slit experiment. At right: diagram of a Michelson interferometer.

The basic formula for interference fringe patterns can be derived for both the temporal and spatial cases. Beginning with the temporal case, the second order effects of light incident on the detector can be described as a function of the path difference, h [24]. Consider the interaction of the field, $\mathbf{u}(t)$ with a delayed version of itself:

$$\begin{aligned} I_D(h) &= \left\langle \left| K_1\mathbf{u}(t) + K_2\mathbf{u}\left(t + \frac{2h}{c}\right)\right|^2 \right\rangle, && (4.1) \\ \text{where } K_1, K_2 &\equiv \text{variable attenuation in each path}, && (4.2) \\ \mathbf{u} &\equiv \text{optical field}, && (4.3) \\ \text{and } c &\equiv \text{speed of light}. && (4.4) \end{aligned}$$

The intensity, I_D, is formed from the average modulus of the field. The brackets $\langle\cdot\rangle$ indicate the expectation or ensemble average operation. Expanding the product inside the expectation results in:

$$\begin{aligned} I_D(h) &= (K_1^2 + K_2^2)I_0 + 2K_1K_2 \operatorname{Re}\left\{\mathbf{\Gamma}\left(\frac{2h}{c}\right)\right\}, && (4.5) \\ \text{where } I_0 &\equiv \left\langle |\mathbf{u}(t)|^2 \right\rangle = \left\langle \left|\mathbf{u}\left(t + \frac{2h}{c}\right)\right|^2 \right\rangle = \mathbf{\Gamma}(0), && (4.6) \\ \text{and } \mathbf{\Gamma}(\tau) &= \langle \mathbf{u}(t+\tau)\mathbf{u}^*(t)\rangle. && (4.7) \end{aligned}$$

The self-coherence function, $\mathbf{\Gamma}(\tau)$, can be written in general form [24]:

$$
\begin{aligned}
\mathbf{\Gamma}(\tau) &= \Gamma(\tau) \exp\left\{-j\left[2\pi \frac{c}{\lambda}\tau - \alpha(\tau)\right]\right\}, && (4.8)\\
\text{where } \Gamma(\tau) &= \left|\mathbf{\Gamma}(\tau)\right|, && (4.9)\\
\text{and } \alpha(\tau) &\equiv \text{general phase function.} && (4.10)
\end{aligned}
$$

The phase function, $\alpha(\tau)$, and self-coherence function can be calculated for a specific source from its power spectral density. Substituting the general form for $\mathbf{\Gamma}(\tau)$ yields the following form for $I_D(h)$:

$$
I_D(h) = (K_1^2 + K_2^2) I_0 + 2K_1K_2\Gamma\left(\frac{2h}{c}\right) \cos\left\{\frac{4\pi h}{\lambda} - \alpha\left(\frac{2h}{c}\right)\right\} \tag{4.11}
$$

Assuming that the path length difference is nearly zero, $\Gamma\left(\frac{2h}{c}\right) \approx I_0$, and $\alpha\left(\frac{2h}{c}\right) \approx 0$. Making these substitutions reduces the interference pattern to a modulated sinusoidal pattern with the argument being a simple phase delay caused by the path length difference h which will be denoted, $\Delta\phi(h)$:

$$
I_D(h) = (K_1^2 + K_2^2) I_0 + 2K_1K_2 I_0 \cos\left\{\Delta\phi(h)\right\}. \tag{4.12}
$$

Similarly, the interference due to spatial separation can be found to be [24]:

$$
\begin{aligned}
I_Q(r_1, r_2) &= K_1^2\Gamma_{11}(0) + K_2^2\Gamma_{22}(0) + && (4.13)\\
&\quad 2K_1K_2\Gamma_{12}\left(\frac{r_2 - r_1}{c}\right) \cos\left\{2\pi\left(\frac{r_2 - r_1}{\lambda}\right) - \alpha_{12}\left(\frac{r_2 - r_1}{c}\right)\right\}, && (4.14)\\
\text{where } \Gamma_{11}(0) &\equiv \text{intensity at point } Q \text{ due to field propagating from pinhole } P_1, && (4.15)\\
\text{and } \Gamma_{12} &\equiv \text{cross-correlation of light from } P_1 \text{ and } P_2. && (4.16)
\end{aligned}
$$

If the separation of the pinholes is very small, then $\Gamma_{12}(\frac{r_2 - r_1}{c}) \approx \sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}$, and $\alpha_{12}\left(\frac{r_2 - r_1}{c}\right) \approx 0$. Making the substitution for Γ_{12} and replacing $\Gamma_{11}(0)$ and $\Gamma_{22}(0)$ with intensity variables I_1 and I_2 produces the following form:

$$
I_Q(r_1, r_2) = K_1^2 I_1 + K_2^2 I_2 + 2K_1K_2\sqrt{I_1 I_2} \cos\left\{\Delta\phi(r_2, r_1)\right\}. \tag{4.17}
$$

Comparing this result to the equation for temporal interference in (4.12), the only differences lie in the relative intensities and whether the phase shift was brought about by relative path delay or a spatial shift. Assuming that $K_1 = K_2 = 1$ and $I_1 = I_2 = I_0$ both expressions reduce to the simplified interference equation:

$$I = 2I_0 \left(1 + \cos\{\Delta\phi\}\right). \tag{4.18}$$

Interferometric wavefront sensors introduce a known shift or delay in a reference wavefront in one path and interfere it with the unaltered beam to generate interference patterns. From the unique characteristics of the interference pattern and the relationship in (4.18) above, the wavefront sensor will attempt to measure ϕ, $\Delta\phi$, or some set of parameters approximating ϕ. The degree of spatial and temporal coherence in a source will determine the amount of calibration and precision necessary to ensure that (4.18) applies and create measurable interference patterns.

Lateral Shear Interferometry. The Lateral Shearing Interferometer (LSI) is designed to provide a measure of the average wavefront slope. The LSI interferes a delayed and shifted portion of a collimated beam with the original beam. The degree of shift between the two beams is called the shear distance. The shear distance is small such that a significant portion of the two beams overlap. An interference pattern is visible within the region of overlap. Analysis of the resulting fringe pattern provides the differential phase, $\Delta\phi/\Delta s$, where Δs represents the shear distance. A simple version of the LSI can be created from a parallel plate as shown in Figure 4.2 [25]. The plate will produce two reflections separated and delayed. More sophisticated shearing interferometers use diffraction gratings as indicated in Figure 4.3 [26]. If a diffraction grating is used, care must be taken to avoid overlapping multiple orders of diffraction. The spacing of grating lines are typically designed such that the +1 and -1 order beams are tangent. LSIs are only capable of providing slope information in the shear direction. LSIs employed in wavefront sensing create shear in two directions to provide two-dimensional slope measurement. This requires a beamsplitter or cross-grating in the case of the diffraction grating style sensor. Splitting the beam by polarization is also a viable technique [27]. The diagrams in Figure 4.4, provide results from a Matlab simulation of lateral shear. LSIs are used extensively in optical system

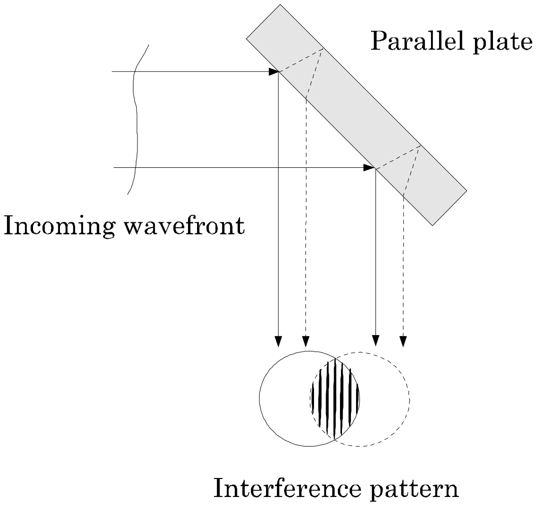


Figure 4.2 Diagram of a parallel plate lateral shear interferometer [26].

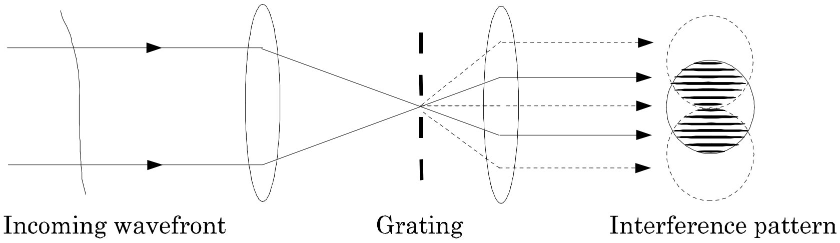


Figure 4.3 Diagram of a diffraction grating lateral shearing interferometer [26].

testing and have been successfully implemented in adaptive optics systems [28]. Shearing interferometers can be used to measure the OTF of an optical system [29]. The OTF is measured by continuously varying the level of shear and path delay. Fast methods exist for recording both the OTF and the Modulation Transfer Function (MTF), the magnitude of the OTF [30], [31].

Software wavefront reconstruction algorithms are used to recover the original phasefront from the interference patterns. The measured slope information is inherently modulo 2π which makes the process of integration nontrivial. This problem is evident in the basic equation for interference:

$$\begin{aligned} I &= 2I_0\left(1+\cos\{\Delta\phi\}\right), && (4.19) \\ \Delta\phi &= \cos^{-1}\left(\frac{I}{2I_0}-1\right), && (4.20) \\ \Delta\phi &\in [-\pi,\pi). && (4.21) \end{aligned}$$

The modulo 2π phase slope is referred to as wrapped phase information. Recovery methods must unwrap the phase and integrate to recreate the original wavefront. Sampling is critical in the phase unwrapping problem because phase changes greater than π radians between adjacent samples can seldom be resolved. Two general categories of wavefront reconstruction include least squares curve fitting methods and path following phase unwrapping algorithms [32], [33].

Point Diffraction Interferometry. Rather than measuring the average wavefront slope, it is possible to measure the wavefront directly by comparing it with an unaberrated reference wavefront. Interferometers that use a reference wavefront generated from the input wavefront are commonly called self-referencing interferometers (SRIs). Of particular importance among SRIs, is the Point Diffraction Interferometer (PDI). The PDI creates a reference wavefront from pinhole diffraction [25]. Smartt proposed a simple construction where the PDI focuses the input wavefront onto a pinhole etched out of a semi-transparent material [34]. The pinhole is sufficiently small (on the order of a few μm) such that it spatially filters out all of the incoming wavefront aberrations and passes only a smooth spherical reference wave. The remainder of the input wavefront is attenuated, but not spatially filtered, by the semi-transparent material. The wavefront must be attenuated

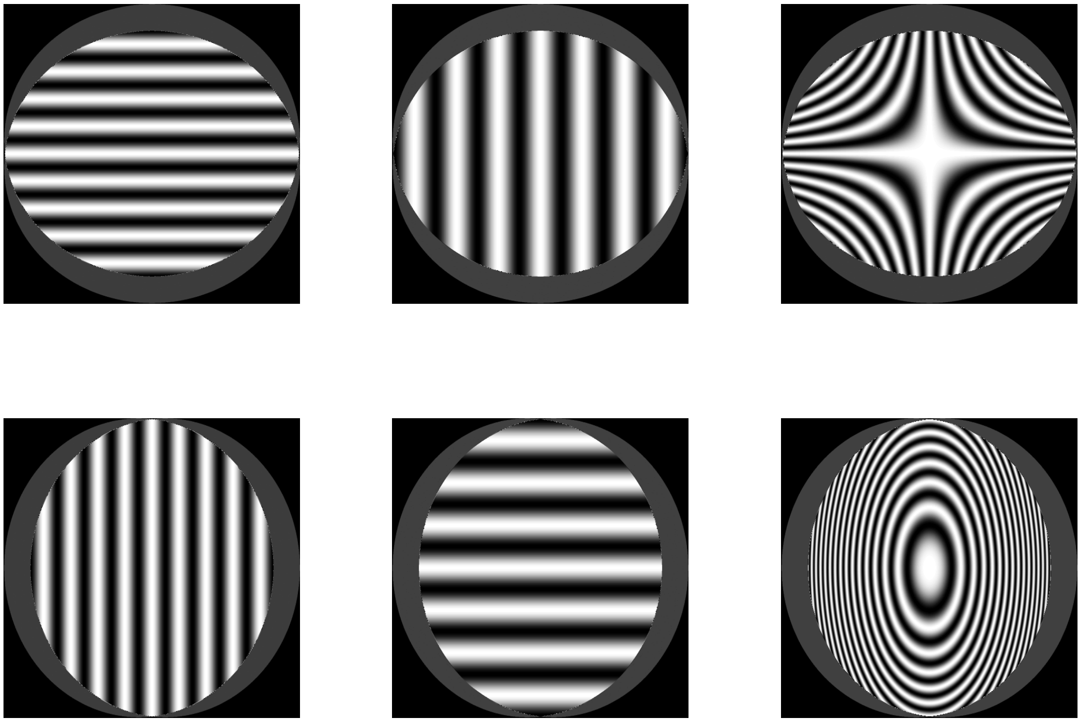


Figure 4.4 Simulated examples of lateral shear interference patterns. Vertical (top) and horizontal (bottom) shear directions for beams with defocus (left), astigmatism (center), and coma (right).

such that it has amplitude on the order of the reference wave in order to create visible interference fringes in the image plane. Visibility of fringes is directly related to the modulation of the sinusoidal pattern [2]:

$$\text{Visibility} \equiv \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}. \tag{4.22}$$

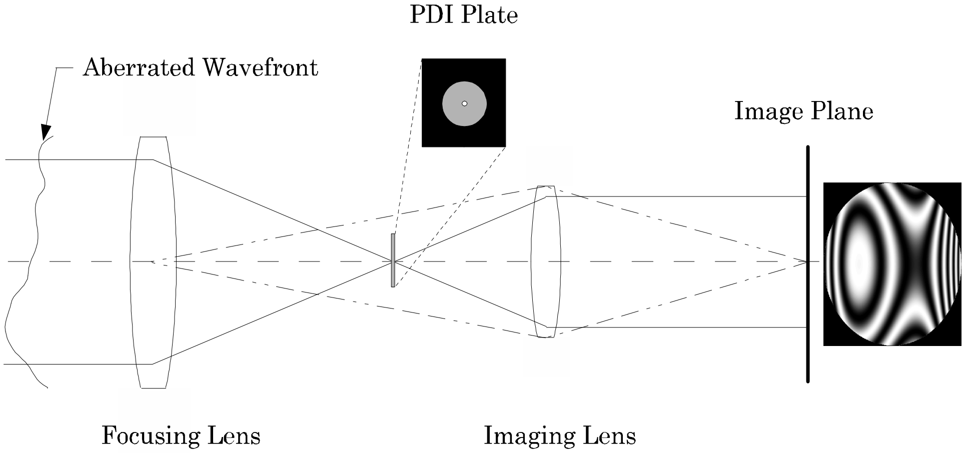


Figure 4.5 Diagram of a common path PDI [26].

As with the LSI, there are many variations on the PDI. In particular, PDIs may have multiple beam paths or a common beam path. Figure 4.5 provides an example diagram for a common path PDI [26]. Multiple beam paths provide for both spatial and temporal dithering of the reference wavefront for improved measurement precision. Common path PDIs, on the other hand, require less calibration and offer increased tolerance to vibrations and harsh operating environments [35]. PDIs may incorporate polarizers or birefringent materials to create orthogonal polarization between the reference beam and the input beam. Orthogonal polarization between the input beam and the reference beam provides for optimal fringe visibility [35]. Figure 4.6 provides simulated PDI interference patterns for the cases of independent defocus, astigmatism and coma.

PDIs, and more generally, self-referencing interferometers, have proven successful in adaptive optics applications [26]. While slightly more complicated than shearing interferometers and other wavefront sensor designs, PDIs offer several benefits. Since PDIs measure the wavefront phase directly, there is less error introduced in the wavefront reconstruction

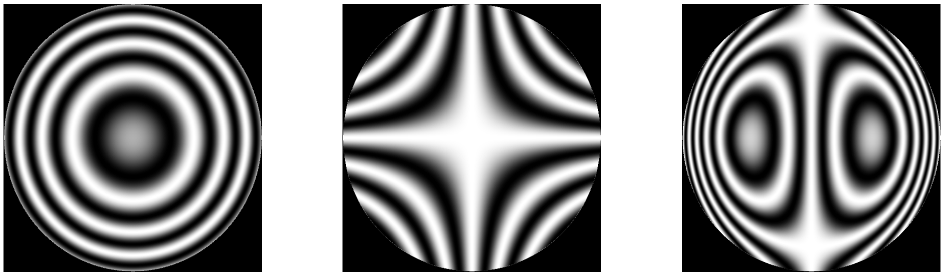


Figure 4.6 Example PDI interference patterns. Left: wavefront with defocus aberration. Center: wavefront with astigmatism. Right: wavefront with coma.

process. PDIs are also less sensitive to field amplitude noise, sometimes called scintillation, than the LSI and the Hartmann sensor to be discussed in Section 4.3 [36]. Common path PDIs are more effective for sources with low temporal coherence than shearing interferometers because the reference wavefront and the aberrated wavefront are not spatially shifted [37].

For all their benefits, PDIs create technical challenges as well. Since the reference wavefront is spatially filtered, the input beam must have enough power to provide sufficient signal to noise ratio (SNR) in the interference pattern. In the case of narrow-band, coherent inputs, Rhoadarmer et. al. describe a fiber laser amplification method which dramatically improves SNR [36]. The presence of large tilt terms in the input wavefront can shift the focus of the input beam away from the spatial filter. For this reason, inputs with large tilt variance force the adaptive optics system to correct for tilt prior to the PDI wavefront sensor, or to somehow incorporate a moving pinhole in the device [38]. Each of these challenges requires an engineering solution which brings with it some set of trade-offs in complexity and efficiency.

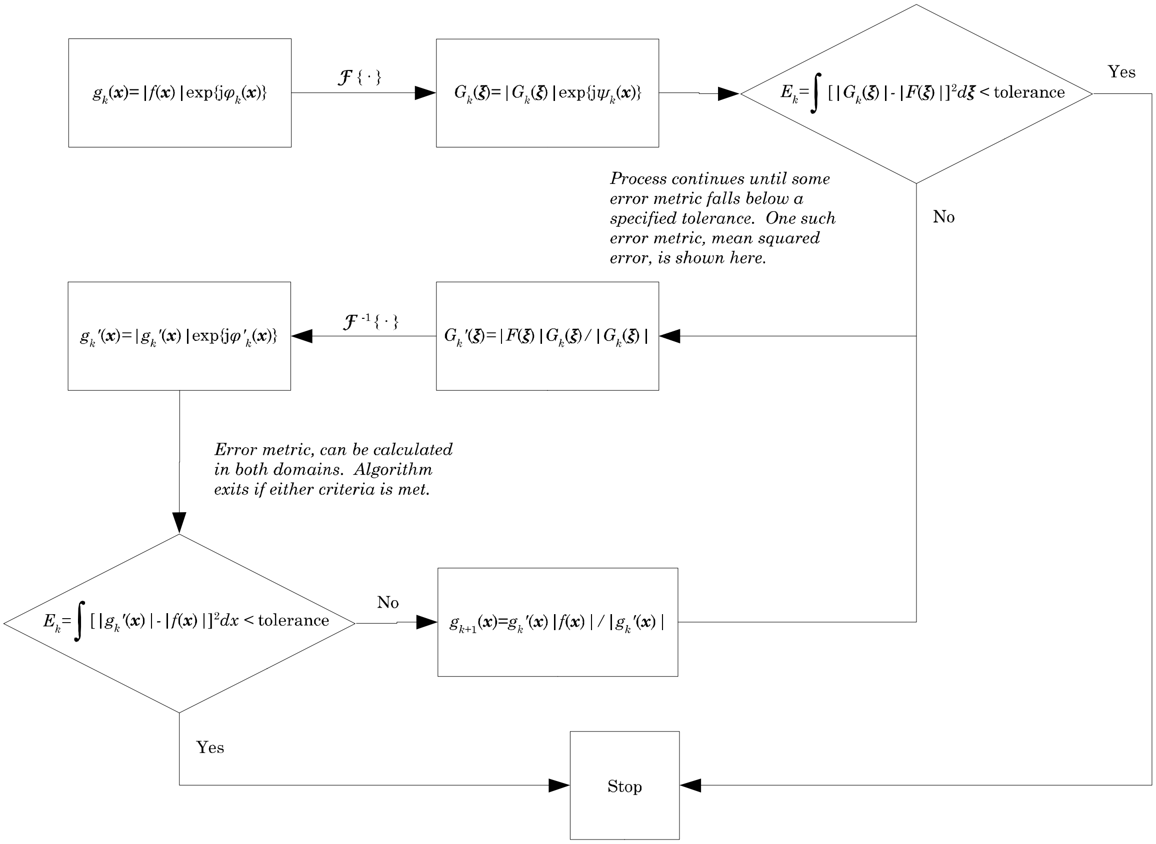
4.2 Phase Retrieval from Intensity Measurements

The problem of wavefront sensing in Adaptive Optics is a single instance of a broader class of problems in electro-optics often referred to as phase retrieval. The need for phase retrieval arises in many other applications such as electron microscopy, x-ray imaging and single-sideband communications [39], [40]. The general phase retrieval problem can be summarized using a Fourier domain model. Consider the complex function and its Fourier transform:

$$
\begin{aligned}
F(\boldsymbol{\xi}) &= |F(\boldsymbol{\xi})| \exp\{j\psi(\boldsymbol{\xi})\}, && (4.23)\\
f(\mathbf{x}) &= |f(\mathbf{x})| \exp\{j\phi(\mathbf{x})\}, && (4.24)\\
F(\boldsymbol{\xi}) &= \int_{-\infty}^{\infty} f(\mathbf{x}) \exp\{-\mathrm{j}2\pi\boldsymbol{\xi}\cdot\mathbf{x}\} d\mathbf{x}. && (4.25)
\end{aligned}
$$

The phase retrieval problem in optics is synonymous with the problem statement: given $|F(\boldsymbol{\xi})|$ and $|f(\mathbf{x})|$, find $\psi(\boldsymbol{\xi})$ and $\phi(\mathbf{x})$. As shown in (2.188), the linear systems diffraction model relates the object and image domains through simple Fourier analysis of the optical system. Thus, through the linear optical system model it is possible to translate the general phase retrieval problem into an optical wavefront sensing problem. In its most general form, the problem of phase retrieval has infinitely many solutions making the problem ill-posed. However, if the problem is constrained by making certain assumptions about $f(\mathbf{x})$, then the infinite set of solutions can become a limited set of solutions, possibly even a unique solution [41]. Gerchberg and Saxton devised an algorithm to solve the phase retrieval problem in electron microscopy [42]. Their approach employs an iterative technique which makes constraining adjustments in both the object and image domains. The simplest set of constraints forces the object and image domain amplitudes to match the measured values at each iteration. The phase is often seeded with a random process to begin the first iteration. The algorithm continues until some minimum error criteria is reached. The block diagram in Figure 4.7 describes the Gerchberg-Saxton (GS) phase retrieval algorithm. This technique is guaranteed to converge in the mean squared sense. However, there is no such guarantee that the resulting phase solution is unique. The non-uniqueness might be tolerable if it were limited to an added constant phase, however, the ambiguity also includes a possible sign change. Although limiting the set of solutions from infinitely many to a sign error is a significant step, the non-uniqueness problem limits the utility of the algorithm in wavefront sensing.

To overcome the issue of non-uniqueness, Misell modified the GS algorithm to incorporate information from two imaging paths [40]. By creating two optical paths with a known phase difference, or phase diversity, between them, the sign in the phase aberration could be resolved. One simple method for creating such a diversity is to purposefully offset



Process continues until some error metric falls below a specified tolerance. One such error metric, mean squared error, is shown here.

Error metric, can be calculated in both domains. Algorithm exits if either criteria is met.

Figure 4.7 Flow diagram of the general Gerchberg-Saxton algorithm.

each image plane thereby introducing a known amount of defocus. A diagram for Misell's algorithm using focus diversity is shown in Figure 4.8.

Fienup proposed a modification to the GS algorithm for image reconstruction [43]. The modified algorithm uses phase diversity much like the Misell approach, however, it is designed for use in post processing of images where only the image modulus is known. In this case, no knowledge is assumed about the pupil amplitude (the object is unknown) making the algorithm estimate both the object and the corrections necessary to improve image quality. The technique was designed for reconstruction of imagery from interferometric data. The key to Fienup's algorithm is that there are two characteristics known about the object: the object is both real and non-negative. These qualities introduce additional constraints into the algorithm. Fienup also proposed a method to increase the rate of convergence through negative feedback. Fienup's input-output version of the GS algorithm is diagrammed in Figure 4.9. The set γ represents all points where the estimated object

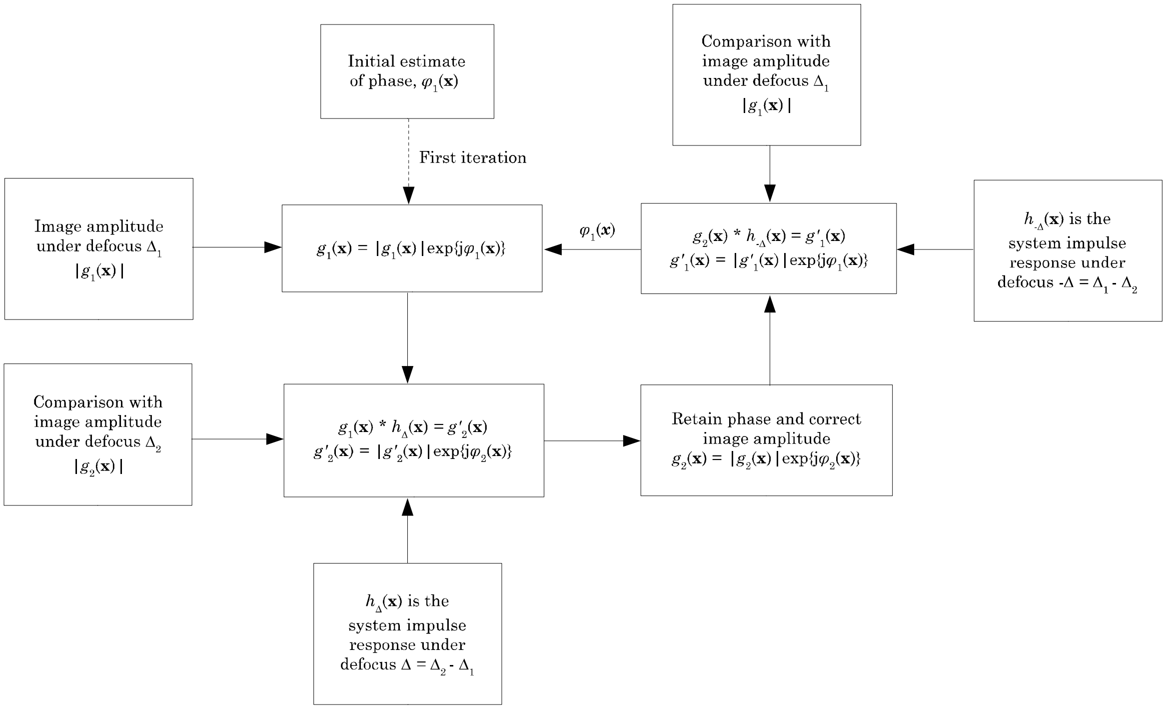


Figure 4.8 Block diagram of Misell's modified GS algorithm [40].

violates the real and non-negative constraints. Only the set of object points which violate the constraints are modified after each iteration. The Fourier domain constraints typically consist of the measured image amplitude.

Shortly after the first published version of the Gerchberg-Saxton algorithm, Gonsalves proposed a parameter searching algorithm for phase retrieval [45]. The method was later improved to include elements of phase diversity and combined wavefront phase and object estimation making it an enticing algorithm for use in a wavefront sensing applications [46]. The approximate pupil phase is parameterized by some suitable set of polynomials such as the first 21 Zernike polynomials. Estimates of the pupil phase are generated in each iteration of the algorithm. Gonsalves' algorithm boasts the capabilities of both Misell's and Fienup's modified GS algorithms. The method is based on a mean squared error estimator for the object and a gradient search routine to minimize error between the observed images and the current iterations estimated images. The phase diversity is defined just as in Misell's algorithm: any arbitrary, but known, phase difference between two image planes. As in Fienup's image reconstructor, the object does not need to be known making the algorithm ideal for extended source objects. Expressed here in the Fourier domain, the

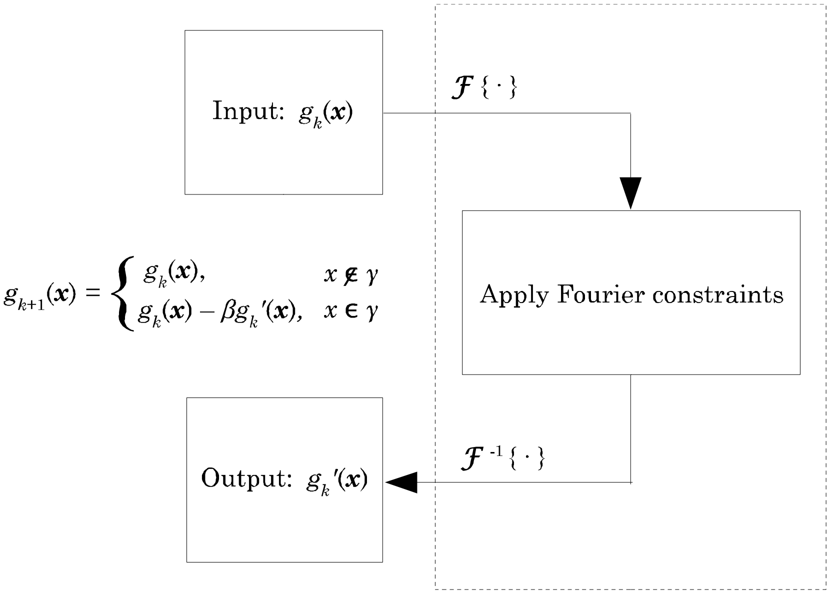


Figure 4.9 Flow diagram of Fienup's input-output modification of the GS algorithm for image reconstruction [44].

object estimator, $\hat{O}(\boldsymbol{\xi})$, is an optimum mean squared error estimator for $O(\boldsymbol{\xi})$ based the prior knowledge provided in the observed images, $G_1(\boldsymbol{\xi})$ and $G_2(\boldsymbol{\xi})$, and the pupil phase estimates $\hat{P}_1(\boldsymbol{\xi})$ and $\hat{P}_2(\boldsymbol{\xi})$:

$$\hat{O}(\boldsymbol{\xi}) = \frac{\hat{P}_1^*(\boldsymbol{\xi})G_1(\boldsymbol{\xi}) + \hat{P}_2^*(\boldsymbol{\xi})G_2(\boldsymbol{\xi})}{\hat{P}_1^*(\boldsymbol{\xi})\hat{P}_1(\boldsymbol{\xi}) + \hat{P}_2^*(\boldsymbol{\xi})\hat{P}_2(\boldsymbol{\xi})} \tag{4.26}$$

The mean squared error metric most often used in versions of Gonsalves' algorithm has

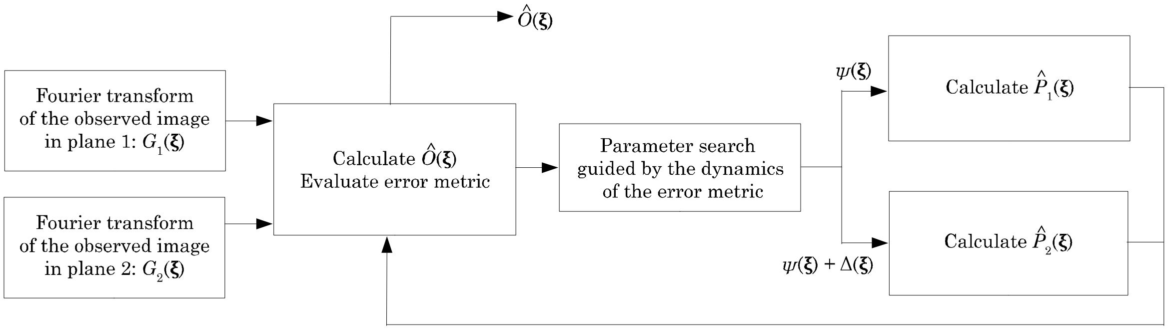


Figure 4.10 Block diagram of Gonsalves' parameter searching phase retrieval algorithm.

come to be known as the Gonsalves metric [47]:

$$E = \int \left| G_1(\boldsymbol{\xi}) - \hat{O}(\boldsymbol{\xi})\hat{P}_1(\boldsymbol{\xi}) \right| + \left| G_2(\boldsymbol{\xi}) - \hat{O}(\boldsymbol{\xi})\hat{P}_2(\boldsymbol{\xi}) \right| d\boldsymbol{\xi} \tag{4.27}$$

Simulated results of Gonsalves' algorithm as well as versions of the GS algorithm are quite remarkable. This is perhaps the reason that these algorithms have been the subject of much research over the past 30 years. Working from nothing more than a pair of distorted images and knowledge of the optical system, these iterative techniques are capable of achieving near diffraction limited performance. Unfortunately, phase retrieval techniques require significant computational power. Each version of the phase diversity algorithm is guaranteed to converge, however, the number of iterations and computational time required for convergence can be too great to be accomplished at the frequency of atmospheric dynamics. Until sufficient computational power becomes available, these algorithms continue to provide only an effective means of post processing recorded image data.

4.3 The Hartmann Wavefront Sensor

The use of interferometric techniques in the manufacture of optics becomes increasingly difficult as the size of the optics increase. For this reason, a special test was developed for manufacture of large telescope optics called the Hartmann test. The test consists of placing a mask of many small subapertures over the optic under test and measuring the focal length of each subaperture. A diagram of the Hartmann test is shown in Figure 4.11. The advantage of the Hartmann test over interferometric processes is that the Hartmann test can be conducted with relatively broadband sources meaning that it circumvents much of the spatial and temporal coherence constraints emphasized in the section on interferometric wavefront sensing. The Hartmann wavefront sensor is an adaptation of the Hartmann test. In the Hartmann wavefront sensor, the mask of subapertures is replaced by a small array of lenslets. The lenslet contribution is attributed to Dr. Roland Shack and, for that reason, the wavefront sensor is often referred to as the Shack-Hartmann wavefront sensor. Figure 4.12 provides a basic diagram of the Hartmann wavefront sensor.

The Hartmann wavefront sensor provides a measure of the wavefront slope much like the shearing interferometer. Each lenslet focuses onto its respective region of a CCD. The

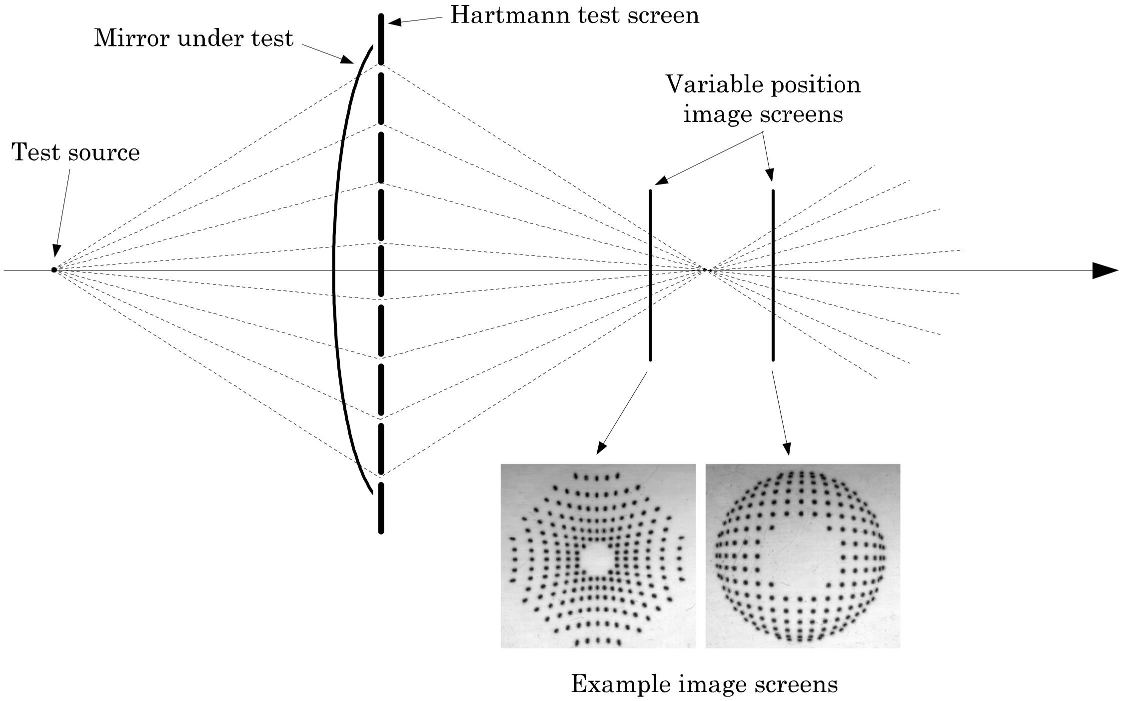


Figure 4.11 Schematic of a Hartmann test setup and example image plane output.

centroid of each subaperture image shifts proportional to the coefficients of Zernike tilt. By locating the centroid of each subaperture image, the sensor provides an estimate of local wavefront slope. If each lenslet image is approximated by a Gaussian shape, then it can be shown that the centroid calculation is a maximum likelihood estimator for tilt. Recall the Poisson conditional density for a detected image **D** given some noiseless image **I:**

$$p_{\mathbf{d}|\bar{\mathbf{a}}}(\mathbf{D}|\bar{\mathbf{A}}) = \prod_{\mathbf{u}\in\mathcal{S}} \frac{\mathbf{I}[\mathbf{u};\bar{\mathbf{A}}]^{\mathbf{D}[\mathbf{u}]}\exp(-\mathbf{I}[\mathbf{u};\bar{\mathbf{A}}])}{\mathbf{D}[\mathbf{u}]!}. \tag{4.28}$$

If **I** depends only on contributions from Zernikes 2 and 3 then the conditional density becomes:

$$p_{\mathbf{d}|a_2,a_3}(\mathbf{D}|A_2,A_3) = \prod_{\mathbf{u}\in\mathcal{S}} \frac{\mathbf{I}[\mathbf{u};A_2,A_3]^{\mathbf{D}[\mathbf{u}]}\exp(-\mathbf{I}[\mathbf{u};A_2,A_3])}{\mathbf{D}[\mathbf{u}]!}. \tag{4.29}$$

In the paraxial region of the image plane, there exists a linear relationship between the Zernike tilt coefficients A_2 and A_3 and independent pixel shifts in the ξ and η directions respectively. An incoming ray making an angle with the optical axis will intercept the image plane offset from the optical axis proportional to the image distance. Using this

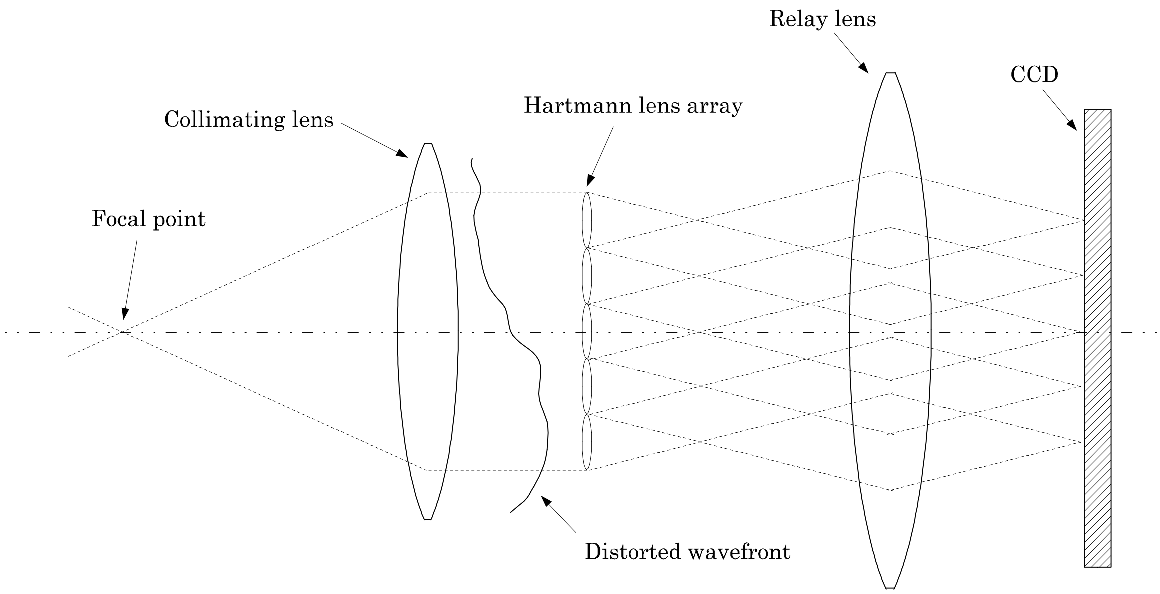


Figure 4.12 Diagram of a Hartmann sensor.

geometric analysis, Zernike tilt relates to the offset distance in the image plane by similar triangles as shown in Figure 4.13. Using this geometric construct, the expression for A_2 in

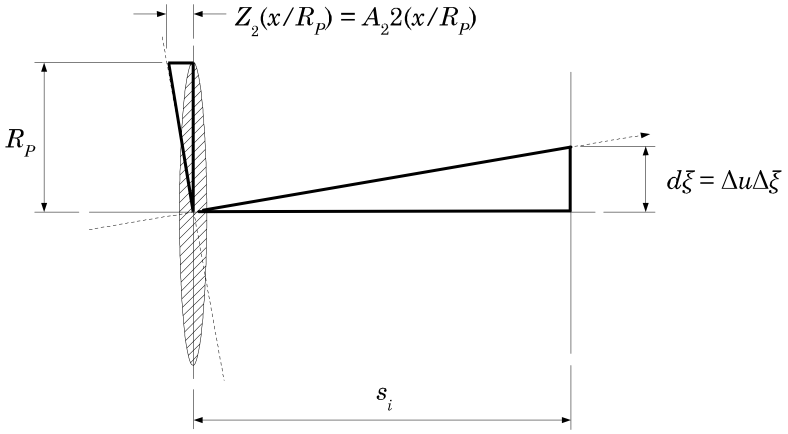


Figure 4.13 Ray optics diagram demonstrates the relationship between a single pixel shift in the image plane and the Zernike tilt parameter.

terms of an arbitrary pixel shift Δu is derived below. Begin with the expression for x-tilt:

$$A_2 Z_2\left(x, \theta\right) = A_2 2\left(\frac{x}{R_P}\right)\cos\left(\theta\right).$$

Substitute in the point along the x axis at the edge of the aperture, $(x = R_P, \theta = 0)$:

$$A_2 Z_2 \left(R_P, 0\right) = A_2 2. \tag{4.30}$$

Now form an expression equating the ratios of the sides of the similar triangles from Figure 4.13, and solve for A_2:

$$\begin{aligned} \frac{A_2 Z_2 \left(R_P, 0\right)}{R_P} &= \frac{d\xi}{s_i}, & (4.31) \\ A_2 &= \frac{R_P d\xi}{2 s_i}. & (4.32) \end{aligned}$$

The distance $d\xi$ has units meters. The index variable Δu which has units in pixels is more useful in the derivation to follow. Recall that the width of a pixel in the image plane given given by $\Delta\xi$ in (3.19). To relate the coefficient A_2 to pixel shifts, substite: $d\xi = \Delta u \Delta\xi$, into the expression for A_2 in (4.32) and solve for the pixel shift:

$$\begin{aligned} A_2 &= \frac{R_P \Delta u \Delta\xi}{2 s_i}, & (4.33) \\ \Delta u &= \frac{2 A_2 s_i}{R_P \Delta\xi}. & (4.34) \end{aligned}$$

Similarly, the expression for Δv is given:

$$\Delta v = \frac{2 A_3 s_i}{R_P \Delta\eta}. \tag{4.35}$$

To compact the notation, denote the pixel shifts as functions of the Zernike parameter and combine them into a vector format: $\mathbf{\Delta u}\left(A_2, A_3\right) = \left(\Delta u\left(A_2\right), \Delta v\left(A_3\right)\right)$. Now return to (4.29) and substitute the shift function, $\mathbf{\Delta u}\left(A_2, A_3\right)$, for the expected image arguments A_2 and A_3:

$$p_{\mathbf{d}|a_2,a_3}(\mathbf{D}|A_2, A_3) = \prod_{\mathbf{u}\in\mathcal{S}} \frac{\mathbf{I}[\mathbf{u};\mathbf{\Delta u}\left(A_2, A_3\right)]^{\mathbf{D}[\mathbf{u}]} \exp(-\mathbf{I}[\mathbf{u};\mathbf{\Delta u}\left(A_2, A_3\right)])}{\mathbf{D}[\mathbf{u}]!}. \tag{4.36}$$

Note that this substitution does not change the expression except to reveal to the reader the direct relationship between the tilt parameters and the xy shift of the expected image in the image plane. Now make the approximation that the expected image, $\mathbf{I}$, has a Gaussian spatial distribution. This approximation is based on the fact that the central lobe of the Fraunhofer diffraction pattern for a diffraction limited optical system resembles the bell shaped curve of a Gaussian distribution. Approximate the noiseless image as:

$$\begin{aligned}
\mathbf{I}[\mathbf{u}-\mathbf{\Delta u}\left(A_2, A_3\right)] &= \frac{K}{2\pi\sigma^2}\exp\left\{-\frac{\left|\mathbf{u}-\mathbf{\Delta u}\left(A_2, A_3\right)\right|^2}{2\sigma^2}\right\}, && (4.37)\\
\text{where } K &= E\left\{\sum_{\mathbf{u}\in\mathcal{S}} \mathbf{D}[\mathbf{u}]\right\}, && (4.38)\\
\text{and } \sigma^2 &= \begin{array}{c}\text{variance parameter based on}\\ \text{the system } f/\#.\end{array} && (4.39)
\end{aligned}$$

The ML estimator log likelihood expression using the input conditional pdf in (4.36) is given by:

$$L_{ml}\left(A_2, A_3\right) = \sum_{\mathbf{u}\in\mathcal{S}} \mathbf{D}[\mathbf{u}] \ln\left\{\mathbf{I}[\mathbf{u};\mathbf{\Delta u}\left(A_2, A_3\right)]\right\} - \mathbf{I}[\mathbf{u};\mathbf{\Delta u}\left(A_2, A_3\right)]. \qquad (4.40)$$

Substituting the Gaussian form of $\mathbf{I}$ into the log likelihood expression:

$$\begin{aligned}
L_{ml}\left(A_2, A_3\right) &= \sum_{\mathbf{u}\in\mathcal{S}} \mathbf{D}[\mathbf{u}] \ln\left\{\frac{K}{2\pi\sigma^2}\right\} - \mathbf{D}[\mathbf{u}]\frac{\left|\mathbf{u}-\mathbf{\Delta u}\left(A_2, A_3\right)\right|^2}{2\sigma^2} - \\
&\quad \frac{K}{2\pi\sigma^2}\exp\left\{-\frac{\left|\mathbf{u}-\mathbf{\Delta u}\left(A_2, A_3\right)\right|^2}{2\sigma^2}\right\}. \qquad (4.41)
\end{aligned}$$

Solving for the ML estimator requires maximizing the log likelihood expression over the range of the shift function, $\mathbf{\Delta u}$, by differentiating with respect to each parameter and setting the result equal to zero.

$$\begin{aligned}
L_{ml}\left(A_2, A_3\right) &= \sum_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v] \ln\left\{\frac{K}{2\pi\sigma^2}\right\} - \mathbf{D}[u,v]\frac{\left(u-\Delta u\left(A_2\right)\right)^2 + \left(v-\Delta v\left(A_3\right)\right)^2}{2\sigma^2} - \\
&\quad \frac{K}{2\pi\sigma^2}\exp\left\{-\frac{\left(u-\Delta u\left(A_2\right)\right)^2}{2\sigma^2}\right\}\exp\left\{-\frac{\left(v-\Delta v\left(A_3\right)\right)^2}{2\sigma^2}\right\}, \qquad (4.42)\\
\frac{d}{dA_2}L_{ml}\left(A_2, A_3\right) &= \sum_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]\frac{u-\Delta u\left(A_2\right)}{\sigma^2}\frac{2s_i}{R_P\Delta\xi} - \\
&\quad \frac{u-\Delta u\left(A_2\right)}{\sigma^2}\frac{2s_i}{R_P\Delta\xi}\frac{K}{2\pi\sigma^2}\times \\
&\quad \exp\left\{-\frac{\left(u-\Delta u\left(A_2\right)\right)^2}{2\sigma^2}\right\}\exp\left\{-\frac{\left(v-\Delta v\left(A_3\right)\right)^2}{2\sigma^2}\right\}, \qquad (4.43)\\
\frac{d}{dA_3}L_{ml}\left(A_2, A_3\right) &= \sum_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]\frac{v-\Delta v\left(A_3\right)}{\sigma^2}\frac{2s_i}{R_P\Delta\eta} - \\
&\quad \frac{v-\Delta v\left(A_3\right)}{\sigma^2}\frac{2s_i}{R_P\Delta\eta}\frac{K}{2\pi\sigma^2}\times \\
&\quad \exp\left\{-\frac{\left(u-\Delta u\left(A_2\right)\right)^2}{2\sigma^2}\right\}\exp\left\{-\frac{\left(v-\Delta v\left(A_3\right)\right)^2}{2\sigma^2}\right\}. \qquad (4.44)
\end{aligned}$$

Assuming that the shift function is distributed Gaussian, the second term in each derivative represents a constant multiplying the central moment of the shift function:

$$\sum_{[u,v]\in\mathcal{S}} \left(u-\Delta u\left(A_2\right)\right)\exp\left\{-\frac{\left(u-\Delta u\left(A_2\right)\right)^2}{2\sigma^2}\right\} = 0. \qquad (4.45)$$

Under the atmospheric model, the central moment for all modes is zero. This makes solving for the parameter estimates straightforward:

$$\sum_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]\frac{u-\Delta u\,(A_2)}{\sigma^2}\frac{d\Delta u\,(A_2)}{dA_2} = 0, \tag{4.46}$$

$$\sum_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]\,(u-\Delta u\,(A_2)) = 0, \tag{4.47}$$

$$\text{therefore } \Delta\hat{u}_{ml} = \frac{\sum\limits_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]u}{\sum\limits_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]}, \tag{4.48}$$

$$\sum_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]\frac{v-\Delta v\,(A_3)}{\sigma^2}\frac{d\Delta v\,(A_3)}{dA_3} = 0, \tag{4.49}$$

$$\sum_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]\,(v-\Delta v\,(A_3)) = 0, \tag{4.50}$$

$$\text{therefore } \Delta\hat{v}_{ml} = \frac{\sum\limits_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]v}{\sum\limits_{[u,v]\in\mathcal{S}} \mathbf{D}[u,v]}. \tag{4.51}$$

Thus, the maximum likelihood estimator for each shift parameter is a centroid calculation. The estimators for Δu and Δv differ from the estimators for the Zernike coefficients by a constant multiplier:

$$\hat{a}_{2_{ml}} = \frac{R_P\Delta\xi}{2s_i}\Delta\hat{u}_{ml}, \tag{4.52}$$

$$\hat{a}_{3_{ml}} = \frac{R_P\Delta\eta}{2s_i}\Delta\hat{v}_{ml}. \tag{4.53}$$

Each lenslet in the Hartmann sensor provides the same information as a single pixel in the shearing interferometer [3]. The Hartmann sensor is preferred over shearing interferometers in low light and low signal to noise ratio viewing [48]. Although its performance is often better than that of shearing interferometers, the centroid calculation in each subaperture of the Hartmann sensor is still prone to error in low SNR environments. Additionally, the Gaussian distribution in the image may not be an accurate assumption when the object is an extended source.

4.4 Wavefront Sensing Using Image Projections

The Hartmann wavefront sensor can be modified to operate on image projections. The short wavelength adaptive techniques (SWAT) wavefront sensor system at MIT Lincoln Labs uses this technique to accelerate the tilt estimation process [49]. The basic configuration of the projection based sensor requires two cameras providing orthogonal image projections as described in Figure 4.14. This configuration divides the amount of light into two paths and reduces the amount of signal available in each camera output. Therefore, what the projection based wavefront sensor loses in signal to noise ratio, it must make up for in speed, efficiency and reduction in read noise. In the SWAT wavefront sensor, projections are formed on the CCD outputting only N locations for an $N \times N$ image space. The background section on image vector projections underlined the importance of vector output from the standpoint of CCD read out time and CCD read noise. Thus, the first design

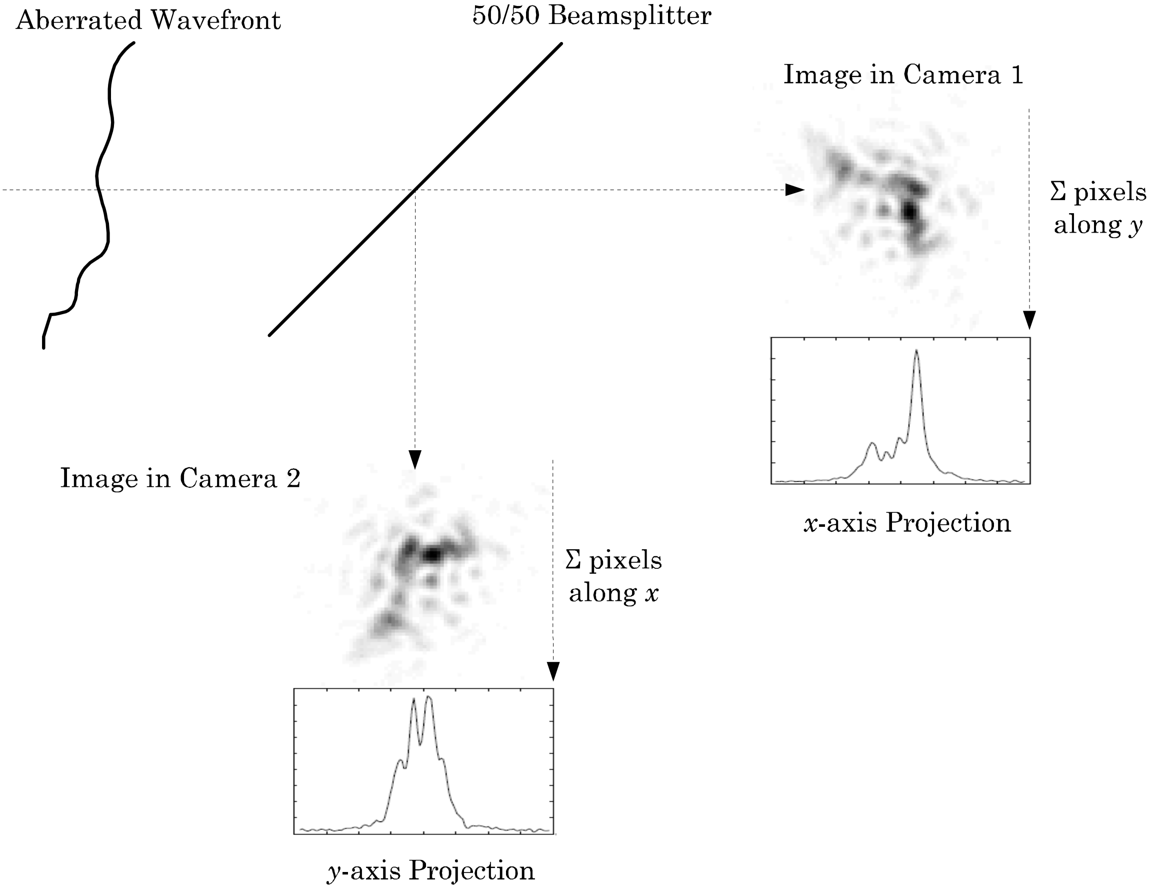


Figure 4.14 Figure demonstrates a conceptual example of two orthogonal projections provided by cameras in a projection correlating tilt sensor.

improvement over a simple two-dimensional centroid sensor is to simply compute part of the centroid operation directly on the CCD prior to read out.

The centroiding sensor may be improved further. To improve the quality of subaperture tilt estimates in low signal to noise environments and provide a more accurate tilt estimator for extended objects, the centroiding calculation should be replaced by an image correlating technique [50]. The cross-correlating tilt estimator provides for greater noise rejection and is better suited for extended source objects [51]. The cross-correlation calculation requires more compuational power than a simple centroid calculation. However, the image projection reduces the calculation time required for correlation style image registration making the calculation feasible. The following derivation outlines the efforts of Cain et. al. [51].

Recall the general projection operator from Section 3.3. Using this notation, the two projections suggested by Cain may be expressed:

$$
\begin{aligned}
& {}_{\mathbf{s}}\mathbf{v}\left(\mathbf{D}_{1,0}\right) && (4.54)\\
\text{and}\quad & {}_{\mathbf{s}}\mathbf{v}\left(\mathbf{D}_{2,90}\right), && (4.55)\\
\text{where}\quad & \mathbf{s} = \left\{(1, N_W)\right\}. && (4.56)
\end{aligned}
$$

The window size, N_W, has been left arbitrary. Consider, once again, the tilt only conditional pdf given in (4.29). Substituting the image projection into (4.29) and separating the conditional pdf into two independent tilt expressions gives:

$$
p_{{}_{\mathbf{s}}\mathbf{v}(\mathbf{d}_{1,0})|a_2}\left({}_{\mathbf{s}}\mathbf{v}\left(\mathbf{D}_{1,0}\right)|A_2\right) = \prod_{l=1}^{N_W} \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{1,0}\left[A_2\right]\right)^{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{1,0}[\mathbf{u}]\right)} \exp\left({}_{\mathbf{s}}\mathbf{v}_l\left(-\mathbf{I}_{1,0}\left[A_2\right]\right)\right)}{{}_{\mathbf{s}}\mathbf{v}_l\left(\{\mathbf{D}_{1,0}[\mathbf{u}]! : u \in \mathcal{S}_1\}\right)}, \tag{4.57}
$$

$$
p_{{}_{\mathbf{s}}\mathbf{v}(\mathbf{d}_{2,90})|a_3}\left({}_{\mathbf{s}}\mathbf{v}\left(\mathbf{D}_{2,90}\right)|A_3\right) = \prod_{l=1}^{N_W} \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{2,90}\left[A_3\right]\right)^{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{2,90}\right)} \exp\left({}_{\mathbf{s}}\mathbf{v}_l\left(-\mathbf{I}_{2,90}\left[A_3\right]\right)\right)}{{}_{\mathbf{s}}\mathbf{v}_l\left(\{\mathbf{D}_{2,90}[\mathbf{u}]! : u \in \mathcal{S}_2\}\right)}. \tag{4.58}
$$

```
$\Delta\hat{u}_{ml} = 0$
while $L_{a_2}\left(\Delta\hat{u}_{ml}\right) < \left\{L_{a_2}\left(\Delta\hat{u}_{ml}-2\right), L_{a_2}\left(\Delta\hat{u}_{ml}-1\right), L_{a_2}\left(\Delta\hat{u}_{ml}+1\right), L_{a_2}\left(\Delta\hat{u}_{ml}+2\right)\right\}$
    if $L_{a_2}\left(\Delta\hat{u}_{ml}+1\right) + L_{a_2}\left(\Delta\hat{u}_{ml}+2\right) > L_{a_2}(\Delta\hat{u}_{ml}-1) + L_{a_2}(\Delta\hat{u}_{ml}-2)$
        $\Delta\hat{u}_{ml} = \Delta\hat{u}_{ml} + 1$
    else
        $\Delta\hat{u}_{ml} = \Delta\hat{u}_{ml} - 1$
    end
end
```

Table 4.1 Pseudo code for the 1-D linear search algorithm.

From (4.57) and (4.58), the log likelihood functions are given:

$$L_{a_2}(A_2) = \sum_{l=1}^{N_W} {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{1,0}\right)\ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{1,0}\left[A_2\right]\right)\right\} - {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{1,0}\left[A_2\right]\right), \tag{4.59}$$

$$L_{a_3}(A_3) = \sum_{l=1}^{N_W} {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{2,90}\right)\ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{2,90}\left[A_3\right]\right)\right\} - {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{2,90}\left[A_2\right]\right). \tag{4.60}$$

Note that the reference image projections, $\mathbf{I}_{1,0}$, and $\mathbf{I}_{2,90}$ no longer have a specified form such as the Gaussian approximation given in the centroid estimation case. The reference image projections may be formed from a known object or from an ensemble average of images in the case where the object is unknown. The reference images can be posed as functions of the arbitrary pixel shift variables from the previous section:

$$L_{a_2}(\Delta u) = \sum_{l=1}^{N_W} {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{1,0}\right)\ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{1,0}\left[\Delta u\right]\right)\right\} - {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{1,0}\left[\Delta u\right]\right), \tag{4.61}$$

$$L_{a_3}(\Delta v) = \sum_{l=1}^{N_W} {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{2,90}\right)\ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{2,90}\left[\Delta v\right]\right)\right\} - {}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{2,90}\left[\Delta v\right]\right). \tag{4.62}$$

Expressions (4.59) and (4.60) are essentially correlation functions. Maximizing these correlation expressions over a carefully selected region of the shift parameters, Δu and Δv, is the job of the correlating tilt wavefront sensor. Cain suggests a linear search method for maximizing the correlation [1]. Examining the case for Δu, update the current estimate $\Delta\hat{u}_{ml}$ via the algorithm in Table 4.1. The diagram in Figure 4.15 describes the search algorithm. The granularity of this approach is limited by the pixel size in the detector, $\Delta\xi$. The shift parameters can be estimated to an accuracy finer than a pixel dimension by

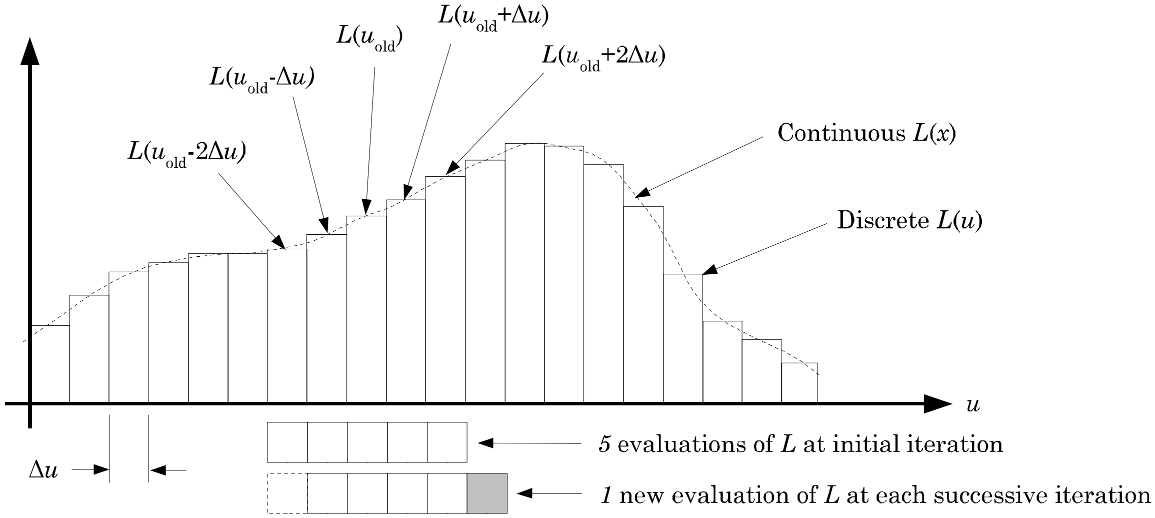


Figure 4.15 Diagram for the 1D log likelihood maximization algorithm.

interpolating the reference image vector, $\mathbf{v}(\mathbf{I}_{1,0})$ [1]. A linearly interpolated vector $\mathbf{v}(\mathbf{I}_{1,0})$ is formed as follows:

$$\begin{aligned}\mathbf{v}\left(\mathbf{I}_{1,0}[u-\Delta u]\right) &= \left(1-\Delta u+\text{floor}\left(\Delta u\right)\right)\mathbf{v}\left(\mathbf{I}_{1,0}[u-\text{floor}(\Delta u)]\right)+ \\ &\quad \text{floor}\left(\Delta u\right)\mathbf{v}\left(\mathbf{I}_{1,0}[u-\left(\text{floor}\left(\Delta u\right)+1\right)]\right), \qquad (4.63)\end{aligned}$$

where Δu is the arbitrary shift parameter, not restricted to integer increments.

4.5 Summary

This chapter reviewed many existing techniques for detecting wavefront phase from intensity measurements. The discussion began with interferometric techniques. The Lateral Shearing interferometer provides a measurement of the wavefront phase slope and the Self-Referencing Interferometer provides a measurement of the wavefront phase. Interferometric techniques have been proven to be effective in adaptive optics systems. Phase retrieval methods were also discussed. These methods are predominantly employed in offline analysis due to the computational complexity involved. The Hartmann sensor is an adaptation of the Hartmann test used for measuring the imperfections in large optics. The Hartmann wavefront sensor is very common due to its speed and effectiveness for measuring localized wavefront tilt. A short derivation demonstrated that the Hartmann wavefront sensor centroid measurement is a closed form solution to the maximum likelihood image shift estimator. The closed form expression for the ML shift estimator comes at the expense of assuming a Gaussian distributed image intensity. The SWAT sensor modifies the Hartmann wavefront sensor to use image projections. The image projection significantly reduces read out noise and CCD read time. Cain's projection correlating wavefront sensor makes use of image projections as well. The projection correlating tilt sensor uses a linear search method for maximizing the likelihood expression rather a closed form expression. Cain's simulation of the projection correlating tilt estimator shows that its performance is better than that of the centroiding tilt estimator under low light conditions. The correlating technique which provides superior tilt estimation performance can be adapted to estimate higher order Zernike terms, specifically the seven Zernikes: Z_4 through Z_{10}. The remainder of this dissertation describes two parameter estimating wavefront sensors designed to estimate and correct for both the tilt polynomials: Z_2 and Z_3, and higher order Zernikes up to Z_{10}.

5. *The Z_{2-4} Wavefront Sensor*

This chapter details the first of two wavefront curvature sensors. Recall that a curvature sensor is a wavefront sensor designed to detect some limited set of higher order aberrations along with x-tilt and y-tilt. The Z_{2-4} wavefront sensor measures Zernike coefficients a_2 through a_4 from point source image vectors. The Z_{2-4} wavefront sensor design builds on the projection correlating maximum likelihood (ML) tilt estimator [1]. This curvature sensor differs from the tilt sensor in several ways. The first, and perhaps the most crucial, difference is the assumption that the sensor will be configured to image a known beacon object. Second, the curvature sensor estimates defocus error within each subaperture. Also, the curvature sensor hardware requires a defocus diversity between the pair of CCD arrays. The estimation algorithm, which I will refer to as the Z_{2-4} estimator, is based on a maximum a posteriori (MAP) estimator versus a maximum likelihood (ML) estimator. Additionally, the expected image lookup has been expanded to account for parameters beyond tilt and the expected images take advantage of the known object assumption. Finally, the likelihood maximization approach has been updated to increase speed and efficiency. While not as effective at registering arbitrary images due to the beacon object assumption, the curvature sensor presented here outperforms the ML and centroid techniques when simulated using a point source input. Below, this chapter provides the details of the Z_{2-4} curvature sensor which include: a description of the hardware considerations, a derivation of the projection based Z_{2-4} estimator algorithm, techniques for fast and efficient likelihood maximization, and a summary of key sensor design variables.

5.1 Sensor Hardware

The focus of this research is on the design of the sensor estimation algorithm rather than sensor hardware design. In keeping with this theme, the purpose of describing the hardware configuration will be limited to identifying key design variables and how they affect the estimation algorithm. The sensor hardware can be broken down into three main components: an array of subapertures, a beamsplitter, and a pair of photon counting CCDs. This configuration is not unique and, in fact, includes the same combination of components discussed in the review of Hartmann sensors and phase diversity techniques. This section will review these three components and highlight any assumptions or requirements specific to the Z_{2-4} estimator algorithm.

The key design variables associated with the subaperture array are the diameter and focal length of the subapertures. The wavefront sensor will be constructed from an array of subapertures. Assume that the diameters of all individual subapertures as well as their focal lengths are identical. Furthermore, assume that the pixel size in the CCDs is chosen according to the Nyquist criteria which is adjusted for $f/\#$. Under this assumption, the focal length becomes transparent to the estimator algorithm design. Therefore, assume that the design of the subaperture array only impacts the estimation algorithm via the ratio of the fixed subaperture diameter to the changing operating environment variables: r_0, L_0, l_0. Since each of the atmospheric variables are estimated parameters provided to the sensor algorithm, it will be important to evaluate the performance of the estimator algorithm over a range of these values and to evaluate the sensitivity of that performance to errors in each estimate.

The beamsplitting device allows the sensor to focus the subaperture array onto two image planes. Lee et. al. demonstrated that the ideal configuration for the beamsplitting device is to provide equal power in both imaging paths in a phase retrieval system [52]. The same performance characteristic holds for the projection based algorithm. Therefore, as a conservative figure, I will assume that the beamsplitter used is a 50/50 beamsplitter with 95% efficiency. The efficiency factor is crucial when comparing the Z_{2-4} curvature sensor to other sensors that do not require beamsplitting.

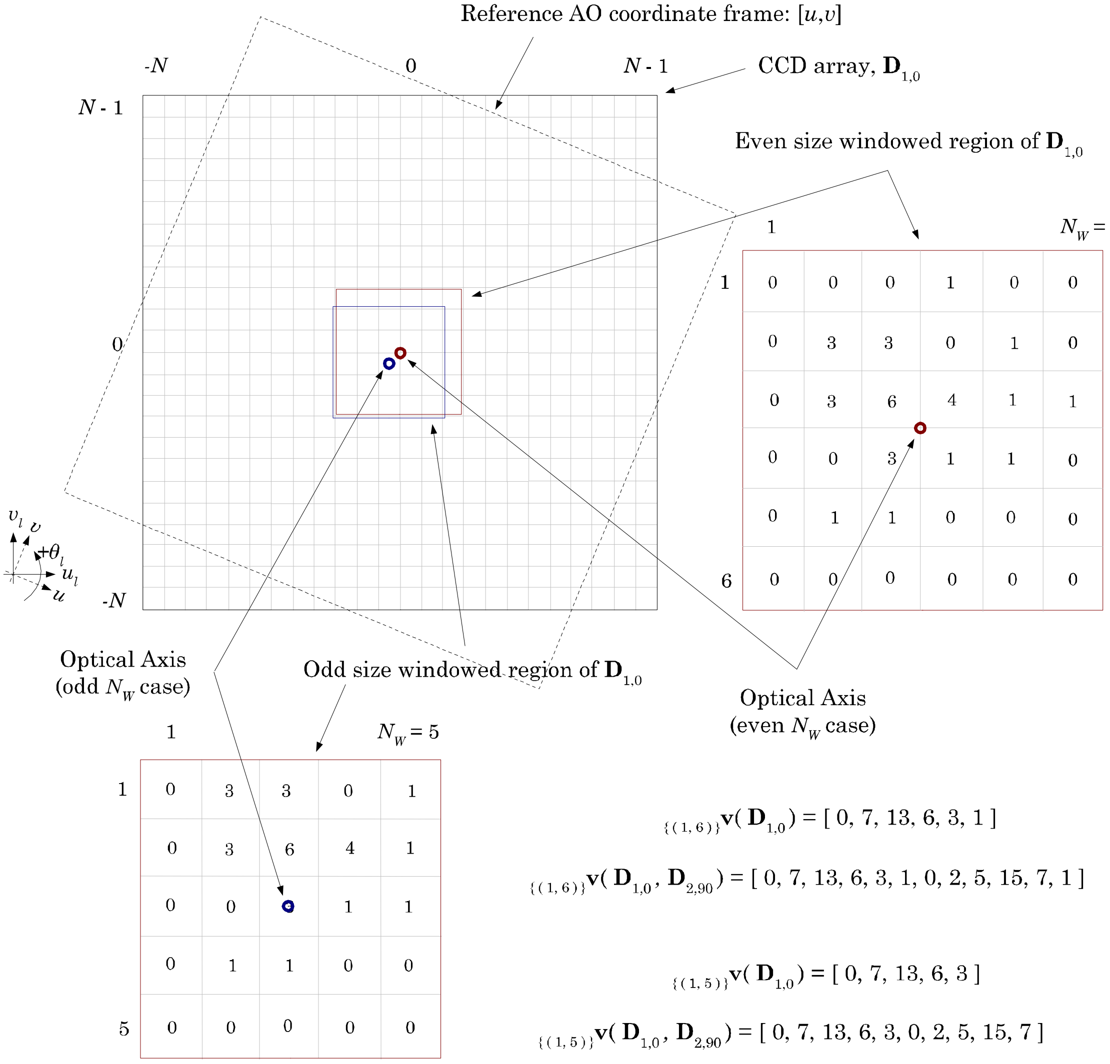
The key design assumptions associated with the CCD arrays are the Nyquist sampled pixel size discussed previously, the ability to produce image projections, and their placement relative to the geometric focus. The first two considerations are straightforward, which leaves the variable of CCD placement open for trade study. Ambiguity in the effects of higher order Zernikes on intensity measurements can be reduced by applying a known phase diversity between the two image planes. In the case of the Z_{2-4} sensor, the phase diversity is necessary because the defocus parameter exhibits identical effects in intensity whether the coefficient is positive or negative. Misell suggested that the simplest method for adding a known phase diversity is to introduce a defocus error by purposefully offsetting the image plane [40]. Lee et. al. showed that the defocus diversity should be applied equally and with opposite sign in each imaging path when using two-dimensional image data for phase retrieval [52]. I have confirmed that the same defocus diversity configuration is ideal for estimating phase from image projections. For this reason, the sensor defocus diversity labeled, $\pm\delta_{a_4}$, will be expressed as an absolute value in units of radians, where it is understood that the defocus will be applied positive in one CCD plane and negative in the other. The ideal choice of diversity will depend largely on both the ratio $\frac{D_P}{r_0}$ and the average photon count per subaperture per exposure, K.

5.2 Image Projections

The Z_{2-4} sensor image projection is a shifted and summed set of pixels from the original image as indicated in Figure 5.1. As shown in Figure 5.1, each CCD has an associated angle of rotation. Applying a rotation to a CCD allows for taking projections in different directions. For convenience of projection notation, the summation is always performed across the v_i direction. Specifying the projection direction is accomplished by associating with each CCD a rotational reference, θ_i, relative to the AO reference image axes. For a description of the variables defining the aperture and image plane, recall the notation for a discrete image formed within the linearized optical model:

$$\begin{aligned} I\left[u\Delta\xi, v\Delta\eta; \bar{\mathbf{a}}\right] &= \left|\frac{\Delta x \Delta y}{\left(\lambda s_i\right)^2} \mathcal{DFT}\left\{\mathcal{P}\left[n\Delta x, m\Delta y; \bar{\mathbf{a}}\right]\right\}\right|^2, && (5.1) \\ \mathbf{I}\left[u, v; \bar{\mathbf{a}}\right] &= I\left[u\Delta\xi, v\Delta\eta; \bar{\mathbf{a}}\right], && (5.2) \\ \mathbf{I}\left[\bar{\mathbf{a}}\right] &= \left\{\mathbf{I}[u, v; \bar{\mathbf{a}}] : u, v \in \mathcal{S}\right\}. && (5.3) \end{aligned}$$

As an example, consider estimating tilt using two orthogonal projections along the reference AO x and y axes as in Cain's ML estimator. This is described in the context of the image projection operator by establishing two CCDs one offset by an angle of 0° and another offset by 90°. The Z_{2-4} sensor uses this same configuration. As such, the CCD images are denoted: $\mathbf{D}_{1,0}$ and $\mathbf{D}_{2,90}$. The pair of projections used in the Z_{2-4} estimator will be referred to as whole plane projections. Whole plane projections are single vectors produced by summing along the entire windowed region v axis. The size of the window should be chosen to provide some minimal residual error and acceptable computation time. For the purpose of this description, the window size will remain variable. Although image projections are



0	0	0	1	0	0
0	3	3	0	1	0
0	3	6	4	1	1
0	0	3	1	1	0
0	1	1	0	0	0
0	0	0	0	0	0

0	3	3	0	1
0	3	6	4	1
0	0		1	1
0	1	1	0	0
0	0	0	0	0

$$_{\{(1,6)\}}\mathbf{v}(\mathbf{D}_{1,0}) = [\,0, 7, 13, 6, 3, 1\,]$$

$$_{\{(1,6)\}}\mathbf{v}(\mathbf{D}_{1,0}, \mathbf{D}_{2,90}) = [\,0, 7, 13, 6, 3, 1, 0, 2, 5, 15, 7, 1\,]$$

$$_{\{(1,5)\}}\mathbf{v}(\mathbf{D}_{1,0}) = [\,0, 7, 13, 6, 3\,]$$

$$_{\{(1,5)\}}\mathbf{v}(\mathbf{D}_{1,0}, \mathbf{D}_{2,90}) = [\,0, 7, 13, 6, 3, 0, 2, 5, 15, 7\,]$$

Figure 5.1 Diagram of the Z_{2-4} sensor's whole-plane image projection operation for even and odd length windows.

read out of the CCDs only once during each exposure, the projection operations used in the estimators: $\hat{a}_2$, $\hat{a}_3$ and $\hat{a}_4$ are distinguished in the notation as if they were separate vectors for mathematical convenience. The three Z_{2-4} sensor projection operations are:

$$_{\{(1,N_W)\}}\mathbf{v}\left(\mathbf{D}_{1,0}\right), \tag{5.4}$$

$$_{\{(1,N_W)\}}\mathbf{v}\left(\mathbf{D}_{2,90}\right), \tag{5.5}$$

$$_{\{(1,N_W)\}}\mathbf{v}\left(\mathbf{D}_{1,0}, \mathbf{D}_{2,90}\right). \tag{5.6}$$

The sensor hardware provides two image projections from each subaperture. Figure 5.2 diagrams the read out and flow of two image projections through the estimator algorithm. Each likelihood expression requires four inputs: a detected image projection, a reference image projection, a set of atmospheric parameter estimates and an estimate of the current photon level, K. Solid lines indicate the flow of real time detected image projections. Dashed lines indicate information used for reference projections which are computed and stored into lookup tables during sensor calibration. The atmospheric parameter estimates and K do not need to be updated at every image frame, but should be updated as often as necessary to ensure some threshold of accuracy. The heavily outlined blocks in Figure 5.2 indicate locations where a likelihood expression is evaluated. Recall the general form for

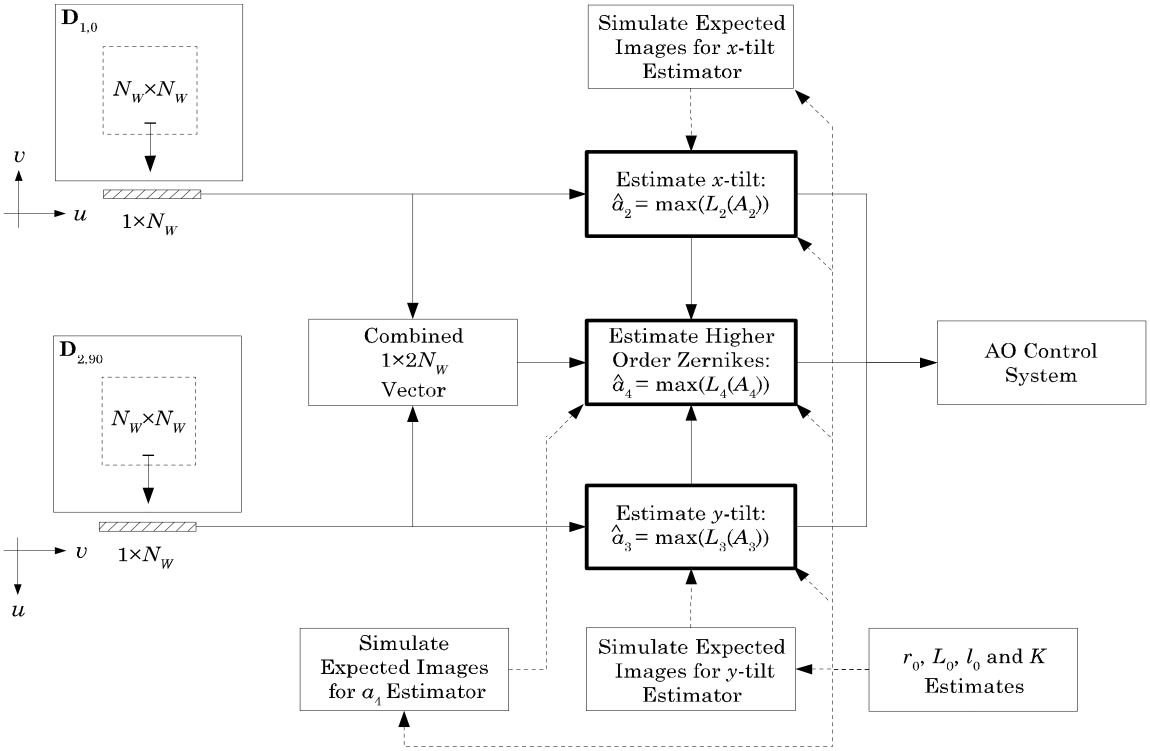


Figure 5.2 Diagram shows the flow of image projections through the Z_{2-4} estimation algorithm.

the MAP estimator:

$$\max_{\bar{\mathbf{A}}} \left\{ \left[\begin{array}{c} \sum_{l=1}^{N_{\mathrm{v}} N_W} \left[{}_{\mathbf{s}}\mathbf{v}_l \left(\mathbf{D}_{\cup} \right) + \sigma_{ro}^2 \right] \times \\ \ln \left\{ {}_{\mathbf{s}}\mathbf{v}_l \left(\mathbf{I}_{\cup} \left[\bar{\mathbf{A}} \right] \right) + \sigma_{ro}^2 \right\} - \\ {}_{\mathbf{s}}\mathbf{v}_l \left(\mathbf{I}_{\cup} \left[\bar{\mathbf{A}} \right] \right) \end{array} \right] - \frac{\bar{\mathbf{A}} \mathbf{\Lambda}_{\bar{\mathbf{a}}}^{-\mathbf{1}} \bar{\mathbf{A}}^t}{2} \right\} \Bigg|_{\bar{\mathbf{A}} = \hat{\mathbf{a}}_{map}} . \tag{5.7}$$

The likelihood expression, denoted $L_{map}(\bar{\mathbf{A}})$, is the expression to be maximized. Extracting the likelihood expression from the MAP estimator equation and tailoring the projection operator specifically for the Z_{2-4} estimator yields:

$$\begin{aligned} L_{map}(\bar{\mathbf{A}}) \quad &= \quad \sum_{l=1}^{N_{\mathrm{v}} N_W} \left[{}_{\{(1,N_W)\}}\mathbf{v}_l \left(\mathbf{D}_{1,0}, \mathbf{D}_{2,90} \right) + \sigma_{ro}^2 \right] \times \\ &\qquad \ln \left\{ {}_{\{(1,N_W)\}}\mathbf{v}_l \left(\mathbf{I}_{1,0} \left[\bar{\mathbf{A}} \right], \mathbf{I}_{2,90} \left[\bar{\mathbf{A}} \right] \right) + \sigma_{ro}^2 \right\} - \\ &\qquad {}_{\{(1,N_W)\}}\mathbf{v}_l \left(\mathbf{I}_{1,0} \left[\bar{\mathbf{A}} \right], \mathbf{I}_{2,90} \left[\bar{\mathbf{A}} \right] \right) - \frac{\bar{\mathbf{A}} \mathbf{\Lambda}_{\bar{\mathbf{a}}}^{-\mathbf{1}} \bar{\mathbf{A}}^t}{2}, \qquad (5.8) \\ \bar{\mathbf{A}} \quad &\equiv \quad \text{the infinite set of atmospheric parameters,} \qquad (5.9) \\ \mathbf{\Lambda}_{\bar{\mathbf{a}}} \quad &\equiv \quad \text{the parameter covariance matrix,} \qquad (5.10) \\ \sigma_{ro}^2 \quad &\equiv \quad \text{CCD read noise variance.} \qquad (5.11) \end{aligned}$$

The likelihood expression in (5.8) has infinite dimensions due to the infinite set of input parameters $\bar{\mathbf{a}}$. Since the estimator is only concerned with providing estimates for a_2, a_3, and a_4, these three dimensions are evaluated independently of all others. Furthermore, to reduce the complexity of the maximization process, the parameters of interest will be estimated independent of each other. Two characteristics of the problem allow maximizing over each parameter independently: the decrease in turbulence power between tilt and defocus, and the separability of Zernike effects in the chosen type of image projections. The random CCD images $\mathbf{d}$ will always be a function of the infinite set $\bar{\mathbf{a}}$, however, if a set of expected image vectors can be precomputed with known amounts of a single parameter then the likelihood can be maximized one dimension at a time. Taking advantage of the 20:1 ratio of tilt power to defocus, the tilt parameters will be estimated first. Tilt reference projections are formed by simulating the beacon image using an OTF containing only the reference tilt aberrations and a long exposure OTF containing contributions from Zernikes Z_4 and higher. The Zernike contributions in the long exposure OTF are based on the estimated atmospheric variables r_0, L_0 and l_0. Recall that the diffraction limited OTF may be formed from the simulated PSF:

$$\mathcal{H} = \frac{\mathcal{DFT}\left\{\mathbf{I}\right\}}{\sum\limits_{\mathbf{u}} \mathbf{I}[\mathbf{u}]}. \tag{5.12}$$

Note that the object is assumed to be a point source and therefore the expected image **I** is a diffraction limited PSF. Any aberrations in **I** will be indicated by including them as arguments of **I**. For instance, an image, which is otherwise diffraction limited, that contains a contribution of x-tilt is denoted:

$$\mathbf{I}[A_2]. \tag{5.13}$$

The corresponding OTF is denoted:

$$\mathcal{H}\left[A_2\right] = \frac{\mathcal{DFT}\left\{\mathbf{I}[A_2]\right\}}{\sum\limits_{\mathbf{u}} \mathbf{I}[\mathbf{u}; A_2]}. \tag{5.14}$$

Similarly, a long exposure OTF can be formed from the discrete Fourier transform of a long exposure PSF. Consider the PSF formed from an ensemble average of images:

$$\begin{aligned} {}_L\mathbf{I} &\equiv E_{\bar{\mathbf{a}}}\left\{\mathbf{i}\left[\bar{\mathbf{a}}\right]\right\}, & (5.15)\\ {}_L\mathcal{H} &= \frac{\mathcal{DFT}\left\{{}_L\mathbf{I}\right\}}{\sum\limits_{\mathbf{u}} {}_L\mathbf{I}[\mathbf{u}]}, & (5.16) \end{aligned}$$

where the expected image **i** occurs here in lower case to emphasize that it is a random quantity. Notice that the point spread function is given a preceding subscript L to distinguish it from a diffraction limited point spread function. Goodman provides an expression for the long exposure OTF in terms of the diffraction limited OTF and an OTF formed from the phase structure function, D_{P_ϕ} [24]:

$$\begin{aligned} {}_L\mathcal{H}\left[\mathbf{n}\right] &= \mathcal{H}_{P_\phi}\left[\mathbf{n}\right]\mathcal{H}\left[\mathbf{n}\right], & (5.17)\\ \mathcal{H}_{P_\phi}\left[\mathbf{n}\right] &= \exp\left(-\frac{1}{2}D_{P_\phi}\left[\mathbf{n}\right]\right). & (5.18) \end{aligned}$$

The phase structure function is evaluated on a spatial grid. The OTF is evaluated on a spatial frequency grid. A factor of λs_i is required to convert between the spatial frequency dimension of $\mathcal{H}_{P_\phi}$ and the spatial dimension of D_{P_ϕ}. The Nyquist relationship between the aperture and image plane sampling grids accounts for the difference in spatial versus spatial frequency dimensions in D_{P_ϕ} and $\mathcal{H}_{P_\phi}$ and, as such, accounts for the use of the same indexing variable $\mathbf{n}$ in both the left and right hand sides of the expression above. D_{P_ϕ} is therefore constructed from the ensemble averaged autocorrelation of the discretized pupil phase over all possible pixel shifts of the discrete pupil function. If the pupil phase does not contain contributions from specific atmospheric parameters then the resulting OTF will be referred to as the long dwell OTF with Z_i removed. For example, define the long exposure OTF for tilt removed turbulence as:

$$_{L23}\mathcal{H}\left[\mathbf{n}\right] = \exp\left(-\frac{1}{2}D_{P_\phi}\left[\mathbf{n}; A_2 = 0, A_3 = 0\right]\right). \tag{5.19}$$

The preceding subscript L is now followed by the numbers 2 and 3 to indicate that the OTF is a long exposure OTF with Zernikes 2 and 3 removed. It follows then that the tilt removed point spread function may be expressed as:

$$_{L23}\mathbf{I} = \mathcal{DFT}\left\{_{L23}\mathcal{H} \cdot \mathcal{H}\right\}, \tag{5.20}$$

where the binary operator $\cdot$ indicates the Hadamard product often referred to as an entrywise or pointwise product:

$$\left(\mathbf{A} \cdot \mathbf{B}\right)_{ij} = \mathbf{A}_{ij}\mathbf{B}_{ij}. \tag{5.21}$$

Furthermore, expected images with arbitrary tilt and long exposure variance contributions from all high order Zernike polynomials are formed by:

$$_{L23}\mathbf{I}\left[A_2, A_3\right] = \mathcal{DFT}\left\{_{L23}\mathcal{H} \cdot \mathcal{H}\left[A_2, A_3\right]\right\}. \tag{5.22}$$

$_{L23}\mathbf{I}$ is representative of the type of expected image used in the MAP estimator. Inserting the appropriate long term expected image projections into the likelihood expression (5.8), the tilt specific likelihood expressions are given by:

$$\begin{aligned} L_{map_2}(A_2) &= \sum_{l=1}^{N_W} \left[{}_{\{(1,N_W)\}}\mathbf{v}_l \left(\mathbf{D}_{1,0} \right) + \sigma_{ro}^2 \right] \times \\ & \ln\left\{ {}_{\{(1,N_W)\}}\mathbf{v}_l \left({}_{L23}\mathbf{I}_{1,0}\left[A_2\right] \right) + \sigma_{ro}^2 \right\} - \\ & {}_{\{(1,N_W)\}}\mathbf{v}_l \left({}_{L23}\mathbf{I}_{1,0}\left[A_2\right] \right) - \frac{A_2^2}{2\sigma_2^2}, \end{aligned} \tag{5.23}$$

$$\begin{aligned} L_{map_3}(A_3) &= \sum_{l=1}^{N_W} \left[{}_{\{(1,N_W)\}}\mathbf{v}_l \left(\mathbf{D}_{2,90} \right) + \sigma_{ro}^2 \right] \times \\ & \ln\left\{ {}_{\{(1,N_W)\}}\mathbf{v}_l \left({}_{L23}\mathbf{I}_{2,90}\left[A_3\right] \right) + \sigma_{ro}^2 \right\} - \\ & {}_{\{(1,N_W)\}}\mathbf{v}_l \left({}_{L23}\mathbf{I}_{2,90}\left[A_3\right] \right) - \frac{A_3^2}{2\sigma_3^2}. \end{aligned} \tag{5.24}$$

Similar to the method for creating tilt reference projections, Z_4 reference images are formed from a combination of a known defocus OTF with a long exposure OTF containing appropriate contributions from Z_5 and higher:

$$\begin{aligned} {}_{L234}\mathbf{I}\left[\mathbf{u}\right] &= E_{\bar{\mathbf{a}}}\left\{ \mathbf{i}\left[\mathbf{u}; \bar{\mathbf{a}} | A_2 = 0, A_3 = 0, A_4 = 0\right] \right\}, & (5.25) \\ {}_{L234}\mathcal{H}\left[\mathbf{n}\right] &= \exp\left(-\frac{1}{2} D_{P_\phi}\left[\mathbf{n}; A_2 = 0, A_3 = 0, A_4 = 0\right] \right), & (5.26) \\ {}_{L234}\mathbf{I}\left[A_4\right] &= \mathcal{DFT}\left\{ {}_{L234}\mathcal{H} \cdot \mathcal{H}\left[A_4\right] \right\}. & (5.27) \end{aligned}$$

The set of Z_4 expected image projections must be preregistered over an array of known tilt values. The estimator will select the preregistered Z_4 projection with the closest matching pair of tilt values:

$$\left(\tilde{A}_2, \tilde{A}_3 \right) = \left(\text{round}\left(\frac{\hat{A}_2}{\Delta A_2} \right) \Delta A_2, \text{round}\left(\frac{\hat{A}_3}{\Delta A_3} \right) \Delta A_3 \right), \tag{5.28}$$

where $\left(\hat{A}_2, \hat{A}_3 \right)$ are formed during tilt estimation by choosing the parameter which maximizes L_{map_2} and L_{map_3} respectively. Note that the function round$(\cdot)$ is a call to the Matlab® rounding function which outputs the nearest integer to the argument. Preregistered Z_4 expected images are given by:

$$
{}_{L234}\mathbf{I}\left[\tilde{A}_2, \tilde{A}_3, A_4\right] = \mathcal{DFT}\left\{{}_{L234}\mathcal{H} \cdot \mathcal{H}\left[A_4, \tilde{A}_2, \tilde{A}_3\right]\right\}. \tag{5.29}
$$

The Z_4 specific likelihood expression is given:

$$
\begin{aligned}
L_{map_4}(A_4) \quad = \quad & \sum_{l=1}^{2N_W} \left[{}_{\{(1,N_W)\}}\mathbf{v}_l \left(\mathbf{D}_{1,0}, \mathbf{D}_{2,90}\right) + \sigma_{ro}^2\right] \times \\
& \ln\left\{{}_{\{(1,N_W)\}}\mathbf{v}_l \left({}_{L234}\mathbf{I}_{1,0}\left[\tilde{A}_2, \tilde{A}_3, A_4\right], {}_{L234}\,\mathbf{I}_{2,90}\left[\tilde{A}_2, \tilde{A}_3, A_4\right]\right) + \sigma_{ro}^2\right\} - \\
& {}_{\{(1,N_W)\}}\mathbf{v}_l \left({}_{L234}\mathbf{I}_{1,0}\left[\tilde{A}_2, \tilde{A}_3, A_4\right], {}_{L234}\,\mathbf{I}_{2,90}\left[\tilde{A}_2, \tilde{A}_3, A_4\right]\right) - \frac{A_4^2}{2\sigma_4^2}.
\end{aligned} \tag{5.30}
$$

5.4 Maximizing the Likelihood Expression

When all the inputs required for each likelihood expression are available, the estimator requires a fast way of evaluating and maximizing the function. Evaluating the likelihood is made more efficient by precomputing and storing banks of expected image projections. However, locating the likelihood maximum can be computationally expensive and, as such, should be accomplished using as few evaluations of the likelihood as possible. The maximization approach and the maximum number of "guesses" used in any gradient search algorithm will be constrained by the operating bandwidth of the wavefront sensing system and its ability to address stored arrays of reference vectors. There are many ways to configure this portion of the estimation algorithm. Here I will offer one possible method of maximization and the rationale behind it.

I will begin by describing the estimator lookup tables. The algorithm starts by generating both tilt estimates independently using a bank of $\mathbf{v}\left({}_{L23}\mathbf{I}\right)$ projections spanning $\pm 4\sigma_{2,3}$ and separated in $A_{2,3}$ by $0.25\sigma_{2,3}$. The pair of tilt estimates are then passed to the defocus estimator which uses a bank of $\mathbf{v}({}_{L4}\mathbf{I})$ projections spanning $\pm 4\sigma_4$ and separated by $0.8\sigma_4$. Recall that each defocus projection must be preregistered over an expected range of tilt values. The performance of the defocus estimator is not significantly affected by tilt estimates accurate within $\pm 0.2\sigma_{2,3}$, therefore the tilt preregistration grid is bounded by $\pm 5\sigma_{2,3}$ with a step size of $0.25\sigma_{2,3}$. These lookup table bounds and step sizes can be summarized as follows:

$$\begin{aligned} \Delta A_{2,3} &= 0.5\sigma_{2,3}, & (5.31) \\ |A_{2,3}|_{\max} &= \pm 4\sigma_{2,3}, & (5.32) \\ \Delta A_4 &= 0.8\sigma_4, & (5.33) \\ |A_4|_{\max} &= \pm 4\sigma_4, & (5.34) \\ \Delta \tilde{A}_{2,3} &= 0.25\sigma_{2,3}, & (5.35) \\ \left|\tilde{A}_{2,3}\right|_{\max} &= \pm 5\sigma_{2,3}, & (5.36) \end{aligned}$$

It is important to note that given these lookup table step sizes and bounds, the maximum number of likelihood evaluations per parameter estimate is 10 for $\hat{A}_2$ and $\hat{A}_3$ and 8 for $\hat{A}_4$. These numbers are based on a search beginning with the three evaluation points about zero and proceeding with a fixed step gradient search.

The parameter estimates are formed using a quadratic curve fit through the three highest points among the available steps. Investigating the nature of the A_2 and A_3 likelihood expressions, reveals that they are very well behaved for the point source case. In fact, given small enough lookup step sizes, the likelihood expression will be nearly quadratic through the three highest points. As an example, consider that the 17 evaluations of the likelihood yield the 3 peak points circled in red in Figure 5.3. In general, the quadratic fit maximum through 3 or more points is given by:

$$\begin{aligned} \mathbf{Xc} &= \mathbf{y}, & (5.37) \\ \mathbf{c} &= \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{X}^t\mathbf{y}, & (5.38) \\ x_{\max} &= -\frac{c_1}{2c_2}. & (5.39) \end{aligned}$$

However, the maximum, assuming exactly 3 points and a fixed step size, Δx, simplifies to:

$$x_{\max} = \frac{(y_1 - y_2)\,\Delta x}{y_1 + y_3 - 2y_2} + x_1 + \frac{\Delta x}{2}. \tag{5.40}$$

This quadratic curve fit is used to estimate Zernike 4. Figure 5.4 demonstrates the quadratic curve fit through an example Z_4 likelihood. The quadratic fit *requires* only 3

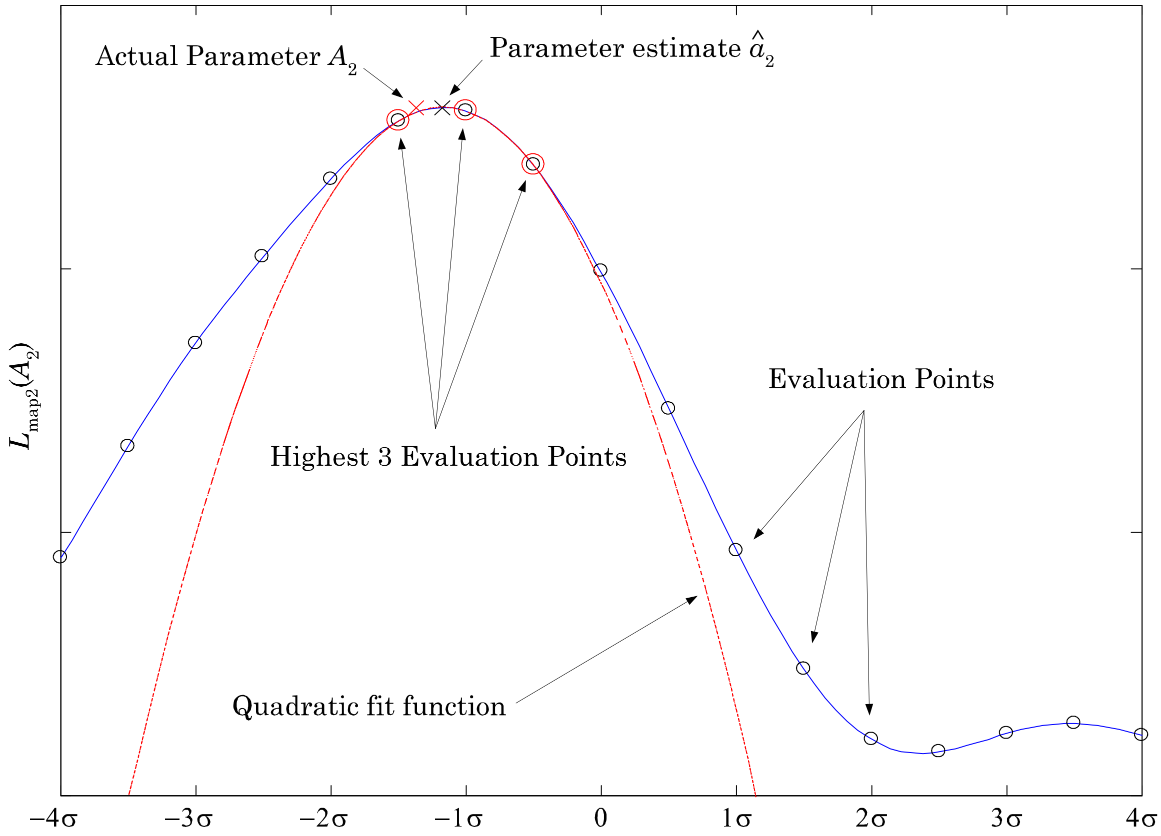


Figure 5.3 Figure provides an example of the evaluation points and the quadratic curve fit used to form each tilt estimate.

points over the parameter range. Judiciously choosing 3 realizations, for instance: $A_x \in \{-3\sigma_x, 0\sigma_x, 3\sigma_x\}$, will produce a faster yet less accurate estimator. Increasing the distance between sample points increases the error in the quadratic fit and increases susceptibility to errors in the r_0 estimate. These types of trade-off considerations force the choice of lookup table design parameters to be specific to each application.

5.5 Sensor Design Variables

The previous paragraphs outlined the general curvature sensor design and suggested some choices for design variables. Next I will discuss the key design variables and how each variable effects sensor performance. Table 5.1 lists the key variables that effect the curvature sensor's performance.

Choice of subaperture diameter will be application specific. In section 2.3, defining the parameter space, I derived a very important fact concerning the Zernike modes present

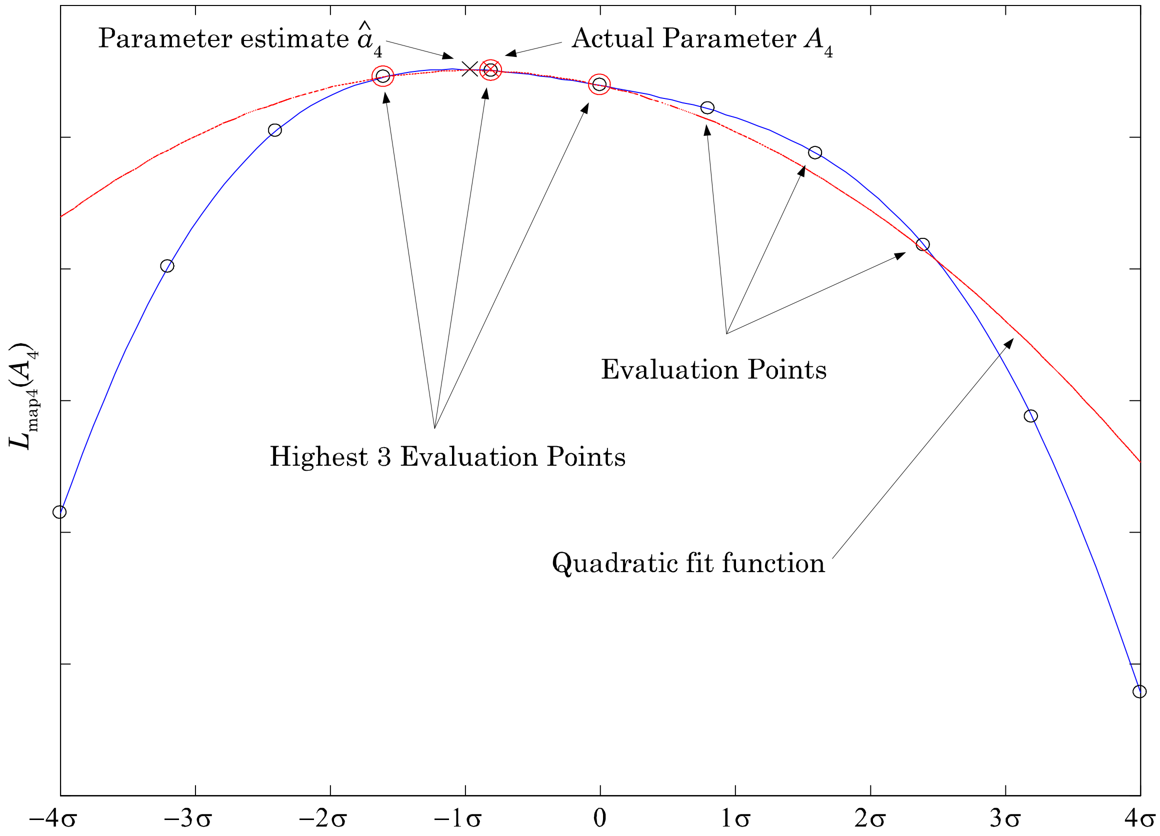


Figure 5.4 Figure provides an example of the evaluation points and the quadratic curve fit used to form each defocus estimate.

in atmospheric turbulence: the expected power in each Zernike mode is based on the ratio of the size of the aperture, D_P, to the characteristic turbulence parameter, r_0. The inner and outer scale parameters: l_0 and L_0 also affect performance, but to a lesser degree. Choice of D_P should therefore be based on the range of atmospheric conditions in which the sensor will operate nominally. The largest ratio $\frac{D_P}{r_0}$ in which the sensor operates will produce the worst case performance. On the other hand, ratios of $\frac{D_P}{r_0} < 1$ place the curvature sensor in an operating environment where the significance of Z_4 contributions in the wavefront are minimal. This condition reduces the sensor's performance beyond that of a tilt only sensor.

As the defocus diversity increases, the sensor's ability to estimate defocus increases. Unfortunately, the opposite is true of the sensor's ability to estimate tilt. If the operating variables are well known, an ideal diversity factor can be selected to provide the proper trade-off between accurate tilt estimates and defocus estimates.

Variable	**Description**
D_P	Aperture diameter
$\pm\delta_{a_4}$	Defocus diversity
θ_i	CCD rotation angles
${}_{\mathbf{s}}\mathbf{v}(\cdot)$	Image projection operation
$\Delta A_i, \lvert A_i\rvert_{\max}$ $\Delta \tilde{A}_{2,3}, \lvert \tilde{A}_{2,3}\rvert_{\max}$	Step size and range of lookup tables

Table 5.1 Curvature sensor design parameters.

The CCD angle θ_i determines the relative angle between the AO reference frame v axis and the image projection direction. It will be demonstrated using a performance bounding measure and via simulated performance that a separation angle, $\theta_1 - \theta_2 = 90$ degrees, is optimal for the Z_{2-4} sensor. Using the same measures, it will also be demonstrated that the performance of the Z_{2-4} sensor is invariant to a change in θ_1 provided that the separation angle is 90 degrees. For this reason, the Z_{2-4} sensor is set up with $\theta_1 = 0^\circ$ and $\theta_2 = 90^\circ$.

The projection operation ${}_{\mathbf{s}}\mathbf{v}(\cdot)$ in the Z_{2-4} estimator accounts for the size of windowing function and the number of vector projections. Consequently, ${}_{\mathbf{s}}\mathbf{v}(\cdot)$ determines the number of pixels read out of each CCD array. As more pixels are read out of the array, there is an increase in information available to the estimator. However, there is also a proportional increase in the amount of read out noise and computational requirements. A design trade-off must be made between information, read out noise, and computational complexity. The increased computational complexity comes from an increase in the number of vector points included in the likelihood equations.

Ideally, the expected projection lookup tables would provide an entry for every possible estimate. This is computationally prohibitive. Instead the estimator uses tables with some finite step size and range. Decreasing the granularity, ΔA_i, and increasing the range, $|A_i|_{\max}$, of the vector lookup tables improves the sensor's performance at the cost of decreased speed and increased memory requirements. The quadratic fit may include more points or be dispensed with entirely for some other approach. Creativity in the design of the maximization routine must balance the speed, accuracy and robustness of the sensor.

5.6 Summary

This chapter outlined the design of a MAP estimator based curvature sensor to include its general hardware requirements and the flow of the software algorithm. The sensor is designed to estimate 3 parameters: x-tilt, y-tilt, and defocus from point source image projections. Key hardware considerations include the use of a beamsplitter to share the incoming optical signal equally between two programmable CCD arrays and applying a defocus diversity in each optical path. The software algorithm forms a MAP likelihood incorporating estimates of the atmospheric conditions: r_0, L_0, and l_0, average photon count, and precomputed image projections. With a maximum of 28 likelihood evaluations, the curvature sensor is capable of estimating Zernikes $2-4$. The next chapter provides performance bounding for the curvature sensor and demonstrates how the performance bound can be used to select ideal design variable settings. Chapter 8 provides simulated performance results. In simulated cases, the curvature sensor is capable of providing improved performance over that of a centroiding tilt sensor and a projection based ML tilt sensor.

6. Wavefront Sensor Performance Bound

The mean squared difference between the compensated wavefront and the desired wavefront is a common performance measure for a wavefront sensor. This chapter provides an analysis of the wavefront residual mean squared error based on the maximum a posteriori estimator described in Chapter 5. This type of performance measure is best suited for simulation because it requires a direct measure of the field input to the optical system which is available only in simulation. Given a known input, it is straightforward to calculate the error in the wavefront sensor response and provide statistics on that error, particularly the bias and the variance. For deeper insight into performance limits, estimation theory provides methods for bounding the error variance. The following sections establish an expression for the residual wavefront mean squared error (MSE) and the Cramér Rao lower bound (CRLB) for estimator variance. These measures will be used to compare the performance of the MAP estimator to existing estimators under various operating conditions. The CRLB will also be useful for determining the ideal design choices for the MAP estimator for a given operating environment.

6.1 Wavefront Mean Squared Error (MSE)

The wavefront MSE must include both error due to the estimator's imperfect response and the error from additional parameters which are not estimated. Recall that a volume of turbulent atmospheric effects can be integrated along the direction of the optical path to form a thin phase screen. The resulting phase screen can be modeled by an infinite series of Zernike polynomials with coefficients, $\bar{\mathbf{a}}$. The infinite set of Zernike coefficients can be divided into a finite set of parameters to be estimated, denoted by S, and an infinite number of higher order coefficients:

$$\bar{\mathbf{a}} \equiv \text{the infinite set of Zernike coefficients,} \tag{6.1}$$

$$\mathbf{a} \equiv \{a_i : i \in S\}, \text{ the estimated set of Zernike coefficients,} \tag{6.2}$$

$$\underline{\mathbf{a}} \equiv \{a_i : i \notin S\}, \text{ Zernike coefficients unknown to the estimator,} \tag{6.3}$$

$$\bar{\mathbf{a}} = \mathbf{a} \cup \underline{\mathbf{a}}. \tag{6.4}$$

While the wavefront sensor attempts to estimate a small set of parameters, $\mathbf{a}$, it will be shown that the remaining parameters $\underline{\mathbf{a}}$ and the noise characteristics of the CCD increase the estimator mean squared error. In the paragraphs to follow, an expression is derived for the residual wavefront error based on the selected set of parameters, $\mathbf{a}$.

Begin the derivation by assuming that the ideal wavefront is a unit amplitude, constant phase, plane wave. Additionally, assume that the optical system is only affected by the piston removed wavefront phase, P_ϕ, so the coefficient a_1 is ignored. Recalling the aperture convention and the relationship between aperture phase and the coefficients $\mathbf{a}$, the general form for the field in the aperture expressed in continuous polar coordinates is given:

$$\mathcal{P}\left(\mathbf{r};\bar{\mathbf{a}},R_P\right) = W_P\left(\mathbf{r};R_P\right)\exp\left\{\mathrm{j}P_\phi\left(\mathbf{r};\bar{\mathbf{a}},R_P\right)\right\}, \tag{6.5}$$

$$\begin{aligned} \text{where } R_P &\equiv \text{aperture radius}, && (6.6)\\ \mathbf{r} &= (r,\theta), && (6.7)\\ 0 &\leq r<\infty, && (6.8)\\ 0 &\leq \theta<2\pi, && (6.9) \end{aligned}$$

$$\text{and } W_P(\mathbf{r};R_P) \equiv \begin{cases} 1, & r\leq R_P\\ 0, & r>R_P \end{cases}. \tag{6.10}$$

Extracting the the piston removed phase expression and expanding it as a series of Zernike polynomials yields:

$$\begin{aligned} P_\phi(\mathbf{r};\bar{\mathbf{a}},R_P) &= W_P(\mathbf{r};R_P)\sum_{i=2}^{\infty} a_i Z_i\left(\mathbf{r};R_P\right), && (6.11)\\ \text{where } Z_i\left(\mathbf{r};R_P\right) &\equiv Z_i\left(\frac{r}{R_P},\theta\right). && (6.12) \end{aligned}$$

Section 2.3 demonstrated that the coefficients, a_i, can be found by projecting each Zernike onto the wavefront phase as follows:

$$a_i \equiv \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho})Z_i(\boldsymbol{\rho})P_\phi\left(\mathbf{r};\bar{\mathbf{a}},R_P\right), \tag{6.13}$$

where $\boldsymbol{\rho}$ represents the scaled polar coordinates:

$$\begin{aligned} \boldsymbol{\rho} &= \left(\frac{r}{R_P}, \theta\right), \\ \text{and } W_Z(\boldsymbol{\rho}) &\equiv \begin{cases} \frac{1}{\pi}, & \frac{r}{R_P} \leq 1 \\ 0, & \frac{r}{R_P} > 1 \end{cases}. \end{aligned}$$

Using this convention, the mean squared error in the compensated wavefront is given by:

$$\left\langle P_{\phi_e}^2 \right\rangle = E_{\bar{\mathbf{a}}} \left\{ \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) \left[P_\phi\left(\mathbf{r}; \bar{\mathbf{a}}, R_P\right) - \sum_{i \in S} \hat{a}_i Z_i(\boldsymbol{\rho}) \right]^2 \right\}, \tag{6.14}$$

where the hatted coefficient variable, $\hat{a}_i$, denotes an estimate of the respective random coefficient a_i. Expanding P_ϕ into a sum of parameters and separating the set of estimated parameters from the higher order parameters gives:

$$\begin{aligned} \left\langle P_{\phi_e}^2 \right\rangle &= E_{\bar{\mathbf{a}}} \left\{ \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) \left[\sum_{i \in S} a_i Z_i(\boldsymbol{\rho}) - \sum_{i \in S} \hat{a}_i Z_i(\boldsymbol{\rho}) + \sum_{i \notin S} a_i Z_i(\boldsymbol{\rho}) \right]^2 \right\}, \quad (6.15) \\ &= E_{\bar{\mathbf{a}}} \left\{ \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) \left[\sum_{i \in S} \left(a_i - \hat{a}_i\right) Z_i(\boldsymbol{\rho}) + \sum_{i \notin S} a_i Z_i(\boldsymbol{\rho}) \right]^2 \right\}. \quad (6.16) \end{aligned}$$

Taking advantage of the fact that the Zernike basis functions are orthonormal, it is possible to expand the square and collect the nonzero terms:

$$\left\langle P_{\phi_e}^2 \right\rangle = E_{\bar{\mathbf{a}}} \left\{ \sum_{i \in S} (a_i - \hat{a}_i)^2 \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i^2(\boldsymbol{\rho}) + \sum_{i \notin S} a_i^2 \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i^2(\boldsymbol{\rho}) \right\}. \tag{6.17}$$

The integral factor $\int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i^2(\boldsymbol{\rho}) = 1$ for all i:

$$\left\langle P_{\phi_e}^2 \right\rangle = E_{\bar{\mathbf{a}}} \left\{ \sum_{i \in S} (a_i - \hat{a}_i)^2 + \sum_{i \notin S} a_i^2 \right\}. \tag{6.18}$$

MSE is in units of rad^2, with the caveat that this measure is intimately tied to the pupil area. Exchange the order of summation and expectation:

$$\left\langle P_{\phi_e}^2 \right\rangle = \sum_{i \in S} E_{a_i} \left\{ (a_i - \hat{a}_i)^2 \right\} + \sum_{i \notin S} E_{a_i} \left\{ a_i^2 \right\}. \tag{6.19}$$

This result shows that the overall wavefront MSE is composed of two summation terms. The first sum represents the estimator MSE. Assuming that the estimator is unbiased, this term represents the estimator variance. The second term represents the total variance of all remaining parameters in the atmospheric model which will be denoted $\left\langle P^2_{\phi_{uncorr}} \right\rangle$:

$$\left\langle P^2_{\phi_{uncorr}} \right\rangle = \sum_{i \notin S} E\left\{a_i^2\right\}. \tag{6.20}$$

Recall the expression for the covariance of Zernike coefficients derived from the von Kármán atmospheric model in (2.138). If $i = i'$ and $m = m'$ the expression simplifies to the variance of the $a_i^{'}$s:

$$\begin{aligned} E\left\{a_i^2\right\} &= 0.4898 \cdot 2^{4/3}\pi \left(\frac{2R_P}{r_0}\right)^{5/3} (n(i)+1)(-1)^{n(i)-m(i)} \times \\ &\quad \int d\kappa \frac{J^2_{n(i)+1}(\kappa)}{\kappa(\kappa^2 + R_P^2\kappa_0^2)^{11/6}} \exp\left(\frac{-\kappa^2}{R_P^2\kappa_m^2}\right). \end{aligned} \tag{6.21}$$

Using this expression, it is possible to approximate the lower bound on MSE for the case of perfect compensation of parameters in the set S by numerically evaluating the integral:

$$\begin{aligned} \left\langle P^2_{\phi_{uncorr}} \right\rangle &= \sum_{i \notin S} 0.4898 \cdot 2^{4/3}\pi \left(\frac{2R_P}{r_0}\right)^{5/3} (n(i)+1)(-1)^{n(i)-m(i)} \times \\ &\quad \int d\kappa \frac{J^2_{n(i)+1}(\kappa)}{\kappa(\kappa^2 + R_P^2\kappa_0^2)^{11/6}} \exp\left(\frac{-\kappa^2}{R_P^2\kappa_m^2}\right). \end{aligned} \tag{6.22}$$

As a reminder, the functions $n(i)$ and $m(i)$ were provided in Table 2.3. This expression relates residual MSE to the atmospheric parameters and the aperture size. Figure 6.1 contains a plot of the residual MSE as it relates to the ratio $\frac{D_P}{r_0}$ for several sets of estimated parameters S. In this case, D_P was fixed at 0.07m while r_0 varied over the range: {0.02m...0.2m}. The plot demonstrates that for a given sensor design, residual MSE will vary as atmospheric conditions change. The plot also reveals that the advantage of estimating additional parameters decreases for higher Zernike modes. Unfortunately, the wavefront sensor will not provide a perfect set of estimates. The vector based curvature sensor must contend with compressed image information and CCD noise. Additionally, the defocus diversity required for higher order modal compensation increases error variance in the tilt estimates. All of these factors will cause an increase in the residual wavefront error

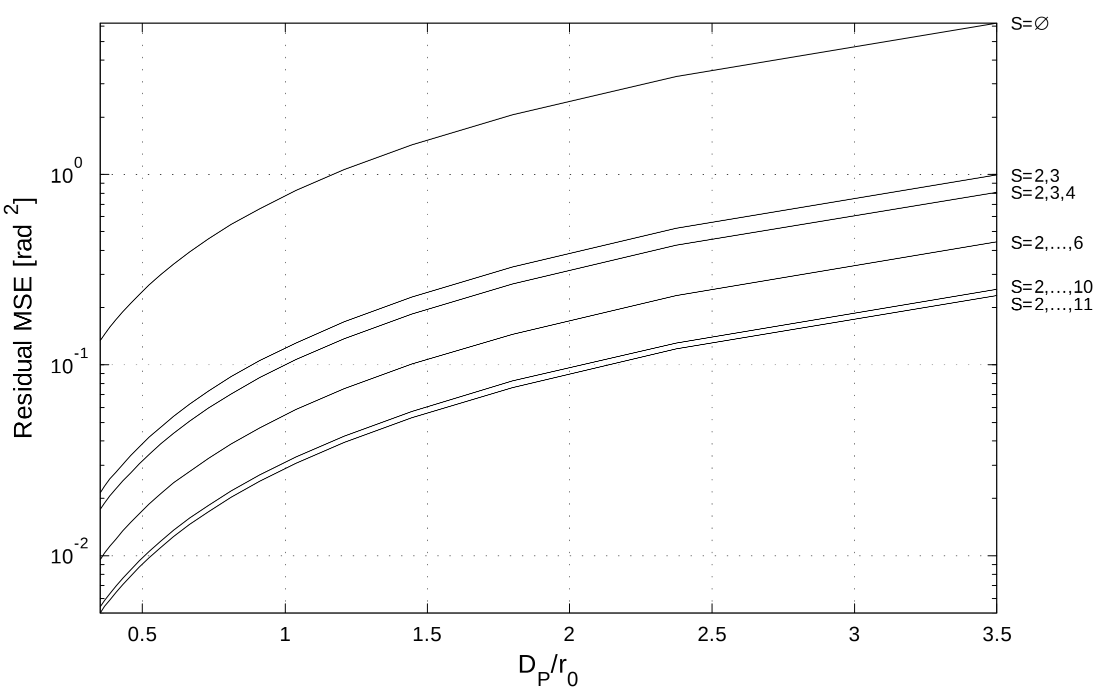


Figure 6.1 Plots of $\left\langle P^2_{\phi_{uncorr}} \right\rangle$ vs. $\frac{D_P}{r_0}$ for several parameter sets S.

by the sum of the error variances in the estimated parameter set:

$$\left\langle P^2_{\phi_e} \right\rangle = \sum_{i \in S} E\left\{(a_i - \hat{a}_i)^2\right\} + \left\langle P^2_{\phi_{uncorr}} \right\rangle . \tag{6.23}$$

Defining how sensor design variables affect the variance term, and consequently how they affect $\left\langle P^2_{\phi_e} \right\rangle$, offers a tool for adjusting the sensor to provide minimum residual error under a given set of atmospheric conditions. The Cramér Rao lower bound is one method for characterizing each error variance term. The next section will derive the CRLB and relate the CRLB to key sensor design variables.

6.2 The Cramér Rao Lower Bound

The following section attempts to bound the performance of the vector based wavefront sensor based on the limits set by the Cramér Rao lower bound. The estimator must be unbiased in order to apply the CRLB. For this reason, assume that the estimator is unbiased or that the bias is not a function of the parameter or system variables and can be removed. Van Trees provides the expression for the CRLB for unbiased estimators of random parameters [8]:

$$E\left\{\left(a_i - \hat{a}_i\left(\mathbf{R}\right)\right)^2\right\} \geq J_T^{ii}. \tag{6.24}$$

Where J_T^{ii} is the iith element of $\mathbf{J}_T^{-1}$. $\mathbf{J}_T$ is the $K \times K$ square matrix formed from $\mathbf{J}_D$, commonly referred to as Fisher's information matrix, and $\mathbf{J}_P$, the matrix of a priori information:

$$\mathbf{J}_T \equiv \mathbf{J}_D + \mathbf{J}_P. \tag{6.25}$$

The ijth element of $\mathbf{J}_D$ is defined [8]:

$$J_{D_{ij}} \equiv -E\left\{\frac{\partial^2 \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A})}{\partial A_i \partial A_j}\right\}. \tag{6.26}$$

The ijth element of the a priori information matrix is defined [8]:

$$J_{P_{ij}} \equiv -E\left\{\frac{\partial^2 \ln p_{\mathbf{a}}(\mathbf{A})}{\partial A_i \partial A_j}\right\}. \tag{6.27}$$

Combining these results, it is easy to recognize that the form of $\mathbf{J}_T$ contains the same internal expression from the previously defined MAP estimator (2.17). The MAP estimator was given as:

$$\max_{A} \left\{\ln p_{\mathbf{r}|a}(\mathbf{R}|A) + \ln p_a(A)\right\}\Big|_{A=\hat{a}_{map}}, \tag{6.28}$$

while the form for $\mathbf{J}_T$ is:

$$\mathbf{J}_T = -E\left\{\frac{\partial^2\left[\ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A}) + \ln p_{\mathbf{a}}(\mathbf{A})\right]}{\partial A_i \partial A_j}\right\}. \tag{6.29}$$

Aside from the vector versus single parameter notation, the function to be maximized for $\hat{a}_{map}$ is the same expression to be differentiated in $\mathbf{J}_T$. This expression is sometimes referred to as the log likelihood expression. The CRLB is in essence a measure of the average curvature or second derivative of the log likelihood expression. Due to its significance and continued recurrence throughout the remainder of this dissertation, I will define the MAP and ML log likelihood expressions:

$$L_{map}(\mathbf{A}) = \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A}) + \ln p_{\mathbf{a}}(\mathbf{A}), \tag{6.30}$$

$$L_{ml}(\mathbf{A}) = \ln p_{\mathbf{r}|\mathbf{a}}(\mathbf{R}|\mathbf{A}). \tag{6.31}$$

In the CCD detector projection model, the observed process $\mathbf{r}$ is replaced by one or more compressed arrays of pixels ${}_{\mathbf{s}}\mathbf{v}_l\ (\mathbf{d}_{\cup})$. Thus the probability density, $p_{\mathbf{r}}(\mathbf{R})$, in van Trees' notation will be replaced by, $p_{{}_{\mathbf{s}}\mathbf{v}_l\,(\mathbf{d}_{\Sigma})}({}_{\mathbf{s}}\mathbf{v}_l\ (\mathbf{D}_{\cup}))$, the pdf for a detected image projection. The final form for the projection based MAP estimator was derived from the assumption that the distribution of $\mathbf{d}_{\cup}$ is a combination of signal dependent Poisson shot noise and Gaussian read noise. The combination of Poisson shot noise and Gaussian read noise was approximated by a biased Poisson process, $\mathbf{d}$ =Poisson$\left\{\mathbf{I}+\sigma_{ro}^2\right\}-\sigma_{ro}^2$ [23]. The prior density $p_{\mathbf{a}}(\mathbf{A})$ is jointly Gaussian. Recall the MAP estimator expression:

$$\max_{\bar{\mathbf{A}}}\left\{\left[\begin{array}{c}\sum_{l=1}^{N_{\rm v}N_W}\left[{}_{\mathbf{s}}\mathbf{v}_l\ (\mathbf{D}_{\cup})+\sigma_{ro}^2\right]\times\\ \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\ \left(\mathbf{I}_{\cup}\left[\bar{\mathbf{A}}\right]\right)+\sigma_{ro}^2\right\}-\\ {}_{\mathbf{s}}\mathbf{v}_l\ \left(\mathbf{I}_{\cup}\left[\bar{\mathbf{A}}\right]\right)\end{array}\right]-\frac{\bar{\mathbf{A}}\boldsymbol{\Lambda}_{\bar{\mathbf{a}}}^{-\mathbf{1}}\bar{\mathbf{A}}^t}{2}\right\}\Bigg|_{\bar{\mathbf{A}}=\hat{\mathbf{a}}_{map}} \tag{6.32}$$

From the MAP estimator, both the MAP and ML log likelihood expressions can be extracted for use in the CRLB calculations to follow:

$$L_{map}(\bar{\mathbf{A}}) = \left[\begin{array}{c}\sum_{l=1}^{N_{\rm v}N_W}\left[{}_{\mathbf{s}}\mathbf{v}_l\ (\mathbf{D}_{\cup})+\sigma_{ro}^2\right]\times\\ \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\ \left(\mathbf{I}_{\cup}\left[\bar{\mathbf{A}}\right]\right)+\sigma_{ro}^2\right\}-{}_{\mathbf{s}}\mathbf{v}_l\ \left(\mathbf{I}_{\cup}\left[\bar{\mathbf{A}}\right]\right)\end{array}\right]-\frac{\bar{\mathbf{A}}\boldsymbol{\Lambda}_{\bar{\mathbf{a}}}^{-\mathbf{1}}\bar{\mathbf{A}}^t}{2}, \tag{6.33}$$

$$L_{ml}(\bar{\mathbf{A}}) = \sum_{l=1}^{N_{\rm v}N_W}\left[{}_{\mathbf{s}}\mathbf{v}_l\ (\mathbf{D}_{\cup})+\sigma_{ro}^2\right]\ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\ \left(\mathbf{I}_{\cup}\left[\bar{\mathbf{A}}\right]\right)+\sigma_{ro}^2\right\}-{}_{\mathbf{s}}\mathbf{v}_l\ \left(\mathbf{I}_{\cup}\left[\bar{\mathbf{A}}\right]\right). \tag{6.34}$$

The infinite parameter vector, $\bar{\mathbf{A}}$, must be reduced to some limited parameter set, $\mathbf{A}$. When evaluating the lower bound, the intent is to model the operating environment as accurately as possible. This requires including as many Zernike modes as possible. Unfortunately, as the number of parameters increases, call that number N, the complexity of the CRLB calculation increases as N^2. As such, the set of parameters must be truncated at a point where error and computation time are both acceptable. Under this approximation, the infinite parameter vector $\bar{\mathbf{A}}$ becomes $\mathbf{A}$. The remainder of the expressions in this section will substitute $\mathbf{A}$ for $\bar{\mathbf{A}}$ in all instances of L_{map} and L_{ml} above. Substituting (6.33) into the expression for $\mathbf{J}_T$ in (6.29) yields:

$$\mathbf{J}_T=-E\left\{\frac{\partial^2}{\partial A_i\partial A_j}\left(\left[\begin{array}{c}\sum_{l=1}^{N_{\rm v}N_W}\left[{}_{\mathbf{s}}\mathbf{v}_l\ (\mathbf{D}_{\cup})+\sigma_{ro}^2\right]\times\\ \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\ (\mathbf{I}_{\cup}\left[\mathbf{A}\right])+\sigma_{ro}^2\right\}-{}_{\mathbf{s}}\mathbf{v}_l\ (\mathbf{I}_{\cup}\left[\mathbf{A}\right])\end{array}\right]-\frac{\mathbf{A}\boldsymbol{\Lambda}_{\mathbf{a}}^{-\mathbf{1}}\mathbf{A}^t}{2}\right)\right\}, \tag{6.35}$$

Evaluating the second partial derivative of the prior gives:

$$\begin{aligned}\mathbf{J}_T &= -E\left\{\frac{\partial^2}{\partial A_i \partial A_j}\left(\begin{array}{c}\sum_{l=1}^{N_{\mathrm{v}}N_W}\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)+\sigma_{ro}^2\right]\times \\ \ln\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)+\sigma_{ro}^2\right\}-{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)\end{array}\right)\right\}+\mathbf{\Lambda}_{\mathbf{a}}^{-1}, \qquad (6.36)\\ &= \mathbf{J}_D+\mathbf{\Lambda}_{\mathbf{a}}^{-1}. \qquad (6.37)\end{aligned}$$

In (6.36), the Fisher information matrix, $\mathbf{J}_D$, contains an expected value operator and partial derivatives. A simpler expression for JD without partials or expectation integrals is desired. Begin by evaluating the two partial derivatives of the conditional density portion of the log likelihood function, $L_{ml}(\mathbf{A})$ in (6.34). Evaluating the first partial derivative yields:

$$\begin{aligned}\frac{\partial}{\partial A_i}L_{ml}(\mathbf{A}) &= \sum_{l=1}^{N_{\mathrm{v}}N_W}\frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)+\sigma_{ro}^2}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)+\sigma_{ro}^2}\frac{\partial}{\partial A_i}\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)\right\}- \\ &\quad \frac{\partial}{\partial A_i}\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)\right\}, \qquad (6.38)\\ &= \sum_{l=1}^{N_{\mathrm{v}}N_W}\frac{\partial}{\partial A_i}\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)\right\}\left[\frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)+\sigma_{ro}^2}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)+\sigma_{ro}^2}-1\right]. \qquad (6.39)\end{aligned}$$

Continuing to evaluate the second partial derivative gives:

$$\begin{aligned}\frac{\partial^2 L_{ml}(\mathbf{A})}{\partial A_i \partial A_j} &= \sum_{l=1}^{N_{\mathrm{v}}N_W}\left[\frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)+\sigma_{ro}^2}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)+\sigma_{ro}^2}-1\right]\frac{\partial^2\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)\right\}}{\partial A_i \partial A_j}- \qquad (6.40)\\ &\quad \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)+\sigma_{ro}^2}{\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)+\sigma_{ro}^2\right]^2}\frac{\partial}{\partial A_j}\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)\right\}\frac{\partial}{\partial A_i}\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)\right\}. \qquad (6.41)\end{aligned}$$

The log likelihood contains a generic projection operator, ${}_{\mathbf{s}}\mathbf{v}_l\left(\cdot\right)$. The projection operator is presented in generic form to indicate that this derivation holds for all possible projection operations. The linear nature of the projection operator allows it to commute with the differentiation operator:

$$\begin{aligned}\frac{\partial^2 L_{ml}(\mathbf{A})}{\partial A_i \partial A_j} &= \sum_{l=1}^{N_{\mathrm{v}}N_W}\left[\frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)+\sigma_{ro}^2}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}\left[\mathbf{u};\mathbf{A}\right]\right)+\sigma_{ro}^2}-1\right]{}_{\mathbf{s}}\mathbf{v}_l\left(\frac{\partial^2\mathbf{I}_{\cup}\left[\mathbf{A}\right]}{\partial A_i \partial A_j}\right)- \\ &\quad \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)+\sigma_{ro}^2}{\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)+\sigma_{ro}^2\right]^2}{}_{\mathbf{s}}\mathbf{v}_l\left(\frac{\partial}{\partial A_i}\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right){}_{\mathbf{s}}\mathbf{v}_l\left(\frac{\partial}{\partial A_j}\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right). \qquad (6.42)\end{aligned}$$

Substituting these results into the expression for $J_{D_{ij}}$ in (6.26):

$$J_{D_{ij}} = -E_{\mathbf{d},\mathbf{a}} \left\{ \begin{array}{c} \sum_{l=1}^{N_{\mathrm{v}} N_W} \left[\frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)+\sigma_{ro}^2}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2} - 1 \right] {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial^2 \mathbf{I}_{\cup}[\mathbf{A}]}{\partial A_i \partial A_j} \right) - \\ \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)+\sigma_{ro}^2}{\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2\right]^2} {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_i} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right) {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_j} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right) \end{array} \right\}.$$

The detected images, $\mathbf{D}_{\cup}$, are the only random quantities with Poisson noise within the derivative expression. Simplify the expression by evaluating the Poisson part of the expectation:

$$\begin{aligned} J_{D_{ij}} &= E_{\mathbf{a}} \left\{ \begin{array}{c} \sum_{l=1}^{N_{\mathrm{v}} N_W} \left[1 - \frac{E_{\mathbf{d}}\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)|\mathbf{A}\right\}+\sigma_{ro}^2}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2} \right] {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial^2 \mathbf{I}_{\cup}[\mathbf{A}]}{\partial A_i \partial A_j} \right) + \\ \frac{E_{\mathbf{d}}\left\{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{D}_{\cup}\right)|\mathbf{A}\right\}+\sigma_{ro}^2}{\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2\right]^2} {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_i} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right) {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_j} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right) \end{array} \right\}, && (6.43) \\ &= E_{\mathbf{a}} \left\{ \begin{array}{c} \sum_{l=1}^{N_{\mathrm{v}} N_W} \left[1 - \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(E_{\mathbf{d}}\left\{\mathbf{D}_{\cup}|\mathbf{A}\right\}\right)+\sigma_{ro}^2}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2} \right] {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial^2 \mathbf{I}_{\cup}[\mathbf{A}]}{\partial A_i \partial A_j} \right) + \\ \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(E_{\mathbf{d}}\left\{\mathbf{D}_{\cup}|\mathbf{A}\right\}\right)+\sigma_{ro}^2}{\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2\right]^2} {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_i} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right) {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_j} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right) \end{array} \right\}, && (6.44) \end{aligned}$$

where $E_{\mathbf{d}}\{\cdot\}$ represents the expectation taken over the CCD noise and $E_{\mathbf{a}}\{\cdot\}$ represents the expectation over the set of random parameters $\mathbf{a}$. To move the expectation operation inside ${}_{\mathbf{s}}\mathbf{v}_l\left(\cdot\right)$, I have once again taken advantage of the linearity of the projection operator. Evaluating the $E_{\mathbf{d}}\{\cdot\}$ operation, the expectation on $\mathbf{D}$ removes the CCD noise resulting in the expected image $\mathbf{I}$:

$$\begin{aligned} J_{D_{ij}} &= E_{\mathbf{a}} \left\{ \begin{array}{c} \sum_{l=1}^{N_{\mathrm{v}} N_W} \left[1 - \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2} \right] {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial^2 \mathbf{I}_{\cup}[\mathbf{A}]}{\partial A_i \partial A_j} \right) + \\ \frac{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2}{\left[{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}[\mathbf{A}]\right)+\sigma_{ro}^2\right]^2} {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_i} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right) {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_j} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right) \end{array} \right\}, && (6.45) \\ &= E_{\mathbf{a}} \left\{ \sum_{l=1}^{N_{\mathrm{v}} N_W} \frac{{}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_i} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right) {}_{\mathbf{s}}\mathbf{v}_l \left(\frac{\partial}{\partial A_j} \mathbf{I}_{\cup}\left[\mathbf{A}\right] \right)}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}_{\cup}\left[\mathbf{A}\right]\right)+\sigma_{ro}^2} \right\}. && (6.46) \end{aligned}$$

Evaluating the CRLB will require a closed form expression for the derivative of the expected image, $\mathbf{I}\left[\mathbf{A}\right]$. The derivation here follows the results provided by Fienup et. al. [53]. Assume the case of a point source object. Compact the pupil notation by removing the explicit dependency on A_P and R_P: assume that the amplitude function for the field in the aperture plane is a constant value of one and $R_P = 1$. The resulting expression for the pupil is:

$$\mathcal{P}\left[\mathbf{n};\mathbf{A}\right] = W_P\left[\mathbf{n}\right] \exp\left\{ \mathrm{j} P_{\phi}\left[\mathbf{n};\mathbf{A}\right] \right\}. \qquad (6.47)$$

Using the linear model of the optical system, the expected image for the case of a point source object is given by:

$$\begin{aligned}
\mathbf{I}\left[\mathbf{u};\mathbf{A}\right] &= K\frac{\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]\mathcal{I}^{*}\left[\mathbf{u};\mathbf{A}\right]}{\sum\limits_{\mathbf{u}}\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]\mathcal{I}^{*}\left[\mathbf{u};\mathbf{A}\right]}, &(6.48)\\
&= \tilde{K}\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]\mathcal{I}^{*}\left[\mathbf{u};\mathbf{A}\right]. &(6.49)
\end{aligned}$$

Normalization by the average photon count K models the SNR in each image plane. The constant $\tilde{K}$ is introduced to compact the notation. An analytical expression for the image derivative requires differentiating the image with respect to the Zernike parameter. Ignoring the constant $\frac{\Delta x\Delta y}{(\lambda s_i)^2}$, which will be replaced by SNR scaling, and differentiating yields:

$$\begin{aligned}
\frac{\partial}{\partial A_i}\mathbf{I}\left[\mathbf{u};\mathbf{A}\right] &= \frac{\partial}{\partial A_i}\left\{\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]\mathcal{I}^{*}\left[\mathbf{u};\mathbf{A}\right]\right\}, &(6.50)\\
&= \frac{\partial}{\partial A_i}\left\{\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]\right\}\mathcal{I}^{*}\left[\mathbf{u};\mathbf{A}\right]+\frac{\partial}{\partial A_i}\left\{\mathcal{I}^{*}\left[\mathbf{u};\mathbf{A}\right]\right\}\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]. &(6.51)
\end{aligned}$$

The image field, $\mathcal{I}$, is calculated via the discrete Fraunhofer diffraction integral of the pupil function:

$$\begin{aligned}
\mathcal{I}\left[\mathbf{u};\mathbf{A}\right] &= \frac{\Delta x\Delta y}{(\lambda s_i)^2}\mathcal{DFT}\{\mathcal{P}\left[\mathbf{n};\mathbf{A}\right]\}, &(6.52)\\
&= \frac{\Delta x\Delta y}{(\lambda s_i)^2}\sum_{\mathbf{n}}\mathcal{P}\left[\mathbf{n};\mathbf{A}\right]\exp\left\{-\mathrm{j}\frac{2\pi}{N}\left[\mathbf{n}\cdot\mathbf{u}\right]\right\}, &(6.53)\\
&= \frac{\Delta x\Delta y}{(\lambda s_i)^2}\sum_{\mathbf{n}}W_P\left[\mathbf{n}\right]\exp\left\{\mathrm{j}P_{\phi}\left[\mathbf{n};\mathbf{A}\right]\right\}\exp\left\{-\mathrm{j}\frac{2\pi}{N}\left[\mathbf{n}\cdot\mathbf{u}\right]\right\}. &(6.54)
\end{aligned}$$

Exchanging the order of summation and differentiation, it is easy to show that $\frac{\partial}{\partial A}\mathcal{I}^* = \left(\frac{\partial}{\partial A}\mathcal{I}\right)^*$:

$$\begin{aligned}
\frac{\partial}{\partial A_i}\mathcal{I}\left[\mathbf{u};\mathbf{A}\right] &= \sum_{\mathbf{n}} W_P(\mathbf{n})\frac{\partial}{\partial A_i}\left[\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\right] \times \\
&\quad \exp\left\{-\mathrm{j}\frac{2\pi}{N}(\mathbf{x}\cdot\mathbf{u})\right\}, && (6.55) \\
\frac{\partial}{\partial A_i}\mathcal{I}^*\left[\mathbf{u};\mathbf{A}\right] &= \sum_{\mathbf{n}} W_P(\mathbf{x})\frac{\partial}{\partial A_i}\left[\exp\left\{-\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\right] \times \\
&\quad \exp\left\{\mathrm{j}\frac{2\pi}{N}\left[\mathbf{n}\cdot\mathbf{u}\right]\right\}, && (6.56) \\
&= \left(\frac{\partial}{\partial A_i}\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]\right)^*, && (6.57)
\end{aligned}$$

Substituting (6.57) and (6.51) into (6.49):

$$\frac{\partial}{\partial A_i}\mathbf{I}\left[\mathbf{u};\mathbf{A}\right] = \tilde{K}\left[\frac{\partial}{\partial A_i}\left[\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]\right]\mathcal{I}^*\left[\mathbf{u};\mathbf{A}\right] + \left(\frac{\partial}{\partial A_i}\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]\right)^*\mathcal{I}\left[\mathbf{u};\mathbf{A}\right]\right]. \tag{6.58}$$

To further define the image derivative, evaluate the derivative of the exponential phase term in (6.56):

$$\frac{\partial}{\partial A_i}\left[\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\right] = \mathrm{j}\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\frac{\partial}{\partial A_i}P_\phi\left[\mathbf{n};\mathbf{A}\right]. \tag{6.59}$$

The phase function, P_ϕ, and its derivative are defined for each of the Zernike modes:

$$\begin{aligned}
P_\phi\left[\mathbf{n};\mathbf{A}\right] &= A_i Z_i\left[\mathbf{n}\right] + \sum_{\substack{l\in S\\ l\neq i}} A_l Z_l\left[\mathbf{n}\right], && (6.60) \\
\frac{\partial}{\partial A_i}P_\phi\left[\mathbf{n};\mathbf{A}\right] &= Z_i\left[\mathbf{n}\right]. && (6.61)
\end{aligned}$$

Substituting (6.61) and (6.59) into (6.55) gives:

$$\begin{aligned}
\frac{\partial}{\partial A_i}\mathcal{I}\left[\mathbf{u};\mathbf{A}\right] &= \mathrm{j}\sum_{\mathbf{n}} W_P\left[\mathbf{n}\right]Z_i\left[\mathbf{n}\right]\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\} \times \\
&\quad \exp\left\{-\mathrm{j}\frac{2\pi}{N}\left[\mathbf{n}\cdot\mathbf{u}\right]\right\}, && (6.62) \\
&= \mathrm{j}\mathcal{DFT}\left\{W_P\left[\mathbf{n}\right]Z_i\left[\mathbf{n}\right]\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\right\}. && (6.63)
\end{aligned}$$

Substituting (6.63) into (6.58) produces the derivative of the expected image:

$$\begin{aligned}\frac{\partial}{\partial A_i}\mathbf{I}\left[\mathbf{u};\mathbf{A}\right] &= \mathrm{j}\tilde{K}\mathcal{DFT}\left\{W_P\left[\mathbf{n}\right]Z_i\left[\mathbf{n}\right]\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\right\}\mathcal{I}^*\left[\mathbf{u};\mathbf{A}\right]+\\ &\quad\tilde{K}\left(\mathrm{j}\mathcal{DFT}\left\{W_P\left[\mathbf{n}\right]Z_i\left[\mathbf{n}\right]\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\right\}\right)^*\mathcal{I}\left[\mathbf{u};\mathbf{A}\right], \quad (6.64)\\ &= -2\tilde{K}\,\mathrm{Im}\left\{\mathcal{DFT}\left\{W_P\left[\mathbf{n}\right]Z_i\left[\mathbf{n}\right]\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\right\}\mathcal{I}^*\left[\mathbf{u};\mathbf{A}\right]\right\}. \quad (6.65)\end{aligned}$$

Using this closed form expression for the image derivative, the Fisher information matrix entry in (6.45) becomes:

$$J_{D_{ij}} = E_{\mathbf{a}}\left\{\begin{array}{c}\sum_{l=1}^{N_v N_W}\frac{4\tilde{K}^2}{{}_{\mathbf{s}}\mathbf{v}_l\left(\mathbf{I}[\mathbf{u};\mathbf{A}]\right)+\sigma_{ro}^2}\times\\ {}_{\mathbf{s}}\mathbf{v}_l\left(\mathrm{Im}\left\{\mathcal{DFT}\left\{W_P\left[\mathbf{n}\right]Z_i\left[\mathbf{n}\right]\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\right\}\mathcal{I}^*\left[\mathbf{u};\mathbf{A}\right]\right\}\right)\times\\ {}_{\mathbf{s}}\mathbf{v}_l\left(\mathrm{Im}\left\{\mathcal{DFT}\left\{W_P\left[\mathbf{n}\right]Z_j\left[\mathbf{n}\right]\exp\left\{\mathrm{j}P_\phi\left[\mathbf{n};\mathbf{A}\right]\right\}\right\}\mathcal{I}^*\left[\mathbf{u};\mathbf{A}\right]\right\}\right)\end{array}\right\} \quad (6.66)$$

Combining this result with the a priori matrix, $\mathbf{J}_P$, yields a straightforward method for evaluating the CRLB. The only nontrivial calculation is the expectation over the parameter set $\mathbf{a}$. This integral cannot be evaluated analytically. Instead it may be approximated using a Monte Carlo simulation. The Monte Carlo simulation requires a sequence of random atmospheric realizations formed using an appropriate distribution of the parameters $\mathbf{a}$. Using random atmospheric realizations, an ensemble average of $J_{D_{ij}}$ values can be computed. The average value of $\mathbf{J}_D$ is then used to form $\mathbf{J}_T$:

$$\mathbf{J}_T = \frac{\sum_k \mathbf{J}_{D_k}}{N_k} + \mathbf{\Lambda}_{\mathbf{a}}^{-1}, \quad (6.67)$$

where $\mathbf{J}_{D_k}$ represents a single Monte Carlo realization of the matrix $\mathbf{J}_D$, and N_k represents the number of Monte Carlo trials. Thus, the lower bound on residual mean squared error is given by:

$$\left\langle P_{\phi_e}^2\right\rangle = \underset{i\in S}{\mathrm{Trace}}\left\{\mathbf{J}_T^{-1}\right\} + \sum_{i\notin S} E\left\{a_i^2\right\}, \quad (6.68)$$

where the trace of the matrix is taken over only the parameters included in the set S.

6.3 Adjusting Design Variables to Minimize CRLB

There are many variables contributing to estimator performance, some are environmental variables over which the system designer has little choice, but others offer flexibility to the designer. Flexible wavefront sensor design variables should be adjusted to some ideal setting for a given set of environment variables. The ideal sensor configuration can be defined as that configuration of flexible design variables which minimizes CRLB for some set of environment variables. Recall the list of key environment and design variables from Section 5.5. The CRLB is calculated by Monte Carlo simulation using atmospheric phase screens that contain a limited set of Zernike polynomials. Due to the limited set of Zernike polynomials present in the simulation, there may be a significant difference between actual performance and the derived lower bound. With this in mind, the designer may view the minimum CRLB settings as a starting point when adjusting the design to optimize performance. In the next two subsections, I will demonstrate using the CRLB to make ideal design choices for a whole plane projection sensor and a half plane projection sensor. The whole plane projection sensor, the Z_{2-4} sensor, was discussed in Chapter 5. The half plane projection includes two vectors from each CCD and is the image information used by the Z_{2-10} sensor to be detailed in Chapter 9.

Whole Plane Projection CRLB. The Z_{2-4} sensor design variables along with the operational variables constitute a multi-dimensional domain space for the CRLB function. An exhaustive search for ideal design settings would involve minimizing the CRLB over the range of all design variables for every location in the operational space. Rather than attempt such a global minimization, I will examine the CRLB at a limited number of operating points and the minimization will be conducted with respect to each design variable independently. Further, to reduce the amount of analysis shown here, the operational space will be limited to a range over the Fried parameter, r_0, and the average photon count per subaperture per exposure period, K. The set of design variables will be limited to:

$$\theta_2 - \theta_1 \equiv \text{projection separation angle,} \tag{6.69}$$

$$\theta_1 \equiv \text{the first CCD projection angle,} \tag{6.70}$$

$$\pm\delta_{a_4} \equiv \text{defocus diversity,} \tag{6.71}$$

$$\text{and } N_W \equiv \text{window size.} \tag{6.72}$$

Within the inertial range, the analytical expression for the phase spectrum relies heavily on the ratio $\frac{D_P}{r_0}$. For this reason, performance within the inertial range is largely a function of the ratio $\frac{D_P}{r_0}$ rather than each variable independently. For this reason, I have chosen to fix D_P and vary r_0. This will provide an indication of how performance varies with the ratio without independently varying both D_P and r_0. For all CRLB plots included in this section, the subaperture diameter will be fixed at 0.07m and the environment variables L_0 and l_0 are fixed at $L_0 = 10$m, $l_0 = 0.01$m. In Figures 6.2 through 6.7 the environment variable r_0 is fixed at 0.05m. Figure 6.2 demonstrates how the CRLB for residual mean squared error varies with respect to separation angle between the two CCD arrays. The CRLB indicates that a separation angle of approximately 90 degrees between image projections is ideal. Using this ideal separation angle, the CRLB can be plotted for varying start angle. For instance, Figure 6.3 shows how CRLB varies with respect to choice of projection angle θ_1 given a separation angle of 90 degrees. The CRLB varies randomly over the range of θ_1 angles in Figure 6.3. The variance of the CRLB is small enough to indicate that no significant change in the lower bound occurs over the range of θ_1 values. This reveals that for a fixed separation angle of 90 degrees between projections, the CRLB is effectively invariant to CCD rotation. Based on these results, the remaining CRLB plots will depict configurations where $\theta_1 = 0°$ and $\theta_2 = 90°$. Figure 6.4 shows how CRLB indicates the ideal defocus diversity for a high SNR case. Examining the plot in Figure 6.4, the CRLB is minimized when the defocus diversity is approximately 0.35 radians. Similarly, Figure 6.5 shows how CRLB indicates the ideal defocus diversity in a low SNR case is approximately 0.15 radians. Given the ideal diversity choices, the CRLB may be used to select a window length. Figure 6.6 demonstrates how CRLB varies over window length in high SNR. Figure 6.7 demonstrates how CRLB varies over window length in low SNR. The CRLB

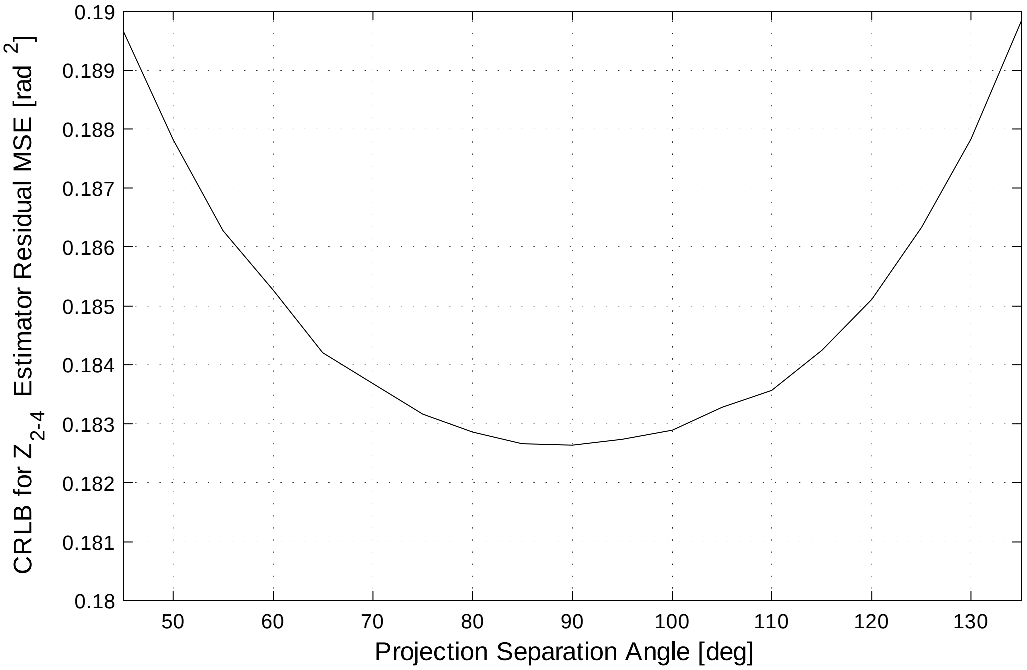


Figure 6.2 Z_{2-4} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ lower bound versus separation angle, $\theta_2 - \theta_1$. $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $W_N = 14$ pixels.

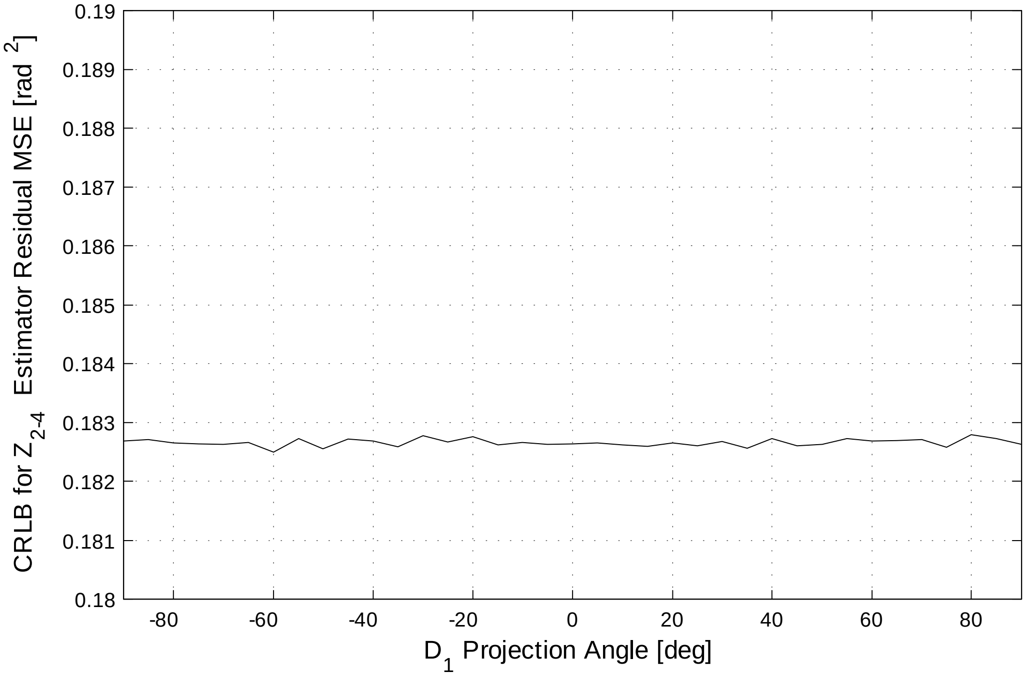


Figure 6.3 Z_{2-4} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ lower bound versus projection angle θ_1 given that $\theta_2 = \theta_1 + 90$. $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $W_N = 14$ pixels.

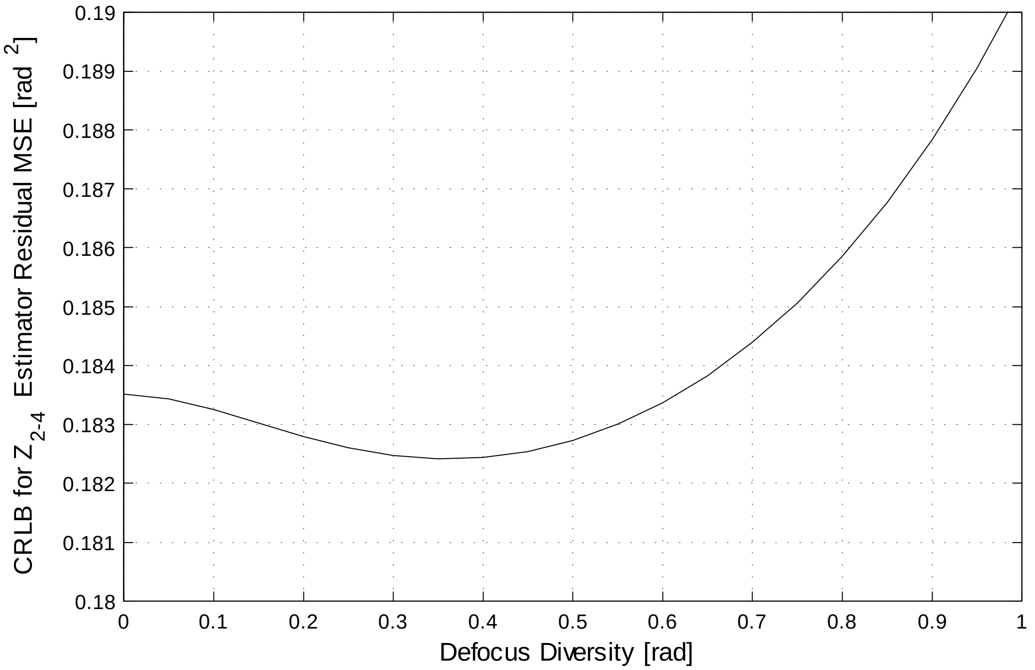


Figure 6.4 Z_{2-4} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ lower bound versus $\pm\delta_{a_4}$ for $K = 1000$ photons per subaperture (high SNR). $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $W_N = 9$ pixels.

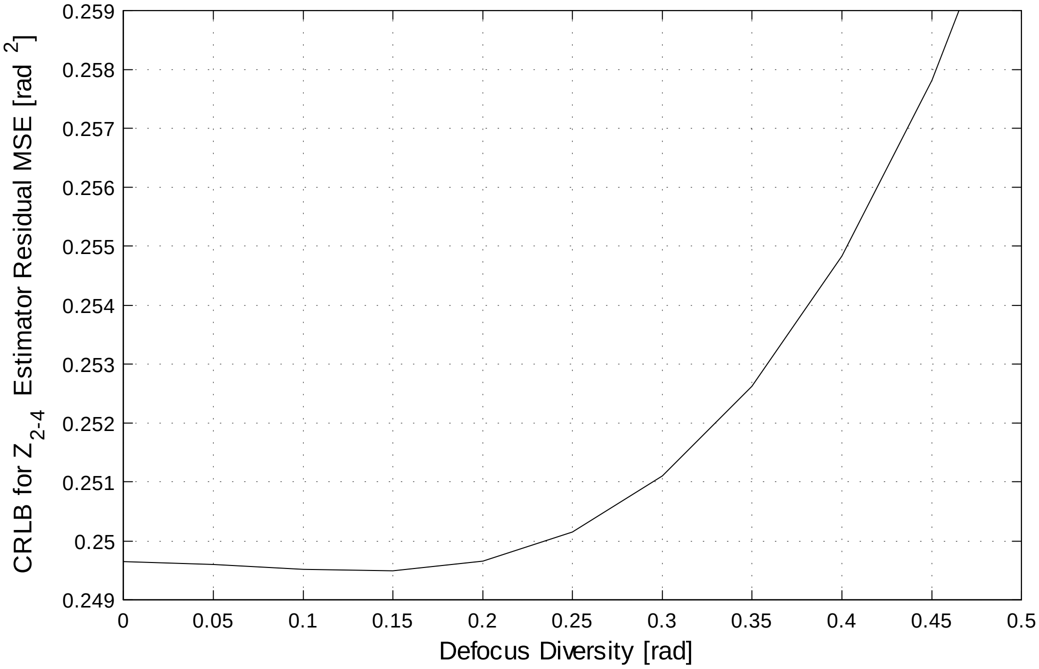


Figure 6.5 Z_{2-4} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ lower bound versus $\pm\delta_{a_4}$ for $K = 100$ photons per subaperture (low SNR). $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $W_N = 9$ pixels.

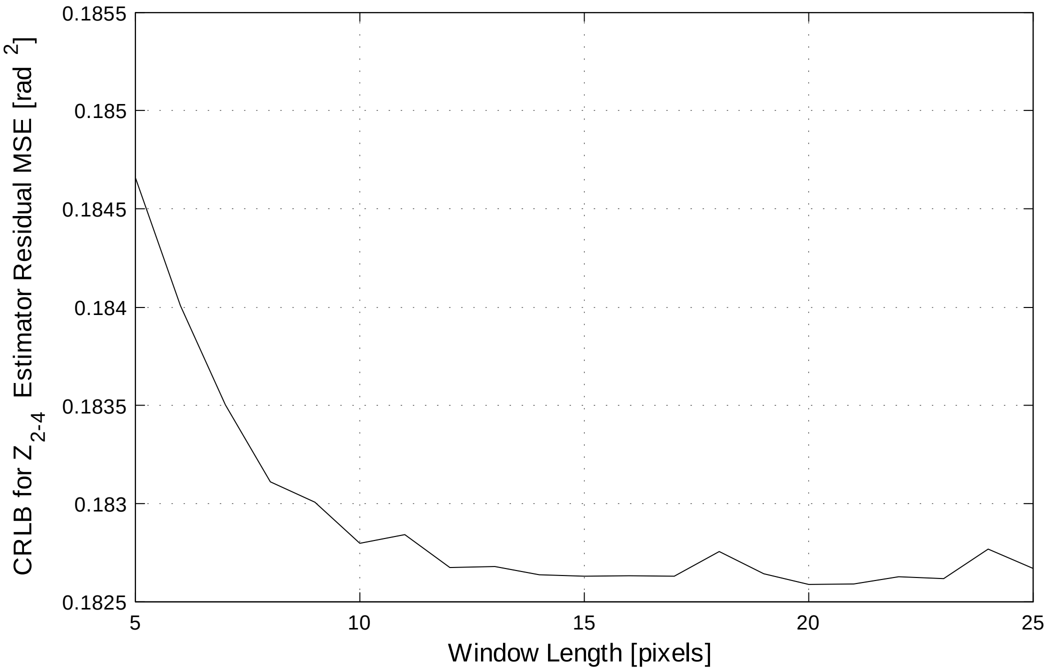


Figure 6.6 Z_{2-4} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ lower bound versus N_W for $K = 1000$ photons per subaperture (high SNR). $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, and $l_0 = 0.01$m.

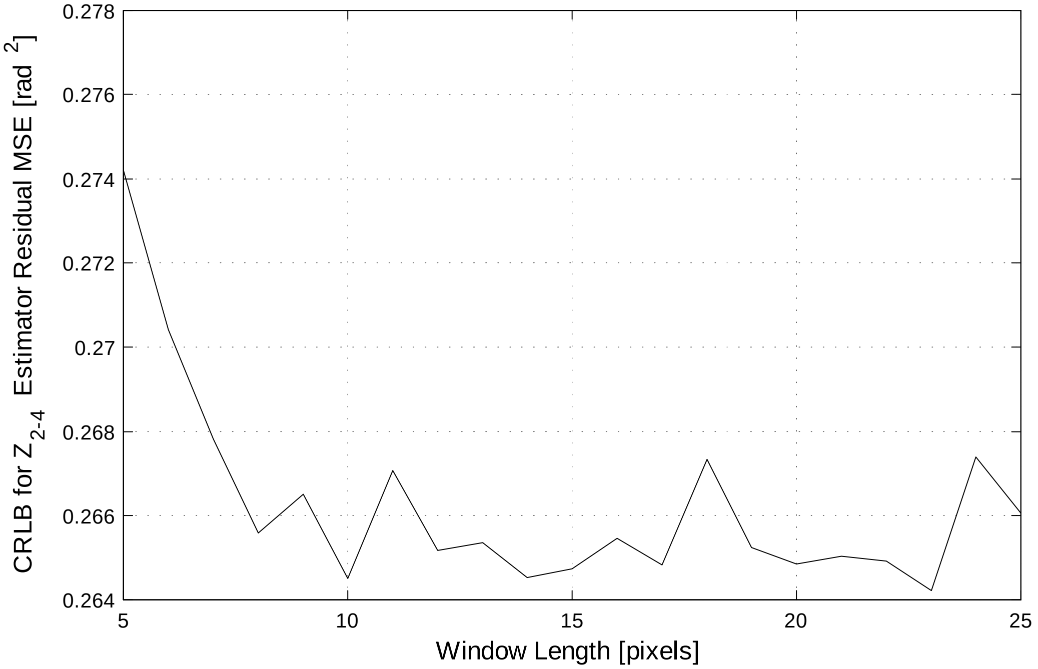


Figure 6.7 Z_{2-4} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ lower bound versus N_W for $K = 100$ photons per subaperture (low SNR). $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, and $l_0 = 0.01$m.

versus N_W results demonstrate that the benefit gained from increasing the window length beyond 11 pixels is minimal. In practice, the ideal window length will be driven largely by the maximum amount of time allotted for CCD read out and the time required to manage the additional pixels in each image projection. The CRLB plots here are useful in that they demonstrate a knee in the performance curve around the $7-11$ pixel range. Using the middle of the knee on the CRLB versus N_W plots, I have selected a 9×9 pixel window and plotted CRLB over a range of r_0 and K values. The CRLB plot in Figure 6.8 demonstrates performance using a 9×9 pixel window. Numbers at each point for the Z_{2-4} CRLB plot indicate the ideal defocus diversity for that point in the operating space.

Define the operating space as a set of environment variables each with some expected range of values. The CRLB can be used to identify a region within the operating space where a particular type of estimator is optimal. As an example, Figure 6.8 provides a comparison of CRLB for the Z_{2-4} sensor for the case where only a_2 and a_3 are estimated, define this as the $Z_{2,3}$ estimator, versus the case where parameters a_2 through a_4 are estimated. Dashed lines indicate $Z_{2,3}$ estimator performance while solid lines indicate Z_{2-4} estimator performance. Both cases are demonstrated over a range of r_0 and K. The ordered pairs in parentheses on the right hand side of the plot region indicate the ratio $\frac{D_P}{r_0}$ and the value of r_0 for the plot lines ending closest to the ordered pair. Once again, numbers at each plot point location for the Z_{2-4} estimator indicate the ideal choice of diversity for that point in the operating space. Note that the ideal choice for diversity for the $Z_{2,3}$ estimator is always 0 radians and therefore the plot line points for $Z_{2,3}$ estimator have no diversity indicators. This figure is included because it provides an indication of the region in the operating space where there is opportunity for increased performance via estimating Z_4. Specifically, the figure indicates that there is a limit, in the lower left of the plot region, as SNR decreases and as r_0 increases (as the ratio $\frac{D_P}{r_0}$ decreases) beyond which there is little or no benefit from estimating a_4.

Figures 6.2 through 6.7 demonstrate that minimum CRLB is a useful measure for determining ideal design variable settings. Figure 6.8 demonstrates that the CRLB is also useful for identifying regions in the operating space where a particular estimator can provide superior performance. Actual performance will only match the CRLB in the event that the estimator is efficient. Thus, the assumption is that actual estimator performance

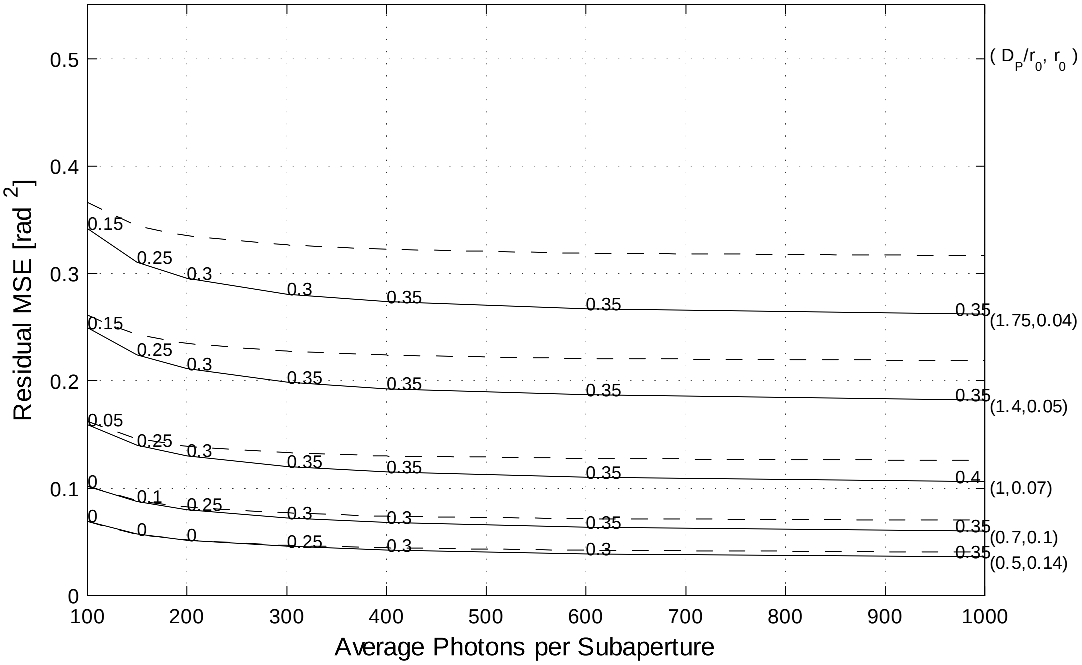


Figure 6.8 Lower bounds on $\left\langle P^2_{\phi_e} \right\rangle$ versus K for several cases of r_0. Dashed lines indicate $Z_{2,3}$ estimator performance bounds. Solid lines indicate Z_{2-4} estimator performance bounds.

trends the same as the CRLB such that the CRLB provides a good starting point for design choices. Along with performance comparisons to other estimators, Chapter 8 will provide performance plots to compare with the CRLB results presented here. Also, the degree to which simulated performance reflects the same ideal design variable settings will determine the effectiveness of using CRLB to make such preliminary design choices. Comparing simulated performance to the CRLB provides a means with which to verify that the sensor simulation is producing valid results. In this manner, simulated performance and CRLB results serve to compliment one another.

Half Plane Projection CRLB. In addition to the whole plane projection sensor discussed in Chapter 5, this dissertation will provide a discussion detailing a half plane projection sensor in Chapter 9. The half plane projection sensor is designed to estimate coefficients for Zernike polynomials Z_2 through Z_{10}. Just as in the previous subsection, a set of figures containing CRLB results for the half plane projection sensor are provided here as a demonstration of how CRLB can be used to choose ideal design settings. Beginning with projection separation angle and projection starting angle, Figures 6.9 and 6.10 show that the results for whole plane projections hold true for half plane projections as well: the ideal separation angle is 90 degrees, and, given that $\theta_2 = \theta_1 + 90°$, the CRLB is effectively invariant over a range of θ_1 values. Figures 6.11 and 6.12 demonstrate that the ideal

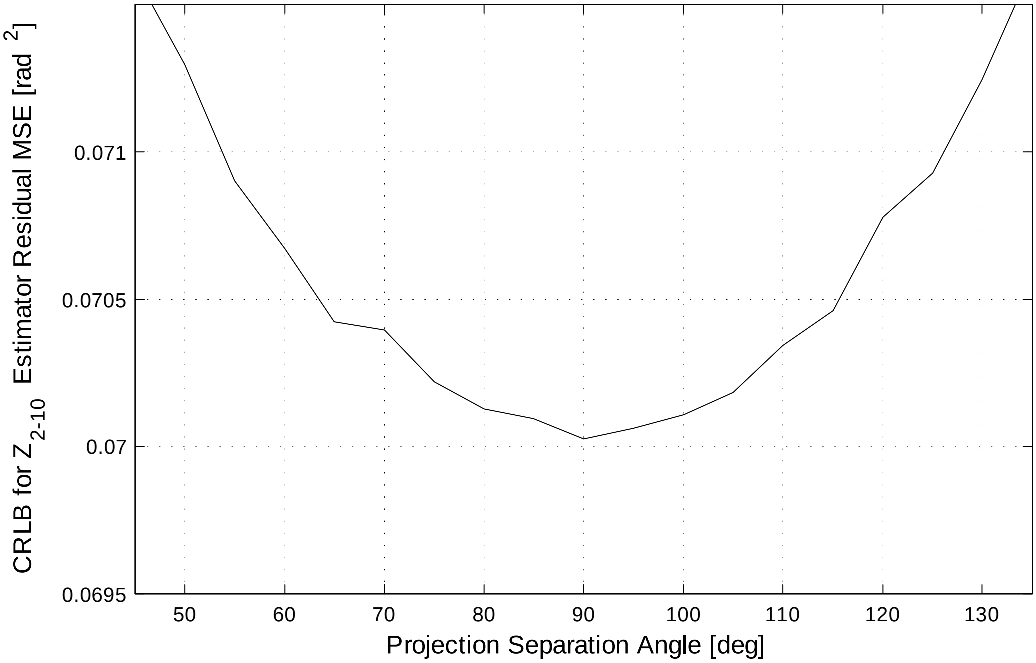


Figure 6.9 Z_{2-10} estimator $\left\langle P^2_{\phi_e} \right\rangle$ lower bound versus separation angle, $\theta_2 - \theta_1$. $D_P =$ 0.07m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $N_W = 14$ pixels.

diversity settings are approximately 0.55 radians in high SNR and 0.4 radians in low SNR. Although the values for the ideal diversity in each case here differ from the Z_{2-4} sensor cases, there is a trend worth emphasizing: ideal diversity decreases as SNR decreases. This is an indication that higher order coefficients are difficult to estimate in low signal situations. In these situations, the sensor performs better by reducing diversity. Reducing the amount of diversity in the sensor improves tilt and lower order coefficient estimates at the cost of higher order coefficients. The CRLB plots in Figures 6.13 and 6.14 suggest that the knee in performance improvement with respect to CCD window size occurs in the $8 - 12$ pixel range. Finally, Figures 6.15 and 6.16 provide plots of CRLB over a region of the operating space. In these figures, several design variables are fixed: $\theta_1 = 0$, $\theta_2 = 90$, and $N_W = 9$ pixels. The environment variables L_0 and l_0 are also fixed at $L_0 = 10$m and $l_0 = 0.01$m. CRLB is plotted as r_0 varies in several steps from 0.04m to 0.14m and as K varies from

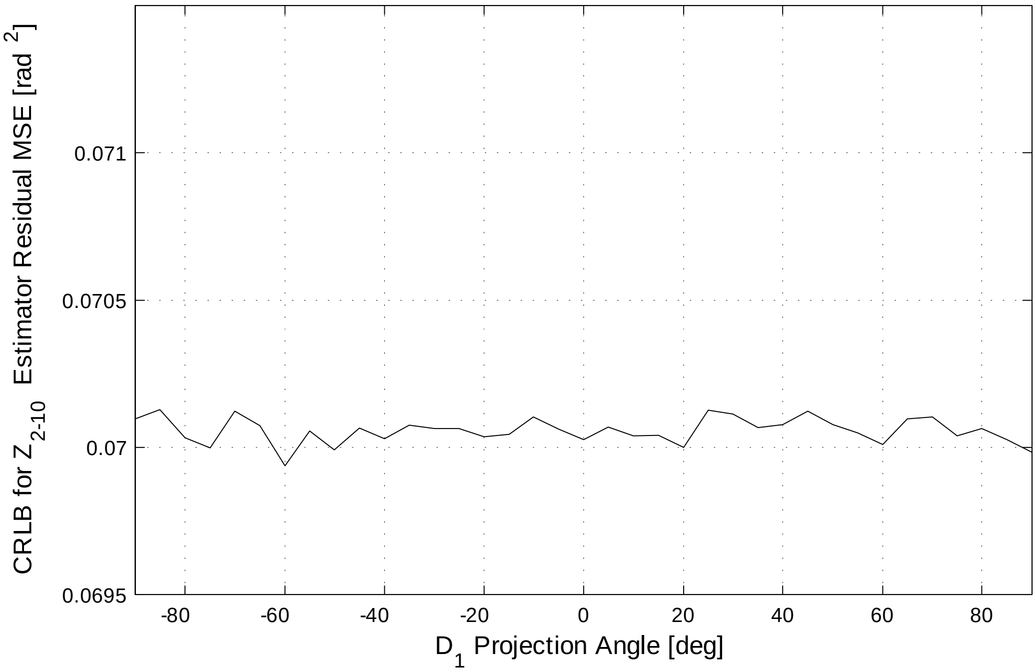


Figure 6.10 Z_{2-10} estimator $\left\langle P^2_{\phi_e} \right\rangle$ lower bound versus projection angle θ_1 given that $\theta_2 = \theta_1 + 90$. $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $N_W = 14$ pixels.

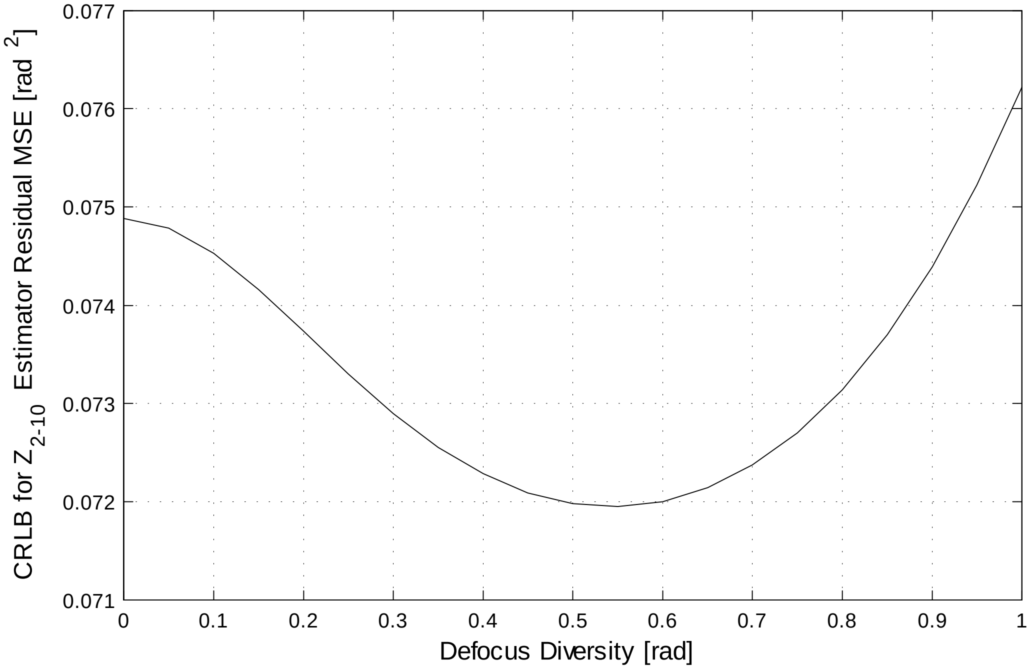


Figure 6.11 Z_{2-10} estimator $\left\langle P^2_{\phi_e} \right\rangle$ lower bound versus $\pm\delta_{a_4}$ for $K = 1000$ photons per subaperture (high SNR). $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $N_W = 9$ pixels.

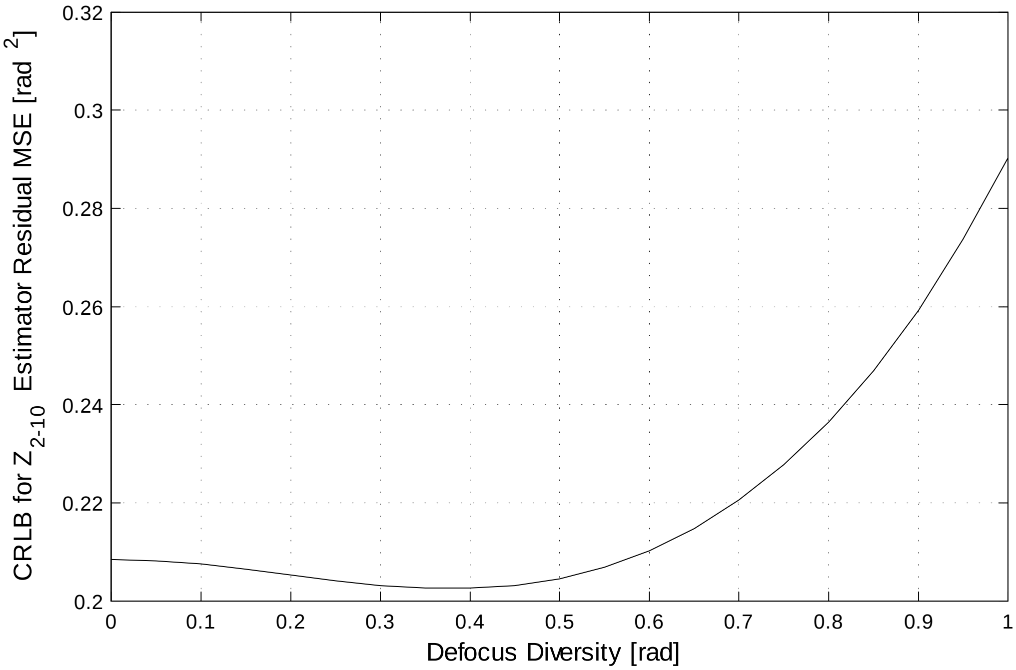


Figure 6.12 Z_{2-10} estimator $\left\langle P^2_{\phi_e} \right\rangle$ lower bound versus $\pm\delta_{a_4}$ for $K = 100$ photons per subaperture (low SNR). $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $N_W = 9$ pixels.

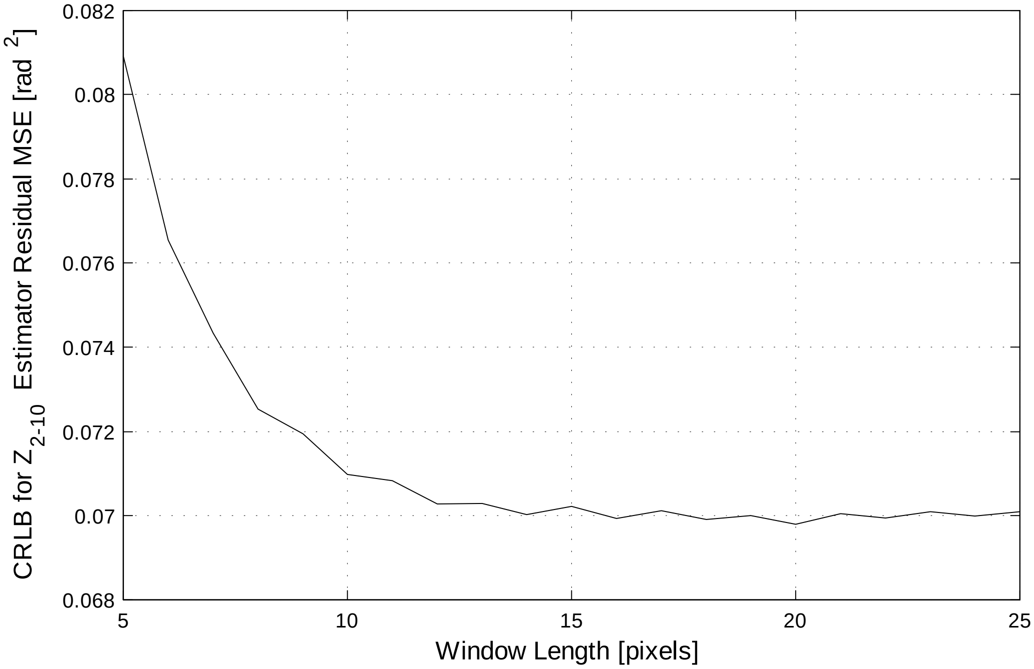


Figure 6.13 Z_{2-10} estimator $\left\langle P^2_{\phi_e} \right\rangle$ lower bound versus N_W for $K = 1000$ photons per subaperture (high SNR). $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, and $l_0 = 0.01$m.

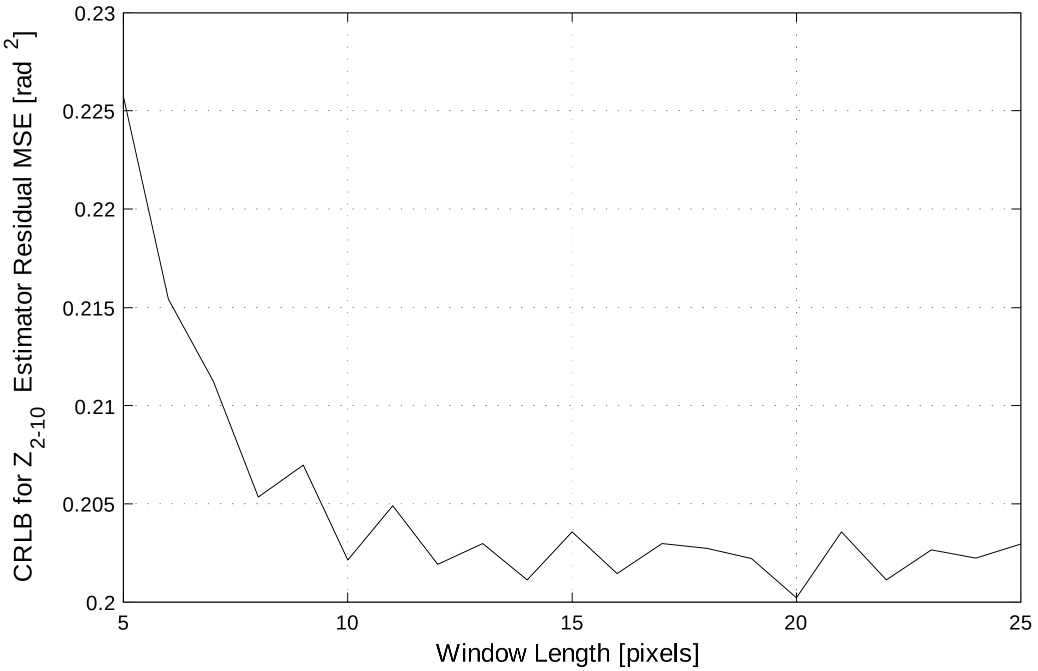


Figure 6.14 Z_{2-10} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ lower bound versus N_W for $K = 100$ photons per subaperture (low SNR). $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, and $l_0 = 0.01$m.

100 to 1000 average photons. The lower bound for the Z_{2-10} sensor is compared to the bound for estimating tilt only (the $Z_{2,3}$ sensor). Dashed lines indicate the $Z_{2,3}$ sensor bound while solid lines indicate the Z_{2-10} sensor bound. Numbers at each plot point location for the Z_{2-10} sensor indicate the ideal choice of diversity for that point in the operating space. Once again, the ideal diversity for estimating tilt only is assumed to be zero.

6.4 Summary

This chapter proposed that the MSE performance measure provides a sound method for comparing simulated wavefront sensor designs. The discussion began with a derivation of the residual MSE in terms of the von Kármán atmospheric modal parameters. A few cases of perfect estimator performance were plotted in order to demonstrate the relationship between residual wavefront MSE, the set of parameters estimated and the size of the aperture. In the section on CRLB, the relationship between wavefront residual MSE and estimator error was highlighted. Considering the wavefront sensor as a parameter estimator, it was easy to see that the portion of the wavefront MSE which can be affected translates directly to the

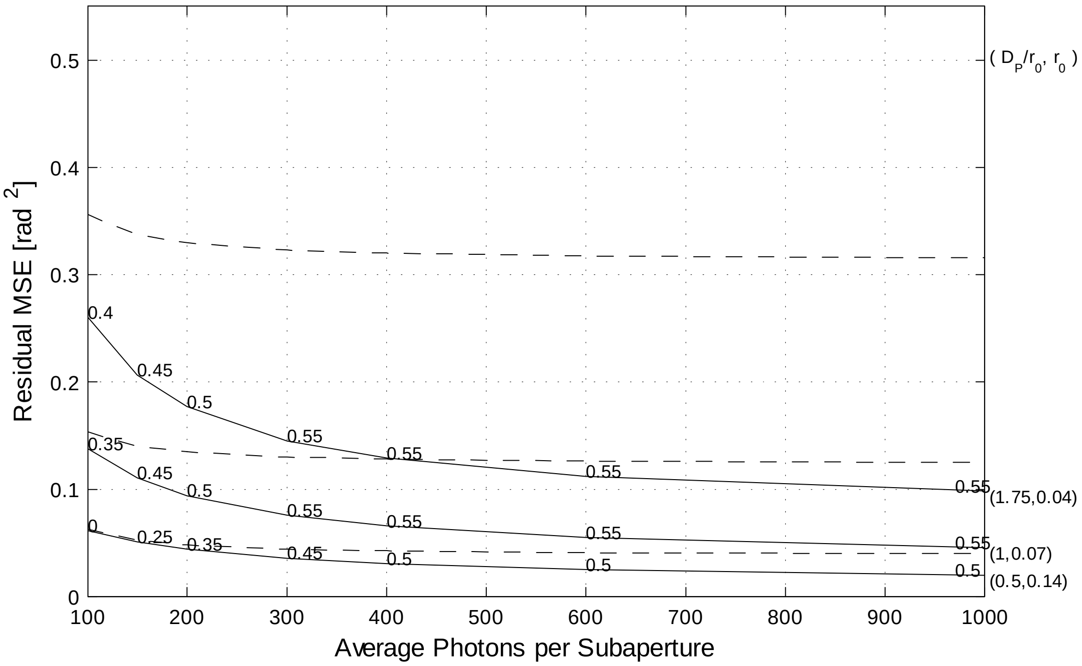


Figure 6.15 Lower bounds on $\left\langle P_{\phi_e}^2 \right\rangle$ versus K for several cases of r_0. Dashed lines indicate $Z_{2,3}$ estimator performance bounds. Solid lines indicate Z_{2-10} estimator performance bounds.

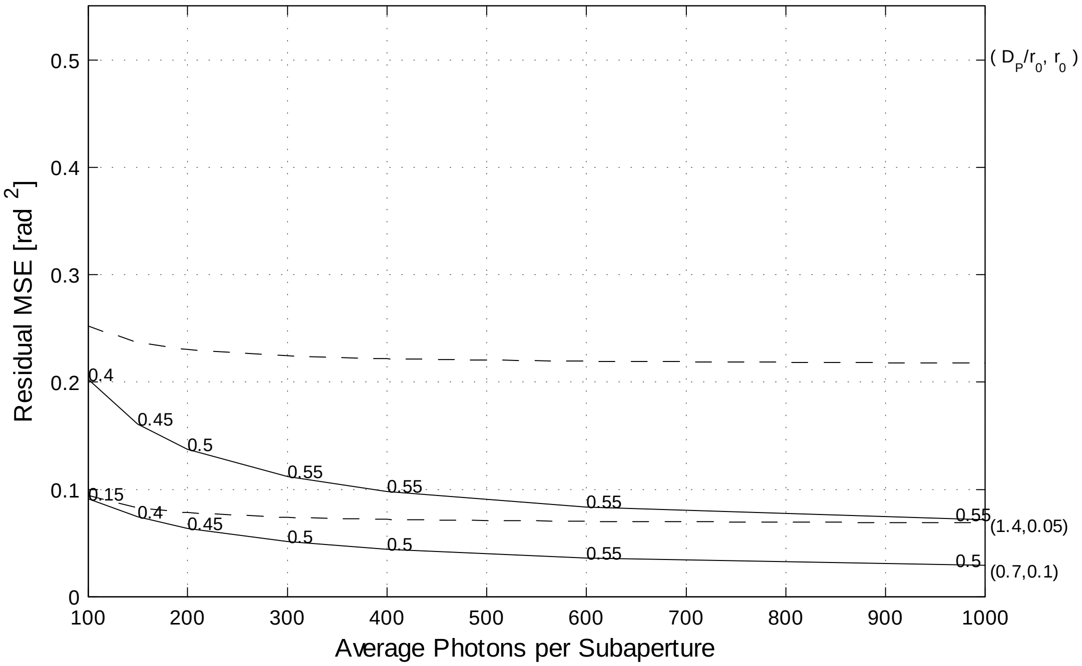


Figure 6.16 Lower bounds on $\left\langle P_{\phi_e}^2 \right\rangle$ versus K for several cases of r_0. Dashed lines indicate $Z_{2,3}$ estimator performance bounds. Solid lines indicate Z_{2-10} estimator performance bounds.

estimator mean squared error. For unbiased estimators, the estimator mean squared error becomes an estimator variance. The Cramér Rao lower bound provides a lower bound on estimator variance. A method was developed for approximating the CRLB for the case where the sensor CCDs are modeled as a Poisson process. This lower bound provides a basis for validating simulated sensor performance and for selecting ideal settings for sensor design variables. The CRLB was computed and example plots were provided for both the whole plane and half plane projection based sensors. In each case, CRLB was used to identify ideal settings for key design variables and regions within the operating space where each sensor has the potential to offer improved performance over the tilt only sensor. It is proposed that the CRLB will provide a complimentary performance measure with which to compare to simulated performance. The combination of CRLB results and simulated performance results should serve to provide confidence in the sensor simulation results and in the use of CRLB as means of establishing initial design variable settings.

7. Simulating the Atmosphere

Simulating the wavefront sensor requires an implementation of the atmospheric model derived in Section 2.2. Recall two of the key assumptions within the atmospheric model: segments of the atmosphere can be compressed into discrete layers, and the atmospheric phase is ergodic. These characteristics serve as guidelines for determining the suitability of random process generation techniques for creating realizations of the statistical phase model. The first assumption allows dividing the atmosphere into discrete layers, referred to as the thin screen approach. The random effects of phase delay in each layer are condensed into an infinitesimally thin phase modifier. The assumption of ergodicity implies that, within each thin screen, spatial statistics equal, in a statistical sense, appropriate temporal statistics [54]. Using this assumption, a temporally evolving atmosphere can be modeled under the assumptions of Taylor's frozen flow. Taylor's frozen flow hypothesis assumes that the random nature of the phase in each layer is frozen except for a constant velocity perpendicular to the optical axis [3]. Therefore, to accommodate both the thin screen and ergodicity assumptions, the simulation must be able to generate thin phase realizations of arbitrary length. The Fourier series method of phase screen generation is a good fit for both of these criteria. Application of the Fourier series technique using an equal spaced Cartesian sampling structure introduces significant computational complexity. Logarithmically spaced sample structures have been suggested to reduce complexity and emphasize lower frequencies [55]. Below, I will investigate the log-Cartesian sampling structure and suggest a log-polar modification to improve performance. To measure the fidelity of each technique, I compare the statistics of the random screens to the derived atmospheric model. Many phase screen generation codes use a visual inspection of the structure function as a verification of performance [56], [57]. I will modify this measure in two ways. First, I will use structure function percent error as a performance measure. Second, I will identify the level of isotropy in the resulting screens by measuring structure function percent error over a range of angles. The curvature sensor simulations discussed in Chapters 8 and 9 include an implementation of the log-polar phase screen technique. Statistical errors in the phase screens have the potential to effect the sensor simulation results. In particular, the Z_{2-4} and Z_{2-10} sensors rely on accurate representation of lower order Zernike variance. For this reason, I include the variance of the estimated Zernike coefficients in the phase screen performance measure. The sections to follow provide a review of Fourier series phase screen generation, a note on the benefits of log-polar frequency sampling, and a comparison of phase screen statistical accuracy when using log-polar versus log-Cartesian sampling.

7.1 Fourier Series Phase Screen Generation

The mathematical foundation for the atmospheric simulation was outlined in Section 2.2 where I presented the derivation of the optical application of Kolmogorov's turbulence model and the von Kármán phase power spectrum, Φ_{P_ϕ} (Note: the Fourier series technique is not limited to any particular power spectrum). All that remains is to formulate a method for creating random phase realizations with the same statistical characteristics as the atmospheric model. The Fourier series method is a common technique for creating random processes from spectral statistics. The following paragraphs provide a review of the derivation of the Fourier series technique. Begin with the well known statistical relationship between the power spectrum, Φ_{P_ϕ}, and the autocorrelation, B_{P_ϕ}, of a wide sense stationary (WSS) random process, P_ϕ [22]:

$$\Phi_{P_\phi}(\mathbf{K}) = \int B_{P_\phi}(\mathbf{R}) \exp\{-\mathrm{j}\mathbf{K}\cdot\mathbf{R}\}\, d\mathbf{R}. \tag{7.1}$$

This relationship along with F_ϕ, the Fourier transform of P_ϕ, is necessary to develop the method. Consider a realization of the pupil phase, P_ϕ, denoted $\hat{P}_\phi$. Under the definition of WSS, $\hat{F}_\phi$ is delta correlated on $\mathbf{K}$ with average value equal to the power spectrum Φ_{P_ϕ} [22]. To demonstrate this property, begin with the Fourier transform of P_ϕ:

$$\hat{F}_\phi(\mathbf{K}) = \int \hat{P}_\phi(\mathbf{R}) \exp\{-\mathrm{j}\mathbf{K}\cdot\mathbf{R}\}\, d\mathbf{R}. \tag{7.2}$$

Forming the correlation yields:

$$E\left\{\hat{F}_\phi(\mathbf{K}_1)\hat{F}_\phi^*(\mathbf{K}_2)\right\} = E\left\{\begin{array}{c} \int \hat{P}_\phi(\mathbf{R}_1) \exp\{-\mathrm{j}\mathbf{K}_1\cdot\mathbf{R}_1\}\, d\mathbf{R}_1 \times \\ \left[\int \hat{P}_\phi(\mathbf{R}_2) \exp\{-\mathrm{j}\mathbf{K}_2\cdot\mathbf{R}_2\}\, d\mathbf{R}_2\right]^* \end{array}\right\}. \tag{7.3}$$

Applying $\left(\int f\right)^* = \int f^*$ gives:

$$E\left\{\hat{F}_\phi(\mathbf{K}_1)\hat{F}_\phi^*(\mathbf{K}_2)\right\} = E\left\{\begin{array}{c}\displaystyle\int\int \hat{P}_\phi(\mathbf{R}_1)\hat{P}_\phi^*(\mathbf{R}_2)\times \\ \exp\left\{-\mathrm{j}\left(\mathbf{K}_1\cdot\mathbf{R}_1 - \mathbf{K}_2\cdot\mathbf{R}_2\right)\right\} d\mathbf{R}_1 d\mathbf{R}_2\end{array}\right\}. \tag{7.4}$$

Exchanging the order of integration and expectation yields:

$$\begin{aligned} E\left\{\hat{F}_\phi(\mathbf{K}_1)\hat{F}_\phi^*(\mathbf{K}_2)\right\} &= \int\int E\left\{\hat{P}_\phi(\mathbf{R}_1)\hat{P}_\phi^*(\mathbf{R}_2)\right\}\times \\ &\quad \exp\left\{-\mathrm{j}\left(\mathbf{K}_1\cdot\mathbf{R}_1 - \mathbf{K}_2\cdot\mathbf{R}_2\right)\right\} d\mathbf{R}_1 d\mathbf{R}_2, \end{aligned} \tag{7.5}$$

where the expectation is the ensemble average over many instances of $\hat{F}_\phi$. Making the substitution: $\mathbf{R}_1 = \mathbf{R}_2 - \mathbf{R}$, and applying the property in 7.1 above:

$$\begin{aligned} E\left\{\hat{F}_\phi(\mathbf{K}_1)\hat{F}_\phi^*(\mathbf{K}_2)\right\} &= \int \exp\left\{-\mathrm{j}\mathbf{R}_2\cdot\left(\mathbf{K}_1 - \mathbf{K}_2\right)\right\} d\mathbf{R}_2 \times \\ &\quad \left[\int B_{\hat{P}_\phi}(\mathbf{R})\exp\left\{-\mathrm{j}\mathbf{K}_1\cdot\mathbf{R}\right\} d\mathbf{R}\right], \qquad (7.6) \\ &= \delta\left(\mathbf{K}_1 - \mathbf{K}_2\right)(2\pi)^4\,\Phi_{P_\phi}(\mathbf{K}_1). \qquad (7.7) \end{aligned}$$

Thus, the correlation of the transform of the random process is related to the power spectrum. This relationship suggests that it is possible to generate a realization of $\hat{F}_\phi$ by filtering a complex white noise process, N, with the power spectrum. For instance, suppose the form $\hat{F}_\phi(\mathbf{K}) = (2\pi)^2\,\Phi_{P_\phi}^{\frac{1}{2}}(\mathbf{K})\,N(\mathbf{K})$. Inverse transforming $\hat{F}_\phi(\mathbf{K})$ would yield $\hat{P}_\phi$, an instance of P_ϕ. The question remains: what are the requirements for the noise process? The noise must have an appropriate mean and variance. Begin by solving for the mean:

$$E\left\{\hat{F}_\phi(\mathbf{K})\right\} = \int E\left\{\hat{P}_\phi(\mathbf{R})\right\}\exp\left\{-\mathrm{j}\mathbf{K}\cdot\mathbf{R}\right\} d\mathbf{R}. \tag{7.8}$$

The expected value of $\hat{P}_\phi$ is zero per the atmospheric model. Substituting the suggested form for $\hat{F}_\phi$ into the left hand side above and evaluating the expectation gives:

$$\begin{aligned} (2\pi)^2\,\Phi_{P_\phi}^{\frac{1}{2}}(\mathbf{K})\,E\left\{N(\mathbf{K})\right\} &= 0, \qquad (7.9) \\ E\left\{\mathrm{Re}\left\{N(\mathbf{K})\right\}\right\} + \mathrm{j}E\left\{\mathrm{Im}\left\{N(\mathbf{K})\right\}\right\} &= 0. \qquad (7.10) \end{aligned}$$

Thus, the real and imaginary parts of N must be zero mean. Now find the second moment by substituting the suggested $\hat{F}_\phi$ into (7.7) and simplifying:

$$\begin{aligned}
E\left\{\begin{array}{c} (2\pi)^2\,\Phi_{P_\phi}^{\frac{1}{2}}(\mathbf{K}_1)\,N(\mathbf{K}_1)\times \\ (2\pi)^2\,\Phi_{P_\phi}^{\frac{1}{2}}(\mathbf{K}_2)\,N^*(\mathbf{K}_2)\end{array}\right\} &= (2\pi)^4\,\delta(\mathbf{K}_1-\mathbf{K}_2)\,\Phi_{P_\phi}(\mathbf{K}_1)\,, && (7.11)\\
(2\pi)^2\,\Phi_{P_\phi}^{\frac{1}{2}}(\mathbf{K})\,(2\pi)^2\,\Phi_{P_\phi}^{\frac{1}{2}}(\mathbf{K})\,E\{N(\mathbf{K})N^*(\mathbf{K})\} &= (2\pi)^4\,\Phi_{P_\phi}(\mathbf{K})\,, && (7.12)\\
E\{N(\mathbf{K})N^*(\mathbf{K})\} &= 1. && (7.13)
\end{aligned}$$

Expanding N into its real and imaginary components and simplifying:

$$\begin{aligned}
E\left\{\begin{array}{c} (\mathrm{Re}\{N(\mathbf{K})\}+\mathrm{j}\,\mathrm{Im}\{N(\mathbf{K})\})\times \\ (\mathrm{Re}\{N(\mathbf{K})\}-\mathrm{j}\,\mathrm{Im}\{N(\mathbf{K})\})\end{array}\right\} &= 1, && (7.14)\\
E\left\{(\mathrm{Re}\{N(\mathbf{K})\})^2\right\}+E\left\{(\mathrm{Im}\{N(\mathbf{K})\})^2\right\} &= 1. && (7.15)
\end{aligned}$$

From (7.10) it is clear that N must be zero mean to create the zero mean process $\hat{P}_\phi$, however, according to (7.15) the variance of N depends on the combination of its real and imaginary parts. If $\hat{P}_\phi$ is a real process, then the real and imaginary parts of $\hat{F}_\phi$ must be considered independently. Implying that both the real and the imaginary parts must each have a variance of 1:

$$\begin{aligned}
E\left\{(\mathrm{Re}\{N(\mathbf{K})\})^2\right\} &= 1, && (7.16)\\
E\left\{(\mathrm{Im}\{N(\mathbf{K})\})^2\right\} &= 1. && (7.17)
\end{aligned}$$

Thus, the form for $\hat{F}_\phi$ is acceptable provided that the real and imaginary parts of N are zero mean, unit variance WSS processes. Combining these results, the inverse Fourier transform provides a method for creating realizations of the phase process as follows:

$$\begin{aligned}
\hat{P}_\phi(\mathbf{R}) &= \frac{1}{(2\pi)^2}\int (2\pi)^2\,\Phi_{P_\phi}^{\frac{1}{2}}(\mathbf{K})\,N(\mathbf{K})\exp\{\mathrm{j}\mathbf{K}\cdot\mathbf{R}\}\,d\mathbf{K}, && (7.18)\\
\hat{P}_\phi(\mathbf{R}) &= \int \Phi_{P_\phi}^{\frac{1}{2}}(\mathbf{K})\,N(\mathbf{K})\exp\{\mathrm{j}\mathbf{K}\cdot\mathbf{R}\}\,d\mathbf{K}, && (7.19)
\end{aligned}$$

Finally, separating $\hat{P}_\phi$ into real and imaginary components, the method produces two independent realizations:

$$\hat{P}_{\phi_1}(\mathbf{R}) = \mathrm{Re}\left\{\hat{P}_\phi(\mathbf{R})\right\}, \tag{7.20}$$

$$\hat{P}_{\phi_2}(\mathbf{R}) = \mathrm{Im}\left\{\hat{P}_\phi(\mathbf{R})\right\}, \tag{7.21}$$

where $\mathrm{Re}\{N(\mathbf{K})\}$, and $\mathrm{Im}\{N(\mathbf{K})\}$ are distributed $\mathcal{N}(0,1)$. This technique is acceptable where many random screens are needed, however, if only a single screen is required for simulation, it is common to take advantage of the spectral symmetry in real functions. Recall that for a real valued function, $x(t)$: $\mathcal{F}\{x(t)\} = \mathcal{F}^*\{x(t)\}$. Using this relationship, it is possible to create N as a Hermitian symmetric Gaussian noise process with real and imaginary parts distributed $\mathcal{N}\left(0, \frac{1}{\sqrt{2}}\right)$:

$$E\left\{(\mathrm{Re}\{N(\mathbf{K})\})^2\right\} + E\left\{(\mathrm{Im}\{N(\mathbf{K})\})^2\right\} = 1, \tag{7.22}$$

$$E\left\{(\mathrm{Re}\{N(\mathbf{K})\})^2\right\} = \frac{1}{2}, \tag{7.23}$$

$$E\left\{(\mathrm{Im}\{N(\mathbf{K})\})^2\right\} = \frac{1}{2}. \tag{7.24}$$

Hermitian symmetry in N guarantees that $\hat{P}_\phi$ is a real random process. The benefit of producing a single real screen is apparent when examining the reduction in computational complexity in the Fourier series approximation.

In a computer application, the continuous Fourier transform must be approximated by a finite Fourier series:

$$\hat{P}_\phi[\mathbf{R}] = \sum_{\mathbf{K}_i \in \mathcal{K}} N[\mathbf{K}_i] \left(\frac{\Phi_{P_\phi}[\mathbf{K}_i]}{\Delta[\mathbf{K}_i]}\right)^{\frac{1}{2}} \exp\{\mathrm{j}\mathbf{K}\cdot\mathbf{R}\}\,\Delta[\mathbf{K}_i], \tag{7.25}$$

$$\text{where } \mathcal{K} = \left\{\mathbf{K}_i : \Phi_{P_\phi}[\mathbf{K}_i]\,\Delta[\mathbf{K}_i] = \int_{\mathbf{\Omega}_i} \Phi_{P_\phi}(\boldsymbol{\xi})\, d\boldsymbol{\xi}\right\}, \tag{7.26}$$

$$\text{and } \Delta[\mathbf{K}_i] = \int_{\mathbf{\Omega}_i} d\boldsymbol{\xi}. \tag{7.27}$$

$\mathbf{\Omega}_i$ is a member of a set of bounded regions in $\mathbf{K}$ such that: $\bigcup \mathbf{\Omega}_i =$ the set of all $\mathbf{K}$, and $\bigcap \mathbf{\Omega}_i = \varnothing$. The substitution of brackets, $[\cdot]$, for parentheses, $(\cdot)$, indicates that continuous valued functions are being evaluated at a set of discrete locations. The Fourier transform has been approximated by a Fourier series evaluated at discrete locations $\mathbf{K}_i$ within the domain $\mathbf{K}$. The domain $\mathbf{K}$ has been divided into a countably infinite set of regions $\mathbf{\Omega}_i$. A single location $\mathbf{K}_i$ within each region in $\mathbf{\Omega}_i$ is chosen such that the integrated power in the region $\mathbf{\Omega}_i$ equals the spectral density function evaluated at that location multiplied by the area of the region. The set $\mathcal{K}$ is the set of discrete $\mathbf{K}_i$ locations to be included in the series approximation. The function $\Delta[\mathbf{K}_i]$ represents the $\mathbf{K}$ domain area included in each region $\mathbf{\Omega}_i$. For the case where Hermitian symmetry is enforced in the complex coefficients N, the expression simplifies to:

$$\hat{P}_\phi[\mathbf{R}] = \sum_{\mathbf{K}_i \in \mathcal{K}'} 2\Delta^{\frac{1}{2}}[\mathbf{K}_i]\,\Phi_{P_\phi}^{\frac{1}{2}}[\mathbf{K}_i] \begin{bmatrix} \mathrm{Re}\{N[\mathbf{K}_i]\}\cos(\mathbf{K}_i \cdot \mathbf{R}) - \\ \mathrm{Im}\{N[\mathbf{K}_i]\}\sin(\mathbf{K}_i \cdot \mathbf{R}) \end{bmatrix}, \tag{7.28}$$

where the set $\mathcal{K}'$ includes a unique half of the Hermitian symmetric set $\mathcal{K}$. Generating a real $\hat{P}_\phi$ requires half the number of computations expended in creating a complex $\hat{P}_\phi$.

If $\hat{P}_\phi$ is to be generated on a set of points where the x and y locations are known in advance, then the trigonometric function evaluations can be precomputed and stored to speed up final evaluation of the Fourier kernel. Expressing the $\mathbf{K}$ domain frequencies into x and y components, the independent kernel computations for a complex $\hat{P}_\phi$ are:

$$\exp\{\mathrm{j}\mathbf{K}\cdot\mathbf{R}\} = \exp\{\mathrm{j}K_xR_x\} \times \exp\{\mathrm{j}K_yR_y\}. \tag{7.29}$$

The kernel components for a real $\hat{P}_\phi$ are:

$$\begin{aligned} \begin{matrix} \mathrm{Re}\{N(\mathbf{K})\}\cos(\mathbf{K}\cdot\mathbf{R}) - \\ \mathrm{Im}\{N(\mathbf{K})\}\sin(\mathbf{K}\cdot\mathbf{R}) \end{matrix} &= \cos(K_xR_x)\begin{pmatrix} \mathrm{Re}\{N(\mathbf{K})\}\cos(K_yR_y) - \\ \mathrm{Im}\{N(\mathbf{K})\}\sin(K_yR_y) \end{pmatrix} - \\ &\quad \sin(K_xR_x)\begin{pmatrix} \mathrm{Re}\{N(\mathbf{K})\}\sin(K_yR_y) + \\ \mathrm{Im}\{N(\mathbf{K})\}\cos(K_yR_y) \end{pmatrix}. \end{aligned} \tag{7.30}$$

7.2 Improving Isotropy and Reducing Kernel Size

Given either formula for generating $\hat{P}_\phi$, the phase screen implementation begins by selecting a finite set of discrete frequency locations to be included in $\mathcal{K}$. The finite set of boundary regions $\mathbf{\Omega}_i$ will not include the entire domain $\mathbf{K}$. The domain $\mathbf{K}$ is instead truncated such that the bulk of the power spectrum is represented in the series approximation. Nyquist sampling requires that the $\mathbf{K}$ domain is bounded by a maximum frequency dictated by the minimum spatial separation in the screen. The spatial resolution is in turn selected to adequately represent the chosen spectrum or to satisfy the requirements of a given simulation, whichever condition is most restrictive. Assume that spectral representation is the driving requirement. In the case of a von Kármán spectrum it is important to adequately represent spectral content within the inertial range. With this requirement in mind, the minimum spatial dimension should be linked to the inner scale, l_0.

$$\begin{aligned} \Delta x &\leq \frac{l_0}{3}, && (7.31) \\ f_{\max} &= \frac{1}{2\Delta x} \geq \frac{3}{2l_0}, && (7.32) \\ \kappa_{\max} &= \frac{\pi}{\Delta x} \geq \frac{3\pi}{l_0} = \frac{3\pi\kappa_m}{5.92}, && (7.33) \\ \kappa &\equiv |\mathbf{K}|\,. && (7.34) \end{aligned}$$

The choice of $\Delta x \leq \frac{l_0}{3}$ ensures that the high frequency cut off lies beyond the inertial range. Recall that beyond κ_m the von Kármán spectrum decreases exponentially, therefore, the bulk of the spectral power occurs below κ_m. Often, a minimum frequency is also specified. This is necessary when the power spectrum is not absolutely integrable, as with Kolmogorov, or when simulation restrictions require. If the von Kármán spectrum is truncated, then the minimum frequency should be related to the outer scale:

$$\begin{aligned} f_{\min} &\leq \frac{1}{10L_0}, && (7.35) \\ \kappa_{\min} &\leq \frac{2\pi}{10L_0} = \frac{\kappa_0}{10}. && (7.36) \end{aligned}$$

Choosing $f_{\min} \leq \frac{1}{10L_0}$ will ensure that $f_{\min}$ is below the inertial range. Once the representative region in $\mathbf{K}$ is established, a method is required for delineating the disjoint regions $\mathbf{\Omega}_i$. If the Fourier series coefficients are evaluated on an equally spaced Cartesian grid to take advantage of Fast Fourier Transform (FFT) algorithms, the grid size becomes computationally prohibitive. For example, the $\mathbf{K}$ grid for the case: $L_0 = 10\text{m}$ and $l_0 = 0.01\text{m}$

		# of **K** locations	
		$\hat{P}_\phi$ real	$\hat{P}_\phi$ complex
Q	$\sqrt{2}$	420	840
	2	760	1520
	4	3120	6240

Table 7.1 The number of **K** grid locations for a given Q: $L_0 = 10\text{m}$ and $l_0 = 0.01\text{m}$

would contain 30000^2 locations. Reducing this number by half for the real versus complex implementation still remains too cumbersome.

Taking advantage of the exponentially decaying power spectrum, it is possible to compress the frequency domain information using a logarithmically spaced Cartesian grid [57]. In this case, the logarithm of the **K** boundaries are equally spaced. This is sometimes referred to as a constant Q frequency spacing:

$$Q = \frac{\kappa}{\Delta\kappa}, \tag{7.37}$$

$$\kappa \equiv \text{sample point}, \tag{7.38}$$

$$\Delta\kappa \equiv \text{sample bandwidth}. \tag{7.39}$$

The sizes of the resulting **K** grids for several Q values are listed in Table 7.1. The tabled grid sizes are many orders of magnitude smaller than the grid size required for equal spaced sampling. The reduction in computational complexity due to logarithmic frequency compression is overwhelming. Of course, these numbers are meaningless if the frequency domain compression leads to unsatisfactory statistical error in the screen realizations. The following section provides a demonstration of structure function percent error and error in lower order Zernike variance for an implementation using these Q values. Before reviewing the performance of the log-Cartesian sampling method, however, I will suggest another sampling scheme.

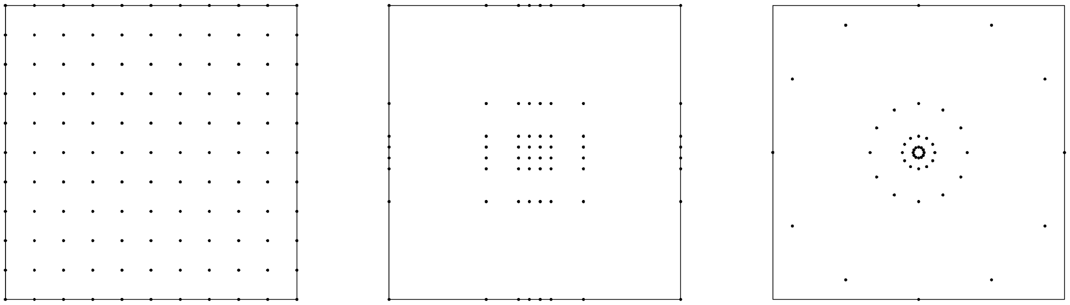


Figure 7.1 (Left) An example of the equispaced Cartesian sample structure. (Center) Log-Cartesian sampling. (Right) Log-polar sampling.

The final form for the phase spectrum Φ_{P_ϕ} is a function of the magnitude κ rather than $\mathbf{K}$. In an effort to take advantage of the radially symmetric nature of the power spectrum, I suggest evaluating the Fourier series over a log-polar grid. On a log-polar grid, the radial or κ axis is sampled logarithmically while the angular axis ψ is sampled equally. This pattern of sampling offers a few advantages. Symmetric spacing of samples in the angular dimension reduces anisotropy in the resulting screen realizations. Φ_{P_ϕ} is symmetric in the angular dimension within each sample region. This symmetry reduces the integral required to find appropriate entries in $\mathcal{K}$ to a single dimension, κ:

$$\mathbf{K}_i = (\kappa_i, \psi_i)\,, \tag{7.40}$$

where the ψ_i's are chosen by equally spacing an arbitrary number of ψ locations from 0 to 2π radians at each distinct κ entry and the κ_i's are selected to satisfy:

$$\begin{aligned}\Phi_{P_\phi}[\kappa_i]\,\Delta\kappa_i &= \int_{\Omega_i} \Phi_{P_\phi}(\xi)\, d\xi, & (7.41)\\ \text{and } \Delta\kappa_i &= \int_{\Omega_i} d\xi. & (7.42)\end{aligned}$$

Ω_i (not bold so as to be distinguished from $\boldsymbol{\Omega}_i$) is a member of a set of one-dimensional bounded regions in κ such that: $\bigcup \Omega_i =$ the set of all κ, and $\bigcap \Omega_i = \varnothing$. This same symmetry reduces the number of unique power, Φ_{P_ϕ}, and area, Δ, calculations. Finally, using the polar sample structure makes the choice of sample density along κ independent of the number of samples in ψ. This is opposed to log-Cartesian sampling where the choice of Q dictates the density of samples in both spectral dimensions. In this way, log-polar sampling provides increased flexibility over the log-Cartesian structure.

7.3 Comparing Sampling Methods

Performance results using both log-Cartesian and log-polar frequency sampling are presented here for comparison. For instance, Figures 7.2 and 7.3 demonstrate percent structure function error at $Q = \sqrt{2}$ for log-Cartesian and log-polar sampling respectively. Each simulation run includes the following set of standardized inputs: $r_0 = 0.088\text{m}$, $L_0 = 10\text{m}$, $l_0 = 0.01\text{m}$, $\kappa_{\min} = \frac{2\pi}{10L_0}$, and $\kappa_{\max} = \frac{3\pi}{l_0}$. Structure function percent error is evaluated at 1000 logarithmically spaced sample points in $|\mathbf{R}|$. To demonstrate isotropy, the direction of each vector of separation, $\mathbf{R}$, is also evaluated at 5° increments in direction from 0° to 45°. The solid plot lines indicate the minimum and maximum structure percent error over the range of directions. The + symbols indicate the spatial distances corresponding to the set of discrete locations $\mathcal{K}$ along the $K_y = 0$ axis. To keep the number of kernel samples consistent between the two methods of sampling, the number of samples in each concentric ring of the log-polar grid is equal to the number of samples in each concentric square in the log-Cartesian grid. The exception is the log-polar case for $Q = 4$ in Figure 7.7. For this case, I have elected to make the number of ψ samples constant for all κ rings at 24 (increments of 15° in ψ). Organizing the sampling grid in this manner decreases the total number of grid points from 3120 to 468. A comparison of the results in Figures 7.6 and 7.7 demonstrates that the log-polar technique can provide reduced structure error and improved isotropy even after an 85% reduction in the number of kernel grid points.

In addition to limiting structure function error, maintaining accurate measure of error in lower order Zernike mode variances is important to the sensor simulation. The accuracy of the Zernike variance within the set of estimated parameters can affect the performance of the estimator by way of introducing error in the environment variable estimates. Table 7.2 contains Zernike coefficient variance results estimated from $100,000$ screen realizations. For this simulation, I maintained the same set of input parameters: $r_0 = 0.088\text{m}$, $L_0 = 10\text{m}$, $l_0 = 0.01\text{m}$, $\kappa_{\min} = \frac{2\pi}{10L_0}$, $\kappa_{\max} = \frac{3\pi}{l_0}$, and added the diameter of a subaperture, $D = 0.07\text{m}$.

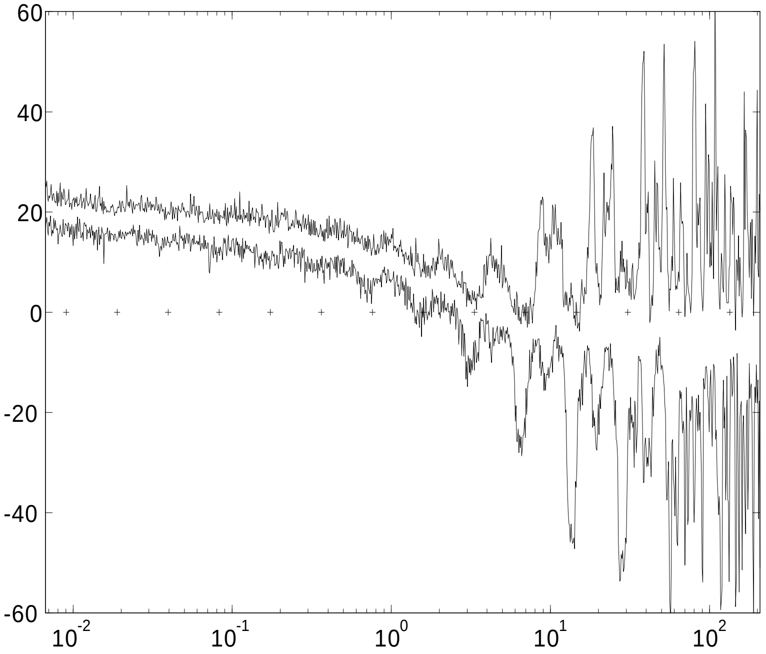


Figure 7.2 Structure function percent error versus separation distance for log-Cartesian sampling with $Q = \sqrt{2}$.

Zernike coefficient:		$a_{2,3}$		a_4		$a_{5,6}$		$a_{7,8}$		$a_{9,10}$		a_{11}	
log sampling method:		cart.	polar	cart.	polar	cart.	polar	cart.	polar	cart.	polar	cart.	polar
Q	$\sqrt{2}$	+16	+8	+21	+4	+21	+4	+8	+6	+11	+9	+27	+39
	2	+6	+3	+10	+5	+10	+4	+5	+3	+8	+5	+11	+12
	4	0	0	+3	+2	+3	+2	+1	-1	+3	+2	-1	-3
	6	-1	0	+2	+3	+1	+1	-2	-1	+2	+2	-2	-2
	8	-2	0	+2	+2	+1	+1	-1	-1	+1	+1	-2	-2

Table 7.2 Percent error in Zernike coefficient variance per varying Q value

Zernike coefficient:	$a_{2,3}$	a_4	$a_{5,6}$	$a_{7,8}$	$a_{9,10}$	a_{11}
$Q = 4$, $\Delta\psi = 15°$	0	+3	+2	-1	+1	-2

Table 7.3 Percent error in Zernike coefficient variance.

Table 7.3 is provided to document Zernike variance for the case where ψ sampling is constant. Recall the case: $Q = 4$ with 24 ψ samples in each κ ring. Table 7.3 shows that, even after reducing the total number of sample points from 3120 to 468, the screen variance statistics yield similar error within the estimated Zernike set.

Reducing the size of the kernel offers an advantage in screen generation time, however, the compression in frequency, while not evident in the average structure or Zernike basis error, becomes evident from visual inspection of individual phase screens. Upon inspection, screens generated using logarithmic compression (of any sampling scheme) exhibit visible patterns as a result of the decreased number of frequency modes. Figures 7.8 and 7.9 offer examples of screens created using log Cartesian and log polar frequency sample grids respectively. In these example screens, the frequency density parameter Q is set to $\sqrt{2}$ resulting in a kernel size of 420 samples. The third example screen (Figure 7.10) combines $Q = 3.6$ and $\Delta\psi = 15°$ using log polar sampling in order to get a kernel size of 420 samples. Each of these examples demonstrates the described pattern effects. This gallery of examples is provided merely to document that the compression becomes visible in the output screens and not to down play the use of log frequency compression techniques. The measures of quality are statistical based rather than based on visual aesthetics. These examples may exhibit visible compression effects, but they guarantee a specific statistical quality threshold.

7.4 Summary

The Fourier series method for random process generation is ideal for creating the type of phase realizations required in thin screen, frozen flow atmospheric simulations. The discussion above included a review of the mathematical background for creating random process realizations using the Fourier series technique. Applying the Fourier series technique to phase screen generation revealed the need to trade reduced accuracy in the Fourier series approximation for reduced computational complexity. The frequency domain sample grid becomes too large when formed using an equally spaced Cartesian sampling structure. This difficulty exposed the importance of using some method of compression in frequency. Constant Q frequency sampling increases sampling density toward lower frequencies and, as such, provides an ideal sample structure for compressing decreasing power law functions. It has been shown that log-Cartesian grids offer an acceptable method for compressing atmospheric spectral models and effectively reducing the complexity in the Fourier series calculations for phase screen generation [57], [14]. I have shown that log-polar sampling further reduces the number of frequency grid points required for some maximum percent error in the structure function while increasing isotropy in the screen realizations. Log-polar sampling also introduces symmetry in the sample structure, which when combined with symmetry in the spectrum, reduces the complexity of calculating the set of frequency domain sample points. These attributes make the log-polar sampled Fourier series the method of choice for the core of the atmospheric simulation.

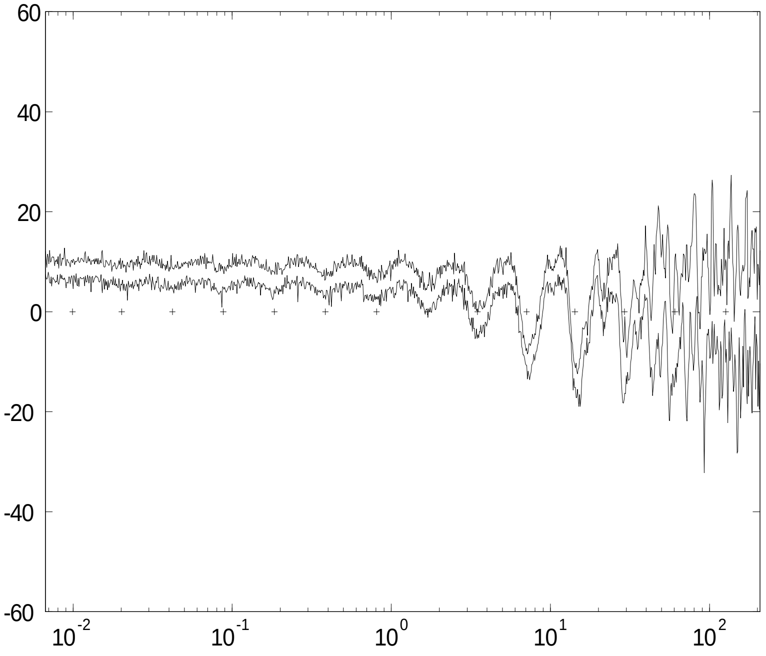


Figure 7.3 Structure function percent error versus separation distance for log-polar sampling with $Q = \sqrt{2}$.

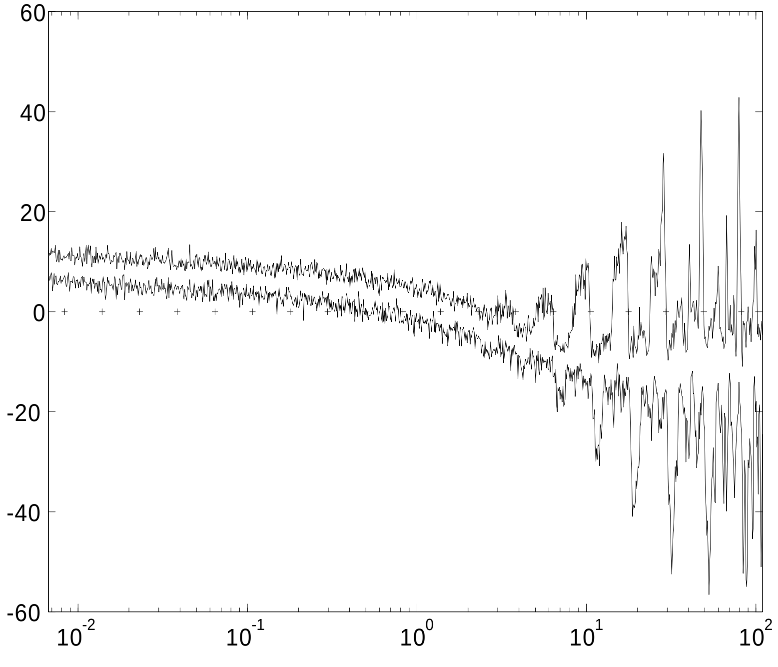


Figure 7.4 Structure function percent error versus separation distance for log-Cartesian sampling with $Q = 2$.

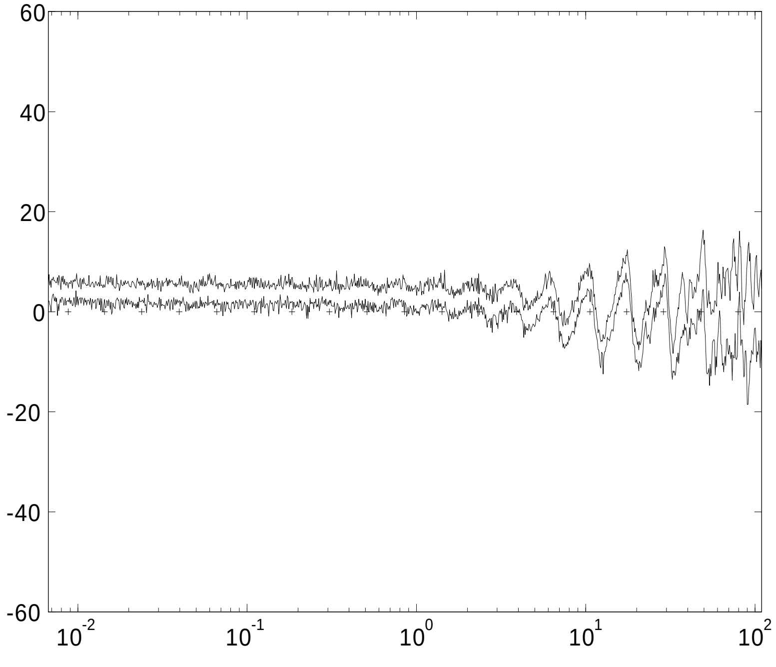


Figure 7.5 Structure function percent error versus separation distance for log-polar sampling with $Q = 2$.

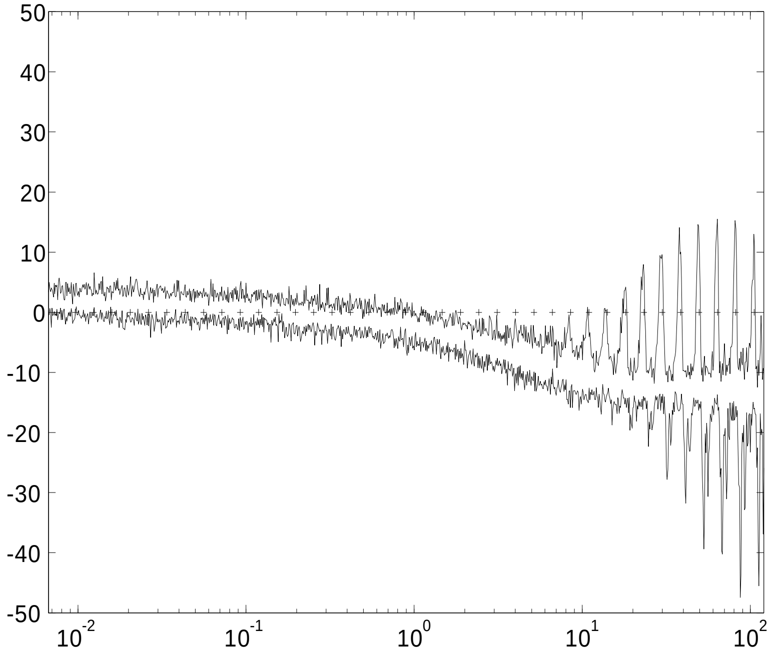


Figure 7.6 Structure function percent error versus separation distance for log-Cartesian sampling with $Q = 4$.

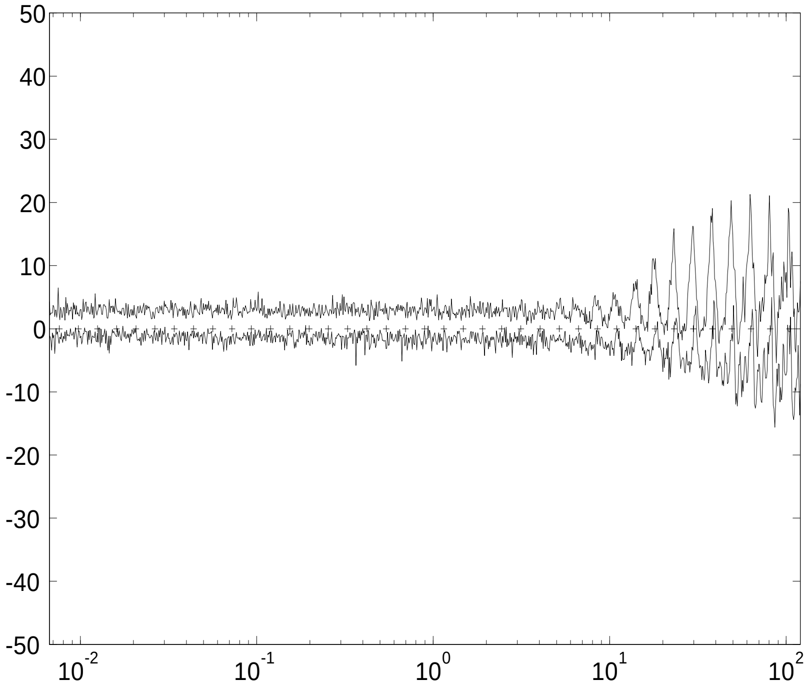


Figure 7.7 Structure function percent error versus separation distance for log-polar sampling with $Q = 4$ and 24 equal spaced ψ samples in each concentric κ band.

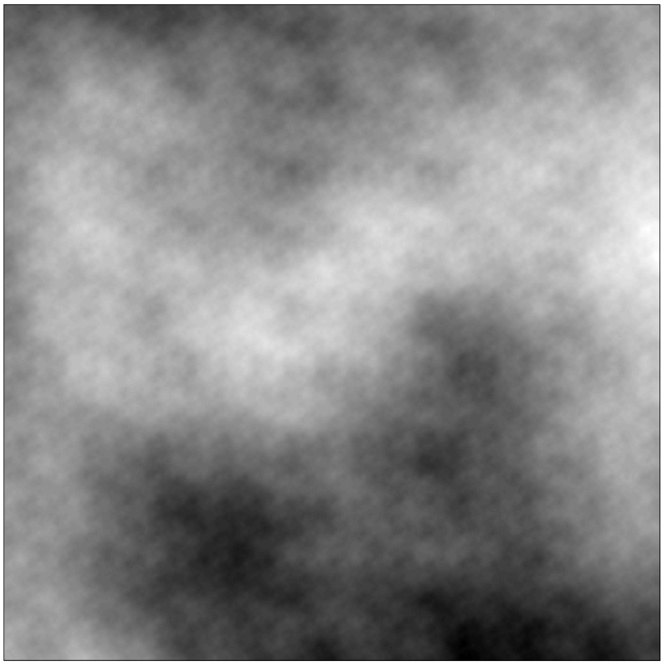


Figure 7.8 Example 1024×1024 pixel phase screen created using log-Cartesian frequency sampling: $Q = \sqrt{2}$, $r_0 = 0.088$m, $L0 = 10$m, $l0 = 0.01$m, $\Delta x = 0.0032$m.

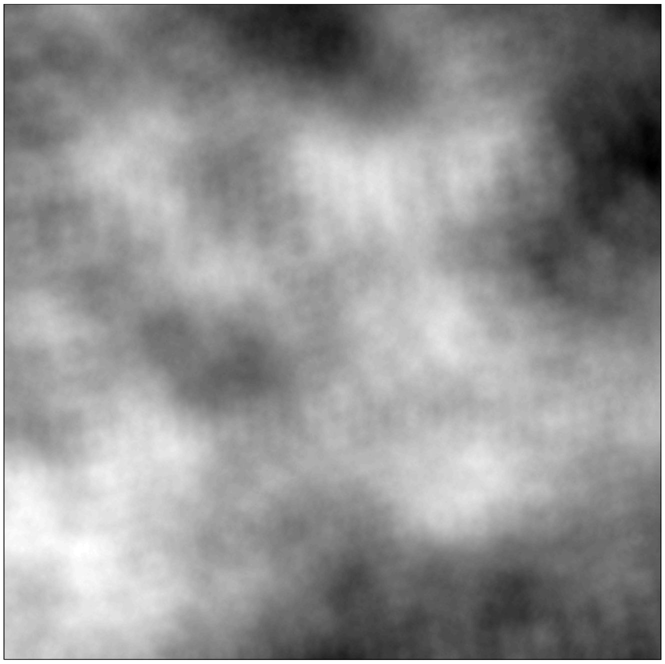


Figure 7.9 Example 1024×1024 pixel phase screen created using log-polar frequency sampling: $Q = \sqrt{2}$, $r_0 = 0.088$m, $L0 = 10$m, $l0 = 0.01$m, $\Delta x = 0.0032$m.

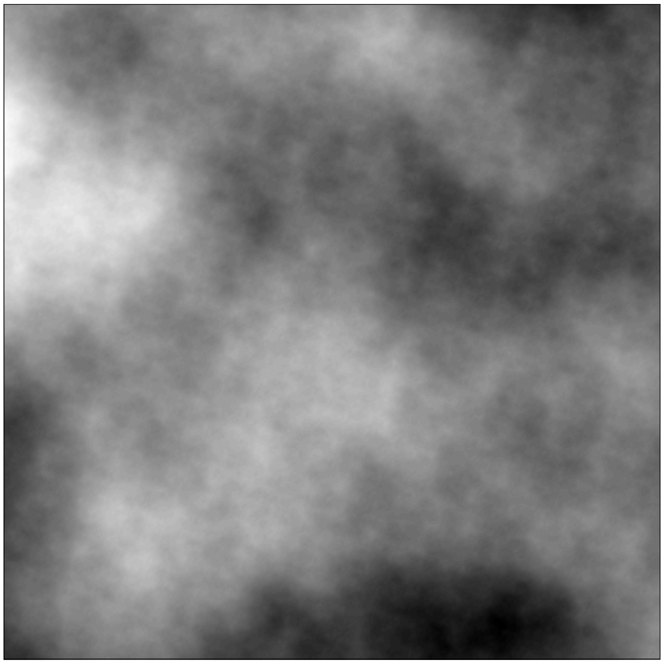


Figure 7.10 Example 1024 × 1024 pixel phase screen created using log-polar frequency sampling: $Q = 3.6$, $\Delta\psi = 15°$.

8. Simulating the Z_{2-4} Wavefront Sensor

The sections to follow provide a brief description of the simulation techniques used and the results of simulated Z_{2-4} sensor performance. The purpose of the simulation is to provide a proof of concept for the sensor design and, as such, should include sufficient rigor to provide insight into whether the sensor algorithm is viable and worthy of further research. The results to follow demonstrate that the Z_{2-4} sensor outperforms ML and centroiding techniques for point source wavefront sensing.

8.1 Constructing the Simulation

The simulation is divided into four segments: the source, the atmosphere, the optical system and the sensor algorithm. The diagram in Figure 8.1 shows these high level segments. With the exception of the sensor algorithm, these segments are intended to model the most significant environmental effects on sensor performance. The algorithm itself is meant to mimic implementation as precise as possible with a computer simulation. In the following sections, I will present each of the segments as they exist for this simulation and discuss any associated assumptions.

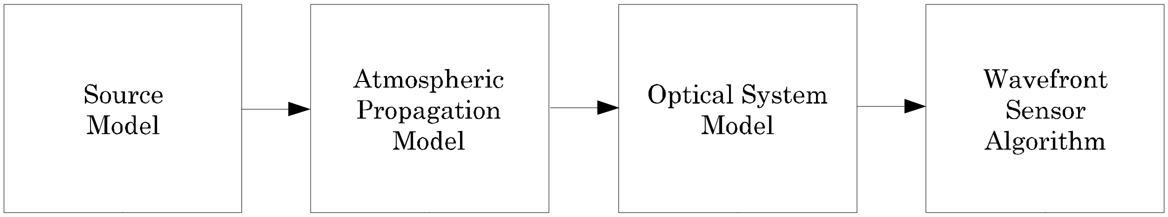


Figure 8.1 Simulation block diagram.

Source and Atmospheric Propagation Models. The source is assumed to be in the far field and is modeled as a plane wave input into the atmospheric model. It is assumed that some object (distant star or beacon) is emitting incoherent light and is spatially unresolvable by the optical system. Assume also that the optical system passes some quasimonochromatic band which will be represented by the center wavelength λ. Signal to noise ratio is regulated within the optical system image plane and therefore the amplitude of the source is arbitrary. The second segment in the simulation implements the atmospheric model derived in Section 2.2. The atmospheric simulation replaces the column of atmosphere between the source and the optical system with a thin phase screen. Within the model, light propagates in a vacuum between the source and the screen which is located at the optical system aperture. The log polar Fourier series method discussed in Chapter 7 is used to generate phase screens with atmospheric variables: r_0, L_0, and l_0, and input parameters: $Q = 4$ and $\Delta\psi = 5°$.

The Optical System. The simulated optical system is assumed to be aberration free. The only aberration in the system is the known defocus diversity required by the wavefront sensor algorithm. The optical system model begins with a discretized aperture. Modeling a circular aperture on the Cartesian grid requires some approximation. The projection operation for calculating Zernike coefficients given in Section 2.3 must be discretized:

$$a_i \equiv \int d\boldsymbol{\rho} W_Z(\boldsymbol{\rho}) Z_i(\boldsymbol{\rho}) P_\psi\left(\rho R_P, \theta\right), \tag{8.1}$$

$$a_i = \frac{D_P^2}{N^2} \sum_{\mathbf{n}} W_Z\left[\mathbf{n}\right] Z_i\left[\mathbf{n}\right] P_\psi\left[\mathbf{n}\right]. \tag{8.2}$$

Here, $W_Z\left[\mathbf{n}\right]$, represents the discrete circular weighting function. If the integrated value of the mask used to represent $W_Z\left[\mathbf{n}\right]$ is not exactly 1, the error will affect Zernike coefficient calculations. For this reason, the pixels intersected by the edge of the circular mask are carefully weighted by calculating the area of the trapezoidal and chordal regions indicated in Figure 8.2. Propagation from the pupil to the image plane is performed via the linear systems approach using a scaled Discrete Fourier Transform:

$$\mathcal{I}\left[\mathbf{u}; \mathbf{A}\right] = \frac{\Delta x \Delta y}{\left(\lambda s_i\right)^2} \sum_{\mathbf{n}} \mathcal{P}\left[\mathbf{n}; \mathbf{A}\right] \exp\left\{-\mathrm{j}\frac{2\pi}{N}\left[\mathbf{n}\cdot\mathbf{u}\right]\right\}, \tag{8.3}$$

$$= \frac{\Delta x \Delta y}{\left(\lambda s_i\right)^2} \mathcal{DFT}\{\mathcal{P}\left[\mathbf{n}; \mathbf{A}\right]\}, \tag{8.4}$$

$$\mathbf{I}\left[\mathbf{u}; \mathbf{A}\right] = K \frac{\left|\mathcal{I}\left[\mathbf{u}; \mathbf{A}\right]\right|^2}{\sum_{\mathbf{u}} \left|\mathcal{I}\left[\mathbf{u}; \mathbf{A}\right]\right|^2}, \tag{8.5}$$

$$= \tilde{K}\left|\mathcal{I}\left[\mathbf{u}; \mathbf{A}\right]\right|^2. \tag{8.6}$$

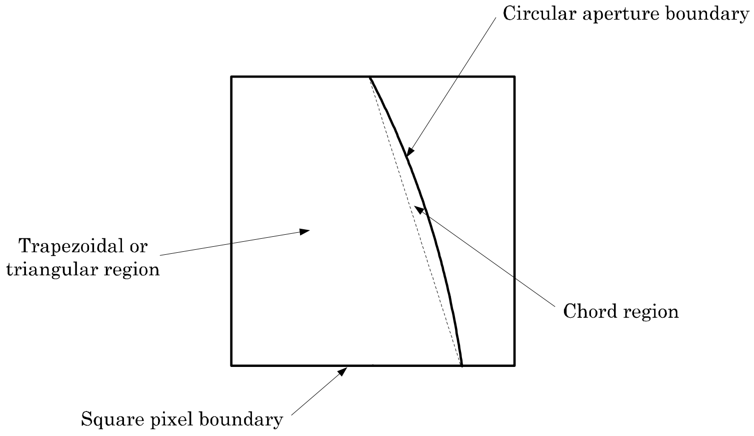


Figure 8.2 Diagram of a single pixel bisected by a circular arc near the perimeter of a circular aperture placed over a Cartesian grid.

The scale factor $\tilde{K}$ accounts for SNR in the CCD array by forcing the total intensity in the image plane to equal some average photon count K. Note that the average photon count, K, is a per subaperture count and, as such, must be divided among the total number of image planes and scaled by the efficiency of the beam splitting device. The D_P diameter circular aperture is inscribed in an $N \times N$ sample aperture grid and the relationship between the sample dimensions in the aperture plane to the image plane grid is: $[\Delta u, \Delta v] = \left[\Delta n \frac{\lambda f}{2D_P}, \Delta m \frac{\lambda f}{2D_P}\right]$. Due to its efficiency, the Fast Fourier Transform (FFT) is used to compute the DFT operation. The FFT requires that the aperture plane and image plane be composed of the same size grid. For this reason, the aperture plane is zero padded to create a $2N \times 2N$ grid prior to the FFT operation.

It is advantageous to position the optical axis at the center of the windowed region in the image plane. Symmetry in positive and negative tilt effects, for instance, is best utilized when the projection window is split evenly along the optical axis. For this reason, the optical axis is centered within the $[0, 0]$ grid location when the image plane window length is odd. When the window length is even, the optical axis is located at $[-0.5, -0.5]$. For the case of Nyquist pixel sizing, the half pixel shift in the optical axis is accomplished by inserting $\frac{-\pi}{8}$ of artificial x and y-tilt in the aperture. Recall the discussion in Section 4.3 and substitute $-1/2$ pixels for Δu in (4.33):

$$A_2 = \frac{-\frac{1}{2} R_P \Delta\xi}{2f}. \tag{8.7}$$

Substituting the Nyquist sampled image plane pixel, $\Delta\xi = \frac{\lambda f}{2D_P}$, yields:

$$\begin{aligned} A_2 &= \frac{-\frac{1}{2}R_P\frac{\lambda f}{2D_P}}{2f}, \\ A_2 &= \frac{-\lambda}{16}. \end{aligned}$$

Converting A_2 to units of radians reveals that an input of $\frac{-\pi}{8}$ radians of tilt is required to shift the image by -0.5 pixels. Examples of the zero padded aperture mask and diffraction limited point spread functions for even and odd length projection windows are provided in Figure 8.3.

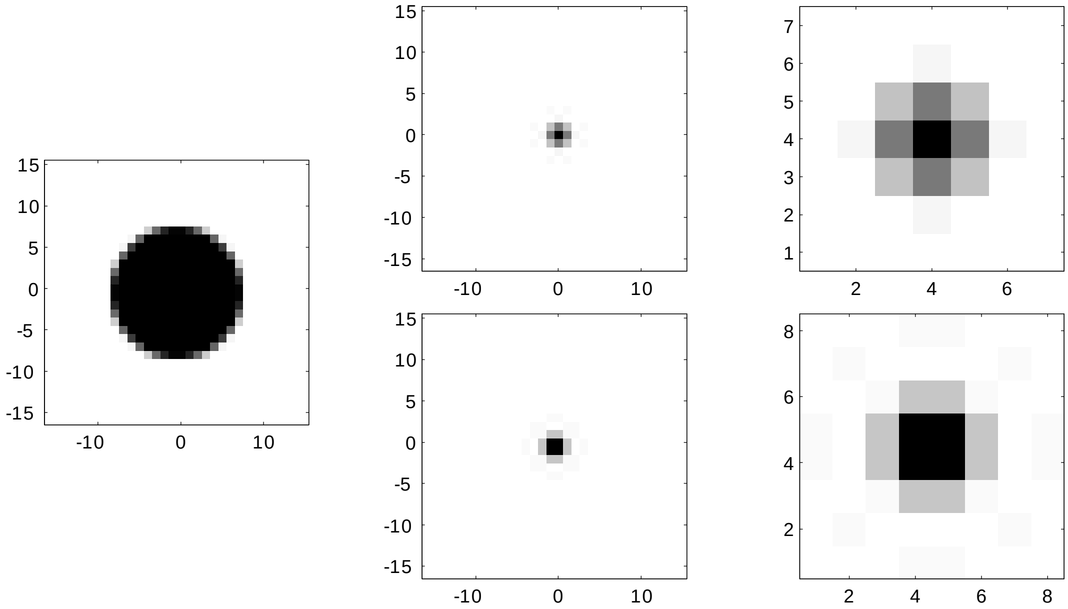


Figure 8.3 (Left) Zero padded aperture mask $W_Z[\mathbf{n}]$. (Center) Diffraction limited PSF: entire image plane after performing FFT. (Right) Diffraction limited PSF: windowed image plane. (Top) Odd N_W. (Bottom) Even N_W.

The expected image $\mathbf{I}$ is combined with noise in the CCD. The details of the CCD noise model are similar to Cain's tilt estimator analysis [1]. CCD noise may be categorized as either signal dependent or independent. Signal dependent noise includes all random light-matter interaction and is modeled as independent Poisson random processes within each detector pixel. This noise term, commonly referred to as shot noise or photon noise, becomes the dominant noise contribution under high light conditions. There are many contributors to signal independent noise. In this simulation, signal independent noise will be comprised of read out noise and A/D conversion noise. Signal independent noise becomes the dominant noise contribution under low SNR conditions. The read out noise term is meant to include effects such as clock noise, camera noise, and amplifier buffer noise. The A/D conversion noise is the result of a scaling and flooring operation. Assuming that the voltage step size in the A/D conversion process is equal to the voltage associated with a single photon detection, the scaling factor is unity and the A/D conversion is modeled as a simple flooring operation. Finally, the detected signal is biased with the read out noise variance as suggested by the discrete model in (3.67) and any negative counts are set to 0. Condensing this description into a convenient mathematical form, the detected image projection may be described by:

$$\begin{aligned}
\mathbf{v}\left(\mathbf{d}_{\cup}\right) &= \max\left(0, \mathrm{floor}\left[\mathbf{v}\left(\mathrm{Poisson}\left\{\mathbf{I}_{\cup}\right\}\right) + \mathbf{n}_{ro}\right] + \boldsymbol{\sigma}_{ro}\right), && (8.8)\\
\text{where } \mathbf{n}_{ro} &\equiv N_{\mathrm{v}} N_W \text{ length vector of } n_{ro} \text{ noise}, && (8.9)\\
n_{ro} &\equiv \mathcal{N}\left(0, \sigma_{ro}\right), && (8.10)\\
\text{and } \boldsymbol{\sigma}_{ro} &\equiv N_{\mathrm{v}} N_W \text{ length vector of } \sigma_{ro}. && (8.11)
\end{aligned}$$

The Sensor Algorithm. The sensor algorithm is implemented just as it would be in an embedded application with the exception of running in the Matlab compiler environment. In essence, the only simulated portion of the algorithm is its interface with the CCD data. All input data are generated from the models described in previous sections. Strategies for evaluation of the likelihood metric and maximizing the likelihood were described in Chapter 5. All parameter estimates are calculated in series in the computer simulation. However, a real system could compute the tilt estimates in parallel. Such a system would be more complex, but would reduce computation time and increase the sensor bandwidth.

8.2 Sensor Performance

The performance plots in this section serve to qualify the use of minimum CRLB as a basis for design variable selection and to compare the performance of the projection curvature sensor to the projection based ML tilt sensor and the more common two-dimensional centroid based tilt sensor. I will begin by presenting a series of simulated performance plots with overlaid CRLB results from Section 6.3, Figures 6.2 through 6.7. Note that, while the performance plots here should trend the same as the CRLB plots, the simulation results will differ from CRLB results for two reasons. The first reason is that the estimator algorithm is not efficient. This means that the variance of the estimator algorithm will approach the lower bound but will not equal the lower bound. The second reason is that the CRLB calculations use Zernike based phase screens to model atmospheric phase turbulence whereas the simulations represented here use polar sampled Fourier series phase screens. The polar sampled phase screens contain higher order phase information which tends to increase the variance of a low order parameter estimator like the one used in the Z_{2-4} sensor.

Figure 8.4 contains a plot of simulated residual MSE versus projection separation angle. The following design variable and environment variable settings were used: $D_P = 0.07\text{m}$, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05\text{m}$, $L_0 = 10\text{m}$, $l_0 = 0.01\text{m}$, $K = 1000$ photons, $\pm\delta_{a_4} = 0.4$ radians, and $N_W = 14$ pixels. The minimum residual error occurs when the separation angle, $\theta_2 - \theta_1$, equals 90 degrees. Based on this result, the remaining simulation plots will contain simulated performance examples with a projection separation angle of 90 degrees. The CRLB result for whole plane projection separation angle is included with an independent y-axis. The CRLB plot line is the dashed line and corresponds to the y-axis on the right hand side of the plot. Notice that although the y-axis scaling indicates different MSE ranges, the separation angle at which the simulation minimum occurs equals the separation angle at the CRLB minimum point. Figure 8.5 demonstrates the performance versus projection angle θ_1. CRLB continues to be plotted as a dashed line scaled to the right hand side y-axis for Figures 8.5 through 8.9. This example suggests that performance is effectively invariant to starting projection angle provided that the separation angle is 90 degrees. A cursory sampling over a two-dimensional range of start angles and separation angles reveals that the ideal separation angle is ±90 degrees and that performance is effectively invariant to starting angle under any fixed separation angle. Due to this result, the projection angle configuration for the remaining performance examples will be limited to cases where $\theta_1 = 0°$ and $\theta_2 = 90°$. Figure 8.6 contains a plot of residual wavefront MSE versus defocus diversity for a high SNR case: $D_P = 0.07\text{m}$, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05\text{m}$, $L_0 = 10\text{m}$, $l_0 = 0.01\text{m}$, $K = 1000$ photons, and $N_W = 7$ pixels. The ideal diversity (minimum point) is approximately the same as the CRLB minimum: CRLB minimum occurs at about 0.375 radians versus the performance minimum at about 0.4 radians. Figure 8.7

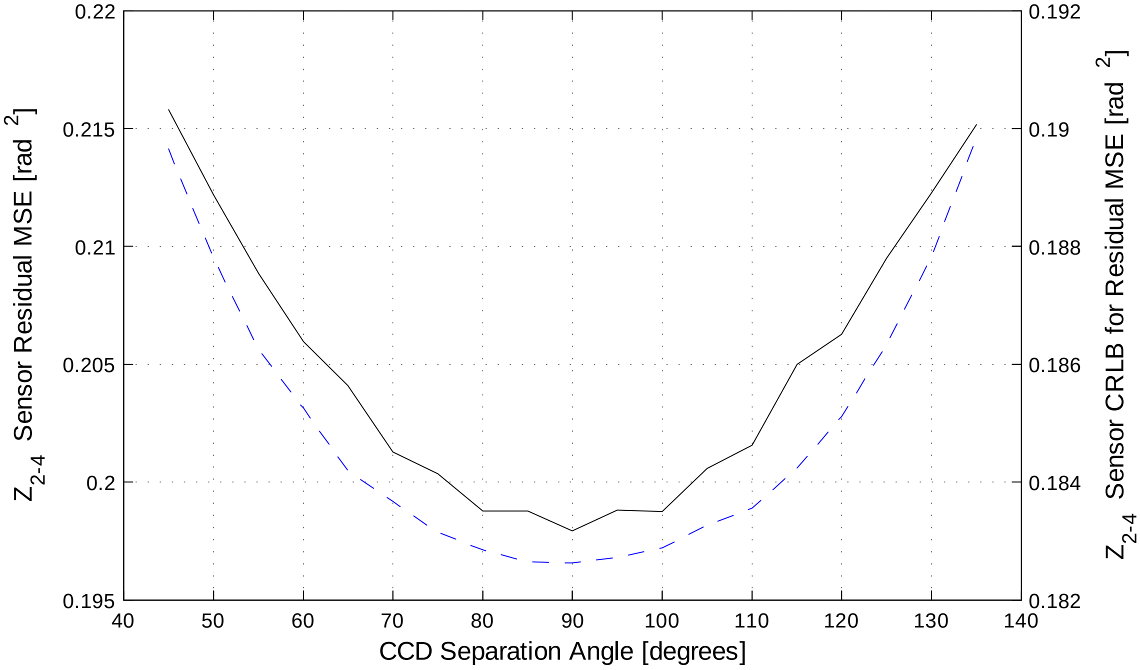


Figure 8.4 Z_{2-4} estimator $\left\langle P^2_{\phi_e} \right\rangle$ versus separation angle, $\theta_2 - \theta_1$. $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, $\sigma_{ro} = 2.13$ counts, and $N_W = 14$ pixels.

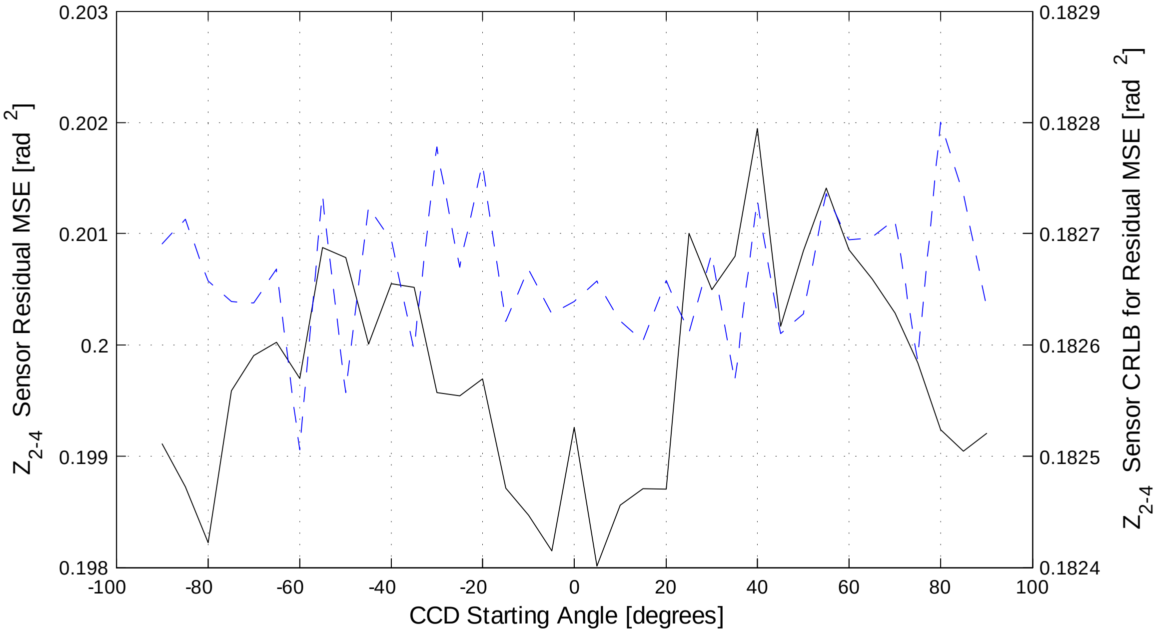


Figure 8.5 Z_{2-4} estimator $\left\langle P^2_{\phi_e} \right\rangle$ versus projection angle θ_1 given that $\theta_2 = \theta_1 + 90°$. $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, $\sigma_{ro} = 2.13$ counts, and $N_W = 14$ pixels.

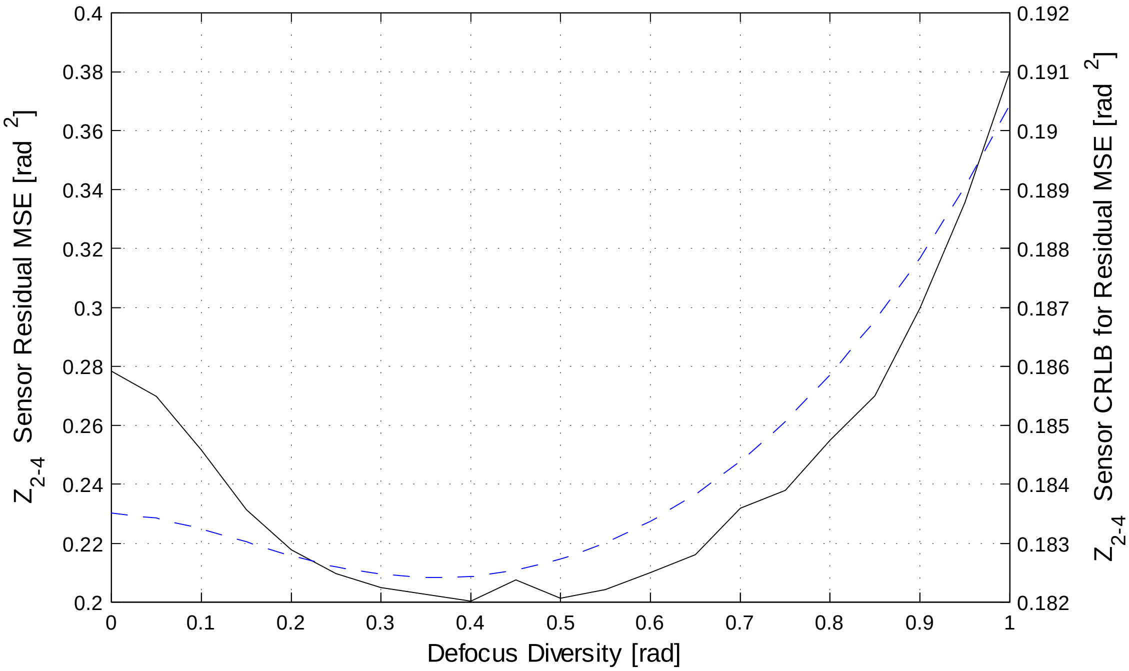


Figure 8.6 Z_{2-4} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ versus $\pm\delta_{a_4}$ for $K = 1000$ photons per subaperture. $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, $\sigma_{ro} = 2.13$ counts, and $N_W = 7$ pixels.

provides a plot of residual wavefront MSE versus defocus diversity for a low SNR case. All simulation parameters are the same except the photon count: $K = 100$. The result in Figure 8.7 suggests that the ideal choice of diversity is very close to that suggested by minimizing the performance bound: ideal $\delta_{a_4} \approx 0.1$. Notice that the plot near the minimum is nearly flat indicating that minimal MSE benefit is gained by applying diversity between the image planes. This is an indication that the benefits of the sensor over tilt only estimation are reduced as SNR decreases. This characteristic will become more apparent in curvature sensor versus tilt only sensor comparison plots to follow. Figures 8.8 and 8.9 contain plots of residual wavefront MSE versus projection window length for high and low SNR cases respectively. The plots of MSE performance versus window length when compared to the CRLB plots demonstrate an increase in the slope of the performance curve as the size of the window is decreased below 10 pixels. This result supports using CRLB as an initial choice for window length. It is worth noting, however, the choice of N_W should not be solely based on minimum MSE. The required system bandwidth may dictate a shorter window length than that which provides the best performance. In cases where bandwidth is the

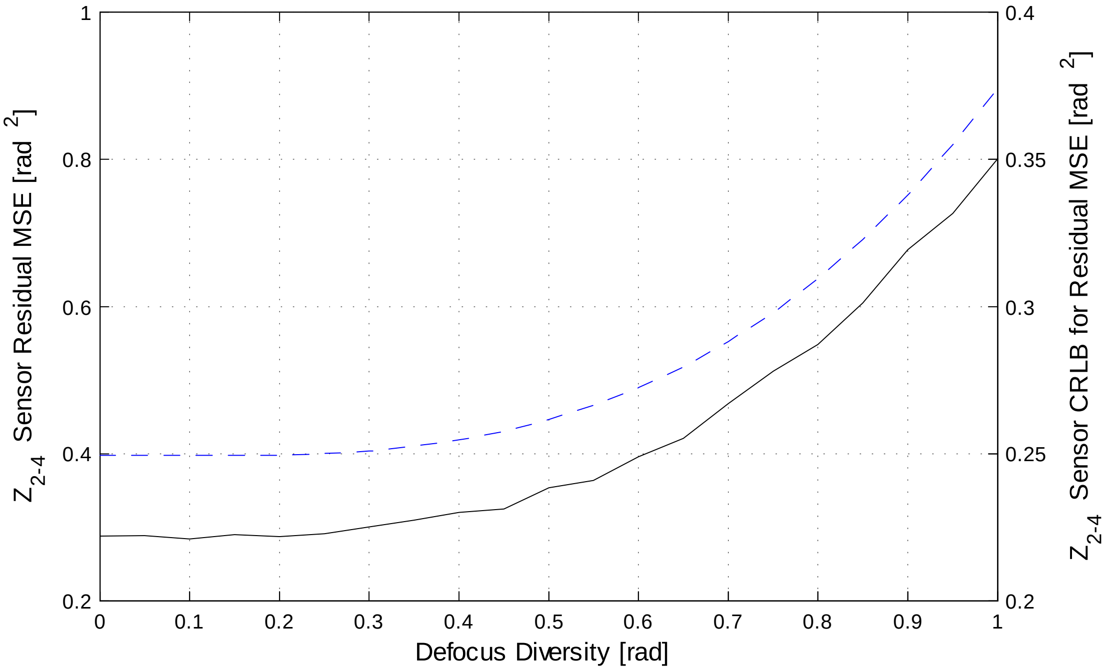


Figure 8.7 Z_{2-4} estimator $\left\langle P^2_{\phi_e} \right\rangle$ versus $\pm\delta_{a_4}$ for $K = 100$ photons per subaperture. $r_0 = 0.05\text{m}$, $L_0 = 10\text{m}$, $l_0 = 0.01\text{m}$, $K = 100$, $\sigma_{ro} = 2.13$ counts, and $W_N = 7$ pixels.

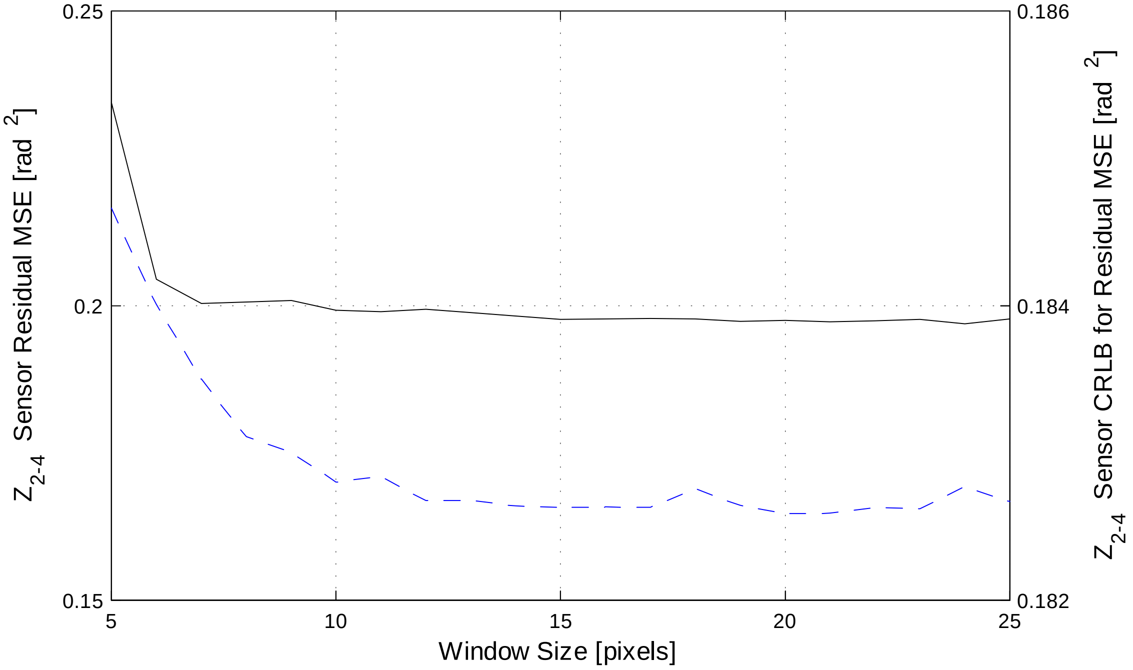


Figure 8.8 Z_{2-4} estimator $\left\langle P^2_{\phi_e} \right\rangle$ versus N_W for $K = 1000$ photons per subaperture. $r_0 = 0.05\text{m}$, $L_0 = 10\text{m}$, $l_0 = 0.01\text{m}$, and $\sigma_{ro} = 2.13$ counts.

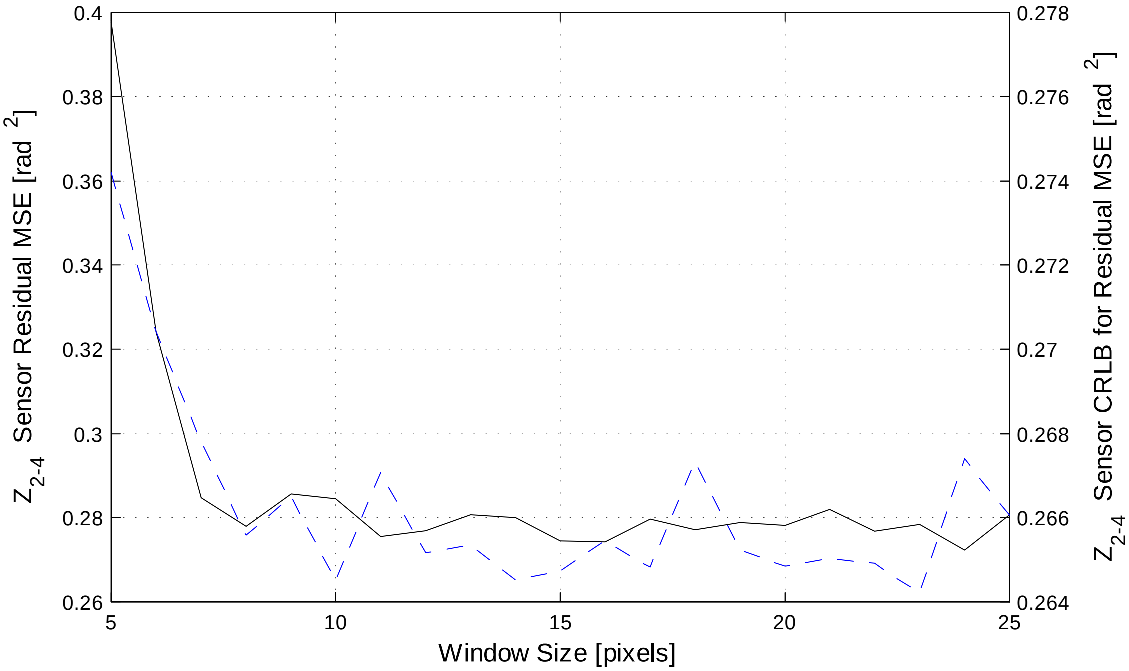


Figure 8.9 Z_{2-4} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ versus N_W for $K = 100$ photons per subaperture. $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $\sigma_{ro} = 2.13$ counts.

limiting factor, performance and CRLB indicate a significant increase in MSE for $N_W < 7$ pixels.

Figure 8.10 provides a demonstration of sensor performance over a range of SNR and r_0 values with all other operating variables fixed: $D_P = 0.07$m, $L_0 = 10$m, $l_0 = 0.01$m, $N_W = 7$ pixels and $\sigma_{ro} = 2.13$ counts. Figure 8.10 compares sensor performance when using the $Z_{2,3}$ estimator (dashed plot lines) to performance when using the Z_{2-4} estimator (solid plot lines). These results suggest that, as SNR and $\frac{D_P}{r_0}$ decrease, the ideal configuration for the estimator requires less defocus diversity, effectively converting the sensor from a Z_{2-4} sensor into a $Z_{2,3}$ sensor. This is consistent with the results from the CRLB analysis. The CRLB suggested operating regions beyond which it is no longer advantageous to estimate a_4. These threshold values can be derived from the CRLB plot in Figure 6.8. The threshold values in Figure 8.10 are slightly more conservative than lower bound threshold values in Figure 6.8. Of course, this does not preclude the use of CRLB for determining the limits on the operating space, but offers evidence that those limits set by CRLB may be optimistic and should be compared with simulated performance.

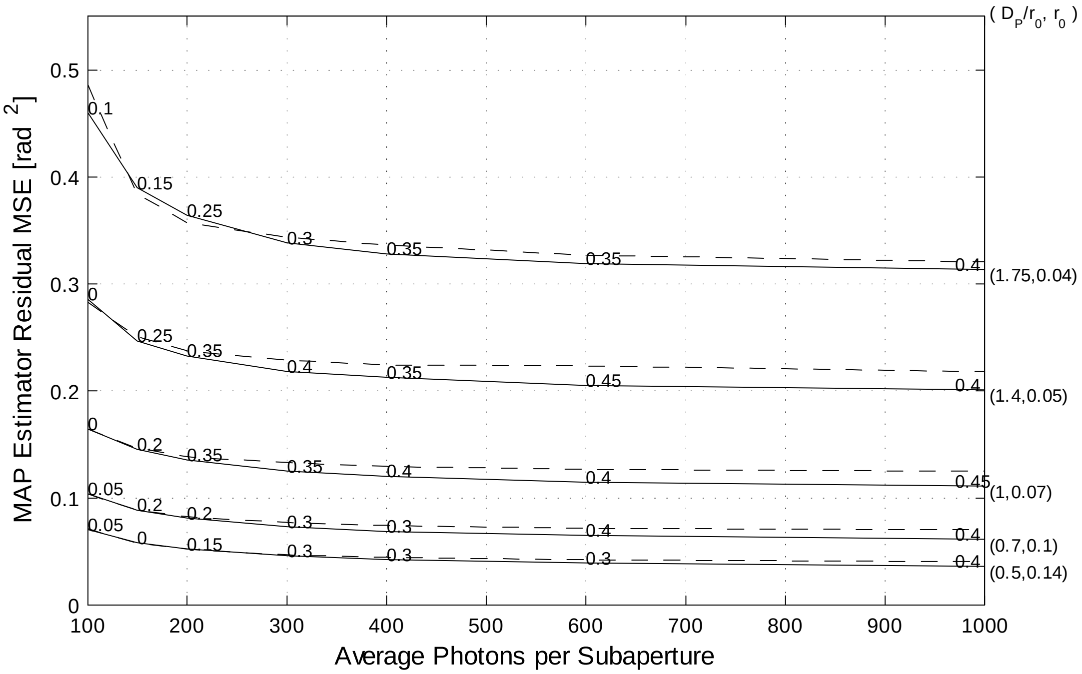


Figure 8.10 Simulated $\left\langle P_{\phi_e}^2 \right\rangle$ versus K for several cases of r_0. Dashed lines indicate $Z_{2,3}$ estimator performance. Solid lines indicate Z_{2-4} estimator performance.

The final set of performance figures offer a comparison of the curvature sensor performance with two existing techniques. The two techniques are the common two-dimensional centroid based estimator and a projection based ML tilt estimator. Figures 8.11 and 8.12 overlay the centroid performance against the results from Figure 8.10. The centroid sensor performance varies over a range of image plane window sizes. To demonstrate how centroid performance changes with increasing window size, colored plot lines {red, green, blue, cyan} correspond to performance over the set $N_W = \{5, 6, 7, 8\}$ respectively. Figure 8.13 overlays the projection based ML tilt estimator performance against the Z_{2-4} sensor results from Figure 8.10. Here the ML tilt sensor operates with minimal prior distribution knowledge: the maximization algorithm uses a range of $\pm 5\sigma_{2,3}$ in the search for the likelihood maximum. The ML tilt uses simulated long dwell images in its expected image lookup table.

Seeing the successful performance of the Z_{2-4} sensor begs the question: why stop estimating parameters at defocus? The simple answer is that estimating parameters higher than a_4 increases the residual MSE of the estimator beyond the performance set by the

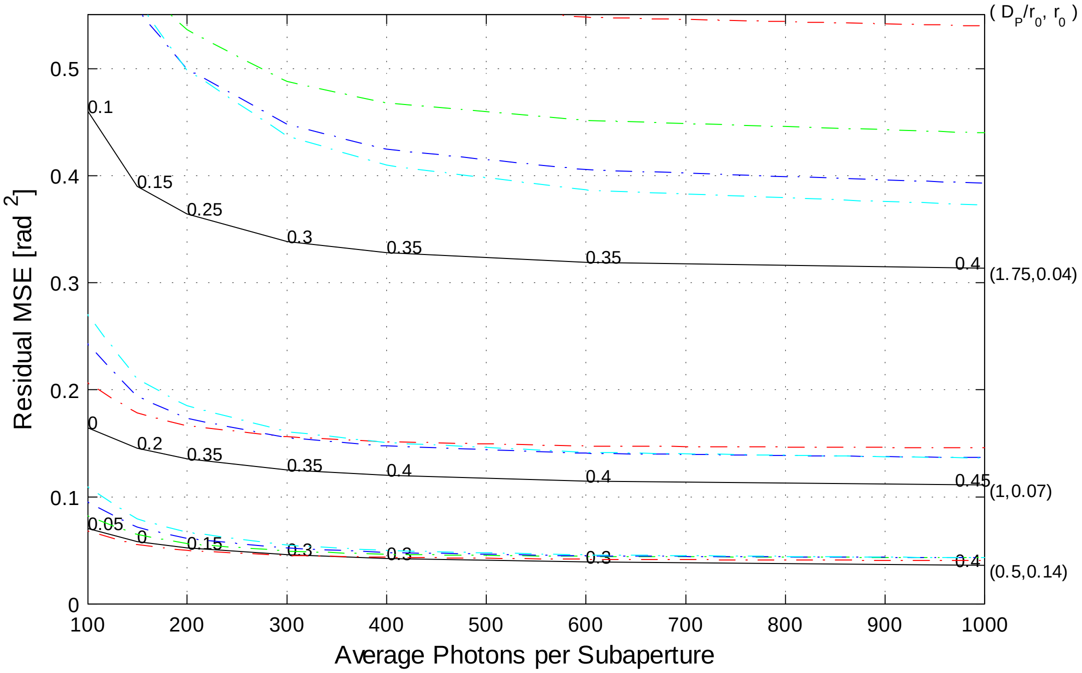


Figure 8.11 Comparison of simulated centroiding tilt estimator performance to the $Z_{2,3}$ and Z_{2-4} MAP estimator over a range of r_0 and K values.

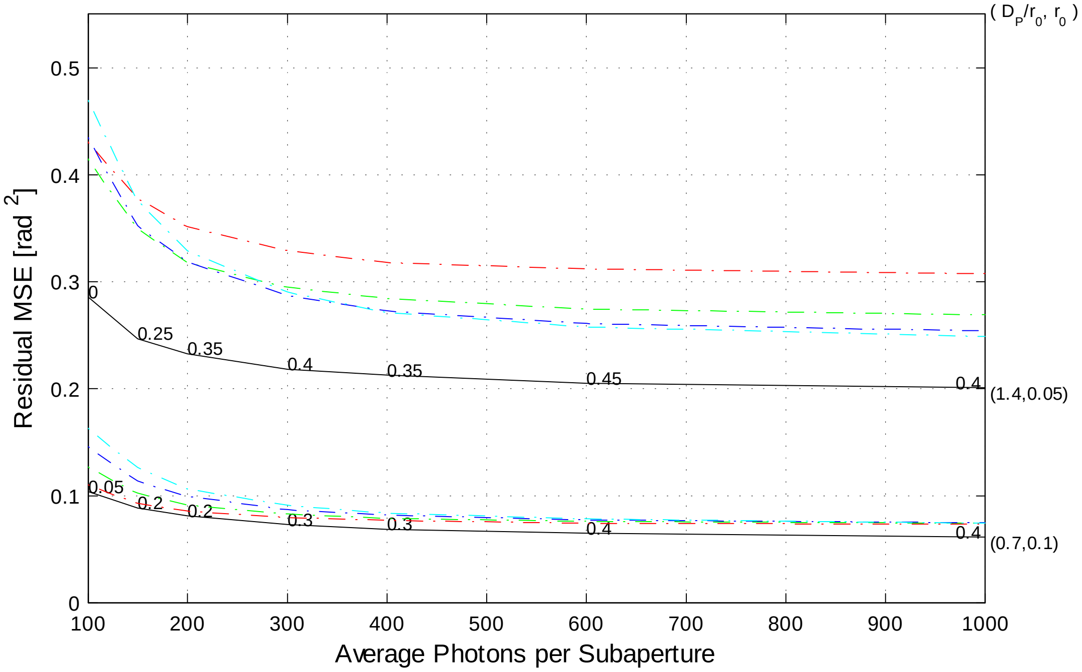


Figure 8.12 Comparison of simulated centroiding tilt estimator performance to the $Z_{2,3}$ and Z_{2-4} MAP estimator over a range of r_0 and K values.

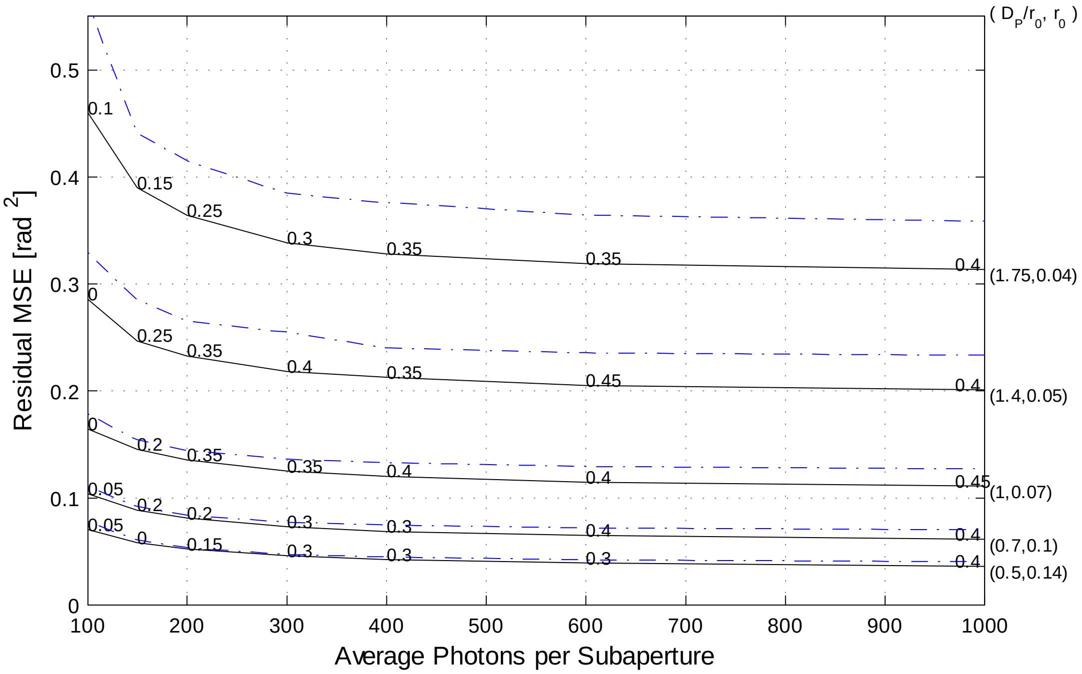


Figure 8.13 Comparison of simulated projection based ML tilt estimator performance to the $Z_{2,3}$ and Z_{2-4} MAP estimator over a range of r_0 and K values.

Z_{2-4} estimator over the majority of the simulated operating space. The performance plot in Figure 8.14 demonstrates this phenomenon. The solid plot lines depict the residual MSE for the Z_{2-4} sensor seen in previous figures. The dashed plot lines indicate the residual MSE for a whole plane projection sensor attempting to estimate parameters a_2 through a_6. The design variables and environment variables used in the simulation are $D_P = 0.07$m, $L_0 = 10$m, $l_0 = 0.01$m, $N_W = 7$ pixels and $\sigma_{ro} = 2.13$ counts. Notice that the whole plane Z_{2-6} sensor either performs on par with or fails to perform better than the Z_{2-4} sensor throughout the operational space. This example suggests that, in order to estimate more parameters, the sensor will require additional information from the CCD.

8.3 Sensitivity Analysis

This section contains performance results from a series of simulations in which the sensor was purposefully given erroneous estimates of the environment variables: r_0, L_0, and l_0. The intent is to provide a demonstration of the sensor's robustness under the influence of inaccurate environment variable estimates. Several design variables remain constant

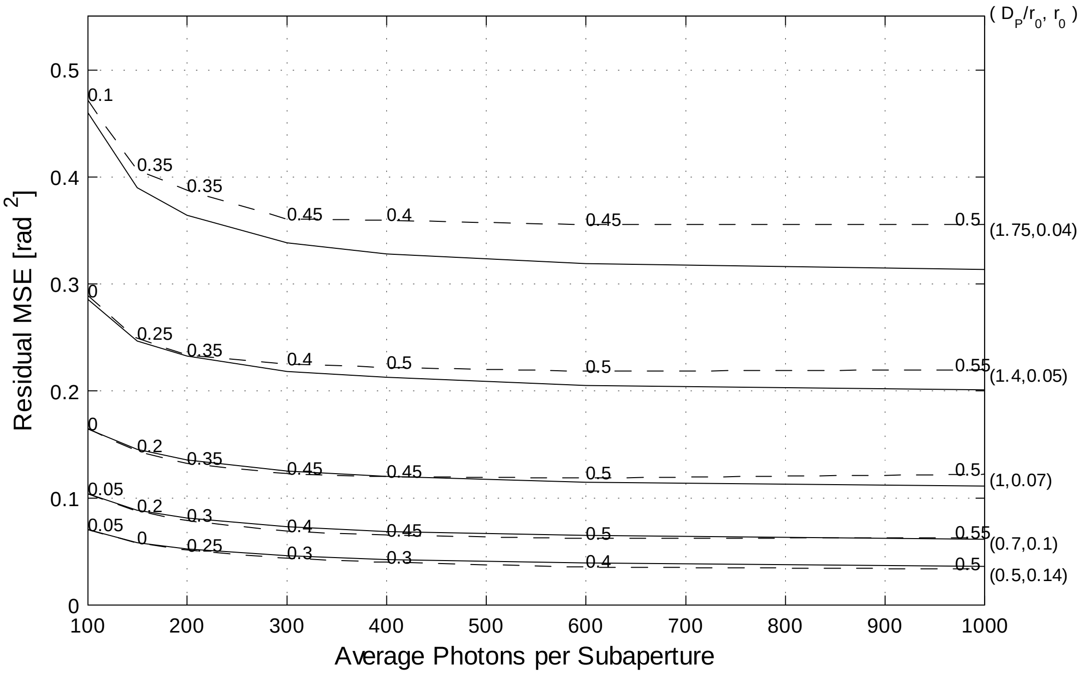


Figure 8.14 Solid plot line depicts Z_{2-4} estimator performance. Dashed plot line indicates Z_{2-6} performance.

throughout each of the figures to follow: $D_P = 0.07$m, $N_W = 7$ pixels and $\sigma_{ro} = 2.13$. Also consistent throughout is the choice of plot line styles and their corresponding data series. Solid lines indicate Z_{2-4} sensor performance, while dashed lines indicate ML tilt sensor performance. The ML performance lines are included as comparison lines to highlight points where poor estimates of the environment variables negate the Z_{2-4} sensor's performance advantage. The true value of the environment variable for each plot line is indicated by the × symbol along the solid line and a ◯ symbol along the dashed line. Each point along the solid performance lines represents an average value from 30 random cases. A confidence interval at each plot point is indicated by a pair of triangles: one pointing upward for $+1\sigma$ and one pointing downward for -1σ.

Figure 8.15 demonstrates r_0 sensitivity in high SNR, $K = 1000$ photons, at three locations in the operating space. The true values of the parameter $r_0 = \{0.04\text{m}, 0.05\text{m}, 0.1\text{m}\}$ correspond to the three pairs of plot lines. As indicated by the x-axis labeling, the estimates of r_0 range from 0.02m to 0.2m. However, the entire range is not tested for each r_0 case. The range of test values is selected based on the true value of r_0. Figure 8.16

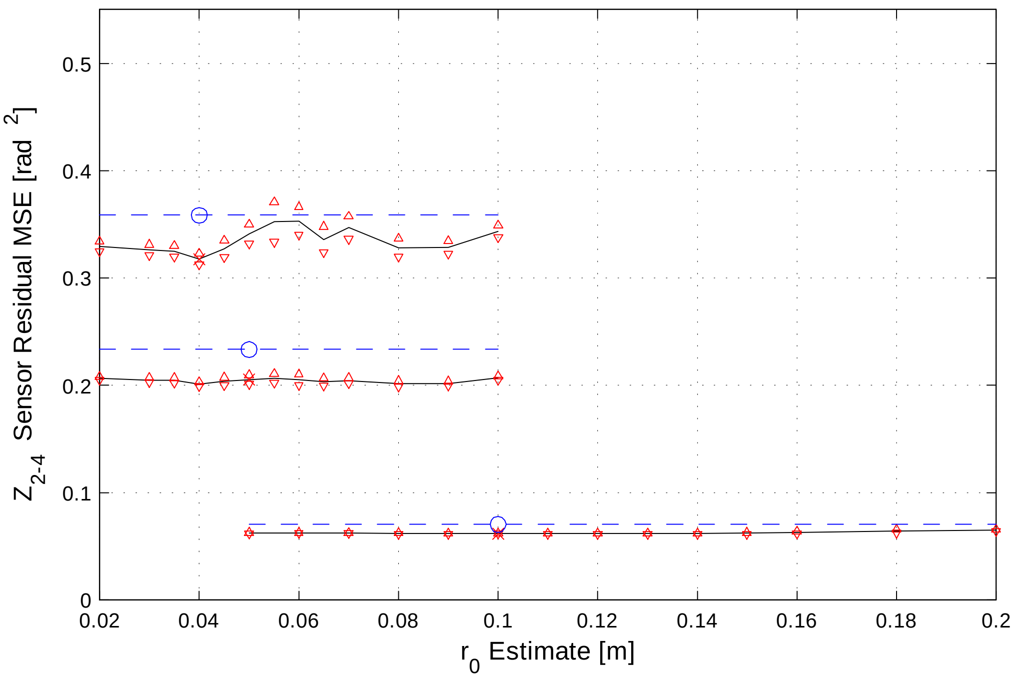


Figure 8.15 Solid lines indicate residual MSE versus r_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of r_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 1000$.

provides the same r_0 sensitivity analysis at low SNR: $K = 200$ photons. Figure 8.17 shows L_0 sensitivity in high SNR: $K = 1000$ photons. The true L_0 value is set at 10m while the estimate of L_0 is in the set {1m, 10m, 100m}. Just as in the r_0 analysis figures, the solid line indicates Z_{2-4} performance while the dashed line indicates ML tilt performance. The true value of L_0 for each performance line is indicated by the × symbol along the solid line and a ◯ along the dashed line. Figure 8.18 provides the same L_0 sensitivity analysis at low SNR: $K = 200$ photons. Figure 8.19 demonstrates l_0 sensitivity in high SNR: $K = 1000$ photons. The true l_0 value is set at 0.01m while the estimate of l_0 is in the set {0m, 0.01m, 0.1m}. Figure 8.20 provides the same l_0 sensitivity analysis at low SNR: $K = 200$ photons.

The results of the sensitivity analysis demonstrate that the sensor performance depends on the r_0 estimate more so than the estimates of L_0 and l_0. In fact, the sensor's performance is nearly invariant to L_0 and l_0 estimates. The simulated cases indicate that when L_0 and l_0 are unknown, the best course of action is to overestimate L_0 (set $L_0 \geq 100$m)

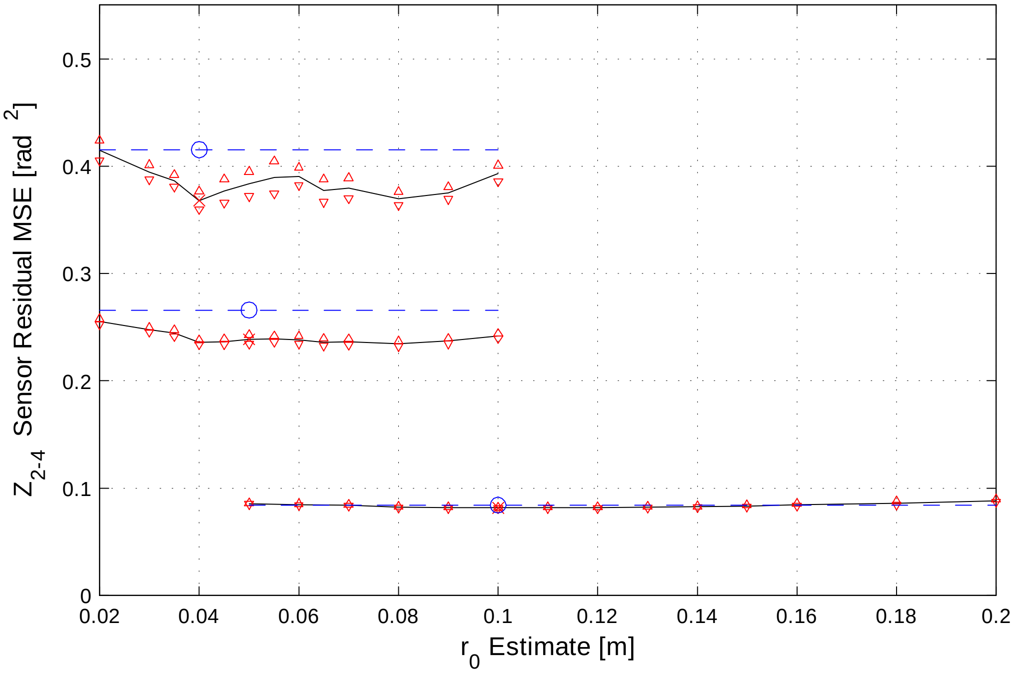


Figure 8.16 Solid lines indicate residual MSE versus r_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of r_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 200$.

and set the l_0 estimate equal to zero. The sensitivity to r_0 is most evident in low SNR and at high $\frac{D_P}{r_0}$. It is worthwhile to note that, although its performance is degraded, the Z_{2-4} sensor performs on par with or better than the ML tilt sensor over the range of simulated r_0 estimates.

8.4 Summary

This chapter began with a brief description of the Z_{2-4} sensor simulation. Simulated performance plots were provided in order to demonstrate the search for ideal design variable settings, and to provide an estimate of sensor performance over a typical range of the operating space. Simulated performance examples used to predict ideal design variable settings show that the computer simulation and the calculated CRLB results compliment each other. The fact that the CRLB results and simulated results derive similar conclusions, increases confidence in the simulation accuracy and suggests that CRLB is an adequate tool for making preliminary design choices. A comparison of the Z_{2-4} sensor performance to

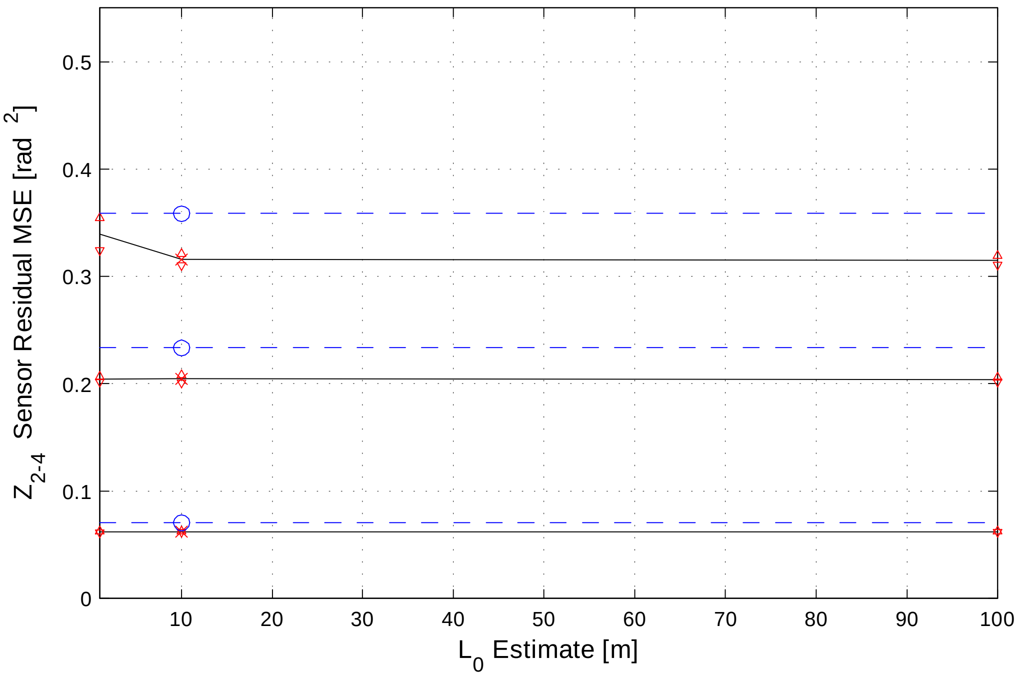


Figure 8.17 Solid lines indicate residual MSE versus L_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of L_0 is indicated by an $\times$ (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 1000$.

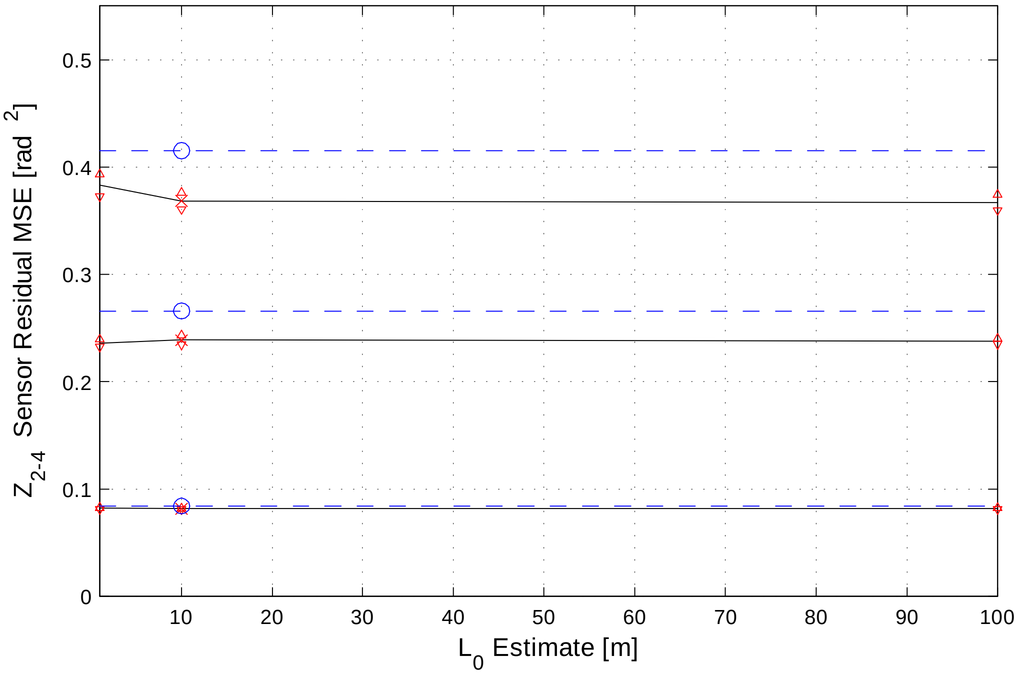


Figure 8.18 Solid lines indicate residual MSE versus L_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of L_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 200$.

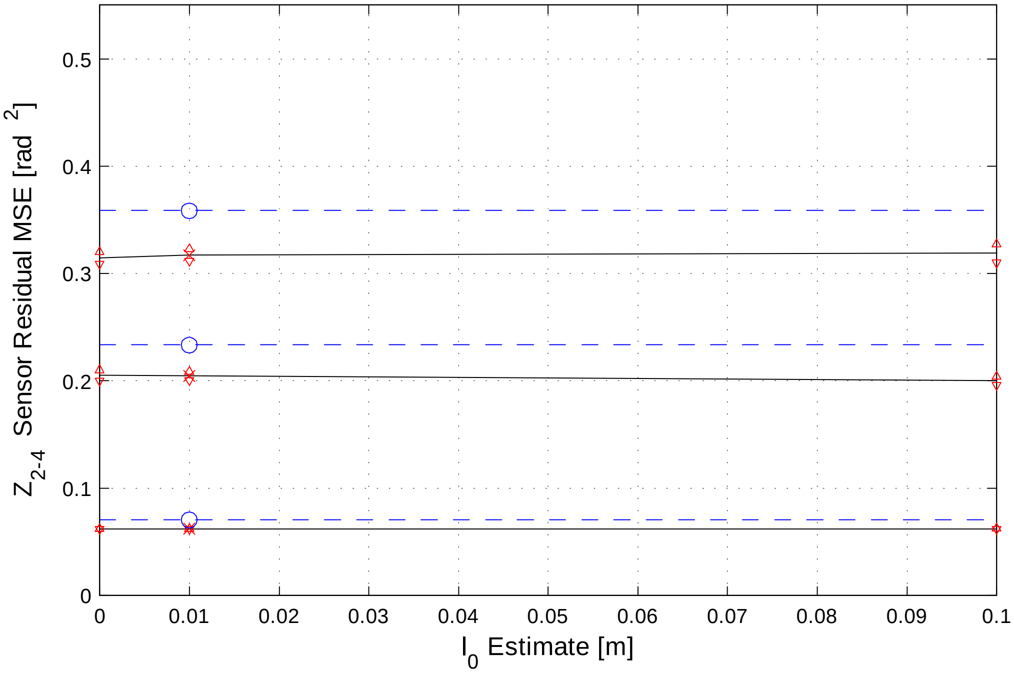


Figure 8.19 Solid lines indicate residual MSE versus l_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of l_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 1000$.

the centroid and tilt-only ML estimator revealed that, under the simulated conditions, the Z_{2-4} sensor performance is on par with or better than the other sensor designs for cases where the average photon count is greater than 100 photons per subaperture and the ratio $\frac{D_P}{r_0}$ is greater than 0.5.

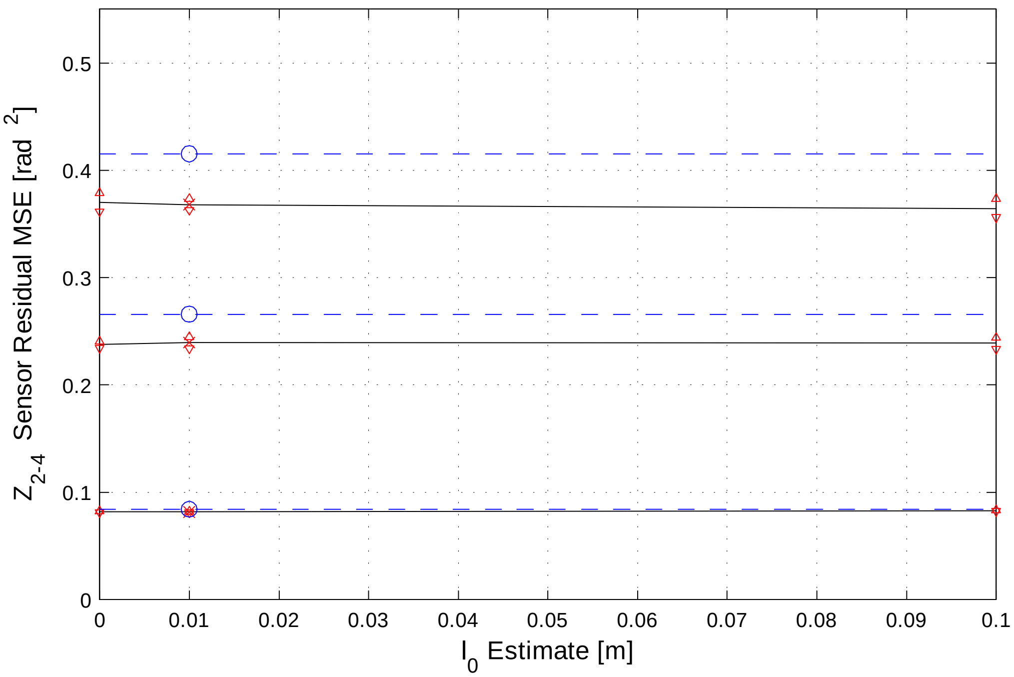


Figure 8.20 Solid lines indicate residual MSE versus l_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of l_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 200$.

9. The Z_{2-10} Wavefront Sensor

This chapter provides the details and simulated performance results for a curvature sensor designed to estimate Zernike polynomial coefficients a_2 through a_{10}. To aid with coefficients higher than a_4, half plane image projections are used. This sensor provides estimates in two stages as did the Z_{2-4} sensor. The tilt estimates are formed in the first stage and all higher order estimates are formed in the second stage making the Z_{2-10} highly parallelizable. If all higher order estimates are computed in parallel, the only additional complexity over the Z_{2-4} sensor comes from managing half plane image projections which increases the computational complexity by about 67%. The sections to follow outline the major differences between the Z_{2-4} sensor and the Z_{2-10} sensor and provide an analysis of Z_{2-10} sensor simulated performance.

9.1 Image Projections

The Z_{2-10} sensor design uses half plane image projections. The half plane image projection is defined as a concatenated pair of projections. The windowed region in the image plane is divided between the two projections. The windowed region is divided evenly if the window length is even. The odd center row of pixels is summed into the second half image projection when the window length is odd. Figure 9.1 provides a diagram of the Z_{2-10} image projections. The projection operations included in the Z_{2-10} estimator expressions are:

$$
\begin{aligned}
& {}_{\{(1,N_1),N_W\}}\mathbf{v}\left(\mathbf{D}_{i,\theta_i}\right), && (9.1)\\
& {}_{\{(N_2,N_W)\}}\mathbf{v}\left(\mathbf{D}_{i,\theta_i}\right), && (9.2)\\
& {}_{\{(1,N_1),N_W\}}\mathbf{v}\left(\mathbf{D}_{1,0},\mathbf{D}_{2,90}\right), && (9.3)\\
& {}_{\{(N_2,N_W)\}}\mathbf{v}\left(\mathbf{D}_{1,0},\mathbf{D}_{2,90}\right), \text{ and} && (9.4)\\
& {}_{\{(1,N_1),(N_2,N_W)\}}\mathbf{v}\left(\mathbf{D}_{1,0},\mathbf{D}_{2,90}\right), && (9.5)\\
\text{where } N_1 &= \text{floor}(N_W/2), && (9.6)\\
\text{and } N_2 &= \text{floor}(N_W/2)+1. && (9.7)
\end{aligned}
$$

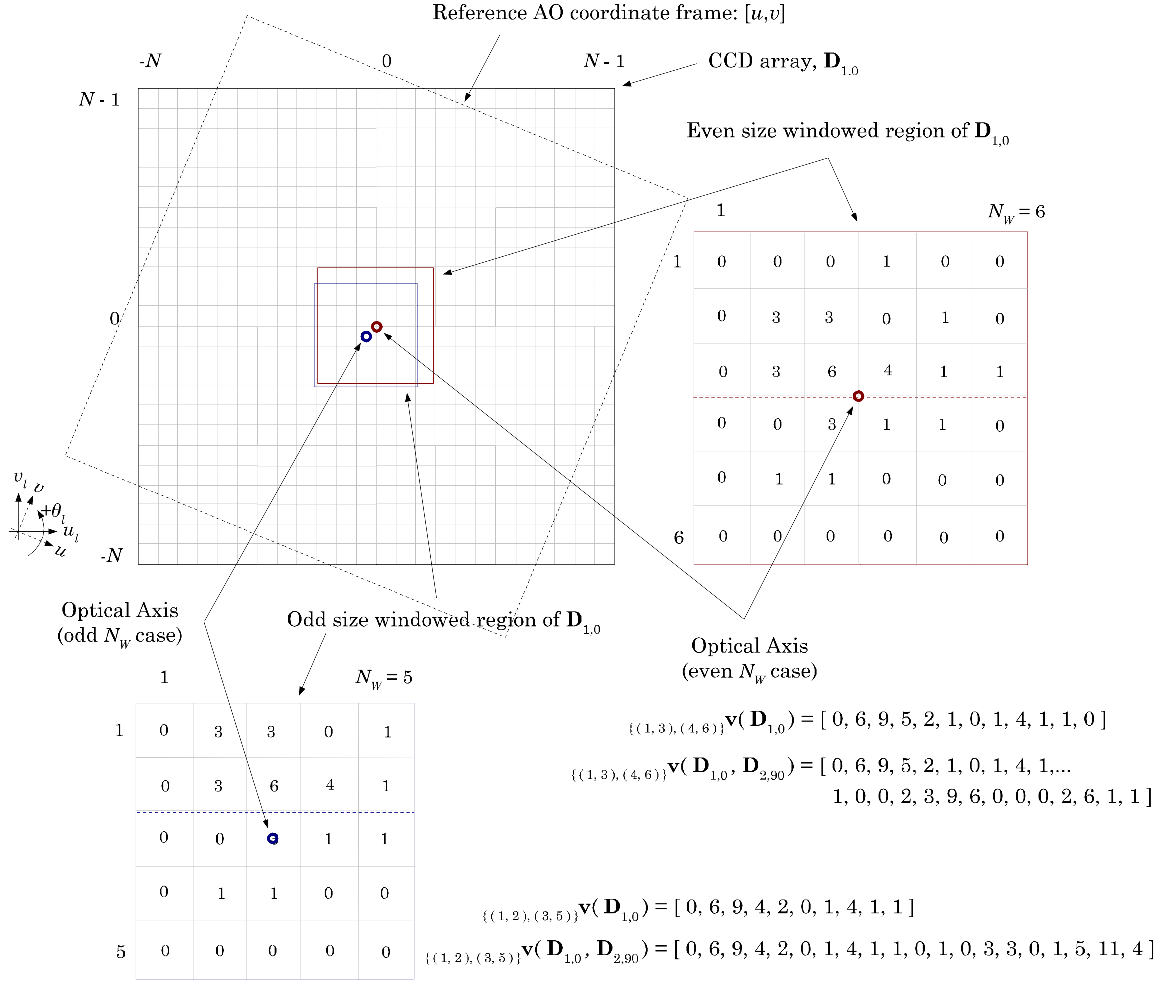


Figure 9.1 Diagram of the Z_{2-10} sensor's half plane image projection operation for 6×6 pixel and 5×5 pixel windows.

Note that each of the half plane projections are read from the CCDs only once. The projection operations are expressed as unique operations here solely for mathematical convenience. The tilt estimator and the a_4 estimator still require whole plane projections. The whole plane projections are constructed from the sum of half plane projections.

Due to the half plane projection requirement, the Z_{2-10} sensor must process more data than the Z_{2-4} sensor. The chart in Figure 9.2 describes the relative difference in computational complexity of estimating a_2 through a_{10} based on the difference in projection data and the level of parallel or serial computation. Thus, any design using half plane image projections accepts the engineering trade-off of increased information in exchange

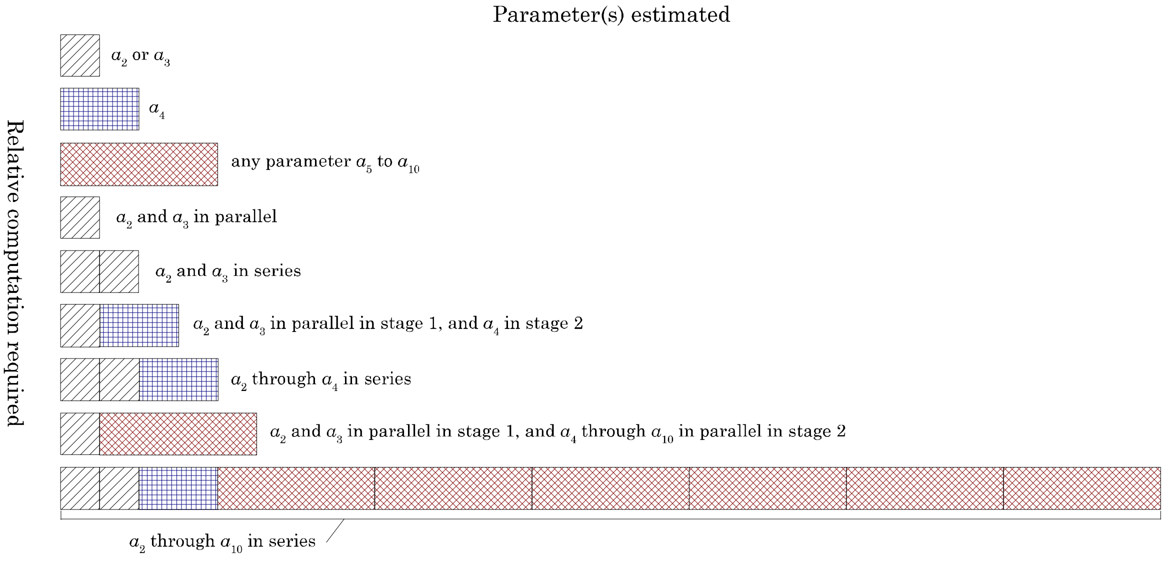


Figure 9.2 Relative computational complexity between serial and parallel estimator configurations.

for increased read noise, increased CCD read time, and increased computation time. The increased information provides the ability to estimate Zernike polynomials beyond Z_4.

9.2 Likelihood Expressions

The sensor hardware provides four half plane image projections from each subaperture. Figure 9.3 diagrams the read out and flow of the four image projections through the estimator algorithm. Each likelihood expression requires four inputs: a detected image projection, a reference image projection, a set of atmospheric parameter estimates and an estimate of the current photon level, K. Solid lines indicate the flow of real time detected image projections. Dashed lines indicate information used for reference image projections which are computed and stored into lookup tables during sensor calibration. The heavily outlined blocks in Figure 9.3 indicate locations where a likelihood expression is evaluated. Each parameter is estimated independently. The tilt parameters must be estimated first.

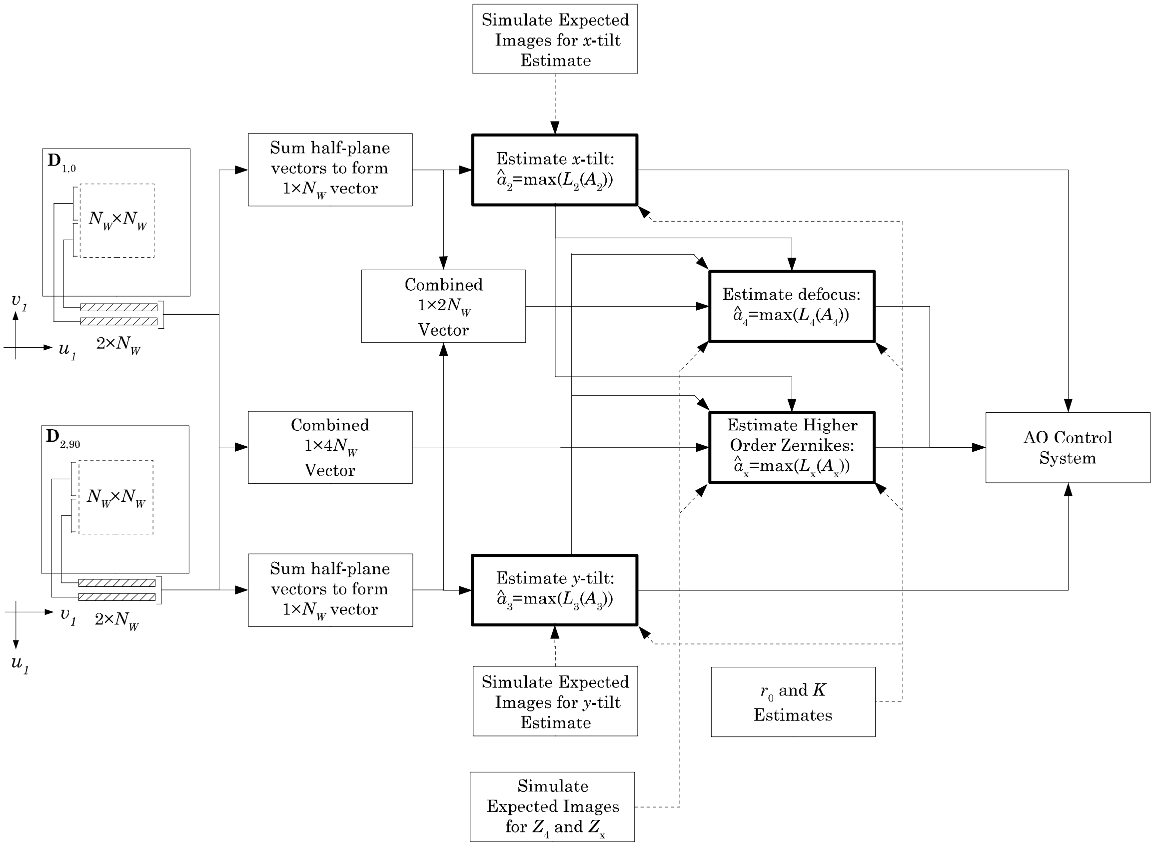


Figure 9.3 Diagram shows the flow of image projections through the Z_{2-10} estimation algorithm.

The tilt likelihood expressions used in the Z_{2-10} estimator are given by:

$$\begin{aligned} L_{map_2}(A_2) \;=\; & \sum_l \left[{}_{\{(1,N_1),N_W\}}\mathbf{v}_l\left(\mathbf{D}_{1,0}\right) + {}_{\{(N_2,N_W)\}}\mathbf{v}_l\left(\mathbf{D}_{1,0}\right) + \sigma_{ro}^2 \right] \times \\ & \ln\left\{ {}_{\{(1,N_W)\}}\mathbf{v}_l\left({}_{L23}\mathbf{I}_{1,0}\left[A_2\right]\right) + \sigma_{ro}^2 \right\} - \\ & {}_{\{(1,N_W)\}}\mathbf{v}_l\left({}_{L23}\mathbf{I}_{1,0}\left[A_2\right]\right) - \frac{A_2^2}{2\sigma_2^2}. \end{aligned} \tag{9.8}$$

$$\begin{aligned} L_{map_3}(A_3) \;=\; & \sum_l \left[{}_{\{(1,N_1),N_W\}}\mathbf{v}_l\left(\mathbf{D}_{2,90}\right) + {}_{\{(N_2,N_W)\}}\mathbf{v}_l\left(\mathbf{D}_{2,90}\right) + \sigma_{ro}^2 \right] \times \\ & \ln\left\{ {}_{\{(1,N_W)\}}\mathbf{v}_l\left({}_{L23}\mathbf{I}_{2,90}\left[A_3\right]\right) + \sigma_{ro}^2 \right\} - \\ & {}_{\{(1,N_W)\}}\mathbf{v}_l\left({}_{L23}\mathbf{I}_{2,90}\left[A_3\right]\right) - \frac{A_3^2}{2\sigma_3^2}. \end{aligned} \tag{9.9}$$

Recall that each parameter estimate is formed by maximizing the likelihood with respect to the parameter:

$$\max_{A_{\mathrm{x}}} L_{map_{\mathrm{x}}}(A_{\mathrm{x}})\Big|_{A_{\mathrm{x}}=\hat{A}_{\mathrm{x}}}. \tag{9.10}$$

Unlike higher order parameters, the tilt estimates are computed from whole plane projections. Therefore, the two half plane projections are summed to create a single whole plane projection prior to evaluation of the likelihood expression. Once tilt estimates are available, all higher order parameters may be estimated using the tilt estimates as indices into the tables of preregistered image projections. The estimator selects the preregistered Z_4 through Z_{10} projections with the closest matching pair of tilt values:

$$\left(\tilde{A}_2, \tilde{A}_3\right) = \left(\mathrm{round}\left(\frac{\hat{A}_2}{\Delta A_2}\right)\Delta A_2, \mathrm{round}\left(\frac{\hat{A}_3}{\Delta A_3}\right)\Delta A_3\right). \tag{9.11}$$

The likelihood for Z_4 is computed from whole plane projections much like the tilt likelihood:

$$\begin{aligned} L_{map_4}(A_4) \quad = \quad & \sum_l \left[{}_{\{(1,N_1),N_W\}}\mathbf{v}_l \left(\mathbf{D}_{1,0}, \mathbf{D}_{2,90}\right) + {}_{\{(N_2,N_W)\}}\mathbf{v}_l \left(\mathbf{D}_{1,0}, \mathbf{D}_{2,90}\right) + \sigma_{ro}^2 \right] \times \\ & \ln\left\{ {}_{\{(1,N_W)\}}\mathbf{v}_l \left({}_{L4}\mathbf{I}_{1,0}\left[\tilde{A}_2, \tilde{A}_3, A_4\right], {}_{L4}\,\mathbf{I}_{2,90}\left[\tilde{A}_2, \tilde{A}_3, A_4\right]\right) + \sigma_{ro}^2 \right\} - \\ & {}_{\{(1,N_W)\}}\mathbf{v}_l \left({}_{L4}\mathbf{I}_{1,0}\left[\tilde{A}_2, \tilde{A}_3, A_4\right], {}_{L4}\,\mathbf{I}_{2,90}\left[\tilde{A}_2, \tilde{A}_3, A_4\right]\right) - \frac{A_4^2}{2\sigma_4^2}. \end{aligned} \tag{9.12}$$

The likelihood for all higher order estimates is given by:

$$\begin{aligned} L_{map_{\mathrm{x}}}(A_{\mathrm{x}}) \quad = \quad & \sum_l \left[{}_{\{(1,N_1),(N_2,N_W)\}}\mathbf{v}_l \left(\mathbf{D}_{1,0}, \mathbf{D}_{2,90}\right) + \sigma_{ro}^2 \right] \times \\ & \ln\left\{ {}_{\{(1,N_1),(N_2,N_W)\}}\mathbf{v}_l \left({}_{L\mathrm{x}}\mathbf{I}_{1,0}\left[\tilde{A}_2, \tilde{A}_3, A_{\mathrm{x}}\right], {}_{L\mathrm{x}}\,\mathbf{I}_{2,90}\left[\tilde{A}_2, \tilde{A}_3, A_{\mathrm{x}}\right]\right) + \sigma_{ro}^2 \right\} - \\ & {}_{\{(1,N_1),(N_2,N_W)\}}\mathbf{v}_l \left({}_{L\mathrm{x}}\mathbf{I}_{1,0}\left[\tilde{A}_2, \tilde{A}_3, A_{\mathrm{x}}\right], {}_{L\mathrm{x}}\,\mathbf{I}_{2,90}\left[\tilde{A}_2, \tilde{A}_3, A_{\mathrm{x}}\right]\right) - \frac{A_{\mathrm{x}}^2}{2\sigma_{\mathrm{x}}^2}. \end{aligned} \tag{9.13}$$

9.3 Sensor Performance

This section contains several examples of Z_{2-10} sensor simulated performance. The simulated performance will be used for design variable selection and it will be compared with the projection based ML tilt sensor and the more common two-dimensional centroid based tilt sensor. To compare the results of using CRLB versus simulation performance as a metric for design variable selection, I will begin by overlaying the series of plots found in Section 6.3 onto simulated sensor performance. Just as with the Z_{2-4} sensor simulation, the performance plots here should trend the same as the CRLB plots but will differ from CRLB results in overall magnitude. The discrepancy between CRLB calculations and the simulation is attributable to variance in the estimator and the difference between the random phase generation method used in the CRLB calculations versus the method used in simulation.

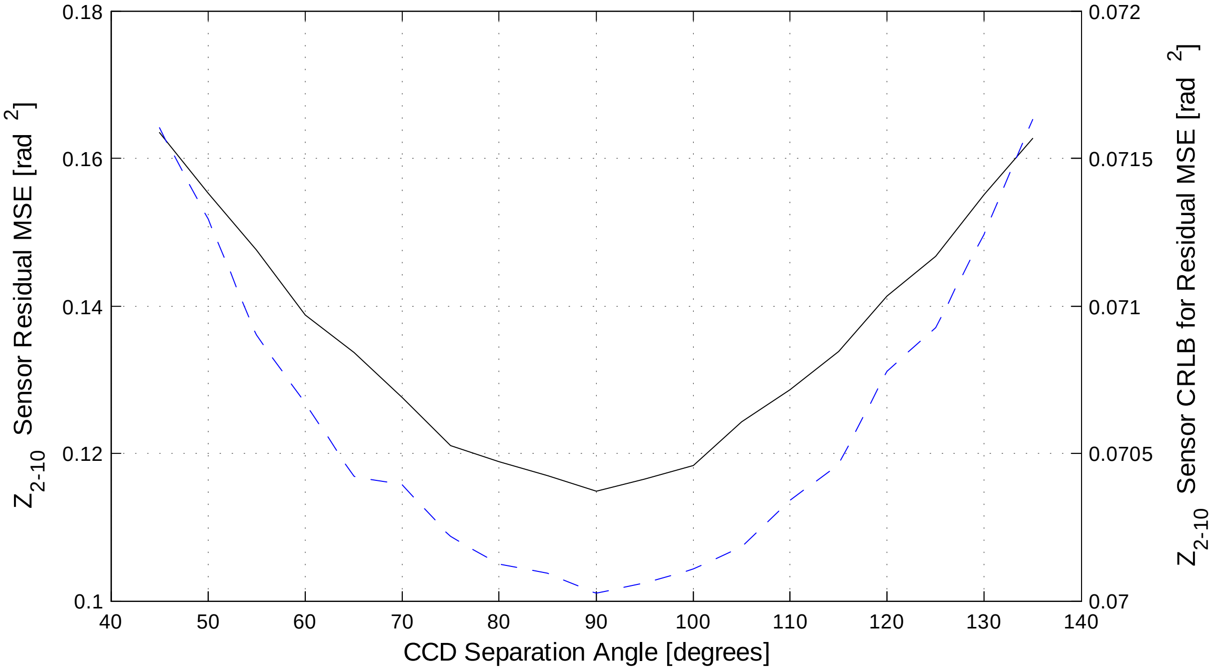


Figure 9.4 Z_{2-10} estimator $\left\langle P^2_{\phi_e} \right\rangle$ versus separation angle, $\theta_2 - \theta_1$. $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $N_W = 14$ pixels.

Figure 9.4 contains a plot of residual MSE versus projection separation angle. The following design variable and environment variable settings were used: $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, $K = 1000$ photons, $\pm\delta_{a_4} = 0.55$ radians, and $N_W = 14$ pixels. The CRLB result for half plane projection separation angle is included with an independent y-axis. The CRLB plot line is the dashed line and corresponds to the y-axis on the right hand side of the plot. The simulated performance results in Figure 9.4 agree with the CRLB results indicating that the ideal choice of separation angle is 90 degrees. Using this result, Figure 9.5 demonstrates simulated performance versus θ_1 for a separation angle of 90 degrees. Figure 9.5 echoes the results seen in CRLB plots for both the Z_{2-4} and Z_{2-10} sensors (Figures 6.3 and 6.10) and in the performance results for the Z_{2-4} sensor (Figure 8.5): residual MSE is effectively invariant over the range of θ_1 given that the difference between projection angles is 90 degrees. Note that the CRLB continues to be represented by a dashed plot line scaled to the right hand side y-axis for Figures 9.5 through 9.9. Figure 9.6 contains a plot of residual wavefront MSE versus defocus diversity for a high SNR case: $D_P = 0.07$m, $\sigma_{ro} = 2.13$ counts, $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, $K = 1000$ photons, and $N_W = 9$ pixels. Performance in Figure 9.6 suggests that the ideal diversity is approximately 0.55 radians, which is the same as the CRLB result. Figure 9.7 provides a plot of residual wavefront MSE versus defocus diversity for a low SNR case. The simulation inputs are the same as those used in the previous example except the photon count: $K = 100$. The result in Figure 9.7 corresponds to the CRLB result once again. Performance suggests that the ideal choice of diversity at low SNR is approximately 0.3 radians. CRLB suggests that the ideal diversity is about 0.4 radians. Figures 9.8 and 9.9 contain plots of residual wavefront MSE versus projection window size for high and low

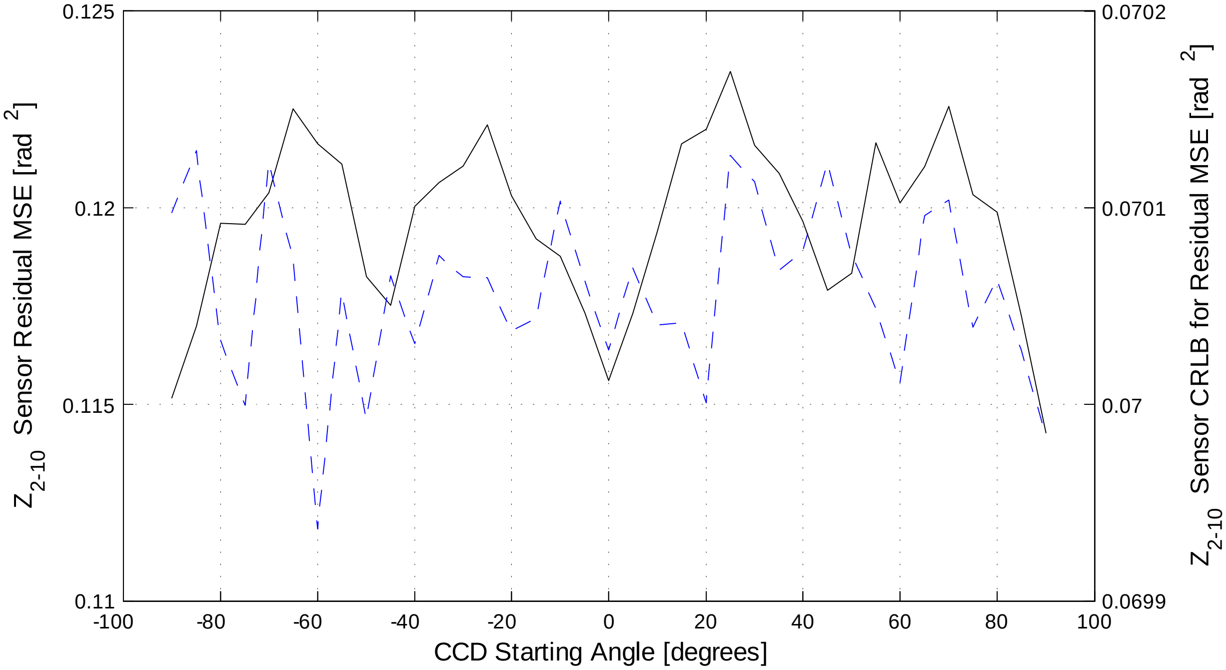


Figure 9.5 Z_{2-10} estimator $\left\langle P^2_{\phi_e} \right\rangle$ versus projection angle θ_1 given that $\theta_2 = \theta_1 + 90°$. $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $N_W = 14$ pixels.

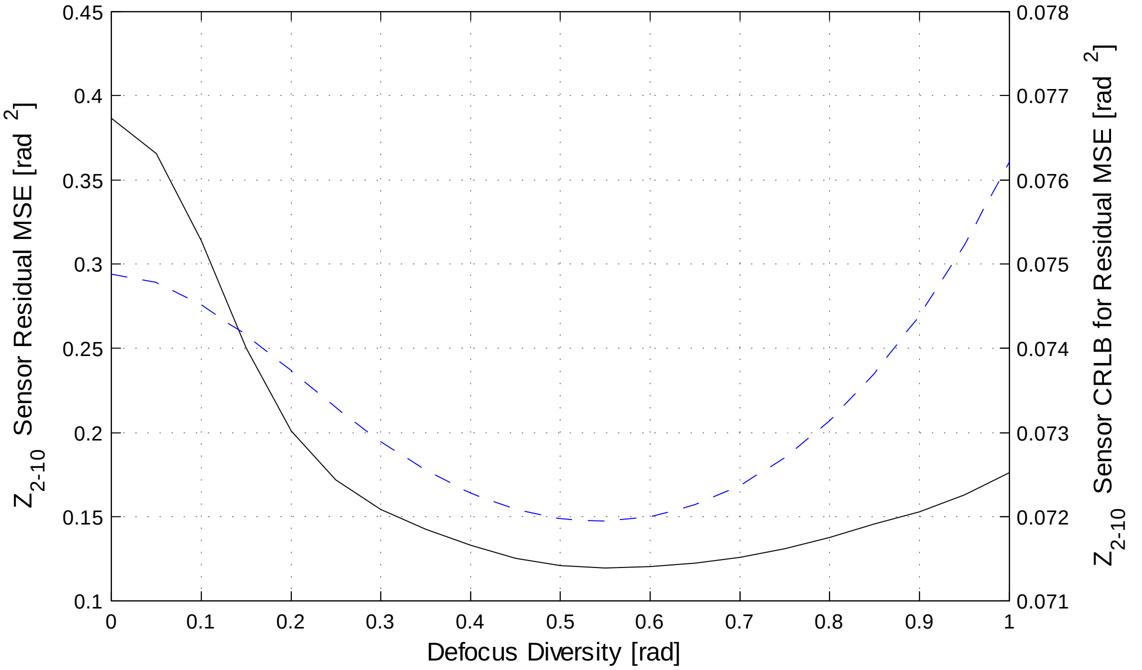


Figure 9.6 Z_{2-10} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ versus $\pm\delta_{a_4}$ for $K = 1000$ photons per subaperture. $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, and $N_W = 9$ pixels.

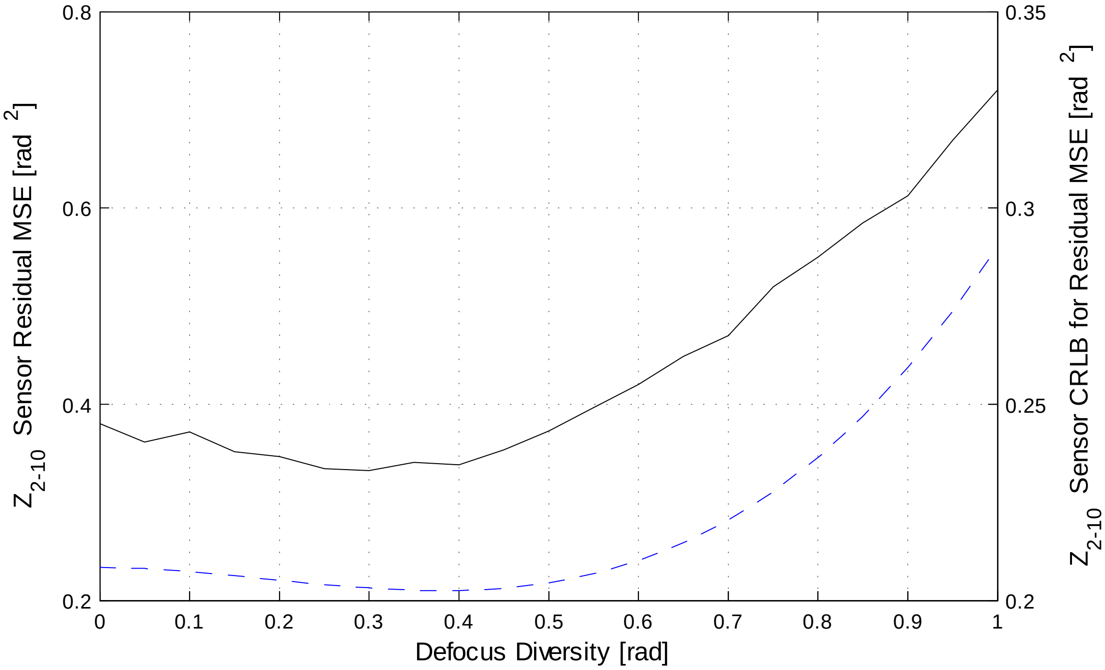


Figure 9.7 Z_{2-10} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ versus $\pm\delta_{a_4}$ for $K = 100$ photons per subaperture. $r_0 = 0.05$m, $L_0 = 10$m, $l_0 = 0.01$m, $K = 100$, and $W_N = 9$ pixels.

SNR cases respectively. The plots of simulated performance versus window size indicate that sensor performance is best when the window size is about 12 × 12 pixels. Required system bandwidth may dictate a shorter window length than that which provides ideal performance. In cases where system bandwidth requires reading out fewer pixels, it is important to note that the CRLB and simulated performance plots indicate a significant reduction in performance for projection vector lengths shorter than 8 pixels.

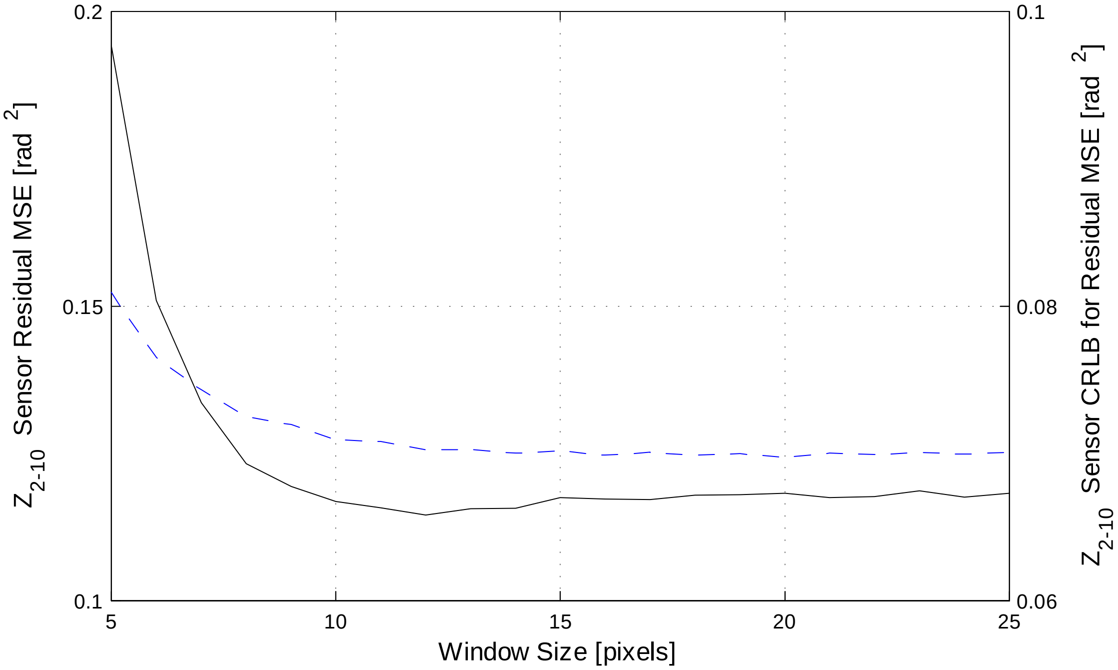


Figure 9.8 Z_{2-10} estimator $\left\langle P^2_{\phi_e} \right\rangle$ versus N_W for $K = 1000$ photons per subaperture.

Figures 9.10 and 9.11 provide a demonstration of the Z_{2-10} sensor performance over a range of SNR and r_0 values with all other operating variables fixed: $D_P = 0.07$m, $L_0 = 10$m, $l_0 = 0.01$m, $N_W = 9$ pixels and $\sigma_{ro} = 2.13$ counts. The solid plot lines represent Z_{2-10} performance. The dashed plot lines represent tilt only performance. The tilt only performance lines are the same dashed plot lines found in the whole plane projection performance plot in Figure 8.10. The tilt only performance lines are included in this figure in order to demonstrate the lower bounds in the operating space on SNR and the ratio $\frac{D_P}{r_0}$ beyond which it is no longer advantageous to implement the higher order estimator. Recall, for instance, that the tilt only performance shown here offers the additional advantage of using a 7 pixel window versus $N_W = 9$ pixels. When compared with the performance of the

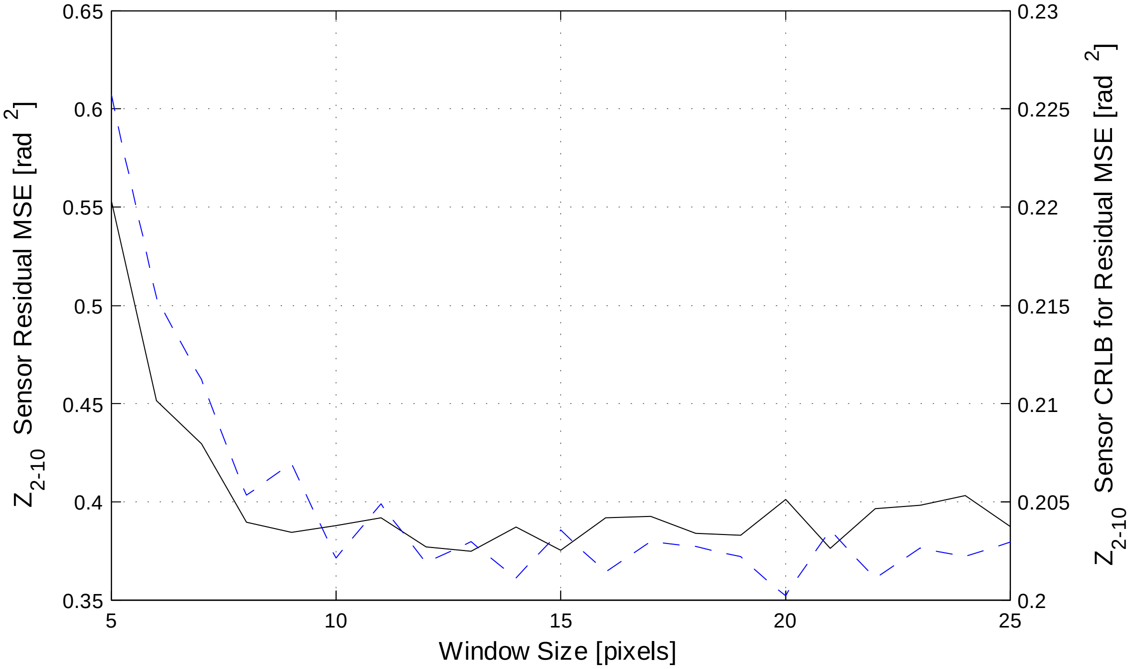


Figure 9.9 Z_{2-10} estimator $\left\langle P_{\phi_e}^2 \right\rangle$ versus N_W for $K = 100$ photons per subaperture.

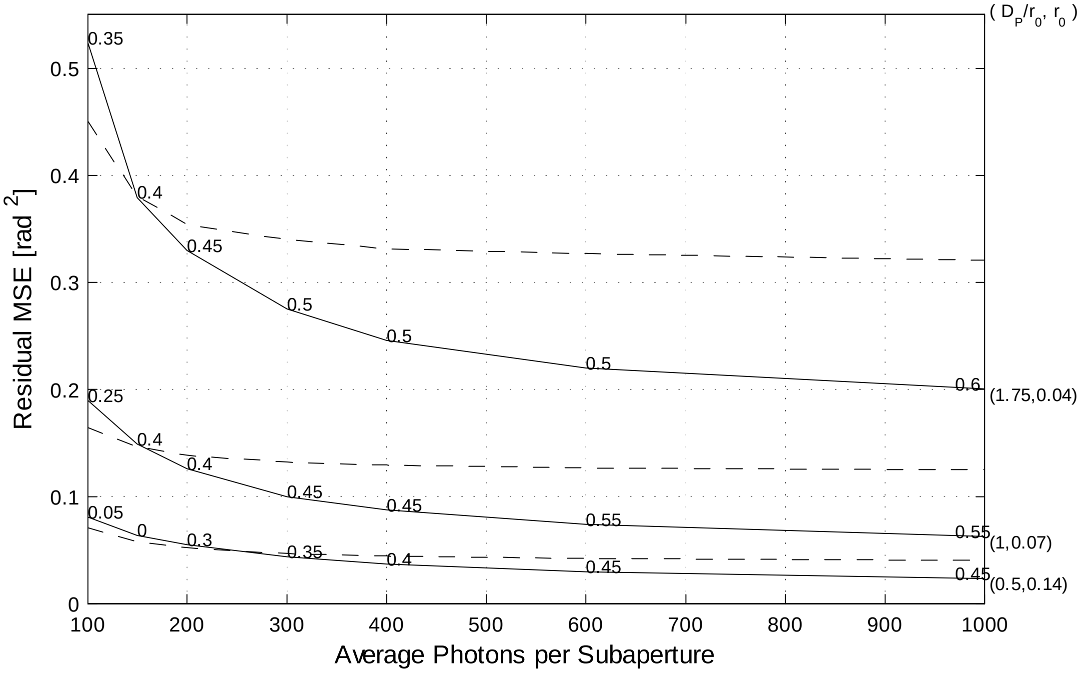


Figure 9.10 Simulated $\left\langle P_{\phi_e}^2 \right\rangle$ versus K for several cases of r_0. Dashed lines indicate $Z_{2,3}$ sensor (whole plane projections) performance. Solid lines indicate Z_{2-10} sensor (half plane projections) performance.

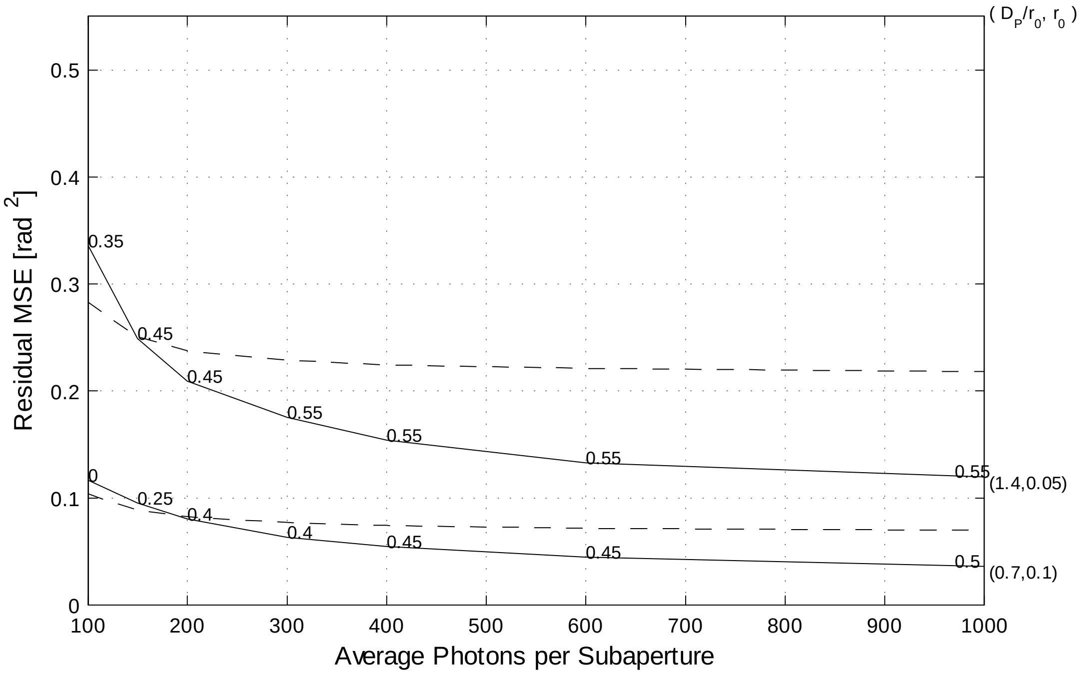


Figure 9.11 Simulated $\left\langle P^2_{\phi_e} \right\rangle$ versus K for several cases of r_0. Dashed lines indicate $Z_{2,3}$ sensor (whole plane projections) performance. Solid lines indicate Z_{2-10} sensor (half plane projections) performance.

$Z_{2,3}$ sensor, it is easy to see the advantages and disadvantages of estimating higher order parameters . The advantage of the Z_{2-10} sensor increases as SNR increases. The Z_{2-10} sensor does not provide significant improvement over the tilt only whole plane projection sensor when $K < 200$ photons per subaperture.

The next set of performance figures offer a comparison of the Z_{2-10} curvature sensor performance with the projection based ML tilt estimator and the two-dimensional centroid based estimator. Figures 9.12 and 9.13 overlay the centroid performance. There is no direct comparison of window length in the centroiding case with a projection length because the centroid estimator uses two-dimensional data. Additionally, the ideal centroiding window size changes over the operational space. For these reasons, I have included several centroid performance plot lines for each case of r_0 to demonstrate centroid performance at different window sizes. The set of colors {red, green, blue, cyan} correspond to the window sizes in the set $\{5, 6, 7, 8\}$ respectively. Figures 9.14 and 9.15 overlay the ML tilt estimator performance. The ML performance lines shown here are the same lines from Figure .

Figures 9.12 through 9.15 reveal that Z_{2-10} sensor performance is on par or better than both tilt estimators throughout the entire operational range.

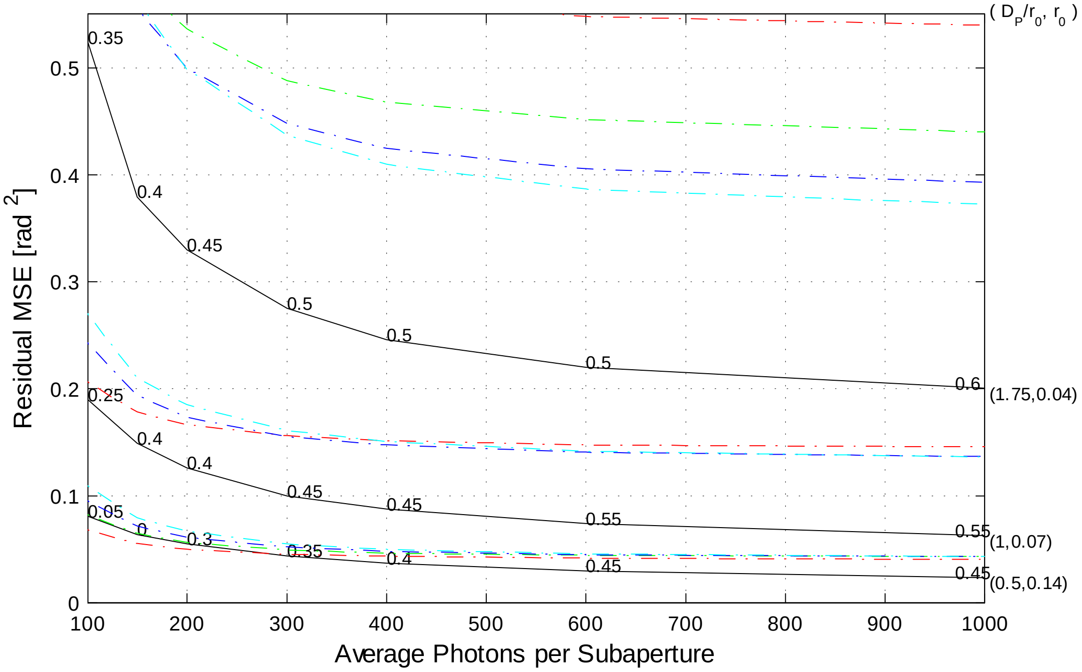


Figure 9.12 Comparison of simulated centroiding tilt estimator performance to the Z_{2-10} estimator over a range of r_0 and K values.

9.4 Sensitivity Analysis

This section contains performance results from a series of simulations in which the sensor was purposefully given erroneous estimates of the environment variables: r_0, L_0, and l_0. The intent is to provide a demonstration of the sensor's robustness under the influence of inaccurate environment variable estimates. Several design variables remain constant throughout each of the figures to follow: $D_P = 0.07$m, $N_W = 9$ pixels and $\sigma_{ro} = 2.13$. Also consistent throughout is the choice of plot line styles and their corresponding data series. Solid lines indicate Z_{2-10} sensor performance, while dashed lines indicate ML tilt sensor performance. The ML performance lines are included as comparison lines to highlight points where poor estimates of the environment variables negate the Z_{2-10} sensor's performance advantage. The true value of the environment variable for each plot line is indicated by the × symbol along the solid line and a ◯ symbol along the dashed line. Each point along the

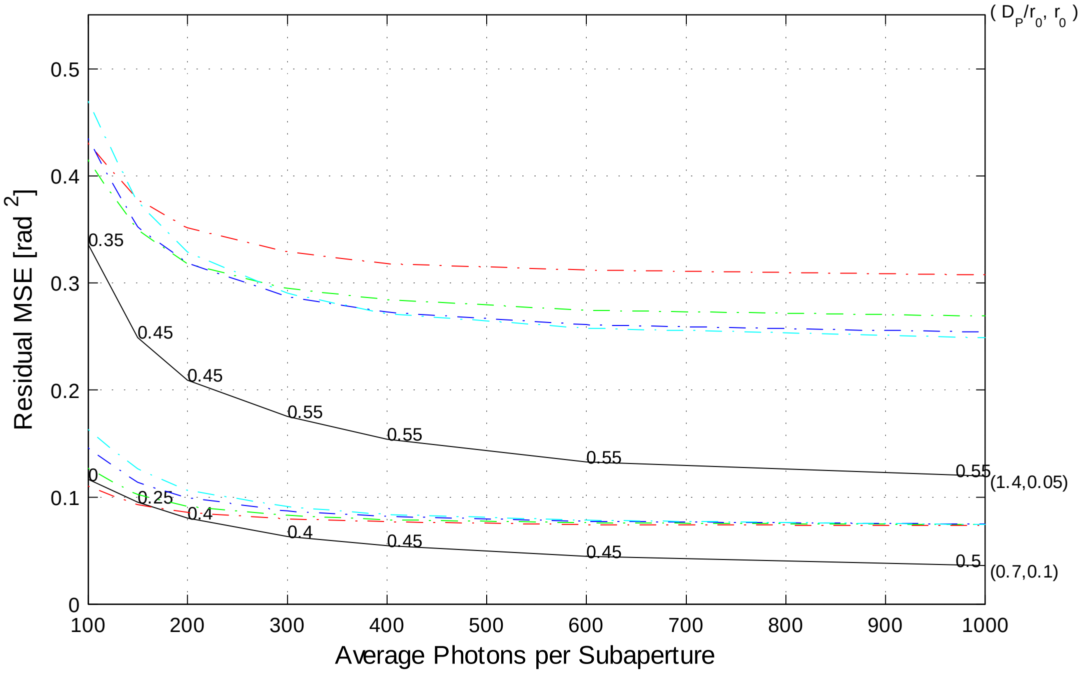


Figure 9.13 Comparison of simulated centroiding tilt estimator performance to the Z_{2-10} estimator over a range of r_0 and K values.

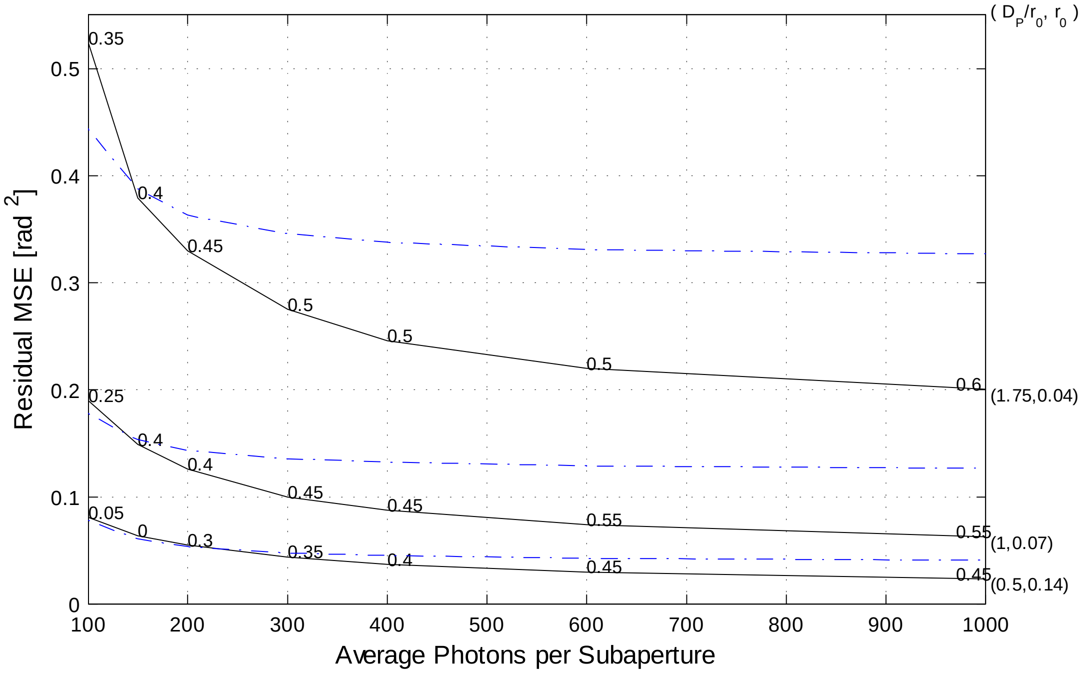


Figure 9.14 Comparison of simulated projection based ML tilt estimator performance to the Z_{2-10} estimator over a range of r_0 and K values.

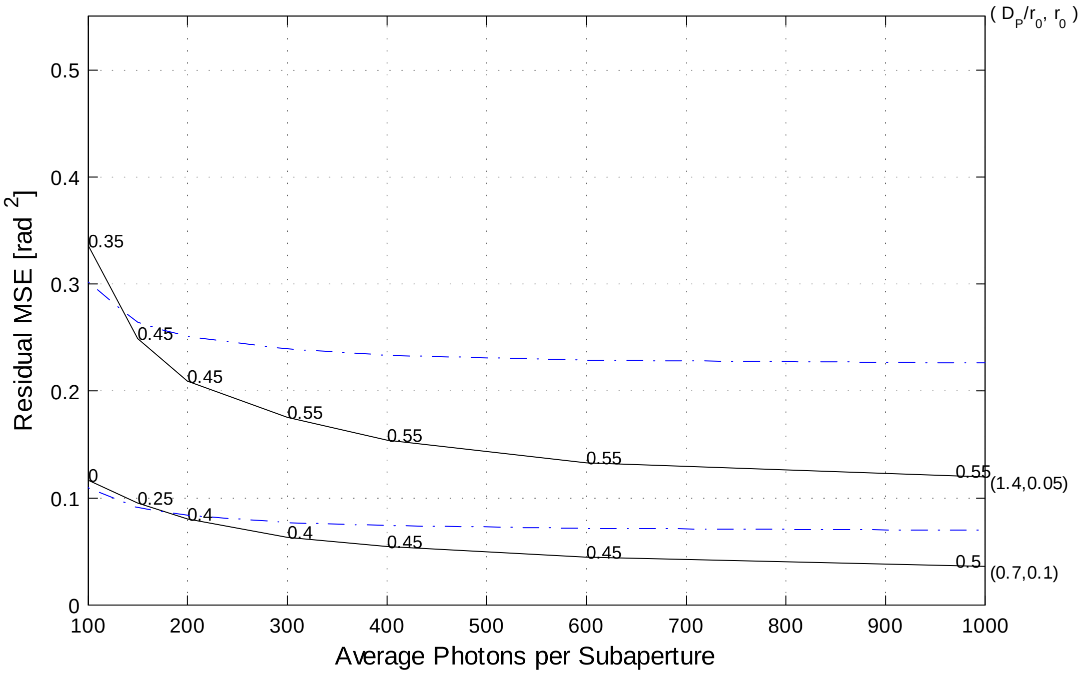


Figure 9.15 Comparison of simulated projection based ML tilt estimator performance to the Z_{2-10} estimator over a range of r_0 and K values.

solid performance lines represents an average value from 30 random cases. A confidence interval at each plot point is indicated by a pair of triangles: one pointing upward for $+1\sigma$ and one pointing downward for -1σ, where σ represents the 30 sample standard deviation.

Figure 9.16 demonstrates r_0 sensitivity in high SNR, $K = 1000$ photons, at three locations in the operating space. The true values of the parameter $r_0 = \{0.04\text{m}, 0.05\text{m}, 0.1\text{m}\}$ correspond to the three pairs of plot lines. As indicated by the x-axis labeling, the estimates of r_0 range from 0.02m to 0.2m. However, the entire range is not tested for each r_0 case. The range of test values is selected based on the true value of r_0. Figure 9.17 provides the same r_0 sensitivity analysis at low SNR: $K = 200$ photons. Figure 9.18 shows L_0 sensitivity in high SNR: $K = 1000$ photons. The true L_0 value is set at 10m while the estimate of L_0 is in the set {1m, 10m, 100m}. The three pairs of lines correspond to the previous r_0 cases: 0.04m, 0.05m, and 0.1m respectively. Just as in the r_0 analysis figures, the solid line indicates Z_{2-4} performance while the dashed line indicates ML tilt performance. The true value of L_0 for each performance line is indicated by the $\times$ symbol along the solid line and a $\bigcirc$ along the dashed line. Figure 9.19 provides the same L_0

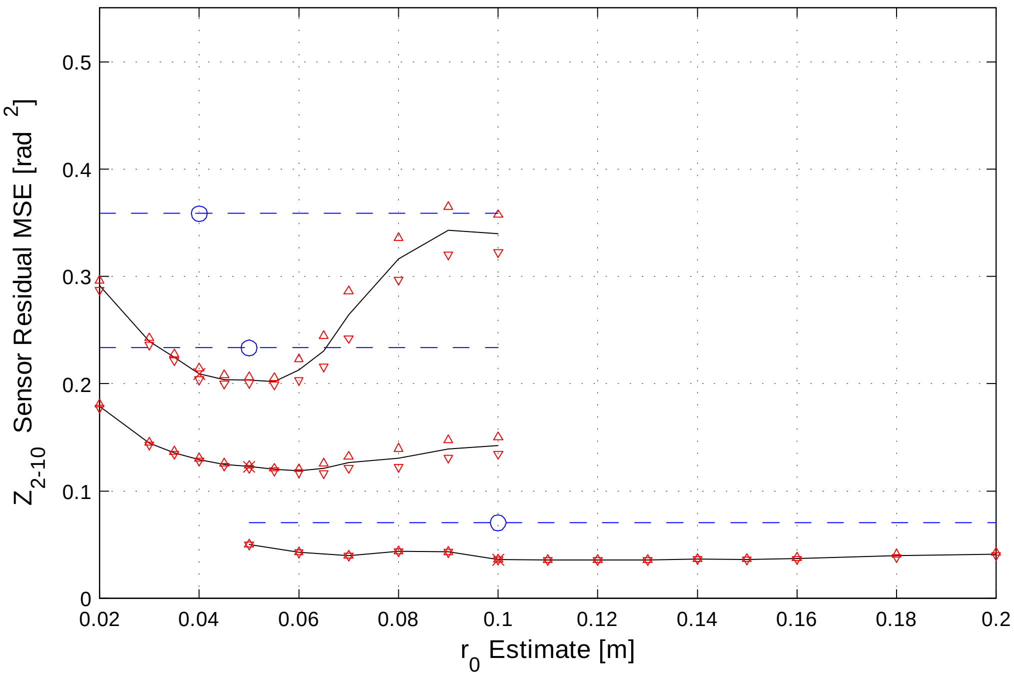


Figure 9.16 Solid lines indicate Z_{2-10} residual MSE versus r_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of r_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 1000$.

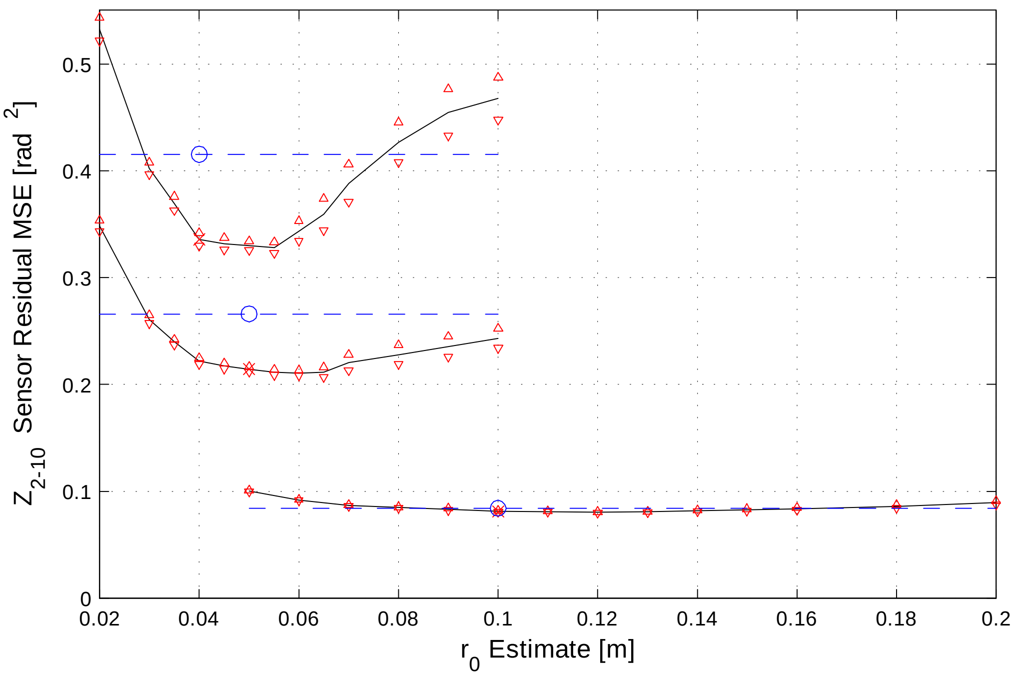


Figure 9.17 Solid lines indicate Z_{2-10} residual MSE versus r_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of r_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 200$.

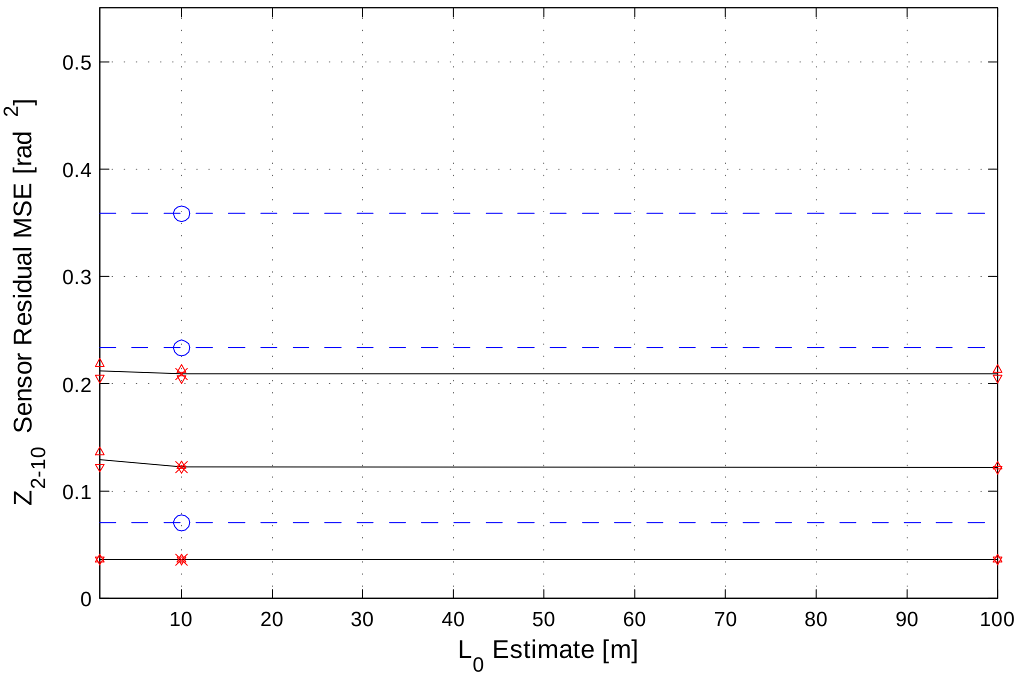


Figure 9.18 Solid lines indicate residual MSE versus L_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of L_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 1000$.

sensitivity analysis at low SNR: $K = 200$ photons. Figure 9.20 demonstrates l_0 sensitivity

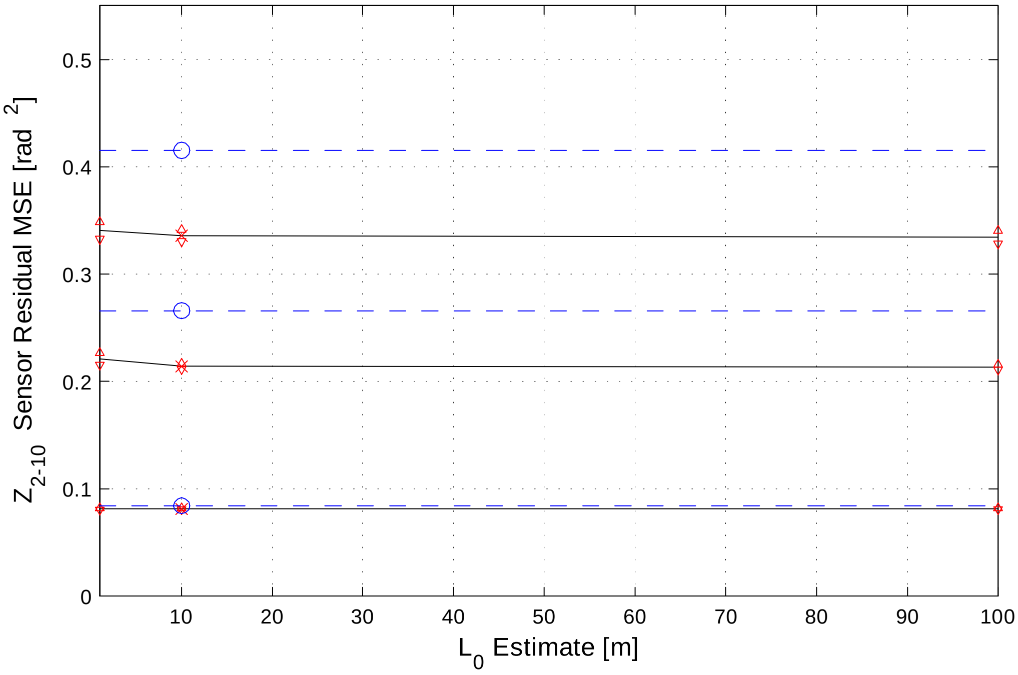


Figure 9.19 Solid lines indicate Z_{2-10} residual MSE versus L_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of L_0 is indicated by an $\times$ (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 200$.

in high SNR: $K = 1000$ photons. The true l_0 value is set at 0.01m while the estimate of l_0 is in the set {0m, 0.01m, 0.1m}. Once again the three pairs of plot lines correspond to the three r_0 cases. Figure 9.21 provides the same l_0 sensitivity analysis at low SNR: $K = 200$ photons.

The results of the sensitivity analysis demonstrate that the sensor performance depends on the r_0 estimate more so than the estimates of L_0 and l_0. In fact, the sensor's performance is nearly invariant to L_0 and l_0 estimates. The simulated cases indicate that when L_0 and l_0 are unknown, the best course of action is to overestimate L_0 (set $L_0 \geq 100$m) and set the l_0 estimate equal to zero. The sensitivity to r_0 is most evident in low SNR and at high $\frac{D_P}{r_0}$.

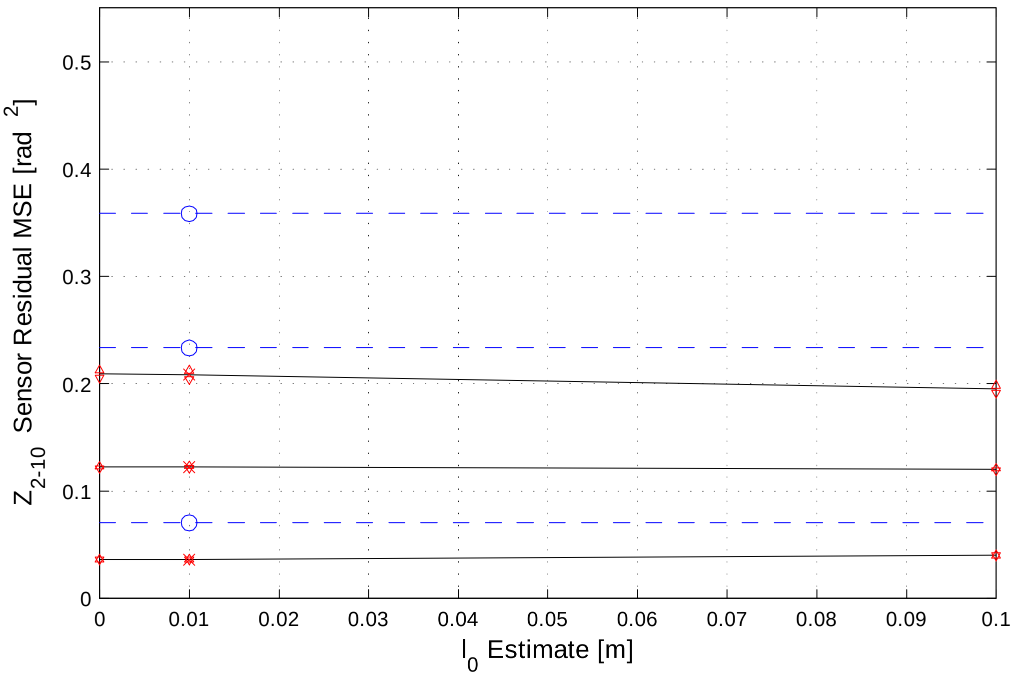


Figure 9.20 Solid lines indicate Z_{2-10} residual MSE versus l_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of l_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 1000$.

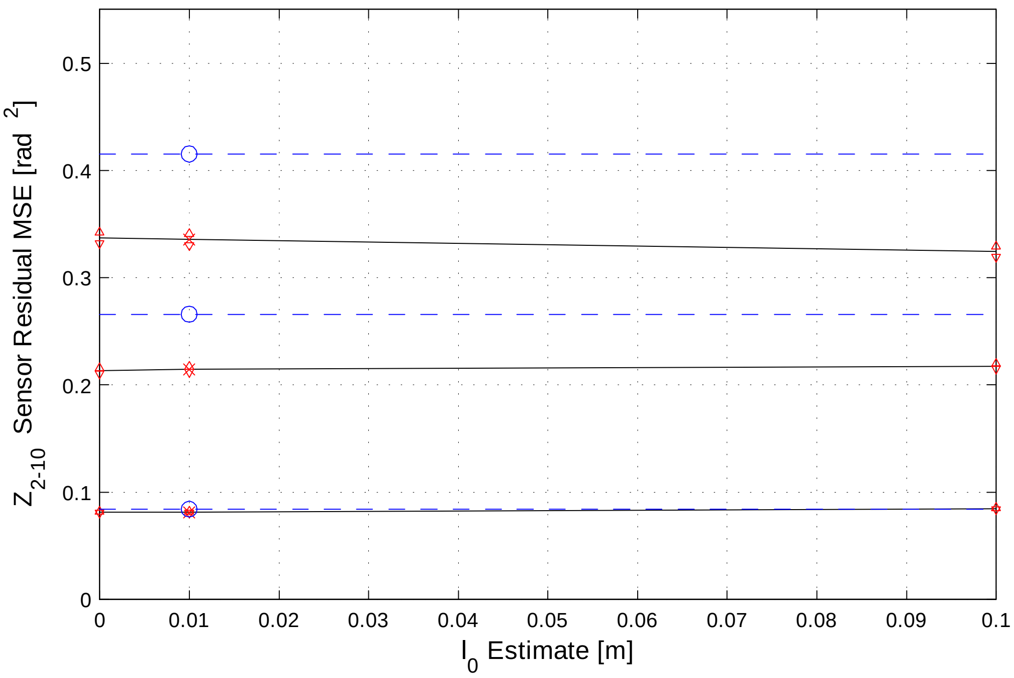


Figure 9.21 Solid lines indicate Z_{2-10} residual MSE versus l_0 estimate. Dashed lines represent the $Z_{2,3}$ ML estimator performance threshold. The true value of l_0 is indicated by an × (solid line) or circle (dashed line). Triangles indicate $\pm 1\sigma$. $K = 200$.

9.5 Summary

This chapter began with a brief description of the Z_{2-10} sensor and the half plane image projection. Simulated performance plots were provided in order to demonstrate the search for ideal design variable settings, and to provide an estimate of sensor performance over a typical range of the operating space. Performance was compared to the CRLB results. As was the case with the Z_{2-4} sensor, simulated Z_{2-10} sensor performance and the calculated CRLB results compliment each other. Once again, the similarity between simulation and CRLB results increases confidence in the simulation accuracy and suggests that CRLB is an adequate tool for making preliminary design choices for the Z_{2-10} sensor. A comparison of the Z_{2-10} sensor performance to the centroid and tilt only ML estimator revealed that, under the simulated conditions, the Z_{2-10} sensor performance is on par with or better than the other sensor designs for cases where the SNR is greater than 200 photons per subaperture.

10. Conclusion

The temperature and pressure throughout the atmosphere change constantly. These factors contribute significantly to the index of refraction which affects the way that light propagates through the atmosphere. In optical imaging systems, the atmosphere's turbulent nature has the effect of introducing an extended lens with randomly fluctuating aberrations into the optical path. Unfortunately, these random aberrations often represent the limiting factor in image resolution for a given optical system. Wherever applications require imaging through the atmosphere, adaptive optics (AO) systems offer the ability to improve the quality of imaging. Adaptive optics improve image quality by sensing and compensating for the random phase fluctuations injected by the atmosphere. The device responsible for detecting atmospheric phase aberrations is called a wavefront sensor. The focus of this research is the design and simulation testing of two new options for wavefront sensing. The AO wavefront sensors simulated in this research estimate lower order phase modes in the aperture from intensity measurements in the image plane. The difficulties involved in detecting wavefront phase from image intensity include: distinguishing lower order modes amid higher order modal effects and maintaining performance under low SNR and the influence of CCD read noise. This research demonstrates two projection based sensor designs which offer improvements in each of these categories over existing sensor designs.

10.1 Research Contributions

The primary research contributions are the development and simulation testing of two projection based wavefront curvature sensors for adaptive optics. The Z_{2-4} and Z_{2-10} wavefront sensors proposed here would replace the traditional physical devices used in interferometric or Hartmann style sensing devices with a beamsplitter, a pair of projection based CCDs and a computer processor. These wavefront sensors are designed to detect several low order harmonics which comprise a large percentage of the atmospheric aberrations. By improving tilt estimates and adding the ability to detect curvature modes, the proposed wavefront sensors are able to significantly reduce wavefront residual mean squared error and expand the operating space of existing sensor designs.

Due to the rate at which the atmospheric turbulence evolves, adaptive optics systems must operate using bandwidths greater than several hundred hertz. A significant portion of the AO loop is dedicated to the control electronics and manipulating the deformable mirror leaving only a few milliseconds to solve the wavefront sensing problem. CCD read out can represent a large portion of the sensor processing time when processing periods are on the order of a millisecond. To save time with CCD read out, previous research has proposed compressing two-dimensional images into vectors of pixels called image projections. The vector based wavefront sensor must detect wavefront phase from the compressed image information. This dissertation demonstrates that curvature modes up to the 10th Zernike polynomial can be estimated from image projections.

This research presents a unique modification to existing image projection techniques. I have developed two MAP estimators: the first of which uses whole plane image projections and the second uses half plane image projections. The MAP estimator requires the wavefront sensor to evaluate and minimize a Bayesian risk function. This process can be reduced to maximizing a likelihood function over multiple parameters. Maximizing the likelihood metric over the multidimensional parameter space inherent to the curvature estimation problem can present a challenging task. I have shown here that sufficient information exists in the image projections to estimate several lower order parameters independently. By estimating parameters independently, the multidimensional problem is reduced to several one-dimensional minimization problems. The accuracy and efficiency of the maximization process can be further enhanced by strategically applying a quadratic curve fit through the likelihood. Under the simulated conditions, Zernike tilt modes 2 and 3 can be estimated independently. Once the tilt estimates are formed, curvature modes Z_4 through Z_{10} can be independently estimated. The proposed sensors' performance in high SNR match the performance of a tilt only sensor while operating with a larger $\frac{D_P}{r_0}$ ratio. For systems currently using the tilt only sensor, this design offers the trade-off of maintaining current performance while increasing the subaperture size and hence reducing the total number of subapertures and the complexity of the AO system or enjoying improved performance at the same subaperture size.

The description of each wavefront sensor also includes a list of key design variables and operating variables. The development of these wavefront sensors included the derivation of a projection based performance bound. The performance bound was shown to provide an effective measure for selecting ideal design variable settings as a function of the operating environment. As an aside to prototype sensor simulation and testing, this dissertation developed the log polar sampling strategy for improving accuracy and isotropy in extensible phase screen generation.

10.2 Future Work

This research shows that it is possible to improve wavefront sensing performance over existing techniques in specific simulated conditions. The next step may be to prove whether a hardware implementation of the sensor design can provide similar benefits. Testing a hardware implementation of the projection based wavefront sensor with a narrowband source would suffice to demonstrate the type of CCD read noise rejection and the relative decrease in CCD read time that could be achieved over existing two-dimensional data sensors. A true test of the projection based sensor may be limited by the availability of specially designed CCD arrays. Although the technology exists to construct a fast, efficient projection based array, the cost of producing CCDs specifically designed for projection based wavefront sensing remains under investigation. In addition to hardware testing, the simulation testing can be improved. Increased complexity in the atmospheric simulation modeling might include temporal simulations using Taylor's frozen flow and multilayered atmospheric models.

The sensor software algorithm can be modified independent of the CCD hardware design. As such, projection based hardware research and testing can begin while enhancements to the curvature sensor algorithm continue to be a subject of future research. Perhaps the most limiting factor in the software algorithm is the need for a point source or guide star reference. Section 4.2 discussed techniques for estimating the object along with the unknown wavefront phase using phase diversity. The drawback was that phase diversity loops are computationally intense. This makes phase diversity difficult to implement in real time AO applications. The bandwidth problem suggests investigating the use of projections in frequency domain MSE estimators for phase diversity applications. For instance, the Gonsalves metric has been proven to work for two-dimensional image data, but it's performance using image projection data should be investigated. The Fourier analysis required for phase diversity applications is computationally expensive. Image projections would allow a one-dimensional transform versus a two dimensional transform significantly reducing the amount of computation required to implement the Gonsalves metric. In addition to reducing complexity via the one-dimensional transform, portions of the Fourier analysis could be precomputed much like the phase screen transform implementation. If a strategic set of OTF frequency points can be determined which provide better ability to estimate lower order Zernikes then the full phase diversity technique may be traded for a low order Zernike estimator capable of operating in higher bandwidth real time applications.

The sensitivity analysis demonstrated an interrelationship between the sensor performance and the quality of atmospheric parameter estimates. The use of this sensor design in conjunction with atmospheric parameter estimation techniques should be investigated. As an example, this sensor design might be used in a feedback loop configuration with existing r_0 estimators to improve the overall AO system performance.

The research here considered performance for a single sensor subaperture. Phase reconstruction using higher order modal estimates continues to be a subject of research. Future work in this area includes the design and performance of a fast wavefront reconstruction method using Zernikes 2 through 10.

In the area of phase screen generation, I suggest trying to reduce the percent error in screen structure outside the inertial range. Large spatial correlation errors created by the inherent periodicity of the Fourier transform might be reduced via implementation using some other transform technique that allows for decorrelating long spatial distances within the screen realizations. It may also be worth while to investigate the use of digital filtering techniques and spectral estimation to further reduce error in phase screen generation via a feedback loop. For instance, the spectrum could be estimated using screen realizations. If the choice of spectral sample points was somehow tied to the screen structure error then the choice of sample locations could be modified to reduce that error. This would provide a measure with which to define the ideal set of frequency sample locations for a given number of frequency sample points.

Bibliography

1. S. Cain. "Design of an image projection correlating wavefront sensor for adaptive optics," *Optical Engineering*, *43*(7):1670–1681 (2004).
2. E. Hecht. *Optics* (4th Edition). Pearson Education, Inc., 2002.
3. M. C. Roggemann and B. M. Welsh. *Imaging through Turbulence*. CRC Press LLC, 1996.
4. I. Newton. "Optics." *Optics 34*. Great Books of the Western World, edited by R. M. Hutchins, Chicago: Encyclopedia of Brittanica, 1952.
5. J. W. Strohbehn, S. F. Clifford, M. Gracheva, A. Ishimaru, J. H. Shapiro, J. L. Walsh, and P. B. Ulrich. *Laser Beam Propagation in the Atmosphere*, *25*. Topics in Applied Physics. Springer-Verlag, 1978.
6. R. K. Tyson. *Adaptive Optics Engineering Handbook*. Optical Engineering, Marcel Dekker, Inc., 2000.
7. R. J. Noll. "Zernike polynomials and atmospheric turbulence," *Journal of the Optical Society of America*, *66*(3):207–211 (March 1976).
8. H. L. van Trees. *Detection, Estimation, and Modulation Theory: Part I. Detection, Estimation, and Linear Modulation Theory*. John Wiley & Sons, Inc., 2001.
9. P. M. B. Vitanyi. "Andrei Nikolaevich Kolmogorov," *CWI Quarterly*, *1*:3–18 (1988).
10. S. Corrsin. "On the Spectrum of Isotropic Temperature Fluctuations in an Isotropic Turbulence," *Journal of Applied Physics*, *22*(4):469–473 (1951).
11. L. C. Andrews and R. L. Phillips. *Laser Beam Propagation through Random Media*. SPIE Press, 1998.
12. J. W. Strohbehn. "Line-of-Sight Wave Propagation Through the Turbulent Atmosphere," *Proceedings of the IEEE*, *56*:1301 (1968).
13. A. Ishimaru. *Wave Propagation and Scattering in Random Media: Multiple Scattering, Turbulence, Rough Surfaces, and Remote Sensing*, *2*. Academic Press, 1978.
14. E. M. Johansson and D. T. Gavel. "Simulation of stellar speckle imaging," *Amplitude and Intensity Spatial Interferometry II*, *2200*(1):372–383 (1994).
15. D. L. Fried. "Optical Heterodyne Detection of an Atmospherically Distorted Signal Wave Front," *Proceedings of the IEEE*, *55*(1):57–67 (January 1967).
16. V. I. Tatarskii. *Wave Propagation in a Turbulent Medium*. McGraw-Hill Book Company, Inc., 1961.
17. I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products* (6th Edition). Academic Press, 2000.
18. D. Malacara. "Mathematical interpretation of radial shearing interferometers," *Applied Optics*, *13*(8):1781 (August 1974).

19. M. Born and E. Wolf. *Principles of Optics* (7th Edition). Cambridge University Press, 1999.

20. J. W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill Electrical and Computer Engineering Series, The McGraw-Hill Companies, Inc., 1996.

21. E. Kreyszig. *Advanced Engineering Mathematics*. John Wiley and Sons, Inc., 1999.

22. A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes* (4th Edition). McGraw-Hill, 2002.

23. D. R. Gerwe and P. S. Idell. "Cramer-Rao analysis of orientation estimation: viewing geometry influences on the information conveyed by target features," *J. Opt. Soc. Am. A*, *20*(5):797 (May 2003).

24. J. W. Goodman. *Statistical Optics*. John Wiley & Sons, Inc., 1985.

25. J. M. Geary. *Introduction to Optical Testing*, *TT15*. Tutorial Texts. SPIE Press, 1993.

26. J. M. Geary. *Intoduction to Wavefront Sensors*, *TT18*. Tutorial Texts. SPIE Press, 1995.

27. J. W. Hardy and A. J. MacGovern. "Shearing Interferometry: a Flexible Technique for Wavefront Measurement." *Interferometric Metrology816*. 180–195. 1987.

28. J. C. Wyant. "Use of an AC heterodyne lateral shear interferometer with real-time wavefront correction systems," *Applied Optics*, *14*(11):2622–2626 (1975).

29. J. C. Wyant. "Double Frequency Grating Lateral Shear Interferometer," *Applied Optics*, *12*(9):2057 (September 1973).

30. J. C. Wyant. "OTF Measurements with a White Light Source: An Interferometric Technique," *Applied Optics*, *14*(7):1613 (July 1975).

31. D. Kelsall. "Rapid Interferometric Technique for MTF Measurements in the Visible or Infrared Region," *Applied Optics*, *12*(7):1398 (July 1973).

32. M. P. Rimmer. "Method for Evaluating Lateral Shearing Interferograms," *Applied Optics*, *13*(3):623 (March 1974).

33. D. C. Ghiglia and M. D. Pritt. *Two-dimensional Phase Unwrapping: Theory, Algorithms, and Software*. John Wiley & Sons, Inc., 1998.

34. R. Smartt and W. Steel. "Theory and Application of Point Diffraction Interferometers," *Japan J. Appl. Phys.*, *14*:351 (1975).

35. J. E. Millerd, S. J. Martinek, N. J. Brock, J. B. Hayes, and J. C. Wyant. "Instantaneous phase-shift point-diffraction interferometer." *Optical Data Storage 2004 5380*, edited by B. V. K. V. Kumar and H. Kobori. 422–429. SPIE, 2004.

36. T. A. Rhoadarmer, J. D. Barchers, J. D. Gonglewski, M. A. Vorontsov, M. T. Gruneisen, S. R. Restaino, and R. K. Tyson. "Noise analysis for complex field estimation using a self-referencing interferometer wave front sensor." *High-Resolution Wavefront Control: Methods, Devices, and Applications IV 4825*, edited by J. D. Gonglewski. 215–227. SPIE, 2002.

37. G. D. Love, T. J. D. Oag, and A. K. Kirby. “Common Path Interferometric Wavefront Sensor for Extreme Adaptive Optics,” *Optics Express*, *13*(9):3491 (May 2005).

38. M. A. Vorontsov, E. W. Justh, and L. A. Beresnev. *Advanced Phase-Contrast Techniques for Wavefront Sensing and Adaptive Optics*. Technical Report CDCSS TR 2001-4, Center for Dynamics and Control of Smart Structures, 2001.

39. W. O. Saxton. *Computer Techniques for Image Processing in Electron Microscopy*. Academic Press, 1978.

40. Misell. “A Method for the Solution of the Phase Problem in Electron Microscopy,” *Journal of Physics D: Applied Physics*, *6*:L6 (1973).

41. R. A. Gonsalves. “Fundamentals of Wavefront Sensing by Phase Retrieval.” *Wavefront Sensing351*. 1982.

42. R. W. Gerchberg and W. O. Saxton. “A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures,” *Optik*, *35*:237 (1972).

43. J. R. Fienup. “Reconstruction of an Object from the Modulus of its Fourier Transform,” *Optics Letters*, *3*(1):27 (July 1978).

44. J. R. Fienup. “Iterative Method Applied to Image Reconstruction and to Computer-Generated Holograms,” *Optical Engineering*, *19*(3):297 (May 1980).

45. R. A. Gonsalves. “Phase Retrieval from Modulus Data,” *J. Opt. Soc. Am.*, *66*(9):961 (September 1976).

46. R. A. Gonsalves. “Phase retrieval and diversity in adaptive optics,” *Optical Engineering*, *21*(5):829–832 (September/October 1982).

47. R. L. Kendrick, D. S. Acton, and A. L. Duncan. “Phase-diversity Wave-front Sensor for Imaging Systems,” *Applied Optics*, *33*(27):6533 (September 1994).

48. B. M. Welsh, M. C. Roggemann, B. L. Ellerbroek, and T. L. Pennington. “Fundamental performance comparison of a Hartmann and a shearing interferometer wave-front sensor,” *Applied Optics*, *34*(21):4186–4195 (1995).

49. H. T. Barclay. “The SWAT Wavefront Sensor,” *Lincoln Lab. J.*, *5*(1):115 (1992).

50. J. Holder, S. C. Cain, and P. Mantica. “Efficient implementation of a projection-based wavefront sensor,” *Advanced Signal Processing Algorithms, Architectures, and Implementations XII*, *4791*(1):116–124 (2002).

51. S. C. Cain, N. M. Hayat, and E. E. Armstrong. “Projection-Based Image Registration in the Presence of Fixed-Pattern Noise,” *IEEE Transactions on Image Processing*, *10*(12):1860–1872 (December 2001).

52. D. J. Lee, B. M. Welsh, and M. C. Roggemann. “Cramer-Rao analysis of phase diversity imaging.” *Image Reconstruction and Restoration II3170*, edited by T. J. Schulz. 161–172. SPIE, 1997.

53. J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin. “Hubble Space Telescope characterized by using phase-retrieval algorithms,” *Appl. Opt.*, *32*(10):1747– (April 1993).

54. K. S. Shanmugan and A. M. Breipohl. *Random Signals: Detection Estimation and Data Analysis*. John Wiley & Sons, Inc., 1988.

55. C. Braccini and A. V. Oppenheim. “Unequal bandwidth spectral analysis using digital frequency warping,” *IEEE Transactions, ASSP*, (22):236–244 (1974).

56. R. G. Lane, A. Glindemann, and J. C. Dainty. “Simulation of a Kolmogorov phase screen,” *Waves in Random Media*, *2*(3):209–224 (1992).

57. B. J. Herman and L. A. Strugula. “Method for inclusion of low-frequency contributions in numerical representation of atmospheric turbulence.” *Propagation of High-Energy Laser Beams through the Earth’s Atmosphere*. 1990.

Vita

Jonathan Buffington graduated Valedictorian from Lamar County Comprehensive High School in 1992. In May 1996, he received a B.S. in Electrical Engineering from the United States Air Force Academy and was commissioned a Second Lieutenant in the United States Air Force. In August of 1996, Buffington was assigned to the Electronic Warfare Directorate at Warner Robins Air Logistics Center, Warner Robins Air Force Base Georgia where he maintained software block cycles on a countermeasures dispenser system. In 1999 he became a spacecraft engineer for Air Force Space Command, 50th Space Wing, Fourth Space Operations Squadron, Schriever Air Force Base Colorado. He quickly rose to Chief of Spacecraft Engineering and then transitioned to Chief of the Orbital Operations Section. While working in his first two assignments, Buffington completed an M.S. in Electrical and Computer Engineering from The Georgia Institute of Technology. In August of 2003, Buffington was assigned to the Air Force Institute of Technology, Wright-Patterson Air Force Base Ohio to pursue a Ph.D. in Space Systems Engineering. Upon graduation, Major Buffington will be assigned to the Washington DC area.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704–0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704–0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD–MM–YYYY)* 21-12-2006	**2. REPORT TYPE** Doctoral Dissertation	**3. DATES COVERED** *(From — To)* September 2003 — December 2006

4. TITLE AND SUBTITLE WAVEFRONT CURVATURE SENSING FROM IMAGE PROJECTIONS	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Buffington, Jonathan C., Major, USAF	**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology Graduate School of Engineering and Management 2950 Hobson Way WPAFB OH 45433-7765	**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT/DS/ENG/07-01
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approval for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research outlines the development and simulation of a signal processing approach to real time wavefront curvature sensing in adaptive optics. The signal processing approach combines vectorized Charge Coupled Device (CCD) read out with a wavefront modal estimation technique. The wavefront sensing algorithm analyzes vector projections of image intensity data to provide an estimate of the wavefront phase as a combination of several low order Zernike polynomial modes. This wavefront sensor design expands on an existing idea for vector based tilt sensing by providing the ability to compensate for additional modes. Under the proposed wavefront sensing approach, the physical wavefront sensor would be replaced by a pair of imaging devices capable of generating vector projections of the image data. Using image projections versus two-dimensional image data allows for faster CCD read out and decreased read noise.

15. SUBJECT TERMS

adaptive optics; wavefront sensing; maximum a posteriori estimator; image projection

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			LtCol Matthew E. Goda, PhD, (AFRL/SND)
U	U	U	UL	230	**19b. TELEPHONE NUMBER** *(include area code)* (937) 904–9819