Theses and Dissertations   Student Graduate Works

6-1997

Optimization Considerations for Adaptive Optics Digital Imagery Systems

Robert T. Brigantic

Recommended Citation

Brigantic, Robert T., "Optimization Considerations for Adaptive Optics Digital Imagery Systems" (1997). *Theses and Dissertations*. 5829.
https://scholar.afit.edu/etd/5829

AFIT/DS/ENS/97-02

OPTIMIZATION CONSIDERATIONS
FOR
ADAPTIVE OPTICS DIGITAL IMAGERY SYSTEMS

DISSERTATION

Robert T. Brigantic, Major, USAF

AFIT/DS/ENS/97-02

DTIC QUALITY INSPECTED 3

Disclaimer Statement

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

OPTIMIZATION CONSIDERATIONS

FOR

ADAPTIVE OPTICS DIGITAL IMAGERY SYSTEMS

DISSERTATION

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University In Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Robert T. Brigantic, B.S., M.S.

Major, USAF

June 1997

OPTIMIZATION CONSIDERATIONS

FOR

ADAPTIVE OPTICS DIGITAL IMAGERY SYSTEMS

Robert T. Brigantic, B.S., M.S.
Major, USAF
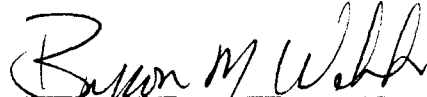
Approved:

_____ $5\text{-}30\text{-}97$
Michael C. Roggemann, Cochairman

_____ $5\text{-}21\text{-}97$
Kenneth W. Bauer, Cochairman

_____ $5\text{-}21\text{-}97$
Byron M. Welsh, Committee Member

_____ $5\text{-}21\text{-}97$
Theodore E. Luke, Dean's Representative

Accepted:

_____
Robert A. Calico
Dean, School of Engineering

Approved for public release; distribution unlimited

Acknowledgments

I am extremely indebted to my dissertation committee cochairmen Dr. Kenneth W. Bauer and Major Michael C. Roggemann and committee member Dr. Byron M. Welsh. In conducting this research and in preparing this dissertation they all provided invaluable help and insight. I greatly enjoyed our weekly meetings and the camaraderie which developed among us, and which contributed to making this venture a real team effort.

I would like to give a special thanks to Dr. Bauer who recommended that I pursue a dissertation topic that would be both interesting and enjoyable to myself -- this research effort certainly proved to do just that. Also, I want to thank Major Roggemann and Dr. Welsh who endured a plethora of questions on atmospheric physics, Fourier optics, complex random phasors, and the like with great patience and earnest response.

I acknowledge sponsorship and funding to attend technical conferences that was provided by the Starfire Optical Range, Kirtland Air Force Base and the Air Force Maui Optical Station.

I want to thank my family who endured all the long hours and time that was taken away from them and devoted to my PhD program, all without complaint. Above all, I must thank my Father who inspired and implanted upon me the importance of education, learning, and teamwork. I love you Dad and this dissertation is dedicated to you and your memory.

Robert T. Brigantic

iii

Table of Contents

List of Figures

List of Tables

List of Most Commonly Used Symbols

AOS	Adaptive Optics System
AMOS	Air Force Maui Optical Station
d	Wavefront Sensor Side Dimension (m)
D	Telescope Pupil Diameter (m)
DM	Deformable Mirror
DOE	Design of Experiments
f_c	Closed Loop Bandwidth (Hz)
f_g	Greenwood Frequency (Hz)
k_g	System Gain (radians/s)
ISR	Instantaneous Strehl Ratio
MSR	Modified Strehl Ratio
MTF	Modulation Transfer Function
m_v	Object Visual Magnitude
NEC	Noise-Effective-Cutoff Frequency
OTF	Optical Transfer Function
RMSE	Root-Mean-Square Error
RSM	Response Surface Methodology
r_0	Fried Parameter (m)
SR	Strehl Ratio
SNR	Signal-to-Noise Ratio
v	Wind Velocity (m/s)
WFS	Wavefront Sensor

Abstract

This dissertation had three objectives. The first objective was to develop image quality metrics that characterize Adaptive Optics System (AOS) performance. The second objective was to delineate control settings that maximize AOS performance. The third objective was to identify and characterize trade-offs between fully and partially compensated adaptive.

For the first objective, three candidate image quality metrics were considered: the Strehl ratio, a novel metric that modifies the Strehl ratio by integrating the modulus of the average system optical transfer function to a "noise-effective-cutoff" frequency at which some specified image spectrum signal-to-noise-ratio level is attained, and the noise-effective-cutoff frequency. It was shown that these metrics are correlated with the root-mean-square error between the detected image and the associated diffraction limited image and that they have traits that make them desirable for AOS performance metrics. For the second objective, optimum closed loop bandwidth settings were determined as a function of target object light levels and atmospheric seeing conditions. A strategy for selecting the closed loop bandwidth to provide robust system performance was also developed. For the third research objective, a qualitative assessment of trade-offs between fully compensated and partially compensated adaptive optics systems was provided.

OPTIMIZATION CONSIDERATIONS

FOR

ADAPTIVE OPTICS DIGITAL IMAGERY SYSTEMS

I. Introduction

General Issue

Ever since its invention by Galileo Galilei in 1610, the performance of the astronomical

telescope has been limited by atmospheric turbulence [1:36]. Turbulence manifests itself by

random distortions of incoming light. In turn, these distortions limit the resolution of optical

systems to less than that of their theoretical design limits [2]. It is a regrettable fact of nature that

light that may have originated from stellar or galactic sources and traveled for millions of light

years in perfect order is jumbled by traveling a mere 100 miles through our atmosphere.

While early telescopes were far from flawless instruments, the effects of the atmosphere on

astronomical images were quickly noted. For instance, Sir Isaac Newton, inventor of the still

popular Newtonian reflector telescope, could foresee no solution to the problem of atmospheric

turbulence. Newton writes [3:3-4]:

> *"If the Theory of making Telescopes could at length be fully brought into Practice,*
> *yet there would be certain Bounds beyond which Telescopes could not perform. For*
> *the Air through which we look upon the Stars, is in perpetual Tremor; . . . . . The only*
> *Remedy is a most serene and quiet Air, such as may perhaps be found on the tops of the*
> *highest Mountains."*

Newton certainly had keen insight, as evidenced by the placement of world renowned

astronomical observatories on mountain tops such as Lick Observatory located on California's

Mount Hamilton and Keck Observatory located on Hawaii's Mauna Kea. The National

Aeronautics and Space Administration Gerard P. Kuiper Airborne Observatory illustrates another

example. In this case, the degrading effects of the atmosphere are significantly reduced by flying a 36-inch diameter aperture telescope aboard a C-141 aircraft up to altitudes 45,000 feet [4]. The Hubble Space Telescope (HST) demonstrates the ultimate in systems designed to overcome atmospheric disturbances by completely eliminating the atmosphere.

Still, the HST has confining factors, such as limits in imaging very dim objects due to the size of its primary mirror compared to that of larger ground based telescopes that can collect more light. The HST's 2.4 meter aperture also implies an theoretical resolution limit that is inferior to ground based systems having larger apertures. Furthermore, due to HST issues such as immense demands and staggering costs, ground-based observatories will certainly continue as the predominant systems used to explore astronomical inquiries. As such, in addition to being located on mountain tops, additional methods to compensate for atmospheric distortions have been devised and are continuing to be developed and refined. One of the most promising class of these methodologies is Adaptive Optics Systems.

Adaptive Optics Systems

In the broadest sense, an Adaptive Optics System (AOS) attempts to improve optical performance by using information about the environment through which an optical signal passes. Human vision is an example of an AOS. As our eyes see, the brain continually interprets the captured image, determines any necessary corrections, and applies that correction through appropriate movement of various eye muscles [3:1-2].

Babcock is acknowledged as being the first to propose using electro-optical components to compensate for atmospheric turbulence affecting telescopic systems. Specifically, he proposed using a sensing device to perceive and compensate for atmospheric distortions via closed-loop control of a deformable mirror inserted in the optical path of the telescope system [3:4]. The

basic premise of adaptive optics is that one can improve the performance of an optical system by modifying the optical elements of that system [3:5]. Due to technological advancements in recent years, Babcock's ideas have been implemented at several of today's astronomical observatories. Most notably and of particular interest are those developed and operated by the United States Air Force at the Air Force Maui Optical Station (AMOS) and the Air Force Starfire Optical Range (SOR).

AMOS is operated by the United States Air Force Materiel Command's (AFMC) Phillips Laboratory. It is situated at the top of Hawaii's Mount Haleakala on the island of Maui. At an altitude of approximately 10,000 feet, the AMOS site has a relatively stable climate of clear, dry air and minimal light pollution, thus making it an ideal location for an astronomical observatory [5]. AMOS is an integral component of the Maui Space Surveillance Site (MSSS), which is also comprised of the Maui Optical Tracking and Identification Facility (MOTIF) and a Ground-based Electro-Optical Deep Space Surveillance (GEODSS) site operated by the Air Force Space Command.

AMOS itself is a "state-of-the-art electro-optical facility" whose mission is to conduct research and development of new and evolving electro-optical sensors and associated equipment [5]. In addition, it also supports operational missions as defined by the United States and Air Force Space Commands. At the heart of AMOS is a 1.6 meter diameter telescope equipped with an adaptive optics system. In addition, construction of a 3.67 meter diameter telescope, called the Advanced Electro-Optical System (AEOS), has recently started and is scheduled for initial operation sometime in 1997. This system will also be equipped with AOS components to compensate for atmospheric turbulence.

The SOR is also operated by the AFMC's Phillips Laboratory. Positioned at an elevation of 6,200 feet, it is located at an isolated part of Kirtland Air Force Base in Albuquerque, New

Mexico. The SOR's primary mission is to develop optical sensing, imaging, and propagation technologies to support Air Force aerospace missions [6]. This includes development of methods to remove atmospheric distortions from images of space objects. In addition to the predetection and correction approach of an AOS, these methods include computer post-processing techniques such as speckle and hybrid imaging [6].

The SOR is equipped with a 1.5 meter aperture telescope and a 3.5 meter aperture telescope, both with AOS capabilities. The 3.5 meter telescope is the world's largest optical instrument capable of tracking low earth-orbiting satellites. When coupled with an AOS, this telescope may be capable of achieving a resolution of about 0.04 arc seconds -- better than that of even the Hubble Space Telescope (HST). This is due to the larger diameter aperture of the SOR's 3.5 meter telescope. The HST has an aperture of 2.4 meters [7]. It is also intended to further develop and refine the use of "laser beacon" adaptive optics using the 3.5-meter telescope. Laser beacon adaptive optics make use of laser beams to excite atoms in the upper-atmosphere to emit light. In turn, this light is scattered back to the telescope to drive the control systems of the AOS [8].

Figure 1 shows a block diagram for a telescope equipped with a typical adaptive optics configuration. Key elements of this AOS are the telescope with its primary mirror, a deformable mirror, a beam splitter, an imaging sensor, an actuator command computer, and a wavefront sensor [9:7-10].

Incident light from a target object enters the telescope and strikes the primary mirror of the telescope. The theoretical resolution of a telescope, the diffraction limit, is determined by the diameter of the primary mirror and wavelength of light. Most modern telescope systems have optics that could approach this limit. The Rayleigh criterion for the theoretical angular resolution of a telescope, θ_R, in radians is [8]

4

$$\theta_R = 1.22 \; \lambda \, / \, D \qquad\qquad (1)$$

where λ is the wavelength of the light in meters and D is the diameter of the primary mirror in meters. Due to the degrading effects of the atmosphere on the incoming light, this limit can seldom be achieved by ground-based systems without atmospheric compensation techniques.
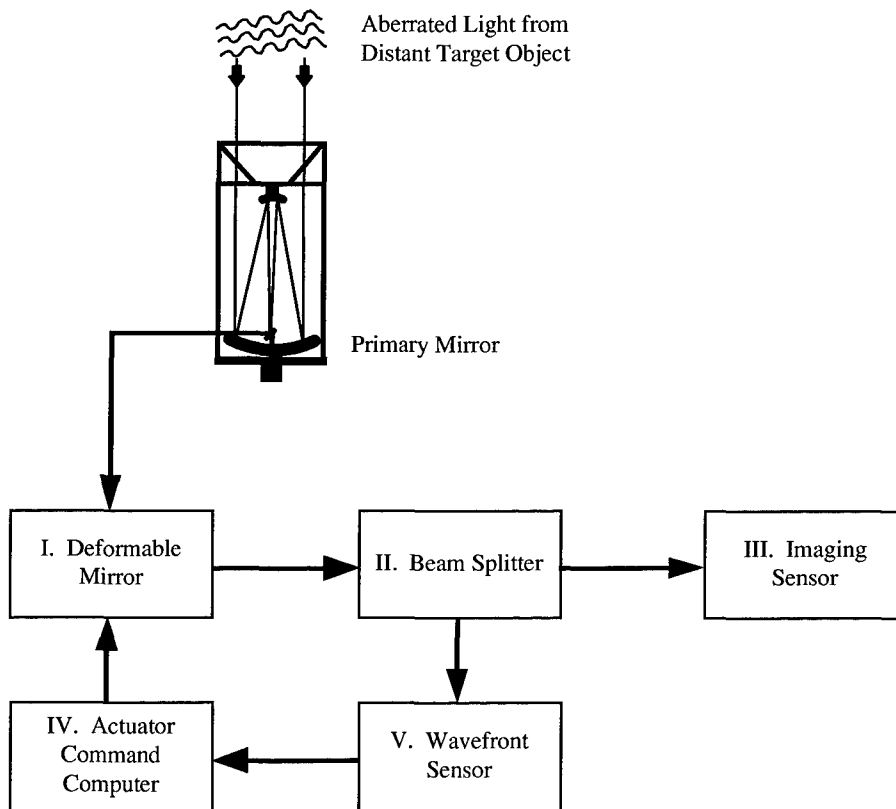
Figure 1. Adaptive Optics System Block Diagram [9:9]

After striking the primary mirror, this light is eventually focused through a series of smaller optical elements onto a Deformable Mirror (DM). The DM is a mirror whose surface can be altered in to match the shape of the incoming light [3:190]. In short, this has the effect of making

light that is aberrated (i.e., non-planar light waves) planar or nearly planar. Since the atmosphere is continuously and randomly changing, if the DM can be altered rapidly and accurately enough, the degrading effects of the atmosphere can be significantly mitigated. Many different DM designs exist. The most common of these are mirror surfaces mounted with piezo-electric actuators that expand or contract as a result of applied voltages to change the shape of the mirror [9:7].

After the DM, the light passing through the AOS system encounters a beam splitter which sends a portion of the light to the imaging sensor, typically a Charged-Coupled Device (CCD) camera, and to the Wavefront Sensor (WFS). The WFS measures gradients in the "phase aberration" across small regions in the telescope pupil, which are called subapertures [9:7]. Typically light from a reference star, often called a beacon, within the same field of view as the target object is used to drive the WFS. Light from the object itself can also be used to drive the WFS. This information is processed by the actuator command computer, which in turn, sends voltage signals to the DM in order to obtain the proper shape of the deformable mirror. One cycle of this closed-loop process must take place at speeds fast enough to keep up with the rate of change of the turbulence-induced phase error [9:7].

Hybrid Imaging Systems

Though the above overview might make adaptive optics seem relatively simple, they are tremendously complicated systems. Moreover, as the aperture of the primary mirror is made larger, such systems become increasingly complicated in terms of sampling requirements, control rates, and data transfer speeds. These issues all translate into increased system costs [9:10-11]. Consequently, simpler adaptive optics systems are being explored. These simpler systems are

known as partially compensated adaptive optics systems. They are characterized by fewer, larger WFS subapertures, and fewer DM degrees-of-freedom than fully compensated systems [9:10].

The above characterization implies that a partially compensated AOS will not yield the same quality images as adaptive optics systems with more WFS subapertures and DM degrees-of-freedom. However, a partially compensated AOS can still measure high spatial frequency information and hence, high resolution information with acceptable signal-to-noise ratios. This information can be extracted using post detection image reconstruction techniques. Systems that combine adaptive optics compensation procedures and post detection techniques are referred to as hybrid imaging systems [9:10]. Studies of hybrid imaging systems have shown that there are trade-offs to be made between predetection compensation and post detection image processing [9:11]. Figure 2 shows a block diagram of a hybrid imaging system.

Blocks I - V are essentially the same as in Figure 1. The only differences being in the number of DM actuators and WFS subapertures as already mentioned. Block VI represents the collection and storage of WFS outputs and calibration data from the reference star used to drive it. Block VII represents deconvolution of the recorded image using information from the wavefront computer. This deconvolution step intends to obtain the object spectrum (i.e., the Fourier transform of the object) by dividing the image spectrum (i.e., the Fourier transform of the detected image) by a quantity know as the Optical Transfer Function (OTF). In simple terms, the OTF contains information about the combined effects of the atmosphere and the optical system itself on the incoming light. Random variations in the atmosphere cause the OTF to also be random. Hence, the need for the recorded data from the WFS. Finally, after deconvolution, the final image can be obtained by taking the inverse Fourier transform of the computed object spectrum.

Aberrated Light from
Distant Target Object

Primary Mirror

VIII. Final
Image

VII.
Deconvolution
Processor

III. Imaging
Sensor

I. Deformable
Mirror

II. Beam Splitter

IV. Actuator
Command
Computer

V. Wavefront
Sensor

VI. Wavefront
Computer
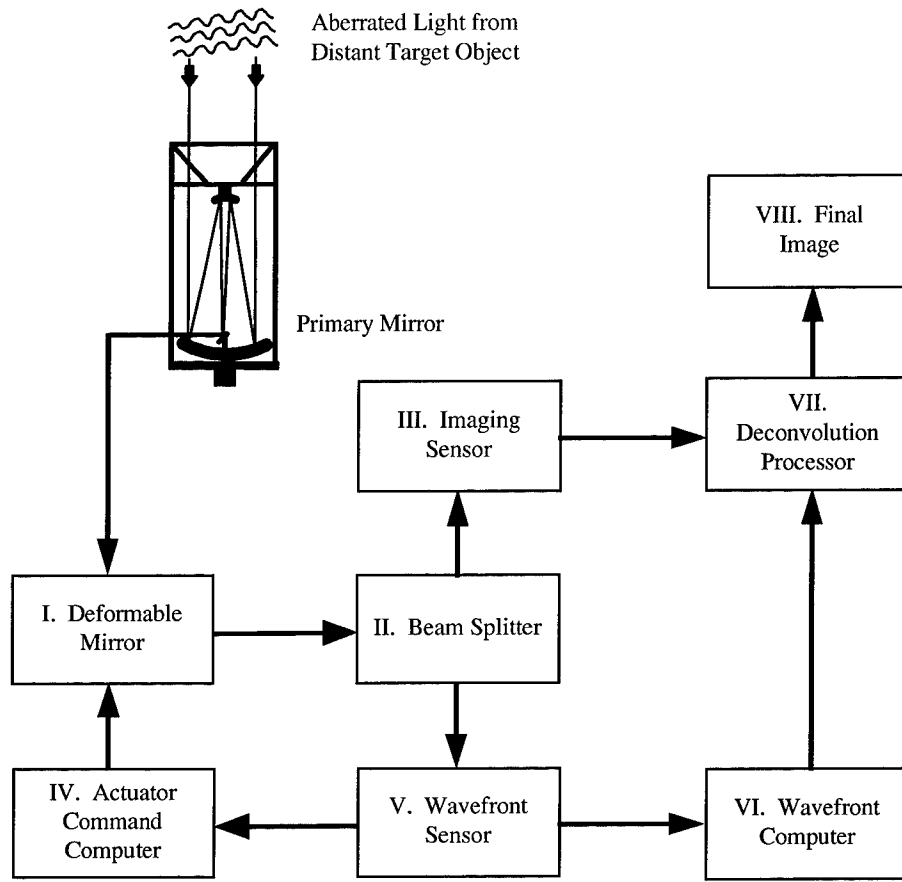
Figure 2. Hybrid Imaging System Block Diagram [9:11]

Another common form of hybrid imaging is to combine the application of speckle imaging

techniques to images measured with adaptive optics systems [9:294]. In speckle imaging, many

short exposures images are obtained that in effect freeze an occurrence of the random

atmosphere. The term speckle comes from the appearance of the resulting short exposure images

which result from a limited number incident photons to form a complete image. By combining

information obtained from these short exposures of the target object with corresponding short

exposures of a reference star, it is possible to obtain both the modulus and the phase of the

Fourier transform of the target object [9:132]. The final image can then be obtained by taking

the inverse Fourier transform of the object spectrum. Generally, the resulting image is much

better than a corresponding long exposure image susceptible to the effects of the atmosphere. It has been shown that by incorporating adaptive optics, speckle imaging methods can yield even better results than if adaptive optics are not used [10].

It should be noted that no constraints have been made in regard to the AOS. Therefore the above hybrid imaging methods can certainly be applied to images obtained from fully compensated adaptive optics systems as well, although the marginal improvement in performance is this case may be small.

Adaptive Optics System Imaging Process

During the past decade many significant advancements in adaptive optics were made by the Air Force. This research was primarily aimed at imaging spacecraft orbiting the earth in order to ascertain potential threats to the United States and its allies. However, with the end of the Cold War, much of this adaptive optics research was declassified, and is now being shared with the civilian astronomy community [7]. Nonetheless, Air Force research is still primarily motivated by imaging earth orbiting space objects in support of United States and Air Force Space Commands' mission requirements. With this in mind, this subsection discusses a typical process that might be used to meet such needs.

Acquiring an image is just the first step of a broader process to analyze and interpret these images. Figure 3 shows a block diagram of such a process. The first block represents the acquisition of digital images by the AOS methods previously discussed -- either a fully compensated adaptive optics system or hybrid imaging system. The resulting digital images usually undergo additional computer processing. Such processing is often designed to make the images more appealing to human observers. This step is intended to improve perceived image

quality by reducing noise, sharpening edges, and increasing contrast. These enhancements are accomplished through a myriad of different digital image processing algorithms.
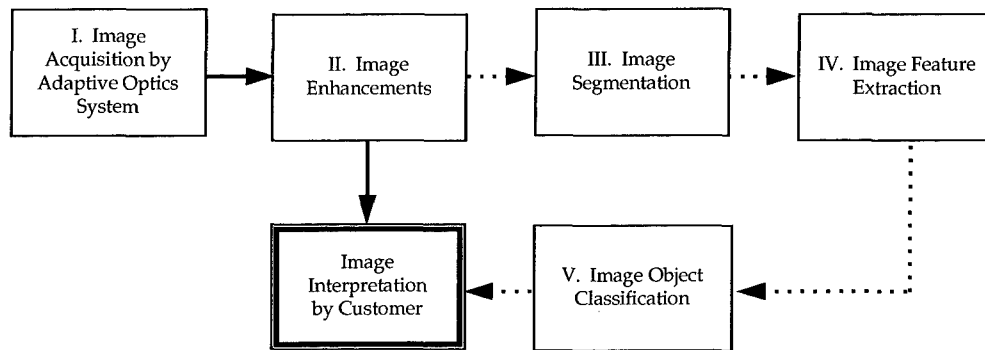
Figure 3. Block Diagram of Typical Adaptive Optics System Image Process

At this point, depending on the nature of the image and analysis requirements, the image may be suitable for subsequent interpretation without any additional processing. At other times, requirements may dictate additional processing such as segmentation, feature extraction, and object classification as related to classical pattern recognition methodologies. These steps may be required for data/dimensionality reduction and computer automated object recognition, all of which may aid in the ultimate interpretation of the image.

Problem Statement

Adaptive optics systems represent a relatively new technology. As such, characterization of their performance under a given set of conditions still has many uncertainties. Even though many aspects of a particular imaging scenario are uncontrollable, such as atmospheric conditions, object brightness, object size and object distance, several other AOS control parameters can be selected by the system operator. These controllable parameters include system integration time,

imaging wavelength, WFS wavelength, sampling considerations, and more. No method for selecting the controllable parameters to optimize performance in order to obtain the best quality image presently exists. Furthermore, an appropriate question to ask is - what does the best quality image mean?

The best looking image to a casual observer might not be the best to an image interpreter. Similarly, the best image to an interpreter might not be the best for computer automated recognition. Many different metrics have been suggested to characterize the quality of an image. Hence, it should be possible to optimize AOS settings to yield the "best" image for a particular set of conditions and imaging requirements by maximizing/minimizing a particular image quality metric. Adaptive optics system optimization is the major goal of this research effort. In addition, investigating different image quality metrics and developing new metrics that characterize the quality of images obtained with an AOS will be explored. Finally, a trade-off analysis will be conducted between fully compensated adaptive optics and hybrid imaging systems with partially compensated adaptive optics.

Research Objectives

The specific research objectives for this dissertation were:

1) Identify and/or develop appropriate image quality metrics that can be used to characterize AOS performance. Some desirable properties for these metrics are a) they should be monotonic, for example, the computed metric continually increases as the image improves, b) they should be simple to compute, c) they should be relatively easy to comprehend, d) they should be able to be computed from readily available system outputs or the corresponding output digital image itself, and e) they should account for overall AOS performance.

2) For a given set of seeing conditions, delineate input settings that maximize AOS performance by maximizing the selected image quality metric(s) identified in Objective 1. The majority of this work was accomplished using the AOS simulation HYSIM which is described in detail in Appendix A. This simulation was used to model the AMOS 1.6 meter AOS. Subobjectives were: a) to identify and understand relationships between controllable system input variables and uncontrollable variables that influence system performance, b) to determine the most important input variables and/or interactions among these variables and to incorporate these into empirical models that relate to AOS performance, and c) to produce robust system solutions. For example, if seeing conditions significantly change during the image acquisition process, control settings that continue to yield a satisfactory image are desired.

3) Identify and characterize trade-offs between fully compensated adaptive optics and hybrid imaging systems with partially compensated adaptive optics.

Scope

This dissertation focused primarily on adaptive optics systems used for imaging earth orbiting space objects. However, it was intended to keep the methodologies used herein generic so that they can be applied to imaging astronomical objects as well. At times, results for space objects (or extended objects) are compared to results for stellar objects (or point sources) to see how conclusions from the two compare with each other. If feasible, optimization results garnered from simulations will be verified using actual adaptive optics systems.

Methodology

The primary methodologies that will be used in this research are Response Surface Methodology (RSM) and Design of Experiments (DOE) [11, 12]. RSM and DOE employ

statistical techniques to construct empirical models of scientific phenomena when analytic

models are unknown or their use is impractical. This is done by relating a system response to the

levels of the input variables that affect it. In this manner, by careful design and analysis of

experiments, it is possible to obtain empirical models without having to resort to an exhaustive

search. Once obtained, these empirical models can be used to decide on input settings to achieve

desirable system results [12:1].

RSM and DOE can help with the following determinations, which parallel some of the

research objectives: a) which variables are most influential on the response, b) where to set

influential controllable variables so the response is near a nominal requirement, c) where to set

influential controllable variables so the variability of the response is small, and d) where to set

influential controllable variables so the effects of the uncontrollable variables are minimized

[11:453].

Montgomery provides additional motivation for DOE, he says [11:454]

.....experimental design methods may be used either in process development or process troubleshooting to improve process performance or to obtain a process that is robust or insensitive to external sources of variability.

Moreover, Montgomery says [11:454]

Experimental design methods can also play a major role in engineering design activities, where new products are developed and existing ones improved. Some applications of statistical experimental design in engineering design include

1. Evaluation and comparison of basic design configurations.

2. Evaluation of material alternatives.

3. Determination of key product design parameters that impact performance.

In this research, the system response will be the AOS image quality metric(s) that are

established in Chapter III. Accordingly, early work concentrated on satisfying the first research

objective of establishing suitable AOS response metrics. As for input variables, in dealing with

an AOS, there are both controllable and uncontrollable factors that affect the system response. It

13

is not possible to arbitrarily set the levels of the uncontrollable variables that affect an AOS. However, by virtue of the use of HYSIM, uncontrollable variables can be set at prescribed levels to accommodate RSM and DOE techniques. Appendix B contains a more detailed review of RSM and DOE procedures and mechanics.

Summary of Key Results

The following key results were obtained from this research:

1) Two novel image quality metrics were developed -- the Modified Strehl Ratio (MSR) and the Noise-Effective-Cutoff (NEC) frequency. It was shown that these two metrics, as well as the commonly used Strehl ratio, are highly correlated with the Root-Mean-Square Error (RMSE) between the detected image obtained with an AOS and the corresponding diffraction limited image. These results were shown to hold across a wide range of atmospheric seeing conditions, object light levels, and AOS settings. These results also hold for imaging both extended objects and point sources. These are important results because the MSR and NEC can be obtained and used in real field experiments to gauge AOS performance whereas the RMSE cannot. Appropriate decisions regarding system settings to maximize AOS performance can be made based on these metrics.

2) For a constant beacon light level, such as might be obtained with the use of a laser guide star, it was shown that the Strehl ratio cannot distinguish between different image qualities. On the other hand, the MSR retains a high degree of correlation with RMSE. This is because the MSR metric can account for both Wavefront Sensor (WFS) performance and measurement noise in the detected image, whereas the Strehl ratio only accounts for WFS performance.

3) As a subsidiary contribution, it was shown why the sample based average image spectrum Signal-to-Noise Ratio (SNR) approaches a limiting value of 0.886 as the normalized

spatial frequency approaches one. This is in contrast to the theoretical average image SNR which approaches a limiting value of zero as the normalized spatial frequency approaches one.

4) The most significant factor affecting AOS performance is the object light level that is used to drive the WFS. This is followed in importance by the system gain, which is akin to the system closed loop bandwidth when the system sample rate is held constant. The next most important factor is the Fried parameter which is a measure of atmospheric seeing conditions. For a constant long exposure time, the least important factor is the short exposure time. This is because about the same number of photoevents will be seen in the long exposure regardless of the short exposure time.

5) An empirical model of the AMOS 1.6 meter adaptive optics system performance based on the MSR metric was constructed. The fitted regression model can be used to accurately predict AMOS performance based on input values for the variables of system sample rate, system gain, object visual magnitude, Fried parameter, and wind speed. The coefficient of multiple correlation for this model was 0.991.

6) Based on the above results, it was decided to optimize the system closed loop bandwidth as a function of light levels and seeing conditions. Optimal closed loop bandwidth settings for the AMOS 1.6 meter system were determined. Moreover, it was shown that the proper choice of the closed loop bandwidth settings are much more critical under low light levels than under high light levels, regardless of the seeing conditions. Under low light levels, these results show that the optimal closed loop bandwidth settings are many times lower than the Greenwood frequency. This is an important result because the widespread rule of thumb specifies a closed loop bandwidth setting that is equal to the Greenwood frequency. Finally, as seeing conditions worsen, the optimal closed loop bandwidth settings shift to slightly higher values and visa versa and seeing conditions improve.

7) A method to produce robust system performance under widely varying atmospheric seeing conditions was developed. That is, a strategy was determined that led to the identification of the closed loop bandwidth setting for a particular light level that will yield performance that is almost always within 95 percent of the maximum, regardless of seeing conditions. This is an important result because the system operator need only know the predicted object visual magnitude that is to be imaged. The preferred system closed loop bandwidth can then be selected without having to worry about seeing conditions. Moreover, these results are important even for systems that incorporate artificial guide star capabilities. This is because such systems will produce beacons with equivalent stellar visual magnitudes of about 8 to 10. In this region, it is shown that selecting the proper closed loop bandwidth to achieve good imaging system performance becomes fairly delicate.

8) Trade-offs between fully compensated and partially compensated adaptive optics systems were identified. The most conspicuous trade-off is of course cost. In addition, fully compensated systems will be able to achieve a higher level of performance via finer wavefront sensing and compensation than partially compensated systems, when no regard is given to light levels. When noise dominates over limitations imposed by sample spacing, then partially compensated systems may actually perform better than fully compensated systems. This is because partially compensated systems can boost SNR levels at the WFS.

9) Based on the above assessment, the concept of clustering smaller WFS subapertures into larger effective subapertures was introduced. An experiment under this notion was conducted by modeling the AMOS 1.6 meter system. It was concluded that clustering of the AMOS subapertures will never benefit system performance, at least up to a very dim visual magnitude of 13. It is believed this is because the AMOS subapertures are already relatively

large, so that by clustering them the resulting wavefront errors due to sample spacing can never overcome the associated reduction in WFS noise.

Organization of this Dissertation

The remainder of this dissertation is organized as follows. Chapter II presents a literature review of pertinent sources that relate to image quality metrics and adaptive optics system performance and optimization. Chapters III, IV, and V present the results of each of the three major research objectives respectively. Chapter VI provides a summary of research contributions and presents recommendations and suggestions for future related research. Appendix A contains a detailed description of the adaptive optics simulation, HYSIM, used in this research. Appendix B presents a thorough review of RSM and DOE methodologies. Finally, Appendix C contains additional documentation for the RSM/DOE analysis conducted in Chapter IV.

Some words from Tyson's text on adaptive optics provide fitting motivation to this research effort, especially in regard to an Operations Research student delving into the world of optical and atmospheric physics. Tyson says [3:ix]:

> *Adaptive optics does not fall neatly into any of the established engineering disciplines; it combines elements from optics, electro-optics, electrical engineering, mechanical engineering, and chemistry. The interdisciplinary nature of adaptive optics may explain why it is such a latecomer to contemporary technology. . . . . . Scientists and engineers with expertise in many disciplines were needed to find the necessary pieces of the adaptive optics puzzle and fit them together. Maybe there is a lesson to be learned here regarding engineering education and practice: Don't specialize so tightly that there is little appreciation of problems and potential solutions in other areas of technology that may at first appear unrelated.*

II. Literature Review

This chapter presents a review of literature pertinent to image quality metrics and the performance/optimization of adaptive optics systems.

Image Evaluation Methods and Metrics

Recall the question posed in the introduction -- what does the best quality image mean? This is an important question because the answer can be used to establish an objective approach to determine what AOS system changes effect image quality and in what ways. As such, this subsection presents a summary of just some of the myriad of methods and metrics that can be used to characterize the quality of digital images.

Image Quality. Human assessment of image quality is subjective in nature. Holst proposed that image quality evaluation is a learned capability. He states that this capability is a perception accomplished by the brain which incorporates sensory system stimuli, emotions, learning, and memory [13:239]. Unfortunately, associated relationships between these factors are not well understood. This is demonstrated by large variations in image quality rankings furnished by different observer's who are tasked to order the same set of images from best to worst [13:239]. Consequently, it is extremely difficult to establish an absolute scale for image quality. Yet despite this assessment, many formulas have been devised to calculate image quality.

Resolution Metrics. One of the most common measures of image quality is resolution. Holst provides four different types of resolution -- temporal resolution, gray scale resolution, spectral resolution, and spatial resolution. The most familiar of these measures is spatial resolution [13:218].

Spatial Resolution. The image quality measure of spatial resolution has many definitions and forms that depend on the particular application of interest. For instance, photo interpreters might be interested in the Ground Resolved Distance (GRD) when dealing with remote sensing system images [13]. Alternatively, astronomers might focus on the Rayleigh criterion or airy disc diameter achieved by a particular optical system [14]. In regard to adaptive optics systems, one might be interested in the finest detail that can be discerned in the resulting images obtained by the overall adaptive optics system. This overall system resolution depends on the severity of atmospheric turbulence, diffraction of light and telescope pupil, optical aberrations, detector resolution and sensitivity, digitization and sampling, system bandwidth, and so forth [13:218].

Table 1 provides an overview of these and a few other spatial resolution metrics [13:222]. In this table, λ is the wavelength of light in meters and D is the primary aperture diameter in meters. MTF is the Modulation Transfer Function which is the modulus of the OTF. GRD and Ground Resolution imply remote sensing systems looking toward the earth. It should be possible however to apply similar methods to systems looking skyward at extended space objects of known size/distance. Both of these measures are a function of target range and both methods rely on experienced photo interpreters to provide estimates of observed feature sizes. Holst references the GRD based Imagery Interpretability Rating Scale developed for military applications and states that it can be modified for environmental remote sensing. This inspires the notion for developing a similar rating scale for imaging space objects.

Holst introduces the concept of sensitivity in his discussion of resolution. He states that sensitivity and resolution are limiting cases for the complete system analysis. When a system is sensitivity limited, overall performance depends on the difference in intensity between the target

Table 1. Spatial Resolution Metrics [13:222]

RESOLUTION	DESCRIPTION	FORUMULA OR METHOD
Rayleigh Criterion	Ability to distinguish two point sources	$1.22\,\lambda\,/\,D$ (rad)
Sparrow Criterion	Ability to distinguish two point sources	$\lambda\,/\,D$ (rad)
Airy Disk	Diffraction limited diameter produced by a point source	$2.44\,\lambda\,/\,D$ (rad)
Blur Diameter	Actual mimimum diameter produced by a point source	Calculated from ray tracing (rad)
Limiting Resolution	Spatial frequency at which MTF = 0.2 to 0.05	Measured or calculated (cycles/rad)
Ground Resolved Distance	Smallest test target (one cycle) that can be distinguished	Measured or calculated (meters)
Ground Resolution	Estimate of limiting feature size that can be distinguished	Measured (meters)
Nyquist Frequency	One-half of the sampling frequency	Calculated (cycles/rad)
Pixels	Number of detector elements or number of digital data points	Counted (unitless)
Effective Instantaneous Field of View (EIFOV)	One-half of the reciprocal of the spatial frequency at which MTF = 0.5	Measured or calculated (rad)

and its background, atmospheric attenuation, and system noise. When a system is resolution limited, the ability to discern the target from its background is dependent only on the target size and system resolution. In the latter case, an estimate that relates the maximum range that a target can be detected is given by

$$\text{Range} \approx \text{target size / resolution} \tag{2}$$

Holst points out that this estimate is most applicable to a single subsystem. For end-to-end

resolution measures, he prescribes Shade's equivalent resolution as a better metric as it can include all the pertinent subsystems [13:221].

Shade's Equivalent Resolution. As a method to compare end-to-end system performance, Holst presents the following form for equivalent resolution (R_{eq}) developed from Shade [13:233]

$$R_{eq} = \frac{1}{2\int_0^\infty [MTF_{sys}(f)]^2 df}$$ (3)

where f represents spatial frequencies and MTF_{sys} is the MTF of the entire system. Since MTF_{sys} is a function of system subsystems, Holst presents a useful form for R_{eq} by applying Equation (3) to each subsystem [13:233]. The simulation used in this dissertation can output a discrete, averaged form of the OTF of the combined AOS and atmosphere. Hence, R_{eq} can be obtained via numerical integration, which essentially comes down to performing a double summation over the modulus of the output average OTF array. Equation (3) shows that the higher the MTF, the smaller or better the resolution of the system. The limiting case would be the MTF of a diffraction limited system. Hence, one could compare the computed R_{eq} to this limiting case, which would be similar to computing the Strehl ratio of the system [15:139]. Also, for simplicity, one could just compute the integral portion of Equation (3) by itself and use the result to compare different systems and system configurations. Finally, as a useful form, Holst gives R_{eq} in radians for a diffraction limited, incoherent optical imaging system as $R_{eq} = 1.845 \, \lambda \, / \, D$ where both λ and D are in meters.

Subjective Image Quality. Resolution by itself does not provide the ideal metric for assessing AOS performance. For instance, by inspection of Equation (3), two differently shaped MTFs could yield the same R_{eq}. Yet, this resolution measure does nothing to describe the

relative amounts of different frequencies used to comprise each image. Perhaps the first image has an MTF with many low frequencies and the second has an MTF with many high frequencies. So even if the two images have the same R_{eq}, they can look very different. Furthermore, there is no single MTF shape that provides the best image quality. In fact, it has been shown that the MTF yielding the best image changes with scene content [13:239]. Nonetheless, tests show that in general images that have higher MTFs and less noise are judged as having better image quality [13:239].

Aside from engineering metrics, another method used to gauge image quality is to use multiple observers to view and rank numerous images containing known amounts of degradations. Equations are then are derived which relate these rankings to the amount of degradation [13:239]. Along these lines, a potential research scenario is to a) produce images under different amounts of atmospheric distortions and AOS control settings, b) have observers rank the quality of these images, c) train a neural network to classify the images, d) extract information from the neural network that can be used to predict image quality, gauge the saliency of input parameters and noise, and perhaps from this develop an AOS image quality metric that can be used to optimize AOS performance. However, this is beyond the scope of this dissertation.

Subjective Quality Factor. Holst presents a metric called the Subjective Quality Factor (SQF) which is based on the sensitivity of the human eye. This form is somewhat similar to Equation (3). Specifically, the MTF of the system is integrated over frequencies that are "very important" to the eye [13:242]. Holst also presents a table that relates SQF values to subjective image quality ratings (e.g., excellent, good, and unsatisfactory).

Perceived Signal-to-Noise Ratio. This metric based on the sensitivity of the human eye and the intensity difference between the target and its immediate background [13:246]. Using

this measure, when the computed Perceived Signal-to-Noise Ratio, SNR_P, is above some threshold value, the target in an image is just perceived. Holst states that as target intensity increases in an image, higher spatial frequencies can be perceived. Accordingly, one would think that system resolution would increase as well. However, most image quality metrics do not directly account for image intensity. Hence, SNR_P may be a useful metric when comparing images of different intensities.

Image Sharpness. Muller and Buffington present the concept of image sharpness such that the value of the sharpness for an image degraded by the atmosphere is always less that that for the true image [16]. They present eight different definitions for image sharpness. They intend to use these different metrics for correction of atmospheric distortions by maximizing the value of the metric in real-time via a feedback control system. These metrics all make use of the irradiance at a point $I(x,y)$ in the image plane of the telescope system. Muller and Buffington show that these metrics all reach a maximum when image distortion has been reached [16]. The first metric they introduce is

$$S_1 = \int I^2(x,y)\,dx\,dy .$$ (4)

For a digital image, one way to implement this metric is to square the gray value for each image pixel for the $I^2(x,y)$ term and then apply the trapezoidal rule of integration to the resulting array [17:406]. If the gray values near the edges of the image are small compared to the interior of the image and the image array is fairly large, then a simple double summation over the array yields nearly identical results as compared with the trapezoidal rule. Another form of this metric is S_5, which replaces $I^2(x,y)$ by $I^3(x,y)$ or $I^4(x,y)$.

The sharpness metric Muller and Buffington designate by S_8 is given by

$$S_8 = -\int |I(x,y) - I_0(x,y)|^n \, dx \, dy \qquad (5)$$

where $I_0(x,y)$ is the undistorted or diffraction limited image. In this equation, if $n = 2$ and the computed value of S_8 is divided by the total number of image pixels, the result can be thought of as the Mean Square Error (MSE) per pixel between the distorted image and the diffraction limited image. This value can be normalized by dividing by the square of the largest possible gray value, for an eight bit digital image this is 255. The minus sign in front of the integral is used so that the maximum value of S_8 is achieved when there is no distortion. In this case $S_8 = 0$. The minus sign could be ignored and then the goal is to minimize MSE. One problem with this metric is that it relies on having the undistorted image available. This is not possible in real field experiments. Another problem is that the undistorted image is a function of the optical system used to obtain it. Hence the use of this metric would only be good for comparing images obtained from the same optical configuration. The other metrics presented by Muller and Buffington are similar to these and all are based on image irradiance. Except S_8, the above metrics are based on the measured image, that is they can be computed from the detected image itself. In contrast, the resolution metrics that were previously discussed are primarily system based, that is they require information about system performance during the image acquisition process. Hence, such system information needs to be collected and stored to compute these type of metrics.

Other Image Based Quality Metrics. Spectral image quality, histogram based image quality, and entropy are just a few of the many other image based metrics discussed in the literature. A brief summary of these follows.

Spectral Image Quality. Lee discusses a method to compare the spectra, that is, the Fourier transform, of images. First, the "energy" of the images are normalized by [18:2-4]

$$f_{norm}(x,y) = \frac{f_{orig}(x,y)}{\sum_x \sum_y f_{orig}(x,y)} \tag{6}$$

where $f_{norm}(x,y)$ is the energy normalized image and $f_{orig}(x,y)$ is the original image. Next, the two-dimensional Fourier transform, $F(u,v)$, of the energy normalized images are computed, where u and v are spatial frequencies in the x and y directions. From each spectra, an Image Quality Metric (IQM) is computed by [18:2-6]

$$IQM = \sum_\theta \sum_{\rho=\rho_{lower}}^{\rho_{upper}} |F(\rho,\theta)| \tag{7}$$

where $|F(\rho,\theta)|$ is the modulus of the Fourier transform in polar coordinates. That is, $\rho = \sqrt{u^2 + v^2}$ is the radial distance from the origin and θ is the angular direction. The parameters ρ_{lower} and ρ_{upper} are selected by the user. Higher values of IQM correspond to better image quality.

Histogram Based Image Quality. Histograms of images can easily be constructed by counting the number of pixels with certain gray levels and then plotting the count versus the different gray levels [18:2-7]. The bin sizes used in the histogram are usually the same as the gray levels in the image (e.g., 8 bit quantization = 256 gray levels = 256 bins). The result can easily be normalized so the area under the histogram equals one, yielding an estimate of a Probability Density Function (PDF) for the pixels in the image. With this, different types of shapes in the histogram or estimated PDF can be used to ascertain image quality. Lee reviews several different histogram related metrics.

Entropy. Analogous to entropy in thermodynamics, entropy in the context of image quality relates to the degree of randomness of an image. An image that displays a lot of variation (e.g., an image with perfect histogram equalization) has high entropy and an image that contains little variation (e.g., an image thresholded to two states) has low entropy. If all pixels have the same value, the entropy of the image will be zero. Hence, by increasing the entropy of an image, one might be able to gain more information from the image.

Lewis presents an equation to compute image entropy under the assumption that pixel values are not correlated with each other [19]. This equation utilizes the probability of occurrence of each pixel value. This is accomplished by counting the number of occurrences of each pixel value and dividing by the total number of pixels. It turns out that the maximum entropy that can be obtained for an image with 256 gray levels is $\log_2(256) = 8$, where the probability of each gray value is equally likely. The central problem with this metric is that a perfectly random image that has an entropy value of 8 may convey no more information than a completely uniform image with an entropy value of 0. A diffraction limited image will have some entropy value less than 8 say. It seems possible then that the entropy values of images obtained under adaptive optics compensation can approach the entropy value of a diffraction limited image from above or below. Hence, this metric is not monotonic and is probably not too useful for AOS performance assessments.

Performance/Optimization of Adaptive Optics Systems

Roggemann, Stoudt, and Welsh explain that some short exposure images acquired with adaptive optics are "better" than others in terms of higher resolution. This is a consequence of the statistical nature of adaptive optics compensation [20]. Their goal is to use image sharpness measures to select the best images from a finite data set of compensated images for subsequent

26

postdetection image processing. They show that this can lead to better image-spectrum SNRs than if the entire image data set is used. They also reference work by Fried that shows there is a finite probability that a short exposure image will be nearly diffraction limited [21]. This probability can be very small, but it implies that some images will be better than others.

In their paper, the authors describe a single-frame image spectrum SNR metric they call SNR_1. If P statistically independent frames are averaged together to form a single image, the resulting image spectrum SNR is increased by the square root of P. That is

$$SNR_P = \sqrt{P}\, SNR_1 \qquad (8)$$

where SNR_P is the resulting image spectrum SNR of the P averaged images. These SNR measures are positive, real, two-dimensional functions. Hence, the authors state that their analysis is simplified by using radial averages of the SNR functions to convert them to one-dimensional functions. They also employ a noise-equivalent cutoff frequency defined as the highest spatial frequency at which the SNR metrics exceeds some threshold. Finally, they compose an SNR gain function as the ratio of the resulting SNR of M averaged images to the SNR of all N images of a complete data set. Hence, if this ratio is greater than one, a relative gain in image quality is achieved by frame selection. The converse is true if the ratio is less than one. The N images are rank ordered using the Muller-Buffington sharpness metrics. Then the best subset of M images are selected from these. The authors apply their methodology to imaging both point sources and extended objects.

To a point, their work shows that the mean, radially averaged image spectrum increases as the number of selected frames, M, is reduced. Image spectrum variance is decreased as M is reduced as well. Also, frame selection increases the noise-equivalent cutoff frequency and narrows the effective Point Spread Function (PSF).

The authors present curves of the gain ratio discussed earlier as a function of the normalized radially frequency. This plot implies an "optimum" frame selection rate of about 60 percent for the setup they used. They conclude that frame selection can improve image quality. Initial work was based on using the S_8 metric discussed earlier. They compared these results to those obtained if different metrics are used for frame selection. There does not appear to be a drastic difference by using metrics other than the S_8 metric presented by Muller-Buffington.

Of even more interest to this dissertation is how these results apply to imaging extended objects. Their results show that frame selection can also lead to improvements in SNR for extended objects. An optimum frame selection rate was seen at a value of bout 70% for the setup they used. In summary, this paper presents very useful information on image quality metrics as well as a methodology for improving image quality via post detection techniques applied to images obtained with adaptive optics systems.

Gardner, Welsh, and Thompson present information on the design and performance of adaptive optics systems that use laser guide stars to drive the WFS [22]. Artificial guide stars for controlling an AOS can be created by using a laser to excite atoms in the upper atmosphere to give off light. This means that adaptive optics systems are not limited to imaging only objects bright enough to drive the WFS. In regard performance investigations of this research, this implies that one could independently vary the brightness of the object to be imaged and the brightness of the source that drives the WFS. In this case then, WFS brightness could be treated as a controllable variable rather than an uncontrollable variable. This could improve performance as the SNR of the WFS can significantly affect the ultimate performance of an AOS.

The authors address imaging performance of an AOS as a function of guide star brightness. They discuss two measures that can be used to assess performance. These are the Residual Mean

Square (RMS) wavefront error over the telescope aperture and the Strehl ratio. Using these measures in their analysis, they conclude that ground-based adaptive optics systems equipped with a laser guide star capability can achieve performance levels similar to that of the HST and can approach diffraction limited imaging. In their presentation, the authors discuss the important issue of laser power requirements and factors that drive these requirements.

Ellerbroek develops first-order analytical models for AOS subcomponents [23]. Using these equations and numerical evaluation methods, he assesses AOS performance using the metrics of mean-square residual phase error, expected OTFs, and Strehl ratios. He presents plots that show the mean-square residual phase error versus D/L for different values of sensor noise, where L is the length of a side of a square subaperture. From these plots, error increases as the ratio of D/L increases, but this eventually flattens out. The error appears to be much more a function of sensor noise that the ratio of D/L.

The author provides a table that gives the short exposure Strehl ratio versus closed loop bandwidth, f_c, for different amounts of sensor noise and for different types of guide stars (i.e., natural or laser). For his setup, natural guide stars and higher values of f_c yield the best performance. Ellerbroek also presents plots of short and long exposure OTFs versus normalized spatial frequency for different types of guide stars. These plots include the diffraction limited case as well as the uncompensated case. Finally, he presents plots of the relative OTF reduction versus normalized spatial frequency for different values of f_c. The values of f_c he uses are 10, 20, and 30 Hz. His results show increasing performance with higher values of f_c. Ellerbroek's work also examines multiconjugate adaptive optics systems (i.e., two guide stars, two DMs, and two WFSs), but these are outside of scope of this research.

For future reference, in his paper the author specifies some typical seeing condition parameters for the AMOS observatory at Mount Haleakala, Hawaii. For $\lambda = 0.5$ μm, these are r_0

= 0.285 m and Greenwood frequency, f_g = 19.7 Hz. The results presented by Ellerbroek are based on using these nominal values of r_0 and f_g. Knowledge of nominal values for these parameters is important for providing a starting point for the dissertation methodology to be implemented in this research.

Gavel, Morris, and Vernon present a paper which discusses optimizing the design of laser-guided adaptive optics systems to minimize laser power requirements [24]. They state that the system design can be optimized through the proper choice of subaperture size, control bandwidth, laser power, and number of guide stars. In their approach, they develop analytical expressions and then use constrained optimization to minimize power. They apply this technique to the Keck Telescope, D = 10 m, to determine optimal values of f_c = 50 and 70 Hz for a WFS with 69 and 241 subapertures respectively. These results are based on nominal values of r_0 = 0.20 m and f_g = 26 Hz for the Keck Observatory. They also give similar results for the Lick Observatory 3 m telescope as f_c = 60 and 50 Hz for a WFS with 69 and 241 subapertures respectively. These results are based on nominal values of r_0 = 0.10 m and f_g = 26 Hz for the Lick Observatory. The results presented here are distinct from this work due to optimization of an image quality metric which considers the light level of the target object.

Tyson discusses how adaptive optics systems can be designed and analyzed using "very effective" system engineering tools [3]. He lists functional trees, performance trees, and interface matrices as some powerful methods. A functional tree is a method to describe an AOS in terms of functions and subfunctions down to their simplest levels. Of course there are various degrees of complexity that can be employed. He provides an example of such a functional tree for a conventional AOS. The functional tree is more or less a qualitative hierarchy. A more quantitative hierarchy is a performance tree. In this regard, Tyson discusses defining the various quantities that affect overall AOS performance. This leads to the development of an "error

budget" or performance hierarchy. He provides an example where nodes of the tree represent relationships between AOS parameters. A tool related to the performance tree is a system interface matrix which represents a cross reference between system functions. The output of one function may be the input to another. Tyson also offers guidelines for AOS design. He lists three major considerations. These are a) determine overall AOS performance requirements, b) consider properties of the optical beam the system is designed to enhance, and c) integrate the system so it performs properly in the environment where it is used. The author provides more details on each of these aspects. This type of information may be useful in the trade-off studies that will be conducted between fully compensated adaptive optics systems and partially compensated, hybrid imaging systems.

Roggemann and Welsh present a qualitative discussion of major factors that influence AOS performance [9]. These include WFS sampling, finite degrees of freedom of the DM, finite temporal system response, finite light levels to the WFS, and anisoplanatism [9:83]. They also provide system component models that can be used characterize "fundamental performance issues." In this area, results focus on a single factor that degrades AOS performance while other parameters are idealized. They also provide a framework that can be used to characterize performance of the entire AOS system. The key element of this framework is an adaptive optics simulation that can accurately model all major system components. This dissertation made heavy use of this simulation. Appendix A is devoted to summarizing important aspects and operating conditions of this simulation.

Max *et al* are studying adaptive optics concepts for the Keck Telescope which uses a wavefront sensor with varying numbers of subapertures so as to respond to changing seeing conditions [25]. The goal is to be able to "gang together" groups of deformable mirror actuators under software control when atmospheric conditions call for larger subapertures. Such can be

studied via the AOS simulation that will be used and is of particular interest in the trade-off studies phase of this research.

Tyler and Fender derive an expression for the imaging wavelength that gives maximum resolution for an AOS [26]. Their analytically predicted results are compared to simulation results and correspondence is shown to be good at widely separated seeing conditions and adaptive optics geometries. Wavelength optimization is an important consideration in AOS design. As such, this information was considered in the trade-off studies phase of this dissertation.

Dayton, Sandven, and Gonglewski note that it been demonstrated that Low Order Adaptive Optics (LOAO) can provide image improvements, especially when used in conjunction with computer post processing methods [27]. In their paper, the authors consider different performance metrics for use in determining expected results from a LOAO system when used for both long exposure and short exposure (speckle) imaging. They compare computer simulations results to data obtained from experiments on the 1.5 meter telescope at the Starfire Optical Range. This work is pertinent to the discussion on hybrid imaging systems and is again related to the aforementioned trade-off study aspects.

Literature Review Summary

The major discovery of this review effort is that there is not a plethora of topics related to optimization of adaptive optics systems in the literature. The limited amount of pertinent sources that are associated with AOS performance usually concentrate on only a single AOS subcomponent, such as the WFS. Moreover, these sources will usually examine the effect of varying only one or at most two system variables at a time. As a result, interactions among all the different variables that affect AOS performance may go unknown and are yet to be

discovered. This means that the potential performance of adaptive optics systems under varying seeing conditions may not be fulfilled. Hence, the objectives of this research that are primarily geared at overall system optimization represent a very important contribution to the field of adaptive optics.

III. Metrics for Characterizing Adaptive Optics System Performance

Introduction

This chapter represents the completion of the first research objective -- establishing metrics to assess adaptive optics system performance. Various metrics to appraise the quality of digital images were introduced in the previous chapter. A detailed analysis of image quality metrics that can be used to characterize adaptive optics system performance under diverse conditions and system settings is now presented. These metrics can be used to improve the performance of operational systems or to optimize the design of new systems -- which are the second and third research objectives and the subjects of the next two chapters.

One desirable trait for image quality metrics is monotonic behavior. For example, the computed metric should continually increase as the image improves due to increased average Modulation Transfer Function (MTF) or increased Signal-to-Noise-Ratio (SNR) levels. Other desirable properties are computational simplicity and the ability to be computed from available system outputs, such as the average Optical Transfer Function (OTF) or the detected image itself. Three such candidate metrics were examined and their relative performances were compared to a slightly modified version of the S_8 metric given in Chapter II. The exact form of this metric will be presented in the next section. Other metrics were explored and were eliminated. For instance, the entropy metric mentioned in Chapter II was eliminated because it does not exhibit monotonic behavior. Subjective quality metrics were eliminated because of the potential for inconsistencies among different observers tasked to rank order the same set of images as was discussed in the previous chapter.

The first metric examined is the commonly used Strehl ratio. The second metric is a novel form that modifies the numerator of the Strehl ratio by integrating the modulus of the average

34

OTF to a noise-effective-cutoff frequency at which some specified image spectrum SNR level is attained. The third metric is simply the noise-effective-cutoff frequency. It will be shown that the Strehl ratio and the modified form of the Strehl ratio can be used as a surrogate for the S_8 metric. This is a great advantage since, as mentioned, it is usually not possible to compute the S_8 metric in the field due to the unavailability of the diffraction limited image.

The basic approach was to produce different sets of images of an extended object and a point source under varying seeing conditions, light levels, and AOS control settings. By varying these parameters, different image qualities were produced. These "detected" images and their associated statistical quantities were obtained and stored using an AOS simulation. From these data, the different image quality metrics and the RMSE for each image were computed. Next, the correlation of each metric with the RMSE across the entire set of images was obtained and used as an indication of metric performance. An assumption is the higher the correlation between a given metric and RMSE, the better its performance.

The results demonstrate that the Strehl ratio is highly correlated with RMSE under most situations and may often serve as a suitable metric to provide information on overall AOS performance. However, situations may arise where the Strehl ratio provides no discrimination between images of very different subjective quality. Such a situation is where an AOS is using a beacon of constant brightness, such as a laser beacon, to drive the wavefront sensor while imaging targets of varying light levels. In this case, the Strehl ratio would remain relatively constant so that its correlation with RMSE is poor as the detected images improve or worsen with increasing or decreasing target brightness respectively. Under these circumstances, the results show that the modified form of the Strehl ratio can provide a high level of correlation with RMSE. This is because the modified Strehl ratio accounts for measurement noise in the detected images. These results hold for both extended objects and a point source. Finally, the third

metric examined, the noise effective cutoff frequency, performs relatively poorly under high light

levels, but its correlation with RMSE increases at low target light levels where it performs

comparable to the Strehl ratio and the modified Strehl ratio.

Candidate Image Quality Metrics

The three metrics discussed in the introduction are described in more detail in this section.

The first metric is the Strehl ratio, SR, defined by [15:139]

$$SR = \frac{\int circ(\vec{f}/f_c) MTF_{sys}(\vec{f}) d\vec{f}}{\int circ(\vec{f}/f_c) MTF_{dl}(\vec{f}) d\vec{f}} \qquad (9)$$

where f_c is the diffraction limited cutoff frequency of the system, MTF_{sys} is the same as in

Equation (3), MTF_{dl} is the modulus of the diffraction limited OTF of the system, and $circ$ is

defined as $circ(\vec{f}/f_c) = 1$ when the normalized frequency $|\vec{f}|/f_c \leq 1$, or 0 otherwise.

The next measure is a novel metric that modifies the Strehl ratio by replacing the diffraction

limited cutoff frequency in the numerator of Equation (9) with a frequency that is based on some

desired image spectrum SNR. Roggemann and Welsh give the theoretical image spectrum SNR

of a detected image as [9:49]

$$\begin{aligned} SNR_D(\vec{f}) &= \frac{|E\{D(\vec{f})\}|}{[var\{D(\vec{f})\}]^{1/2}} \\ &= \frac{\bar{K}|O_n(\vec{f})E\{H(\vec{f})\}|}{\sqrt{\bar{K} + (\bar{K})^2 |O_n(\vec{f})|^2 var\{H(\vec{f})\}}} \end{aligned} \qquad (10)$$

where $E\{D(\vec{f})\}$ and $\mathrm{var}\{D(\vec{f})\}$ are the mean and variance of the Fourier transform of the detected image, $E\{H(\vec{f})\}$ is the expected value of the OTF, $O_n(\vec{f})$ is the object spectrum normalized so that $|O_n(0)| = 1$, and $\overline{K}$ is the average number of photoevents in the detected image. The second expression given in Equation (10) holds under conditions of a doubly stochastic Poisson random process [9:53].

A sample based estimate of the image spectrum SNR of detected image is computed as follows. An average long exposure detected image is obtained by averaging n short exposure detected images. By taking the Fourier transform of each short exposure detected image, $D_i^S(\vec{f})$, the sample based estimate of the mean of the Fourier transform of the average long exposure detected image is computed as

$$\tilde{D}(\vec{f}) = \frac{1}{n} \sum_{i=1}^{n} D_i^S(\vec{f}). \tag{11}$$

Likewise, the sample based estimate of the variance of the Fourier transform of the average long exposure detected image is computed as

$$\tilde{\sigma}_D^{\,2}(\vec{f}) = \frac{1}{n} \sum_{i=1}^{n} |D_i^S(\vec{f})|^2 - |\tilde{D}(\vec{f})|^2 \,. \tag{12}$$

Hence, a sample based estimate of the image spectrum SNR of detected image is

$$\tilde{\mathrm{SNR}}_D(\vec{f}) = \sqrt{n}\,\frac{|\tilde{D}(\vec{f})|}{[\tilde{\sigma}_D^{\,2}(\vec{f})]^{1/2}}\,. \tag{13}$$

The $\sqrt{n}$ term is included in Equation (13) to account for the fact that averaging n independent frames improves $\tilde{SNR}_D(\vec{f})$ by a factor of $\sqrt{n}$ because the variance of the estimate $\tilde{D}(\vec{f})$ decreases with increasing n.

Returning to Equation (10), note that since $E\{H(\vec{f})\} \to 0$ as $(|\vec{f}|/f_c) \to 1$, the theoretical $SNR_D(\vec{f}) \to 0$ as $(|\vec{f}|/f_c) \to 1$ within the pass band of the optical system. However, photon noise is not band-limited in the Fourier transform domain of the detected image. Outside the effective pass band of the optical system we shall see that $\tilde{SNR}_D(\vec{f})$ takes on a limiting value which is now computed. The effective pass band of the optical system as the region in frequency space where the product $\left| O_n(\vec{f})E\{H(\vec{f})\} \right|$ has appreciable value. Within the pass band the theoretical calculation leading to the second line of Equation (10) is valid. However, outside the effective pass band $\tilde{D}(\vec{f})$ is essentially pure noise.

In this analysis we are particularly concerned with the statistical behavior of the phasor $\tilde{D}(\vec{f})$ outside the effective pass band. $\tilde{D}(\vec{f})$ arises from the sum of a large number of image spectra $D_i^S(\vec{f})$. Hence, an appropriate statistical model for $\tilde{D}(\vec{f})$ outside the effective pass band is the random phasor sum described by Goodman [29]. The phasors representing the spectral components of individual $D_i^S(\vec{f})$ are well-modeled as having real and imaginary parts which are zero-mean, with common variance denoted by $\sigma_{R,I}^2$ at each frequency [30]. As shown below, it is not necessary to compute $\sigma_{R,I}^2$ to compute the limiting value of $\tilde{SNR}_D(\vec{f})$ outside the effective pass band.

This analysis is based on an extension of Goodman's results. Outside the effective pass band, we obtain the following theoretical expressions for the sample based mean and variance of the image spectra

$$\tilde{D}(\vec{f}) = \frac{1}{\sqrt{n}} \sqrt{\frac{\pi}{2}} \sigma_{R,I} \tag{14}$$

and

$$\tilde{\sigma}_D^2(\vec{f}) = \frac{1}{n} \sum_{i=1}^{n} |D_i^S(\vec{f})|^2 - \frac{\pi}{2n} \sigma_{R,I}^2$$
$$= \sigma_{R,I}^2 \left(2 - \frac{\pi}{2n} \right). \tag{15}$$

When the fact that averaging n independent frames improves $\tilde{SNR}_D(\vec{f})$ by a factor of $\sqrt{n}$ is

accounted for, the following theoretical expression for $\tilde{SNR}_D(\vec{f})$ is obtained

$$\tilde{SNR}_D(\vec{f}) = \sqrt{n} \frac{\frac{1}{\sqrt{n}} \sqrt{\frac{\pi}{2}} \sigma_{R,I}}{\sigma_{R,I} \sqrt{2 - \frac{\pi}{2n}}}. \tag{16}$$

Hence, in the limit of large n, $\tilde{SNR}_D(\vec{f}) = \sqrt{\pi} / 2 \approx 0.886$.

In the simulation results presented later in this chapter $SNR_D(\vec{f})$ was computed at every

point in the two dimensional sampled Fourier transform domain of the detected image. The two

dimensional $SNR_D(\vec{f})$ function was then radially averaged by computing the average of

$SNR_D(\vec{f})$ around circles of constant radius to allow the two-dimensional $SNR_D(\vec{f})$ function to

be presented as a one dimensional plot. The fact that $SNR_D(\vec{f})$ takes on a limiting value

outside the effective pass band is clearly evident in these radially averaged plots. By using the

simulation, the limiting value of $SNR_D(\vec{f})$ was confirmed outside the effective pass band of the

system is 0.886. Note that the values of $\mathrm{SNR}_D(\vec{f})$ used in this chapter were always larger than 0.886. Hence the fact that $\mathrm{SNR}_D(\vec{f})$ takes this limiting value does not affect the results presented here.

Figure 4 shows a sample $\mathrm{SNR}_D(\vec{f})$ function that was radially averaged to yield a one dimensional plot as discussed above. From this, an SNR level is selected and a corresponding spatial frequency is determined. This spatial frequency, the "noise-effective-cutoff" frequency, f_{nec}, replaces f_c in the numerator of Equation (9). Subsequently, this metric will be referred to as the modified Strehl ratio, *MSR*. We have

$$MSR = \frac{\int circ(\vec{f}/f_{nec})\,MTF_{sys}(\vec{f})\,d\vec{f}}{\int circ(\vec{f}/f_c)\,MTF_{dl}(\vec{f})\,d\vec{f}} \ . \tag{17}$$

For notational simplicity, the SNR level used will be appended to the end of the modified Strehl ratio. For example, *MSR1* and *MSR2* are the modified Strehl ratios obtained using SNR levels of 1 and 2 respectively.

The third metric considered is simply the noise-effective-cutoff frequency described above. We have

$$NEC = f_{nec} \ . \tag{18}$$

Again, for notional simplicity, the SNR level used will be appended to the end of the noise-effective-cutoff frequency. For example, *NEC1* and *NEC2* are the noise-effective-cutoff frequencies obtained using SNR levels of 1 and 2 respectively.

40
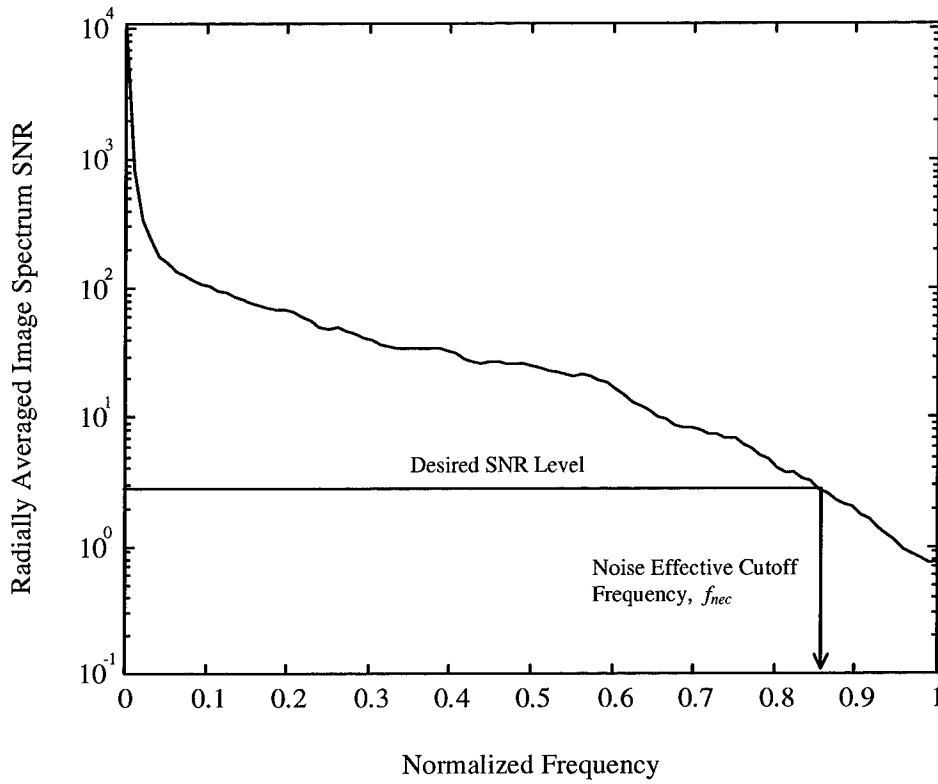
Figure 4. Sample Diagram Demonstrating How to Determine f_{nec}

All three of the above metrics are compared to the RMSE per pixel between the detected image and the associated diffraction limited image. In this case, RSME is computed by using the S_8 metric given by Equation (5) in the previous chapter with $n = 2$ and taking the square root of the integral portion divided by the total number of image pixels. The minus sign in front of Equation (5) is dropped to avoid confusion. Hence, as the image gets worse, RMSE will increase. Adopting a discrete pixel notation, RMSE is given by

$$RMSE = \left\{ \frac{1}{RC} \sum_{x=1}^{R} \sum_{y=1}^{C} [I^{N}(x,y) - I_0^{N}(x,y)]^2 \right\}^{1/2} \tag{19}$$

41

where the superscript N indicates the "energy normalized" image and R is the number of rows and C is the number of columns in the image array. Images were normalized to have unit energy using

$$I^N(x,y) = \frac{I(x,y)}{\left[\sum_{x=1}^{R}\sum_{y=1}^{C} I(x,y)^2\right]^{1/2}}. \tag{20}$$

The three metrics presented in this section were applied to images of extended objects and point sources obtained via a simulated AOS. Diverse system input control settings and seeing conditions were investigated. The next section provides more details on the simulation, input conditions, and control settings.

Experimental Setup

This research incorporated an adaptive optics simulation called the Hybrid Imaging Simulator or HYSIM developed by Michael Roggemann [20]. HYSIM is a FORTRAN based program that allows different atmospheric conditions and AOS configurations to be specified. For user provided input parameters, HYSIM models and computes the average long exposure image, the image spectrum SNR, and the average system OTF for both an extended object and a point source. For simulated imaging of extended objects, a computer generated rendition of a satellite which orbits the earth at an altitude of 650 kilometers was used and is shown in Figure 5. The input object is assumed to be approximately 12.5 meters across at its widest dimension. The object is embedded in an array of 512 x 512 pixels. For simulated imaging of point sources, an input array with a one at the center and zeros elsewhere was used.
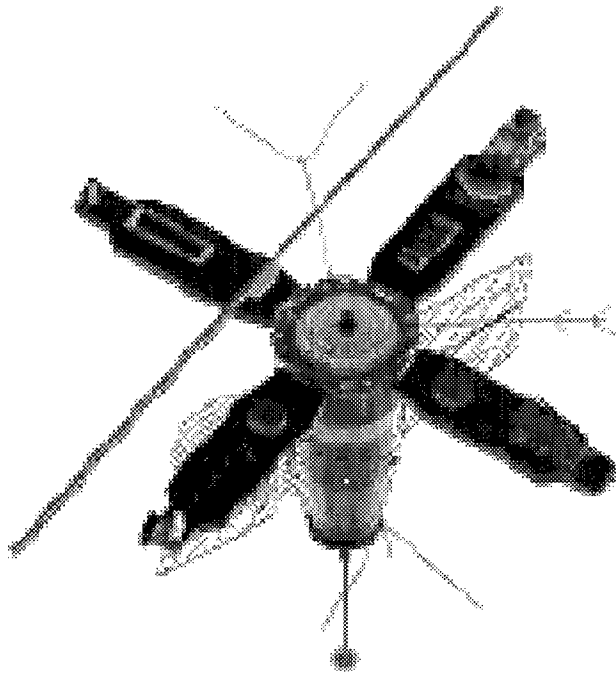
Figure 5. Simulated Satellite for Extended Object Imaging

Figure 6 shows the HYSIM input variables and their corresponding values that were used in this chapter. The primary controllable input variables are those that are easy to change in near realtime. Secondary controllable input variables are considered variables that are relatively difficult to change once a system is designed and built. Uncontrollable input variables that characterize seeing conditions and target light levels are also shown. The values used for the input variables represent typical settings and conditions that might be encountered in the field.

The stated visual magnitudes are an overall integrated value for the target object. Each of these values give rise to a different photon flux from the object. An overall energy efficiency of 10 percent, from the top of the atmosphere through the output of both the WFS and imaging detectors, was assumed. Finally, a total average image exposure time of 5 seconds was used in

all cases. More specific details of the simulation are provided by Roggemann *et al* and are also

presented in Appendix A [20].

Primary   Secondary

- Closed loop bandwidth: 50 Hz
- Frame exposure time: 0.01 & 0.1 s
- Number of frames: 500 & 50

- Primary mirror diameter: 1.0 m
- Obscurration diameter: 0.2 m
- Imaging wavelength: 7.0E-07 m
- WFS wavelength: 6.0E-07 m
- Number of WFS subapertures: 12
- DM actuator separation: 0.12 m

Controllable

Uncontrollable

Inputs

AOS
(Model by HYSIM)

Outputs

- Average Long Exposure Image
- Image Spectrum SNR
- Average System OTF

- Target Distance: 6.5E+05 m
- Object and WFS visual magnitudes: 4.25, 6.0, 8.0, & 10.0
- Fried parameter: 0.09 & 0.25 m
- Wind speed: 5.21, 10.42, 15.63, 31.25 m/s
- Physical size of array holding input object: 40.0 m

Figure 6. AOS Inputs and Outputs Modeled by HYSIM

Results

Using the input settings shown in Figure 6, images of an extended object and a point source

with varying qualities were produced. For both situations, the diffraction limited image was also

produced so that the RMSE could be computed for each detected output image. The input

settings were divided into five cases. Case 1 consists of combinations under relatively bright

extended object conditions. Case 2 consists of combinations under dimmer extended object

conditions. Case 3 combines Case 1 and Case 2 together. In Case 4 the object visual magnitude,

m_v, was varied over a wider range while keeping the WFS light level constant at a visual

magnitude of 6 and then a visual magnitude of 8. This allows us to examine how these metrics

perform under conditions of a relatively constant system OTF such as might be encountered with

the use of a laser guide star. Finally, in Case 5 the target was a point source. Table 2 displays a

summary of these different cases.

Table 2. Summary of Different Cases Examined in Chapter III. Different combination of the parameters in this table were used to produce a wide diversity of different image qualities.

Case Number	Target	Fried Parameter (m)	Wind Speed (m/s)	Object Light Level (m_v)	WFS Light Level (m_v)	Number of Images Produced
1	Extended Object	0.09 and 0.25	5.21, 10.42, 15.63, and 31.25	4.25 and 6	4.25 and 6	16
2	Extended Object	0.09 and 0.25	5.21, 10.42, 15.63, and 31.25	8 and 10	8 and 10	16
3	Extended Object	0.09 and 0.25	5.21, 10.42, 15.63, and 31.25	4.25, 6, 8, and 10	4.25, 6, 8, and 10	32
4	Extended Object	0.12	10.42	0, 2, 4, 6, 8, 10, 12, and 14	6 and 8	16
5	Point Source	0.09 and 0.25	5.21, 10.42, 15.63, and 31.25	8 and 10	8 and 10	16

For all cases, the correlation between the examined metrics and RMSE was computed.

Specifically, the correlation coefficient of each metric with RMSE was computed as follows

[31:102]

$$r = \frac{\sum_{i=1}^{k}(X_i - \overline{X})(Y_i - \overline{Y})}{\left[\sum_{i=1}^{k}(X_i - \overline{X})^2 \sum_{i=1}^{k}(Y_i - \overline{Y})^2\right]^{1/2}} \tag{21}$$

where k is the number of images in the data set and for notational simplicity, X represents RMSE and Y represents the metric under consideration. The correlation coefficient given in Equation 15 is a common byproduct of simple linear regression via the method of least squares [31:100-101]. The range of r is $-1 \leq r \leq 1$. The closer r is to -1 or 1, the greater the degree of linear association between the metric under consideration and RMSE [31:101]. A value of -1 indicates a perfect negative correlation and a value of 1 indicates a perfect positive correlation [31:101]. Since RMSE increases as any of the examined metrics decrease, all correlation coefficients given in this chapter are actually negative. However, for convenience the minus sign on all values of r were dropped.

Extended Object Results. This subsection presents results for extended object imaging. Figure 7 (a) shows the diffraction limited image of the simulated satellite for the parameters given in Figure 6. Figure 7 (b)-(f) display a sampling of images output from the AOS simulation from Cases 1 and 2. Specific input parameters that were varied to produce these images and the remainder of the images shown in this chapter are shown in Table 3. Other parameters were set as shown in Figure 6.

Figure 7 also displays the corresponding RMSE and *MSR2* value for each image shown. One can see that as the value of RMSE increases, the respective image shows decreasing clarity and detail. The same holds for decreasing values of *MSR2*. Figure 8 (a) shows the modulus of the diffraction limited OTF for the optical system modeled in this research. Figure 8 (b)-(f) show the corresponding radially averaged image spectrum SNR, the modulus of the average compensated system optical transfer function, |<OTF>|, and the Strehl ratio. As discussed earlier, Figure 8 also serves to demonstrate how the $SNR_D(\vec{f})$ function takes on a limiting value of 0.886 outside the effective pass band. As verified by simulations, this result holds regardless of the number of individual frames that are run.

46

Table 3. Summary of Parameters Used for Images Displayed in Chapter III

Image Label	Fried Parameter (m)	Wind Speed (m/s)	Object Light Level (m_v)	WFS Light Level (m_v)	Frame Exposure Time (s)
Figure 7 (b)	0.25	15.63	4.25	4.25	0.10
Figure 7 (c)	0.25	15.63	6.0	6.0	0.10
Figure 7 (d)	0.25	31.25	8.0	8.0	0.10
Figure 7 (e)	0.25	15.63	10.0	10.0	0.10
Figure 7 (f)	0.25	31.25	10.0	10.0	0.10
Figure 14 (a)	0.12	10.42	6.0	0.0	0.055
Figure 14 (b)	0.12	10.42	6.0	12.0	0.055
Figure 14 (c)	0.12	10.42	6.0	14.0	0.055
Figure 14 (d)	0.12	10.42	8.0	0.0	0.055
Figure 14 (e)	0.12	10.42	8.0	12.0	0.055
Figure 14 (f)	0.12	10.42	8.0	14.0	0.055
Figure 15 (b)	0.25	15.63	4.25	4.25	0.10
Figure 15 (c)	0.25	15.63	6.0	6.0	0.10
Figure 15 (d)	0.25	31.25	8.0	8.0	0.10
Figure 15 (e)	0.25	15.63	10.0	10.0	0.10
Figure 15 (f)	0.25	31.25	10.0	10.0	0.10

Figures 9, 10, and 11 show scatter plots of the different metrics versus RMSE for Cases 1, 2, and 3. In addition to the Strehl ratio, the modified Strehl ratio and noise-effective-cutoff frequency that provided the best correlation with RMSE are plotted in these figures. In Case 1, the Strehl ratio and modified Strehl ratio exhibit high correlation with RMSE. For both SR and $MSR1$, $r = 0.82$. To a much lesser degree, there is some correlation between the noise-effective-cutoff frequency and RMSE, but it is not as high as the Strehl ratio and modified Strehl ratio.
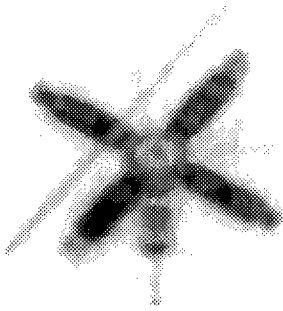
(a) Diffraction Limited Image　　(b) RMSE=0.00129, MSR2=0.70　　(c) RMSE=0.00147, MSR2=0.46
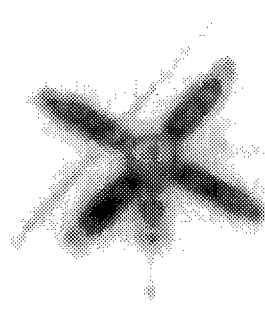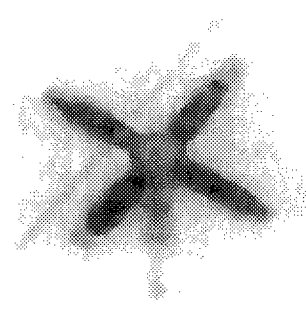
(d) RMSE=0.00178, MSR2=0.28　　(e) RMSE=0.00280, MSR2=0.04　　(f) RMSE=0.00310, MSR2=0.03
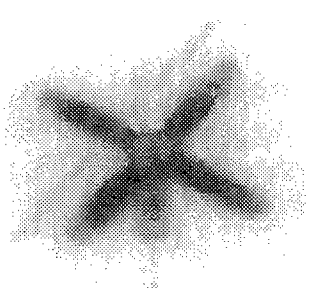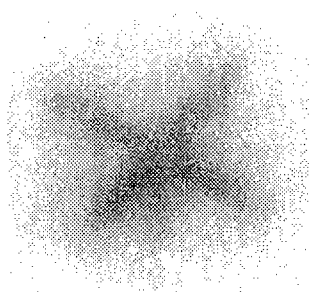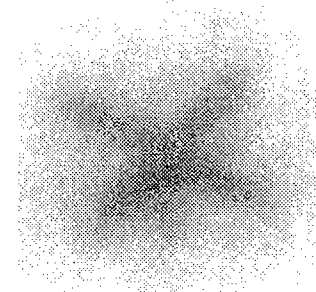
Figure 7. Diffraction Limited Image and Sample of AOS Output Images for Cases 1 and 2

The highest correlation between the noise-effective-cutoff frequency and RMSE in this case was $r = 0.52$ for $NEC2$. The reason that the noise-effective cutoff performs poorly under these relatively bright object conditions is explained below.

Under bright conditions, the image spectrum SNR does not approach one until nearly the cutoff frequency of the optical system. At these higher frequencies the image spectrum SNR plots for different images all begin to converge. This means that the noise-effective-cutoff frequency under bright conditions would tend to indicate that different images are nearly the same, even if they are not. On the other hand, even under bright conditions, there can be a clear

(a) Diffraction Limited OTF

(b) SR = 0.713

(c) SR = 0.498

(d) SR = 0.364
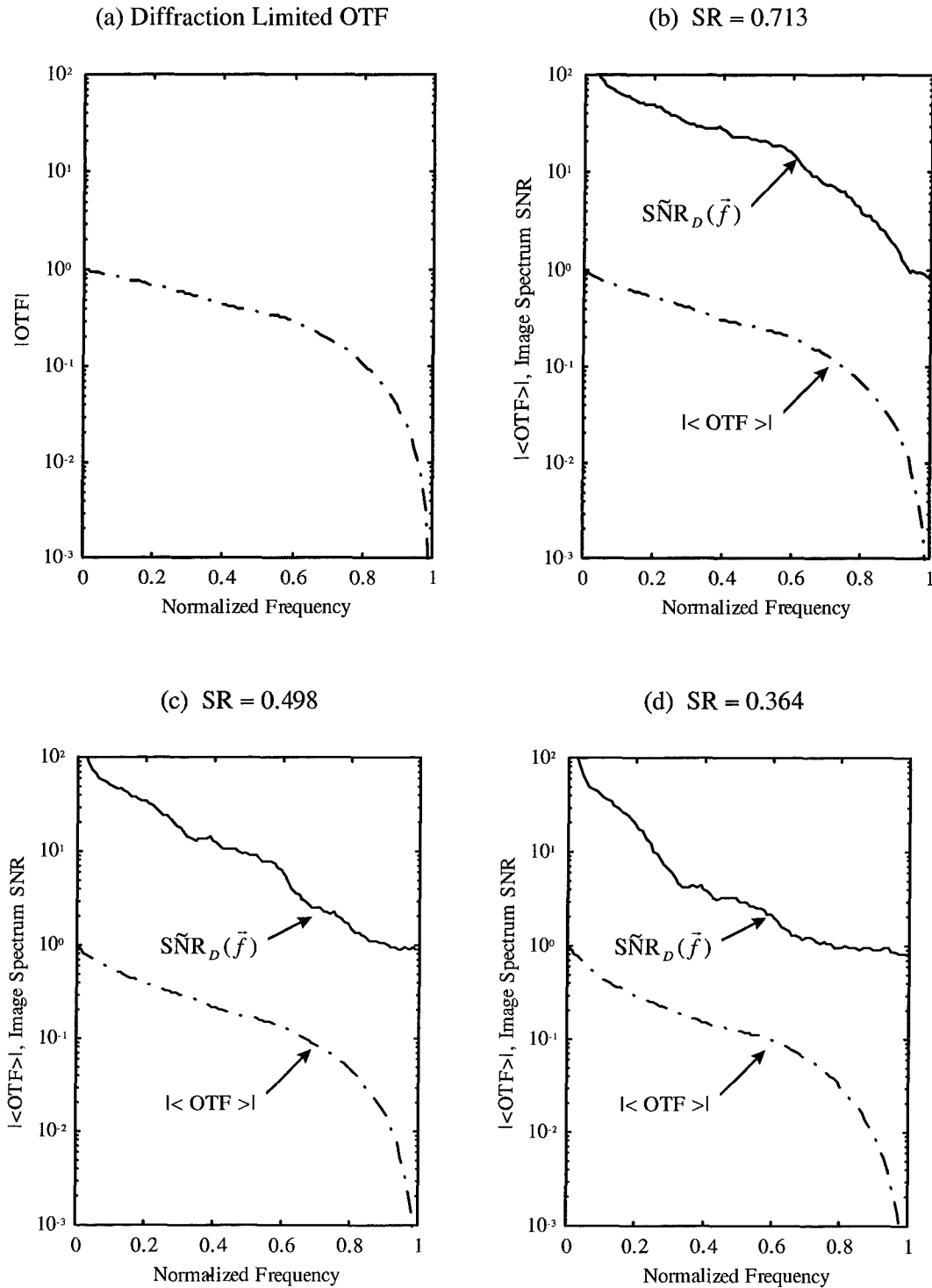
Figure 8. Modulus of OTF, Image Spectrum SNR, and *SR* Metric for Each Image Displayed in Figure 7. In this figure, the dashed line represents the OTF and the solid line represents the image spectrum SNR.

49

(e) SR = 0.127

(f) SR = 0.102
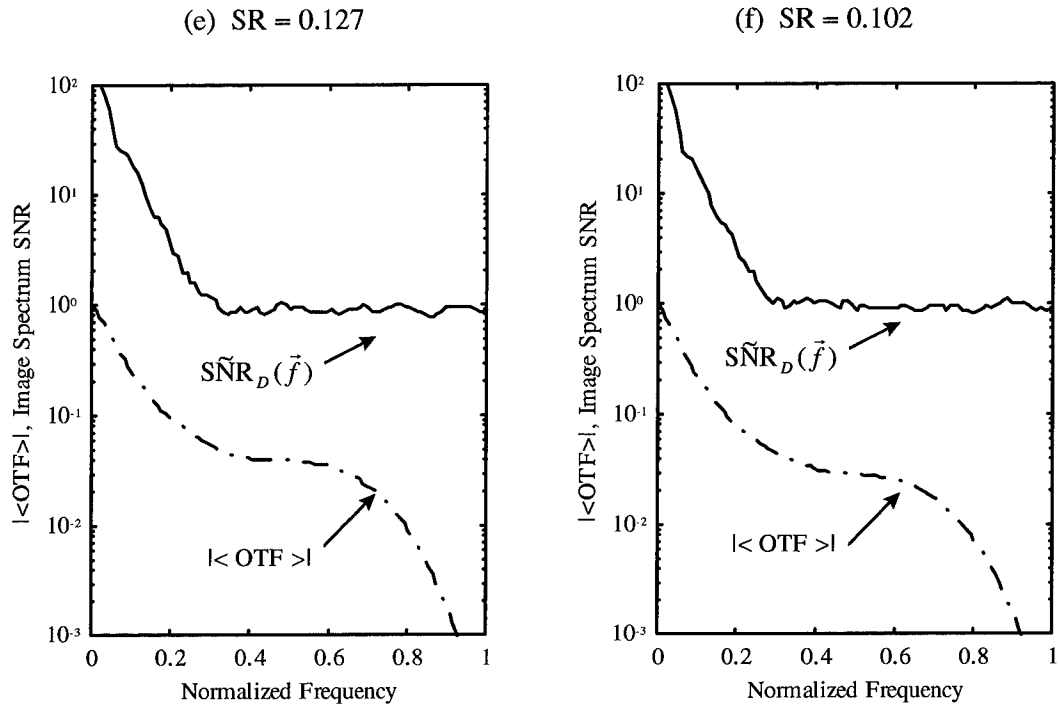
Figure 8 (continued)
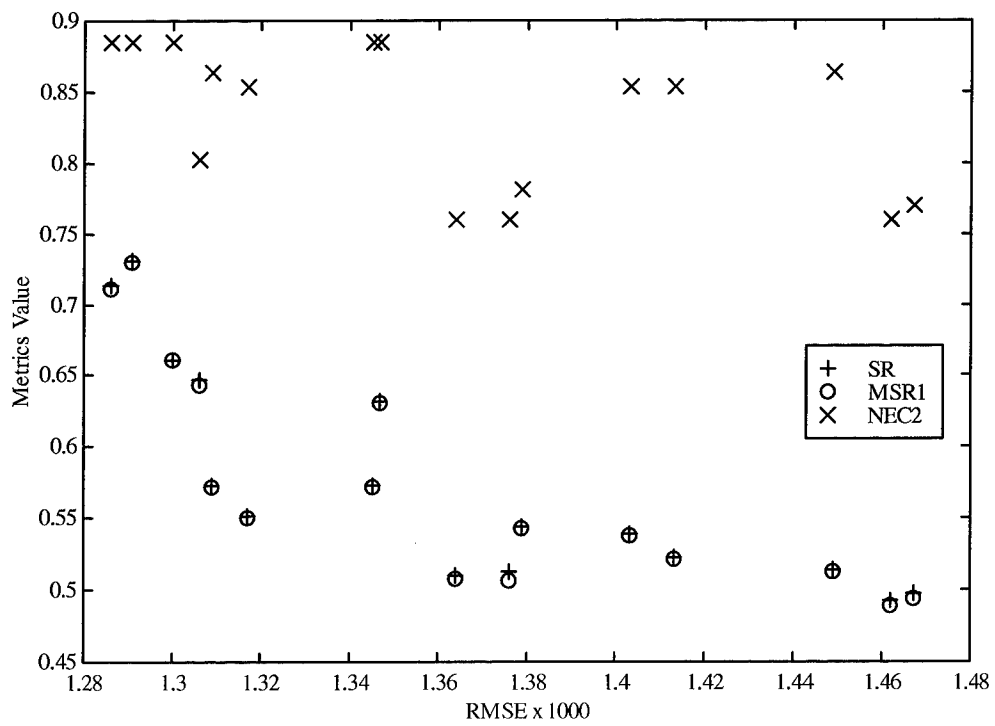
Figure 9. Metrics versus RMSE for Case 1

50
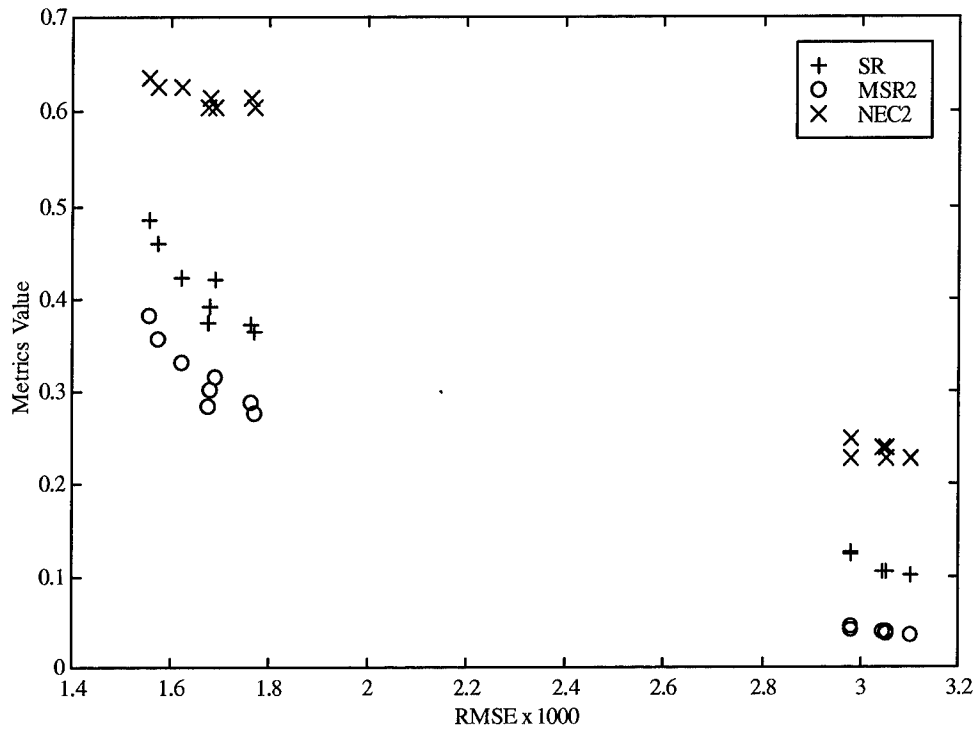
Figure 10. Metrics versus RMSE for Case 2

difference between respective image OTFs as influenced by the Fried parameter, wind speed, or other factors. The computed Strehl ratios or modified Strehl ratios will therefore be different so that they can be used to discern differences in image qualities.

In Case 2, the Strehl ratio and modified Strehl ratio still show strong correlation with RMSE. For both *SR* and *MSR2*, $r = 0.99$. However, this correlation starts to depart from a linear relationship. In this low light level case, the noise-effective-cutoff frequency shows much higher correlation with RMSE with $r = 0.99$ for *NEC2*.

For Case 3, curvature of the Strehl ratio and modified Strehl ratio versus RMSE is apparent. Based on this, nonlinear forms of these metrics versus RMSE were examined. A nonlinear form increases correlation because as defined the correlation coefficient can be thought of as how well

51
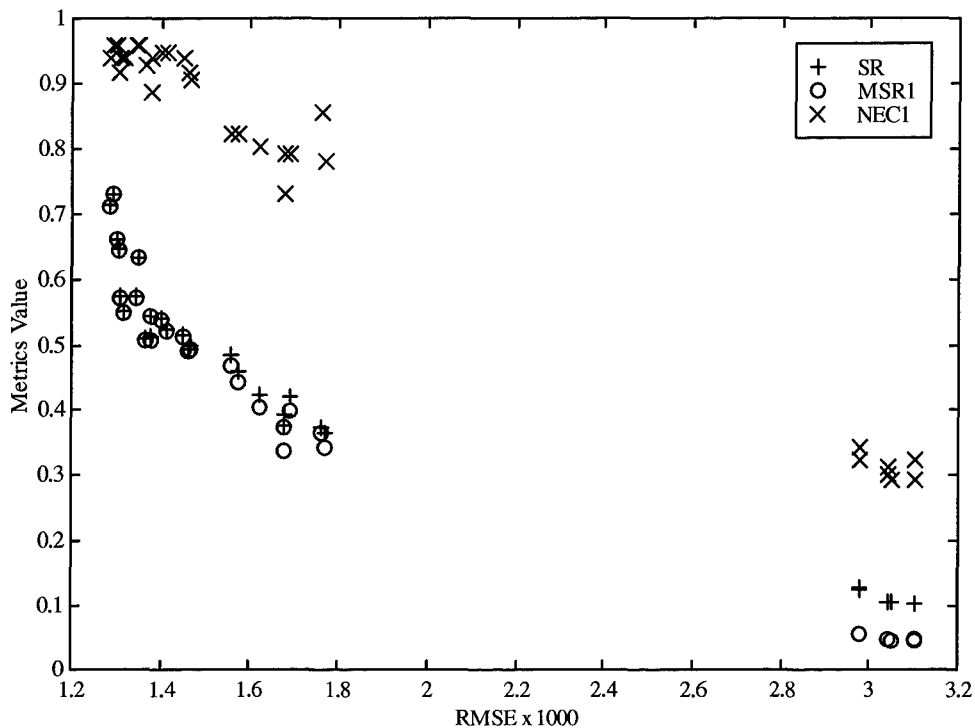
Figure 11. Metrics versus RMSE for Case 3

a straight line can be fit to the data. As curvature between values of the metrics plotted against

RMSE increase, a straight line will not fit the data very well. However, a nonlinear form, such as

the natural log of *SR*, results in a more linear relationship so that the correlation coefficient

between ln(SR) and RMSE is higher than that between *SR* and RMSE. Figure 12 shows the

natural logarithm of the Strehl ratio versus RMSE for Case 3 and a fitted regression line. This

results in a very high correlation between ln(*SR*) and RMSE, with $r = 0.99$. The performance of

ln(*MSR2*) is nearly identical to that of ln(*SR*).

In Figure 11, there is a discernible "gap" between the original Case 3 data. This gap is

primarily a result of a drastic decrease in performance as the object light level is reduced.

Specifically, the gap occurs between an object light level of $m_v = 8$ and 10. Consequently, additional validation runs were made between these light levels to fill in the gap and to demonstrate that the fitted regression line could be used to accurately predict RMSE as a function of SR. Specifically, four additional runs each were conducted at $m_v = 8.5, 9.0,$ and 9.5 using different combinations of the Fried parameter and wind speed as shown in Table 2 for Case 3. RMSE and the SR metric for each of these runs were computed and added the results to obtain the validation points as shown Figure 12. One can see that these points fall very close to the regression line fitted to the original Case 3 data. This indeed confirms that the natural logarithm
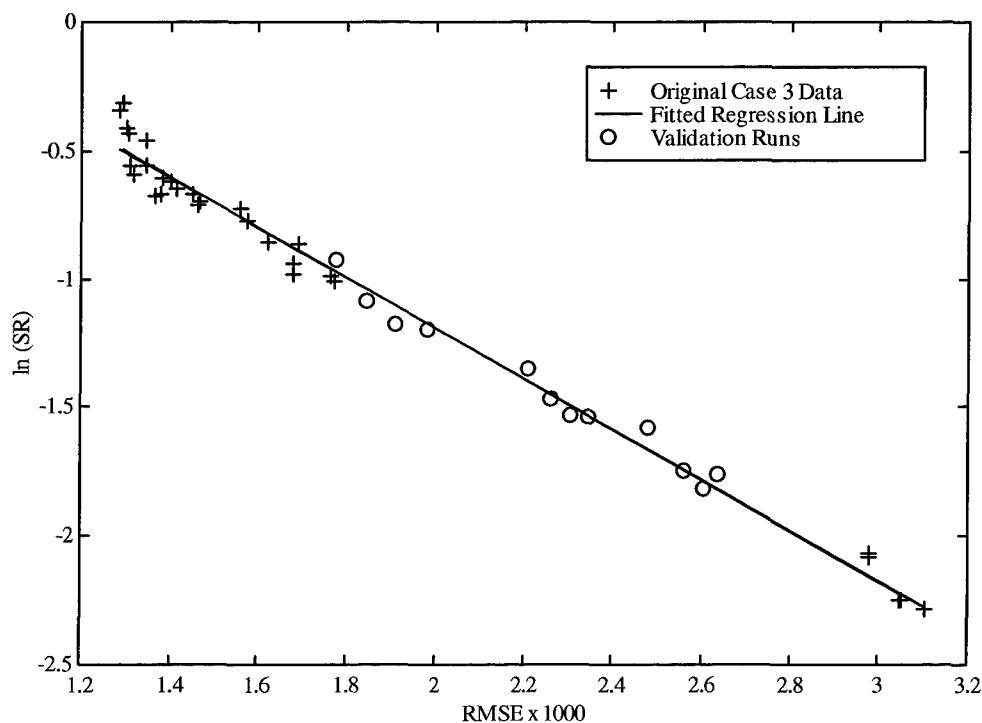
Figure 12. Linear Regression of ln(SR) versus RMSE for Case 3. Validation runs were added to fill in the apparent "gap" in the data and to verify the validity of the regression model. The computed regression model yields $ln(SR) = 0.774 - 984.6 \times RMSE$ with $r = 0.99$.

53

of SR could be used to accurately predict RMSE and that these results hold across a wide range of image qualities.

Figure 13 visually displays the performance of each metric for each case, the performance measure being the correlation coefficient. It is evident that the Strehl ratio and the modified Strehl ratio using an SNR level of 1 or 2 provide the highest correlation with RMSE across all cases, except Case 4 which will be discussed below. As an example, the correlation coefficients for *SR* are 0.82, 0.99, and 0.96 for the first three cases respectively. Similarly, the correlation coefficients for *MSR2* are 0.81, 0.99, and 0.93. In contrast, the correlation coefficients for *NEC2* are 0.52, 0.99, and 0.97 for the first three cases respectively. This indicates the noise-effective-cutoff frequency metric is less consistent across diverse image qualities than the former two metrics. However, since its performance is very comparable to the Strehl ratio and modified
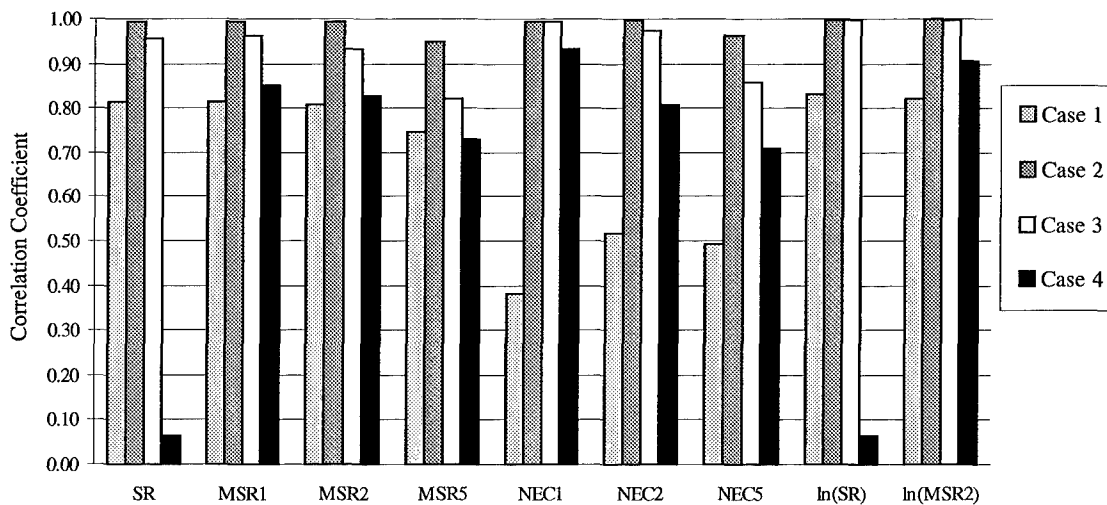
Figure 13. Diagram Depicting Performance of Different Metrics For Cases 1, 2, 3, and 4. This figure displays the correlation coefficient between the examined metrics and RMSE for Cases 1 through 4. The SNR level used for each metric is appended to the end of the metric. For example, MSR1 is the modified Strehl ratio that uses an SNR level of 1.

Strehl ratio at low light levels, the noise-effective-cutoff frequency should not be dismissed altogether. If working under a low light level regime, the noise-effective-cutoff frequency may serve as an adequate metric to quantify different image qualities. Figure 13 also shows how the performance of the *SR* and *MSR2* metrics improve by incorporating their natural logarithms to account for curvature.

Based on the above assessment, it appears that the Strehl ratio would be a sound choice for characterizing different image qualities. However, in dealing with adaptive optics systems, situations exist where the Strehl ratio by itself cannot be used to distinguish images of different quality. Specifically, under conditions that yield a constant average system OTF, such as would be obtained from an artificial guide star of constant brightness or laser beacon, the Strehl ratio will remain constant from image to image. Yet, image quality can vary as a function of target light levels. In this case, the modified Strehl ratio and the noise-effective-cutoff frequency, can still account for differences in image quality. To demonstrate this, this work studied extended object brightness of $m_v = 0, 2, 4, 6, 8, 10, 12$, and 14 and WFS beacon levels of $m_v = 6$ and 8. Other HYSIM input variables were set at the midpoint of their ranges given in Figure 6. To portray reality, a different random seed was used for each simulation run so that the associated output Strehl ratio varies slightly from image to image. It is noted that in this illustration of a laser beacon, the issue of estimating tilt separately from a natural guide star is not addressed, nor is the issue of focus anisoplanatism addressed.

Figure 14 displays a sampling of output images for this case along with the RMSE, *SR*, and *MSR2* value for each image. The specific values of the parameters used to obtain each image are shown in Table 3. As expected, one can see how the error increases and the *MSR2* value decreases as the object brightness decreases across each row, whereas *SR* remains constant. Similarly, for the same object brightness, both RMSE and *MSR2* behave as expected. That is,

(a) RMSE=0.00132, SR=0.58, MSR2=0.56   (b) RMSE=0.00288, SR=0.58, MSR2=0.18   (c) RMSE=0.00520, SR=0.56, MSR2=0.12
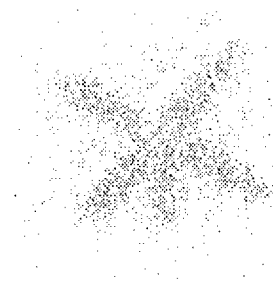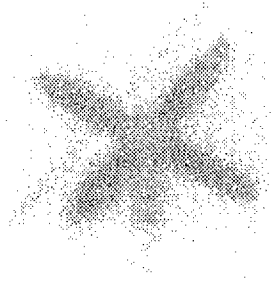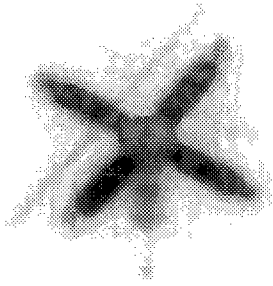
(d) RMSE=0.00153, SR=0.44, MSR2=0.43   (e) RMSE=0.00308, SR=0.44, MSR2=0.12   (f) RMSE=0.00532, SR=0.43, MSR2=0.09
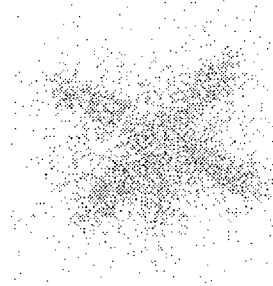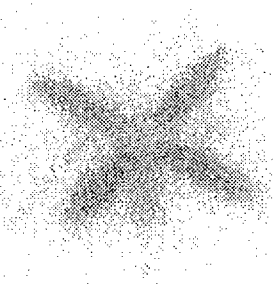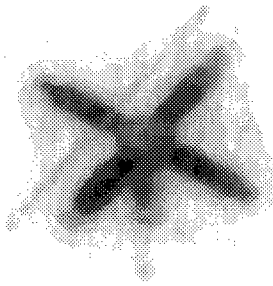
Figure 14. Sample of AOS Output Images for Case 4. The first row has WFS $m_v = 6$ and second row has WFS $m_v = 8$. The columns represent object light levels of $m_v = 0, 12,$ and 14 respectively.

comparing panels (a), WFS $m_v = 6$, and (d), WFS $m_v = 8$, for instance, the image in panel (a) has

a smaller error and larger *MSR2* value than the image in panel (d). This is because the WFS will

see more photoevents in (a) than (d) and therefore it will be able to achieve better compensation.

A summary of the performance of each metric for this case are displayed in Figure 13 under

Case 4. One can see how the correlation of the Strehl ratio with RMSE completely deteriorates

in this case, with a correlation coefficient of $r = 0.065$. This is because the computed Strehl ratio

is essentially constant except for slight variations due to different random seeds. However, the

image quality clearly gets worse with increasing visual magnitude as was shown in Figure 14.

This is due to increased noise in the detected image as the object gets dimmer. The Strehl ratio does not describe the image degradation for this situation. On the other hand, the modified Strehl ratio and the noise-effective-cutoff-frequency still perform relatively well for this case. As an example, $MSR1$ has a correlation coefficient of $r = 0.85$ and $NEC1$ has a correlation coefficient of $r = 0.93$. This is because these metrics can account for noise in the detected image. Furthermore, it is noted that the Strehl ratio cannot account for differences in exposure times of the detected image. This is because exposure time does not influence the Strehl ratio at all. Yet, the modified Strehl ratio and the noise-effective-cutoff-frequency can be influenced by exposure times through differences in the image spectrum SNR of the detected image.

Point Source Results. This subsection presents results for point source imaging. The images were produced using the same exact AOS input settings as in Case 2. Recall, in this case $m_v = 8$ and 10. Figure 15 (a) shows the diffraction limited image of the point source. Figure 15 (b)-(f) display a sampling of point source images output from the AOS simulation for this case, along with the RMSE and $MSR2$ value for each image. The specific input parameters used to produce each of these images are shown in Table 3.

As before, it can be seen that as the value of RMSE increases, the respective image shows decreasing clarity and detail. Likewise, the same also holds for decreasing values of $MSR2$. The computed metrics where again correlated with RMSE. The resulting correlation coefficients are displayed graphically in Figure 16. The same high correlation levels are seen as in Case 2 with a correlation coefficient of r = 0.99 for SR, $MSR1$, and $NEC1$. Hence, the demonstrated results hold for not only extended objects, but also for point sources as well.

(a) Diffraction Limited Image   (b) RMSE=0.00106, MSR2=0.49   (c) RMSE=0.00134, MSR2=0.40

(d) RMSE=0.00160, MSR2=0.36   (e) RMSE=0.00389, MSR2=0.13   (f) RMSE=0.00434, MSR2=0.10

Figure 15. Diffraction Limited Image and Sample of AOS Output Images for Case 5

Conclusions

The goal of this chapter was out to identify image quality metrics that can be used to

characterize adaptive optics system performance under a wide variety of imaging environments.

The results presented demonstrate that the Strehl ratio and modified Strehl ratio can serve as

suitable image quality metrics under most situations. Both of these metrics were shown to

provide a high level of correlation with the RMSE between images obtained with a simulated

AOS and the diffraction limited image of the modeled adaptive optics system. This high level of

correlation was seen under diverse conditions and system settings. The performance of *SR* and
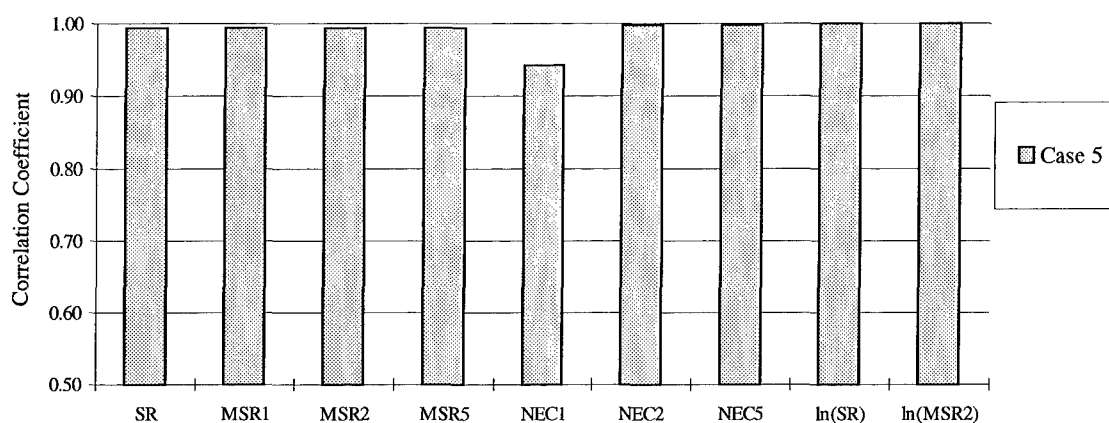
Figure 16. Diagram Depicting Performance of Different Metrics For Case 5. This figure displays the correlation coefficient between the examined metrics and RMSE for Cases 5. The SNR level used for each metric is appended to the end of the metric. For example, MSR1 is the modified Strehl ratio that uses an SNR level of 1.

MSR1 and *MSR2* in particular were seen to be the best across the different cases examined. However, it was also shown that the performance of the Strehl ratio is severely impacted under conditions that yield a constant average system OTF. In this case, the ability of *SR* to distinguish between different image qualities is essentially useless. On the other hand, the modified Strehl ratio retained a high degree of correlation with RMSE under this case. This is because the modified Strehl ratio can account for measurement noise in the detected image. Finally, it was shown that the modified Strehl ratio can account for differences in image exposure times whereas the Strehl ratio cannot.

The noise-effective-cutoff frequency metric provided good performance under low light level conditions. It also had the ability to distinguish between different image qualities under conditions that yield a constant OTF. However, its performance across the different cases examined did not yield the same consistency as did the modified Strehl ratio. This is primarily

59

because it performs poorly under bright light level conditions as explained earlier. Hence, it is

concluded that the modified Strehl ratio is the best metric of the ones examined.

IV. Adaptive Optics System Performance Analysis and Optimization

Introduction

This chapter addresses the second research objective -- delineate input settings that maximize AOS performance for a given set of seeing conditions. In the previous chapter image quality metrics were developed that can be used to characterize adaptive optics system performance. These results are now used to examine the affects that changing the value of both controllable and uncontrollable system variables have on overall AOS performance.

The initial parts of this research concentrated on Response Surface Methodology (RSM) and Design of Experiments (DOE) techniques. These techniques were utilized to gain a coarse understanding of the relationships between controllable and uncontrollable variables, including identification of the most salient input variables and interactions between examined factors. This information was then used to pare down the examined variables and to conduct a more thorough analysis of the remaining factors.

This research continues to make use the adaptive optics simulation, HYSIM. However, the system parameters have been modified from those presented in Chapter III to match those of the AMOS 1.6 meter adaptive optics system.

RSM/DOE Analysis Summary

This section serves to summarize the RSM and DOE analysis of the AMOS system modeled in this research.

Background and Setup. The RSM/DOE analysis portion of this research was primarily aimed at gaining a coarse understanding of the relationships between controllable and uncontrollable variables, including identification of the most salient input variables and

interactions between examined factors. This information was then used to pare down the examined variables and to conduct a more thorough analysis of the remaining factors.

The approach used in this section was to first use system experts to identify pertinent controllable and uncontrollable factors that may influence AOS performance. The pertinent factors identified are shown in Figure 17. The capital letters indicate the coded value of each factor using the coding scheme provided in Appendix C.

- A = sample rate, s (Hz)
- B = gain, k_g (rad/s)   ⟸ Controllable factors
- C = short frame exposure time, t_s (s)

- D = object visual magnitude, m_v
- E = Fried parameter, r_0 (m)   ⟸ Uncontrollable factors
- F = wind speed, v (m/s)
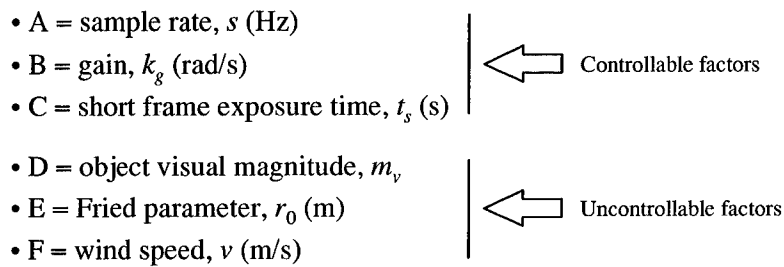
Figure 17. Controllable and Uncontrollable Factors Examined in RSM/DOE Analysis

The system experts also helped to identify typical values and ranges for the controllable and uncontrollable variables used in the RSM/DOE analysis. These values are summarized in Figure 18. This figure also provides parameters for the modeled AMOS 1.6 meter adaptive optics system. With this, as described in Appendix B, a 2^{6-1} resolution VI design was used to examine the effects of the six initial factors. This design ensures that no first or second order interactions are confused, or aliased, with any other first or second order interactions. This design can also be used to examine third order interactions, although some of these may be aliased with other third order interactions. Table C1 of Appendix C summarizes the 2^{6-1} resolution VI design incorporated in this section.

Controllable Parameters   AMOS System Parameters

- Number of frames: 100/1000
- System sample rate: 500/1000 Hz
- System gain: 0.05/0.10
- Frame exposure time: 0.002/0.02 s
- Average long exposure time: 2 s

- Primary mirror diameter: 1.6 m
- Obscurration diameter: 0.335 m
- Imaging wavelength: 7.0E-07 m
- WFS wavelength: 6.0E-07 m
- Number of subapertures across pupil diameter: 16
- DM actuator separation in pupil plane: 0.11 m

System Variables

- - - - - - - - - - - - - - - - - - Inputs - - | AOS (Model by HYSIM) | Outputs

Uncontrollable
Variables

- Average long exposure image
- Average long exposure OTF
- Image Spectrum SNR

- Extended object target distance: 6.5E+05 m
- Object and WFS visual magnitudes: 6/9
- Fried parameter: 0.10/0.20 m
- Average local wind velocity: 12.5/25.0 m/s
- Physical size of array holding input object: 40.0 m

Figure 18. Summary of Parameters Used to Model AMOS 1.6 Meter Adaptive Optics System and Values of Variables Used in RSM/DOE Analysis

Results. Three replications of the design matrix given in Table C1 of Appendix C were made using different random number seeds and the *MSR2* metric was computed for each of the 32 different design points. These data were used to construct a plot of main effects as shown in Figure 19. This figure shows that by far the most significant factor affecting AOS performance is the object visual magnitude. This is followed by the system gain setting and the Fried parameter. This basic result is built on further in the next section when the system closed loop bandwidth is optimized as a function of seeing conditions and light levels. In the meantime, this section summarizes a few other results from the RSM/DOE analysis.
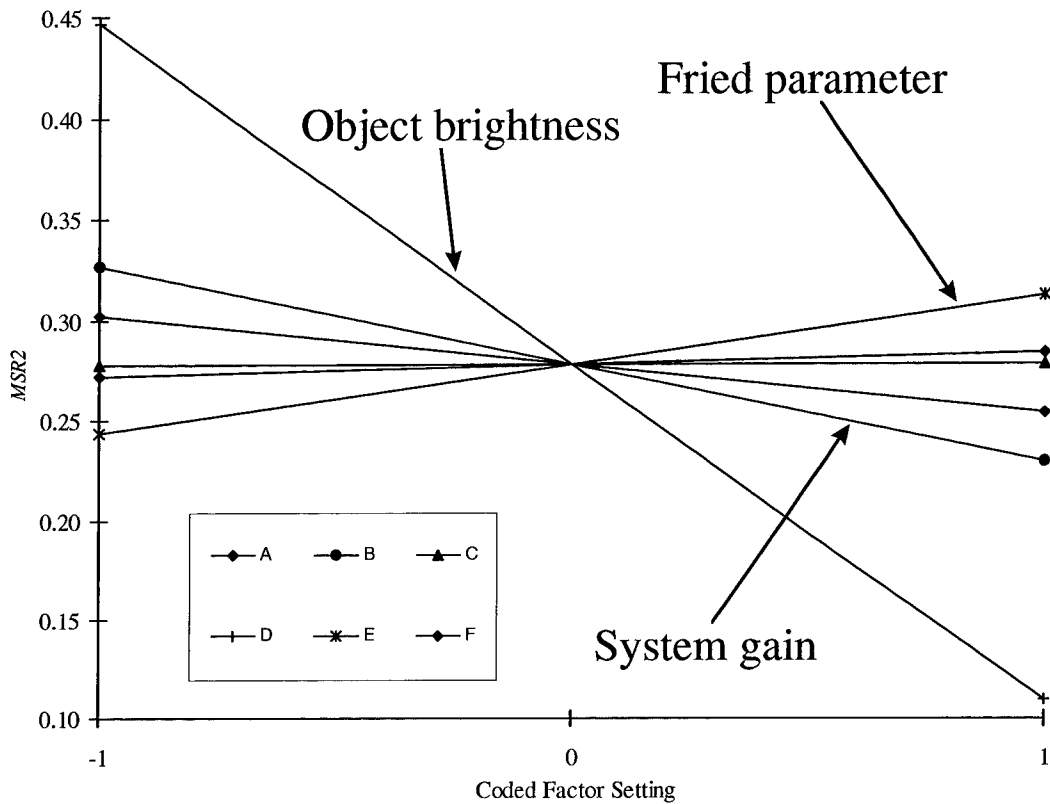
Figure 19. Main Effects of Preliminary Factors Considered in RSM/DOE Analysis

A normal probability plot was constructed to ascertain the presence of any significant interaction effects that should be included in a regression model. This plot is shown Figure C2 of Appendix C. This figure reveals potential second order interactions that should be included in a regression model. There are no significant third order interactions. With this knowledge, a linear regression model was fit to the *MSR2* response metric. Two center point runs were also included to facilitate a formal lack of fit test. The results of this regression are shown in the analysis of variance table, Table C2, in Appendix C. Table C3 of Appendix C summarizes the lack of fit analysis that was accomplished. This test shows second order terms may need to be included in the regression model. Accordingly, axial points were added to the original design

matrix to ascertain the presence of second order terms. Based on including predictor variables

with a P-value of less than 0.05, the data show that the second order term k_g^2 and m_v^2 should

be included in the regression model. The results of this regression are shown in the analysis of

variance table, Table C4, in Appendix C for the coded variables. Equation (22) provides the

equation for the fitted $MSR2$ metric, $\hat{MSR2}$, as a function of the pertinent controllable and

uncontrollable variables in their uncoded form. This function is valid over the ranges for the

factors given in Appendix C. However, tests show that this function is still fairly accurate over a

wider range than those provided. Figure 20 provides a visual comparison between the observed

$MSR2$ values and the fitted $MSR2$ values.
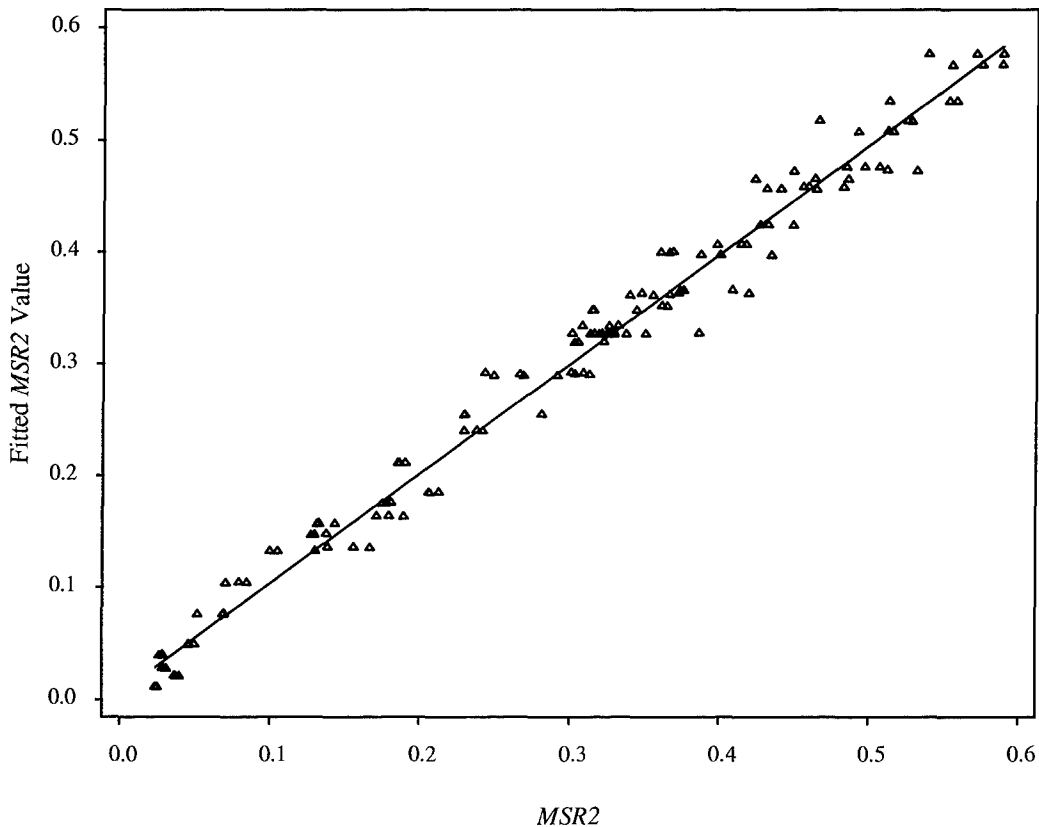
Figure 20. Fitted $MSR2$ Value versus Observed $MSR2$ Value Using Regression Model Provided in Table C4 of Appendix C

$$M\hat{S}R2 = 2.76\,r_0 - 2.28 \times 10^{-4}\,s - 0.0225\,v + 1.203\,k_g + 0.141\,m_v + 0.00142\,m_v v - 0.957\,k_g{}^2 -$$
$$0.0106\,m_v{}^2 - 5.71 \times 10^{-5}\,s\,m_v + 1.069 \times 10^{-5}\,s\,v - 0.0815\,k_g\,m_v - 0.276\,m_v\,r_0 - 0.0855 \tag{22}$$

Conclusions. It was shown how RSM/DOE can be used to develop a model to predict AOS performance. In particular, the regression model given in Equation (22) can be used to accurately estimate the performance of the AMOS 1.6 meter system as a function of system gain, sample rate, wind speed, Fried parameter, and object visual magnitude. Based on realtime observed values of the uncontrollable factors, prudent choices could be made as to how to set the controllable factors of system gain and sample rate.

This analysis has shown that the most significant factor affecting system performance is object visual magnitude. In addition, atmospheric seeing as characterized by the Fried parameter was seen to be the next most significant uncontrollable factor. System gain was the most important controllable factor. Short frame exposure time is not significant. This will be discussed further shortly. Based on discussion with the system experts, the system gain is the most commonly varied controllable factor during realtime AOS operations. System sample rate is often a fixed quantity. Hence, based on all of these considerations, a more thorough analysis of optimizing the system closed loop bandwidth* as a function of object light levels and seeing conditions is presented in the next section. [*Note: As will be presented in more detail later, the system closed loop bandwidth is a function of system sample rate and gain, but the approach used in the next section was to fix the sample rate and vary the gain. So in reality, optimizing the closed loop bandwidth is akin to optimizing the system gain.]

As shown in the previous chapter, the real virtue of the modified Strehl ratio is that it allows insight into performance at the WFS detector, as indicated by the Strehl ratio, and overall AOS performance, as indicated by the image spectrum SNR. It was shown that this is particularly

applicable to an AOS using a laser beacon so that the light levels to the WFS and imaging detectors are independent. However, optimizing the performance of laser guided systems is beyond the scope of this research. Moreover, the RSM and DOE analysis have led us to conclude that for a fixed long exposure time, the short frame exposure time is not that significant. Hence, the goal for the remainder of this research is to simply maximize the Strehl ratio, and in particular, the instantaneous Strehl ratio which is defined shortly. This goal is accomplished by optimization of the system closed loop bandwidth for a given set of seeing conditions and light levels.

Instantaneous Strehl Ratio

The Strehl ratio given by Equation (9) in the previous chapter is for the Optical Transfer Function (OTF) averaged over the long exposure time. The Strehl ratio for the compensated system OTF associated with each incremental update of the adaptive optics system Deformable Mirror (DM) can also be calculated. This quantity is defined as the Instantaneous Strehl Ratio (ISR). In equation form [15:139]

$$ISR = \frac{\int MTF_{ist}(\vec{f})\,d\vec{f}}{\int MTF_{dl}(\vec{f})\,d\vec{f}} \tag{23}$$

where MTF_{ist} is the modulus of the instantaneous system OTF, and MTF_{dl} is the modulus of the diffraction limited OTF of the system.

By virtue of the use of an AOS simulation, the instantaneous OTF is relatively easy to obtain. In this case, the instantaneous OTF corresponds to the OTF of the combined optical system and residual atmospheric perturbation after each update of the DM. The duration of each

instantaneous OTF is the reciprocal of the system sample rate. This metric allows continuous insight into AOS performance. The sample mean of the observed *ISR* values during a particular simulation run can also easily be computed and used as an indication of overall AOS performance. Note that higher *ISR* or mean *ISR* values correspond to improved resolution and improved imaging performance.

It was previously shown that the Strehl ratio and the modified Strehl ratio are both highly correlated with the RMSE between the detected image and the corresponding diffraction limited image. As a check of validity for the *ISR* metric, over 100 long exposure images under different seeing conditions and light levels were produced. The *MSR2* metric and the mean *ISR* value were computed for each image. Next, the correlation coefficient between these metrics was calculated using the formula given in Equation (21) in Chapter III. As expected, this analysis shows that the mean *ISR* value is positively correlated with *MSR2*, with a correlation coefficient of about 0.97. Hence, maximizing the *ISR* will lead to improved imaging system performance.

Closed Loop Bandwidth Optimization

This section presents results of research aimed at optimizing adaptive optics closed loop bandwidth settings to maximize system performance. The optimum closed loop bandwidth settings are determined as a function of target object light levels and atmospheric seeing conditions. This work shows that for bright objects, the optimum closed loop bandwidth is near the Greenwood frequency. However, for dim objects, the preferred closed loop bandwidth settings are many times lower than the Greenwood frequency. In addition, under low light levels, selection of the proper closed loop bandwidth is much more critical for achieving maximal performance than under high light levels.

Background. In operating an Adaptive Optics System (AOS), proper selection of the closed

loop bandwidth, f_c, is a crucial step in achieving successful performance. As stated by Gavel *et*

al [24], "Once an adaptive-optics system is installed on the telescope, much of the flexibility in

choosing the optimum parameters is lost. The number of subapertures, and therefore, *d*

[subaperture side dimension], is fixed by the choice of deformable mirror. The cost of laser

power is mostly in building the laser, so varying the total laser power during operation is not of

great benefit. However, the control-loop bandwidth can certainly be made adjustable."

As will be demonstrated later, if f_c is too low, the system will lag too far behind the

changing atmosphere to provide the highest level of performance. On the other hand, if f_c is too

high, the system can become unstable due to low SNR in the wavefront sensor arising from the

necessarily short sample times, and performance can even become worse than that of an

uncompensated system. The challenge is how to select the proper closed loop bandwidth for a

given set of light levels and seeing conditions in order to maximize overall imaging system

performance. This chapter provides a strategy for addressing this question and presents results

obtained by modeling a 1.6 meter adaptive optics system using a closed loop adaptive optics

simulation. First, a short review of other work that relates to closed loop bandwidth optimization

is presented.

Ellerbroek [23] develops first-order analytical models for AOS subcomponents. Using

these equations and numerical evaluation methods, he provides a table that gives the short

exposure Strehl ratio versus closed loop bandwidth for different amounts of sensor noise and for

different types of guide stars (i.e., natural or laser). For his setup, natural guide stars and higher

values of f_c yield the best performance. The values of f_c he uses are 10, 20, and 30 Hz

respectively. His results show increasing performance with higher values of f_c. In every

instance, the higher f_c values yield better performance for any given amount of sensor noise.

Later, it is shown that as wavefront sensor noise increases, due to decreasing target object brightness, the optimal value of f_c decreases. Ellerbroek's work is based on nominal atmospheric conditions for the AMOS observatory at Mount Haleakala, Hawaii. For a wavelength of $\lambda = 0.5$ μm, these are a Fried parameter of $r_0 = 0.285$ m and a Greenwood frequency of $f_g = 19.7$ Hz.

Harrington and Welsh [32] provide a frequency-domain analysis of the temporal response of adaptive optics systems. Their analysis accounts for DM piston and tilt removal and spatial bandwidth limitations due to the finite size of the WFS subapertures and DM actuator separation. They examine the mean square value of the residual phase error of corrected wavefronts. The primary objective of their work is to describe performance effects of an adaptive optics system's temporal response by accounting for the above factors. Their work does not account for a varying Signal-to-Noise Ratio (SNR) to the WFS as a function of the temporal response, but the significant conclusions of their research is that finite sample spacing of WFS subapertures and DM actuators has little effect on system performance for large temporal response times. As the temporal response time of the system becomes faster, finite sample spacing does become the limiting factor of the adaptive optics system performance.

Gaffard and Boyer [33] present an adaptive optics optimization scheme for a given optics diameter and a set of different values for the Fried parameter. They then try to optimize performance by proper choice of the number of DM actuators and the influence diameter, "a parameter which characterizes the size of the required corrections," by maximizing the optical transfer function. They assume that the WFS is distortion free and that residual distortions are due only to a limited number of DM actuators. Hence, no consideration is given to varying light levels.

Greenwood [34] derives an expression for the bandwidth of an AOS used for astronomical imaging. For what he terms "near-worst-case" atmospheric turbulence conditions for the Air

Force Maui Optical Station (AMOS), Greenwood concludes that the necessary bandwidths are less than 200 Hz. Some of the results presented later are in agreement with this finding.

Ellerbroek *et al* [35] present a paper on optimizing the closed loop performance of adaptive optics systems that incorporate multiple control bandwidths. That is, systems that use more than one wavefront sensing beacon. They discuss how time varying distortions seen by and adaptive optics system are best corrected by using a high control bandwidth. At the same time they acknowledge how noise in the WFS measurements are decreased by reducing the control bandwidth and "temporal smoothing" of the wavefront correction to be applied to the DM. Indeed, this research supports this hypothesis. For the single bandwidth case, they present a plot that shows how the normalized residual phase variance reaches a minimum as the normalized servo lag is increased as WFS noise increases. The authors conclude that the use of multiple control bandwidths significantly reduces allowable WFS noise levels for the system to operate at performance limits associated with fitting error. They also show how the performance of a single bandwidth system can be improved by optimizing both the control bandwidth and wavefront reconstruction algorithm.

Gavel *et al* [24] describe a technique for optimizing the design of laser-guided adaptive optics to maximize the system Strehl ratio for a given amount of laser power, or to minimize laser power requirements to achieve a specified Strehl ratio. They state that the system design can be optimized through the proper choice of subaperture size, control bandwidth, laser power, and number of guide stars. They apply their technique to the Keck Telescope, $D = 10$ m, to determine optimal values of $f_c = 50$ and 70 Hz for a WFS with 69 and 241 subapertures respectively. Their results are based on nominal atmospheric conditions of $r_0 = 0.20$ m and $f_g = 26$ Hz for the Keck Observatory. They also give similar results for the Lick Observatory 3 m telescope as $f_c = 60$ and 50 Hz for a WFS with 69 and 241 subapertures, respectively. This

outcome is based on nominal values of $r_0 = 0.10$ m and $f_g = 26$ Hz for the Lick Observatory. For both the Keck and Lick results, the optimum closed loop bandwidths are approximately twice the Greenwood frequency. However, in this case the systems were modeled to have guide stars of constant light levels via laser beacons. This work examines how the optimum closed loop bandwidth changes with beacon brightness.

Gavel *et al* also provide a general rule of thumb in regard to adaptive optics correction. That is, the spatial scale of the correction should be at least as small as r_0 and the closed loop bandwidth should be at least as large as f_g. In these results, this rule of thumb applies when there are sufficient light levels at the WFS. However, for dim objects without the use of a laser beacon, the preferred closed loop bandwidth settings are a small fraction of the Greenwood frequency. As f_g increases under low light levels, the optimal bandwidths also increase but still remain many times lower than the Greenwood frequency. In addition, under low light levels, it is shown that the selection of the proper closed loop bandwidth is more critical for achieving maximum performance than under high light levels.

A strategy is also presented for selecting the closed loop bandwidth to provide robust system performance for different target object light levels. The major conclusion of this research is that the predominant factor in determining optimal bandwidths is light level to the WFS, even under very poor seeing conditions as indicated by a very high Greenwood frequency.

<u>Experimental Setup</u>. The work in this section again used the HYSIM adaptive optics simulation. Recall, that for user provided input parameters, HYSIM models and computes an average system OTF and average long exposure image of an input object. In addition, the Strehl ratio corresponding to the instantaneous OTF of each DM update is also computed and can be stored for subsequent analysis.

Work in this section modeled the Air Force Maui Optical Station (AMOS) 1.6 meter

adaptive optics system. Figure 21 shows the input values used by HYSIM to model the AMOS

system parameters. This figure also shows the values of other system controllable parameters

and uncontrollable variables that were incorporated in this research.

Controllable Parameters   **AMOS System Parameters**

- Number of frames: 200
- System sample rate: 1000 Hz
- System gain: 0.005 - 0.15
- Frame exposure time: 0.001 s
- Average long exposure time: 0.2 s

- Primary mirror diameter: 1.6 m
- Obscurration diameter: 0.335 m
- Imaging wavelength: 7.0E-07 m
- WFS wavelength: 6.0E-07 m
- Number of subapertures across pupil diameter: 16
- DM actuator separation in pupil plane: 0.11 m

System Variables

Inputs

Uncontrollable
Variables

AOS
(Model by HYSIM)

Outputs

- Average long exposure image
- Average long exposure OTF
- Vector of instantaneous
  Strehl ratio (*ISR*) values

- Extended object target distance: 6.5E+05 m
- Object and WFS visual magnitudes: 0 - 9
- Fried parameter: 0.05 and 0.40 m
- Average local wind velocity: 12.5 m/s
- Physical size of array holding input object: 40.0 m

Figure 21. Summary of Parameters Used to Model AMOS 1.6 Meter Adaptive Optics System
and Variables Used in Closed Loop Bandwidth Optimization Study

HYSIM allows the user to define atmospheric seeing conditions by inputting desired values

for the Fried parameter, r_0, and local wind velocity, $\vec{v}$. These two quantities allow computation

of a corresponding value for the Greenwood frequency as given by [9:219]

$$f_g = 0.426 \frac{|\vec{v}|}{r_0} \ . \tag{24}$$

Feedback Control Model. Application of feedback control theory as used in HYSIM provides for computing the current command signal to a particular DM actuator in discrete form as [23, 9:196-197]

$$C_j = C_{j-1} + k_g e_j \tag{25}$$

where C_{j-1} is the previous command signal, k_g is the gain in units of radians per second, and e_j is the error control signal arising from the closed loop slope measurement. In equation form

$$e_j = MS_j - C_{j-1} \tag{26}$$

where M is the reconstruction matrix and S_j is the open loop slope measurement. Using Equation (26), Equation (25) can be rewritten in a more interpretable form as

$$C_j = C_{j-1}(1 - k_g) + k_g S_j \ . \tag{27}$$

Equation (27) shows that the command signal at any time step is strictly a function of the current and all previous slope measurements. The gain serves to implement the relative importance of each of these measurements. Simple algebra shows the relative weight, ρ_n, of each slope measurement in constructing the command signal is

74

$$\rho_n = k_g (1 - k_g)^n \qquad (28)$$

where $n = 0, 1, 2, \dots$ is the discrete time step behind the current slope measurement. By inspection of Equations (27) and (28), in constructing the current command signal, high values of k_g will put more emphasis on the current measured slope than on past measurements. In contrast, low values of k_g will put more emphasis on past measurements than on the current slope measurement.

The above assertion lends credence to an observation made during the review of the paper by Ellerbroek *et al* [35]. That is, in this paper there is an interaction between the optimum closed loop bandwidth and the number of subapertures. In particular, for the Keck Observatory, the optimum value for f_c increased with the number of subapertures and for the Lick Observatory, the optimum value for f_c decreased with the number of subapertures. It is surmised that this interaction is primarily due to the significant differences in the side lengths of the subapertures between Keck and Lick and the associated light levels seen by these subapertures. That is, for the same number of total subapertures, the Keck subapertures have roughly 11 times greater area than the Lick subapertures. In their study, the Lick Observatory laser had twice as much power (40 W) as the Keck Observatory laser (20 W). So in effect, for the same number of total subapertures, each of the Keck subapertures will see about 5.5 times more incident photons than the Lick subapertures. This might imply that as the number of subapertures is increased from 69 to 241, the Lick system does not have enough light to accurately measure the incident perturbed wavefronts under high closed loop bandwidths. Therefore, f_c must be decreased to boost signal-to-noise levels at the WFS. On the other hand, the Keck system may still have enough light for accurately determining the incident wavefronts when the number of subapertures is increased

75

from 69 to 241. Therefore, increasing the closed loop bandwidth does not hamper performance in this case. This explanation is consistent with results presented later in the next section.

To further demonstrate the above points, Figure 22 shows the relative weights that are applied to past slope measurements in constructing the new DM command signal as a function of each measurement's time lag. This plot shows that for low gains, past measurements remain about as important as more recent measurements. For high gains, the importance of past measurements rapidly diminish as the their time lag increases behind the current command. Similarly, Figure 23 shows the cumulative sum of the weights used to determine the current command at each time lag behind the current command. This plot shows that for low gains, many previous measurements are needed to fully determine the current command signal. For high gains, such as $k_g = 0.6$, the current command signal will be practically completely determined by using only about the first five previous slope measurements.

This discussion implies that large values of gain will be favored when the measured slope can be determined accurately. On the other hand, when the measured slope cannot be determined accurately, then large gain values may in turn lead to command signals to the DM that are also inaccurate. These assertions coincide with observed simulation results that will be presented in more detail in the next section. These results show that low gain settings are optimal under low light level conditions -- when the measured slopes are prone to error due to insufficient signal-to-noise levels at the WFS. On the other hand, under high light levels, when signal-to-noise levels at the WFS are sufficient to accurately gauge the measured slopes, high gain settings are optimal. Also, under a high light level scenario, to a point, low gain settings will not significantly impair performance. This is especially true when seeing conditions are good so that the Greenwood or characteristic atmospheric frequency is low. When the Greenwood frequency is high, low gain
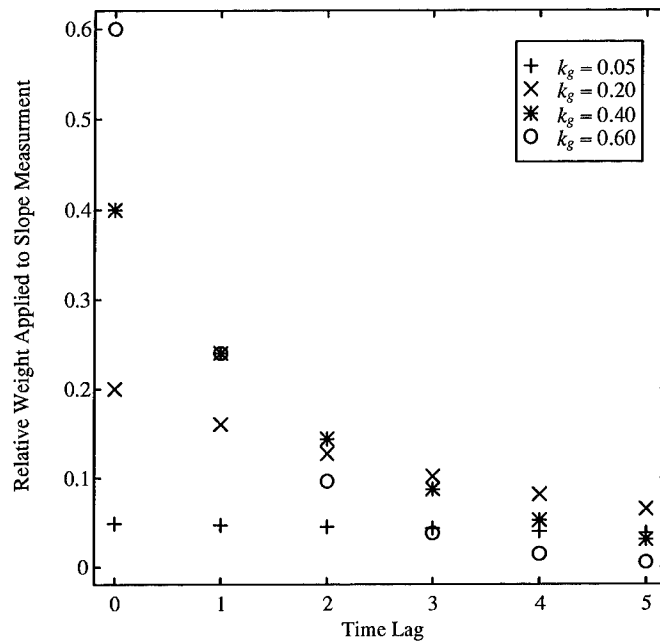
Figure 22. Relative Weights Applied to Past Slope Measurements Used to Determine Current
DM Command Signal as a Function of Time Lag, n, and System Gain, k_g
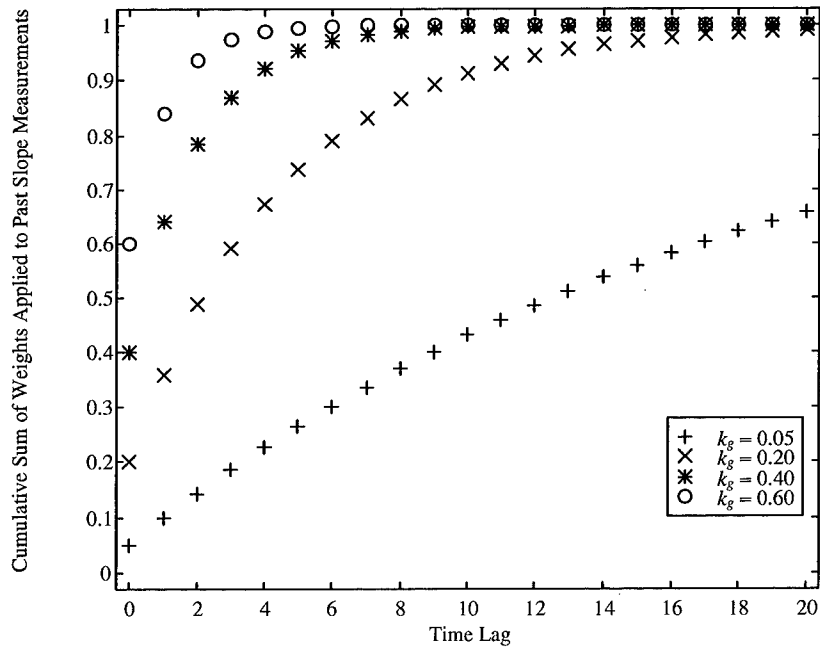
Figure 23. Cumulative Sum of Weights Applied to Past Slope Measurements Used to Determine
Current DM Command Signal as a Function of Time Lag, n, and System Gain, k_g

settings will begin to hamper performance because closed-loop control of the DM configuration will start to lag too far behind the changing atmosphere to provide good compensation.

Figure 24 compares the effect of three different gain settings on the ability of an AOS to overcome startup transients and its subsequent performance. In this figure, Panel (a) represents a bright object ($m_v = 0$) and Panel (b) represents a dim object ($m_v = 9$). Consistent with the previous discussion, under bright conditions, a very low gain setting, $k_g = 0.0314$, shows that the system takes a long time to start to increase in performance as indicated by the slow rise in the instantaneous Strehl ratio. Here, overall system performance is also degraded because the low gain setting is causing the system to lag too far behind the changing atmosphere. In this case the system will never perform as well as it will with the mid to high gain settings.

When the gain is increased to $k_g = 0.314$, the system overcomes startup transients after only about seven or eight DM updates. Similarly, when $k_g = 0.628$, the system also "locks in" after about the same number of DM updates. In this case the greater oscillations in the instantaneous frame Strehl ratio during the startup phase are due to the increased emphasis on the most recent wavefront slope measurement. As indicated in Panel (a) and as will be shown further later, under bright conditions there is virtually no difference is overall system performance between mid to high gain settings such as $k_g = 0.314$ or 0.628. In short, these gain settings are preferred over low gains when the target object is bright.

In Panel (B), it is clearly evident that mid to high gain settings are inferior to a low gain setting. This is because under the low light levels, the wavefront slope measurements are prone to increased error. Hence, placing to much emphasis on only a few recent measurements also induces errors in the commands sent to the DM. When the gain is low, these wavefront slope measurements are averaged over many more incremental updates so that the effect of their errors

.

78
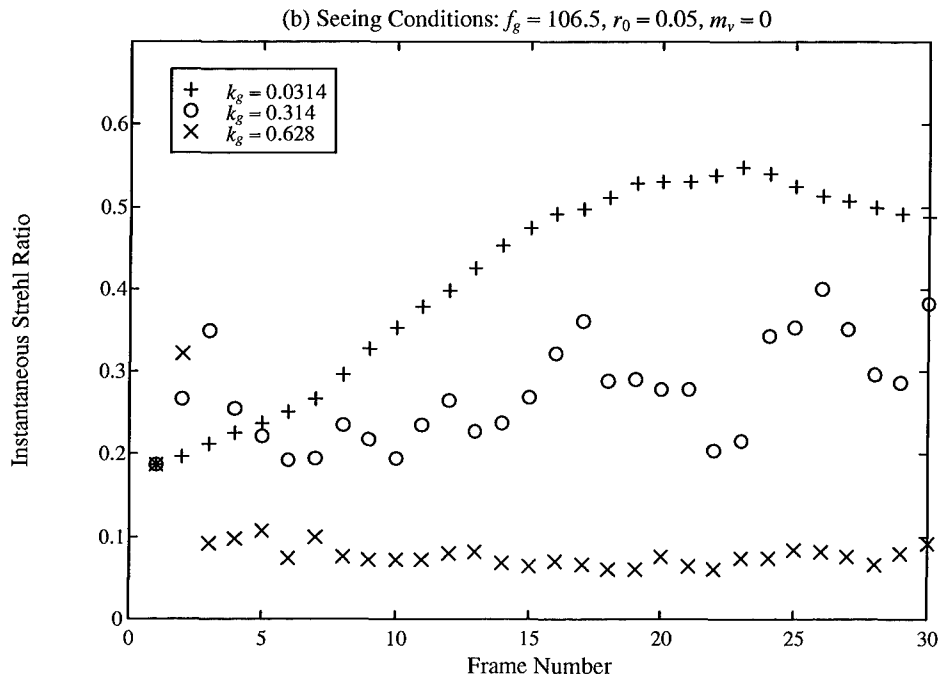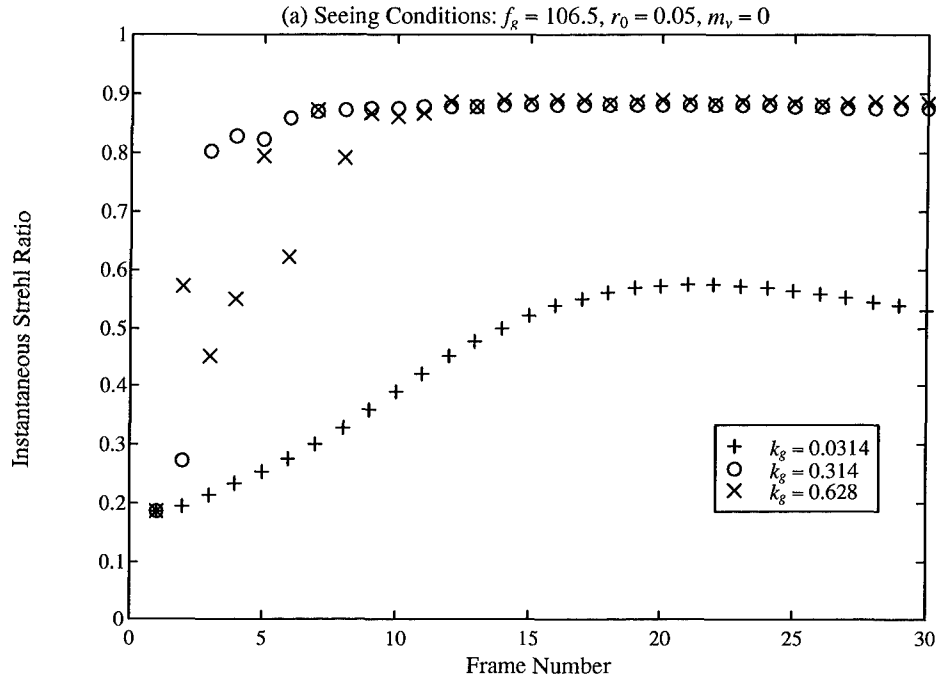
Figure 24. Comparison of Different Gain Settings on AOS System Startup. The sample rate is 1000 Hz in all cases. Panel (a) has the object visual magnitude, $m_v = 0$ and Panel (b) has $m_v = 9$. The system gain is given by k_g.

79

on the computed DM commands is lessened. Hence, low gain settings are preferred when the

target object is dim

Before leaving this subsection, Equation (27) also shows that as k_g approaches one, the

current command will be based solely on the measured slope. This can cause the system to start

to become unstable. Values of k_g greater than one are not logical. This implies that there is an

upper limit to the gain that can be used to control the system. Likewise, the closed loop

bandwidth, f_c, is a function of the gain, k_g, and system sample rate, s. That is [9:197],

$$f_c = \frac{k_g s}{2\pi} .$$

(29)

Hence, for a given sample rate, there will be an upper limit to achievable closed loop

bandwidths. Once this limit is reached, higher sample rates must be incorporated to achieve

higher closed loop bandwidth settings. Indeed, some of these results show that for a sample rate

of 1000 Hz, the highest practical closed loop bandwidths are about 100 - 125 Hz. This means

that if the Greenwood frequency is significantly above this range, higher sample rates will be

necessary in order to achieve optimal results. This is especially true if light levels are

sufficiently high to provide accurate slope measurements so that the rule of thumb mentioned

earlier by Gavel *et al* is appropriate.

Figure 25 shows the instantaneous Strehl ratio subject to an changing atmosphere for

different closed loop bandwidths as computed using Equation (29). In this figure, one can see

how increasing the closed loop bandwidth improves performance until about 100 to 125 Hz. As

the closed loop bandwidth is increased beyond this point, the system starts to become unstable.

This is especially noticeable when the closed loop bandwidth is equal to 150 Hz.
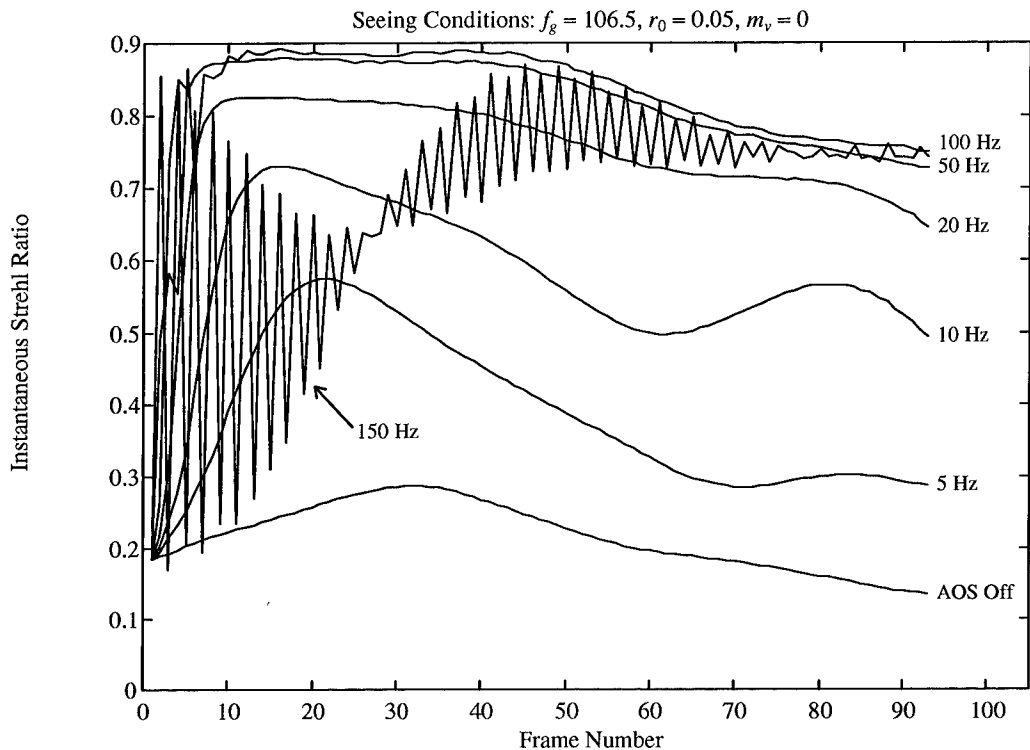
Figure 25. Effect of Increasing Closed Loop Bandwidth on AOS Performance. The closed loop bandwidth is given in Hz. The instantaneous Strehl ratio for the case where the AOS is turned off is also shown. In all cases the sample rate was 1000 Hz.

Results. This subsection presents a more thorough analysis of optimal closed loop bandwidths as a function of seeing conditions and light levels. The basic approach was to select a wide range of target object light levels, as indicated by visual magnitude. These light levels were used to compute an average incident number of photons to both the imaging and WFS detectors under the assumptions given in earlier. Next, three different cases were considered to model very diverse seeing conditions. Case 1 had a Fried parameter of 0.05 m and wind velocity, of 12.5 m/s. Case 2 had a Fried parameter of 0.40 m and wind velocity of 12.5 m/s. Case 3 had a Fried parameter of 0.05 m and wind velocity of 25.0 m/s. With these parameters selected, the Greenwood frequency was computed using Equation (24). Table 4 summarizes the parameters

81

for each of these three cases. Lastly, the system sample rate was fixed at 1000 Hz and different gain settings were selected to achieve desired system closed loop bandwidth settings using Equation (29).

Table 4. Summary of Parameters for Different Cases Studied in Chapter IV

Case	r_0 (m)	v (m/s)	f_g (Hz)
1	0.05	12.5	106.5
2	0.40	12.5	13.3
3	0.05	25	213.0

Figure 26 shows the mean instantaneous Strehl ratio versus closed loop bandwidth for different object light levels under Case 1, poor seeing conditions. The mean *ISR* value for the atmosphere for the case where the adaptive optics system is turned off is also displayed. This plot shows that for high light levels (e.g., m_v = 0 or 3), the optimal closed loop bandwidth is near the Greenwood frequency. It is noted that the optimal closed loop bandwidth is considered to be the bandwidth that provides the greatest mean *ISR* value for a particular object light level. For lower light levels (e.g., m_v = 8 or 9), the optimal closed loop bandwidth is many times lower than the Greenwood frequency. In addition, this figure shows that for the higher light levels the proper choice for f_c is much less critical than it is for low light levels. For low light levels, performance rapidly diminishes if the selected closed loop bandwidth is not near the optimal setting.
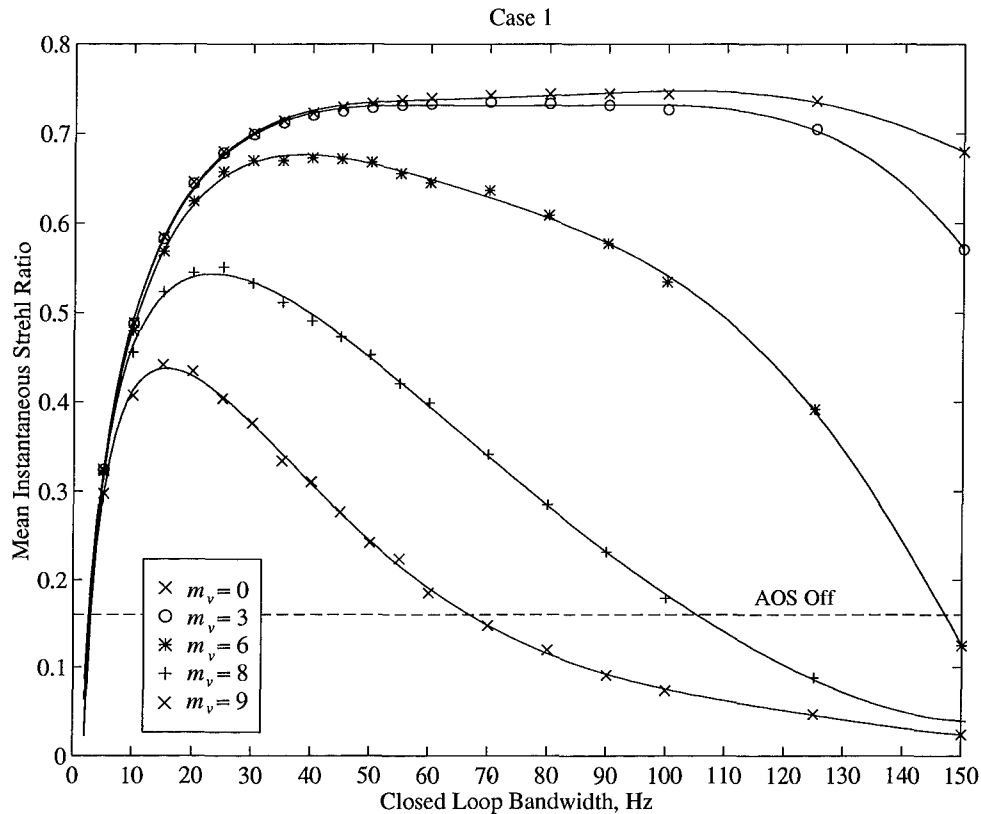
Figure 26. Mean Instantaneous Strehl Ratio versus Closed Loop Bandwidth for Different Object Visual Magnitudes for Case 1. Seeing conditions were $r_0 = 0.05$ and $f_g = 106.5$ Hz. The object visual magnitude is given by m_v.

Figure 27 is similar to Figure 26 except that the Fried parameter has been increased from 0.05 meters to 0.40 m (Case 2), representing good seeing conditions. The results are similar to those seen for Case 1. However, the optimal closed loop bandwidths have shifted to slightly lower frequencies. This corresponds to a lower Greenwood frequency. Under the high light levels, a broad range of f_c settings that yields high mean *ISR* values is still seen. Under low light levels, a decrease in system performance is seen to occur even more rapidly as the closed loop bandwidth is increased.
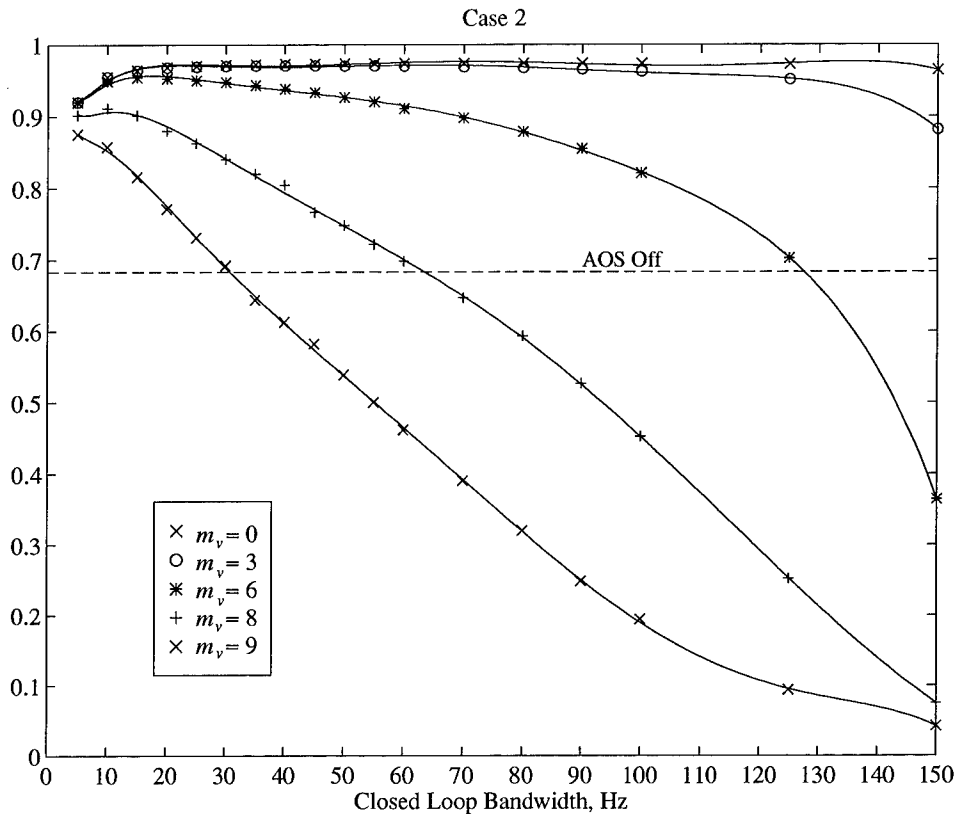
Figure 27. Mean Instantaneous Strehl Ratio versus Closed Loop Bandwidth for Different Object Visual Magnitudes for Case 2. Seeing conditions were $r_0 = 0.40$ and $f_g = 13.3$ Hz. The object visual magnitude is given by m_v.

In Figure 28 the Fried parameter is again 0.05 m but the wind speed has been doubled to 25.0 m/s (Case 3), representing very poor seeing conditions. In this case, for all light levels, the optimums have shifted to slightly higher closed loop bandwidths. This shift is more noticeable for the lower light levels than the higher light levels. This is certainly in part this due to the fact that for the high light levels, the optimum closed loop bandwidths are near the maximum feasible bandwidths at a sample rate of 1000 Hz. Hence, they cannot be increased much further. Even under the very high Greenwood frequency modeled in this case, the optimal bandwidths under
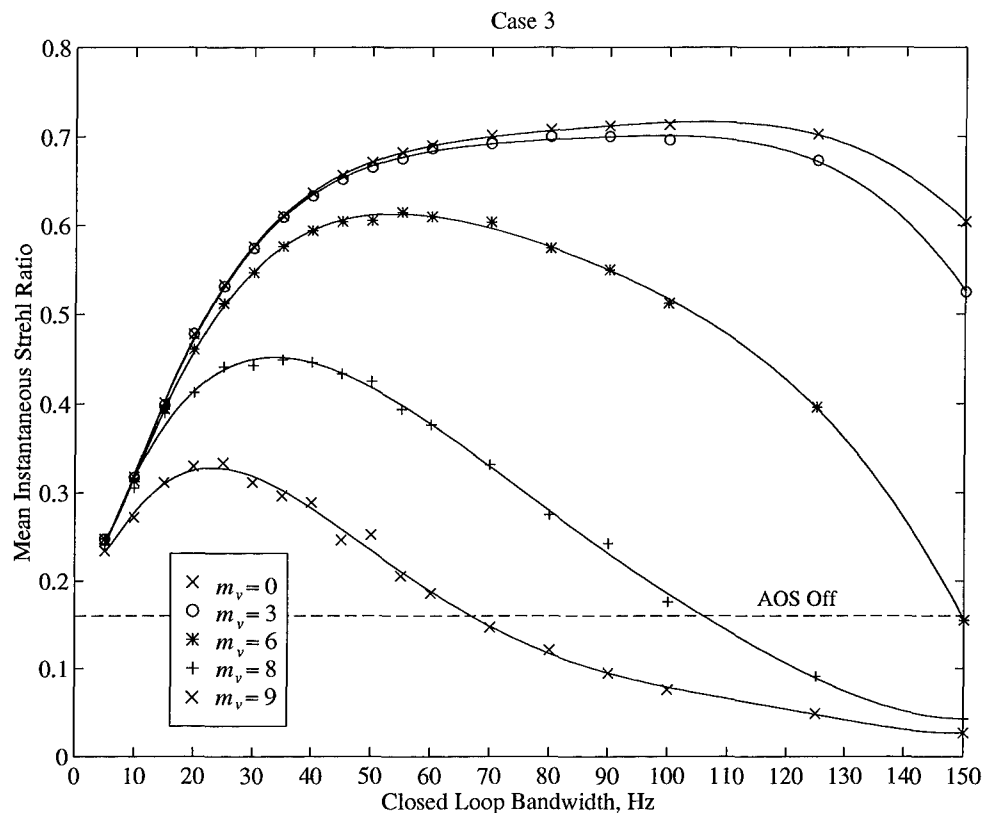
Figure 28. Mean Instantaneous Strehl Ratio versus Closed Loop Bandwidth for Different Object Visual Magnitudes for Case 3. Seeing conditions were $r_0 = 0.05$ and $f_g = 213.0$ Hz. The object visual magnitude is given by m_v.

low light levels are still many times lower than the Greenwood frequency. Also, the region of optimality is still very narrow compared to those under bright conditions. For the high light levels the bandwidth intervals that are within 95 percent of the optimum have narrowed from Case 1 and Case 2 when the seeing conditions were better.

Using the above results, Figures 29, 30, and 31 show the optimal closed loop bandwidth frequencies as a function of light level levels for each of the three cases examined. Also shown are vertical bars that illustrate the range of f_c settings that yield mean *ISR* values that are within 95 percent of the observed maximums for each light level. These plots clearly show that under

high light levels the system will be much more robust in achieving a high level performance for various f_c settings than under low light levels. Hence, properly choosing the closed loop bandwidth under low light levels is much more crucial to successful AOS operation than when light levels are high.
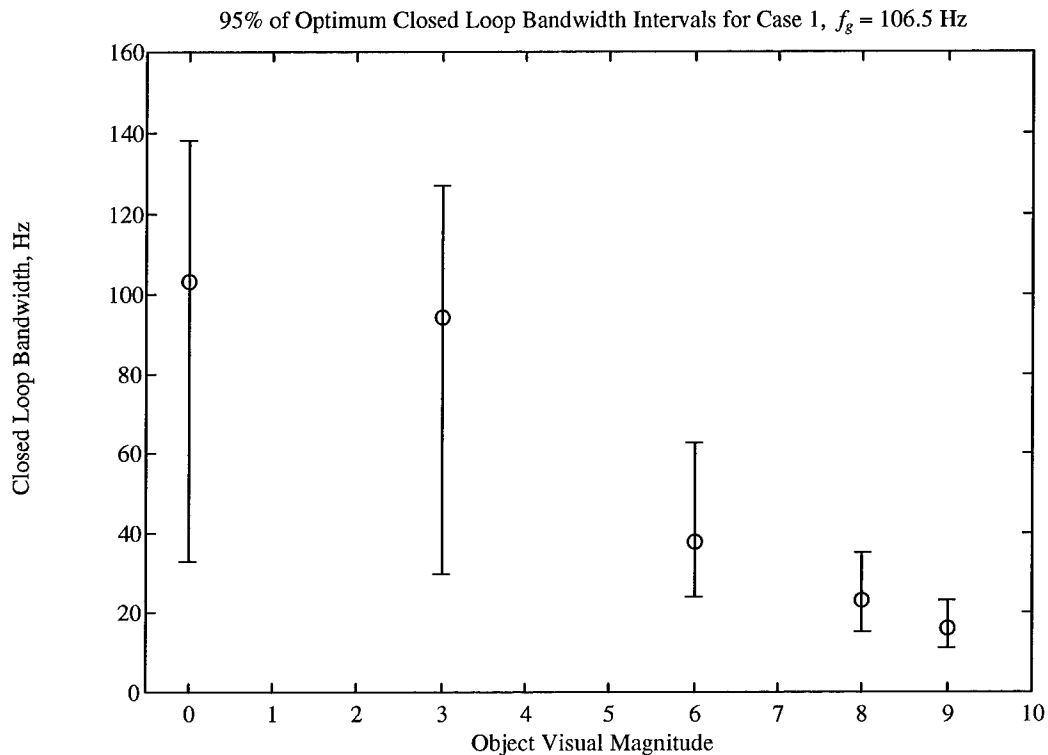
Figure 29. Optimal Closed Loop Bandwidth versus Object Visual Magnitude for Case 1. The O's indicate optimal bandwidths that provide the maximum mean instantaneous Strehl ratio and the vertical bars indicate ranges of f_c settings that yield 95 percent of these maxima respectively.

In Figure 32, the data from Figures 29, 30, and 31 is combined into a single plot. This plot suggests an appropriate strategy for robust performance in selecting f_c settings for any light level. That is, by examination of Figure 32, for a given light level a prudent choice would be to

95% of Optimum Closed Loop Bandwidth Intervals for Case 2, $f_g = 13.3$ Hz
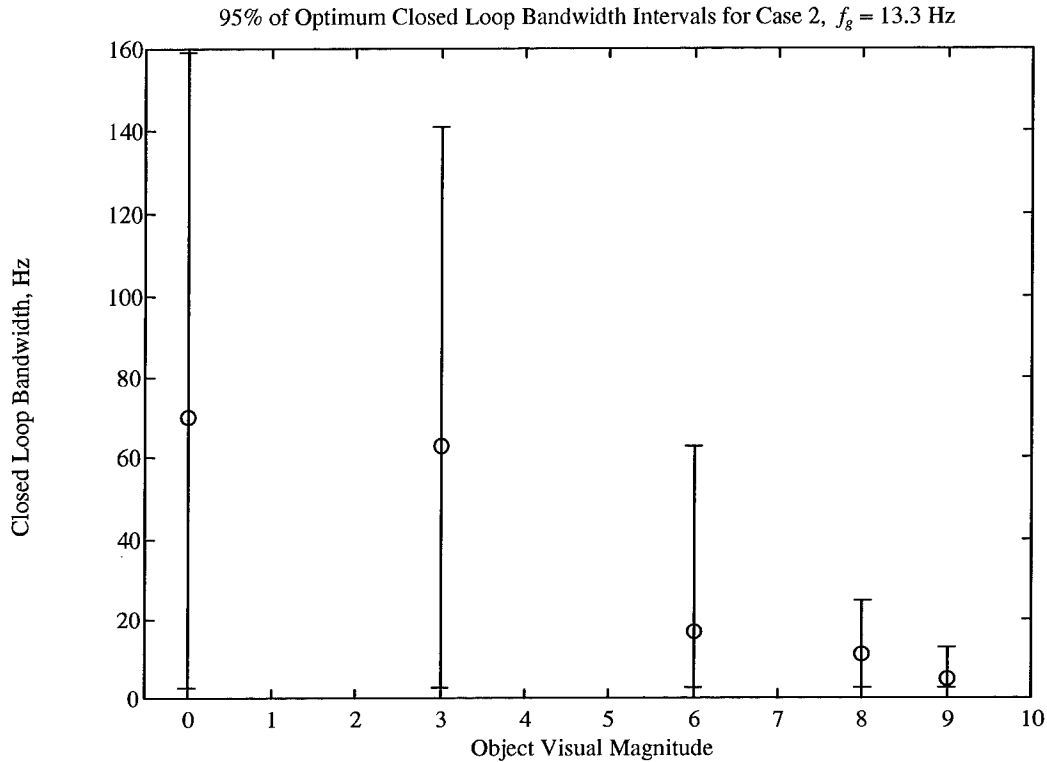
Figure 30. Optimal Closed Loop Bandwidth versus Object Visual Magnitude for Case 2. The O's indicate optimal bandwidths that provide the maximum mean instantaneous Strehl ratio and the vertical bars indicate ranges of f_c settings that yield 95 percent of these maxima respectively.

set f_c at the optimum closed loop bandwidth setting for the poor seeing conditions case, Case 1. Under this strategy, if the seeing conditions improve or worsen over the course of AOS operations through temporal changes in either the Fried parameter or wind velocity, then the optimum closed loop bandwidth intervals show that the system will still be within 95 percent of peak performance for almost of the different light levels and seeing conditions examined. For instance, for a light level of $m_v = 0$, 3, or 6, or 8 the optimum f_c setting for Case 2 ($f_g = 106.5$) is contained by both the best ($f_g = 13.3$) and worst case ($f_g = 213.0$) bandwidth intervals. Hence, for these light levels the system would always perform at 95 percent of optimum by this strategy. For a light level of $m_v = 9$, this strategy shows that the optimum f_c setting for Case 1 is within the

87

95% of Optimum Closed Loop Bandwidth Intervals for Case 3, $f_g = 213.0$ Hz
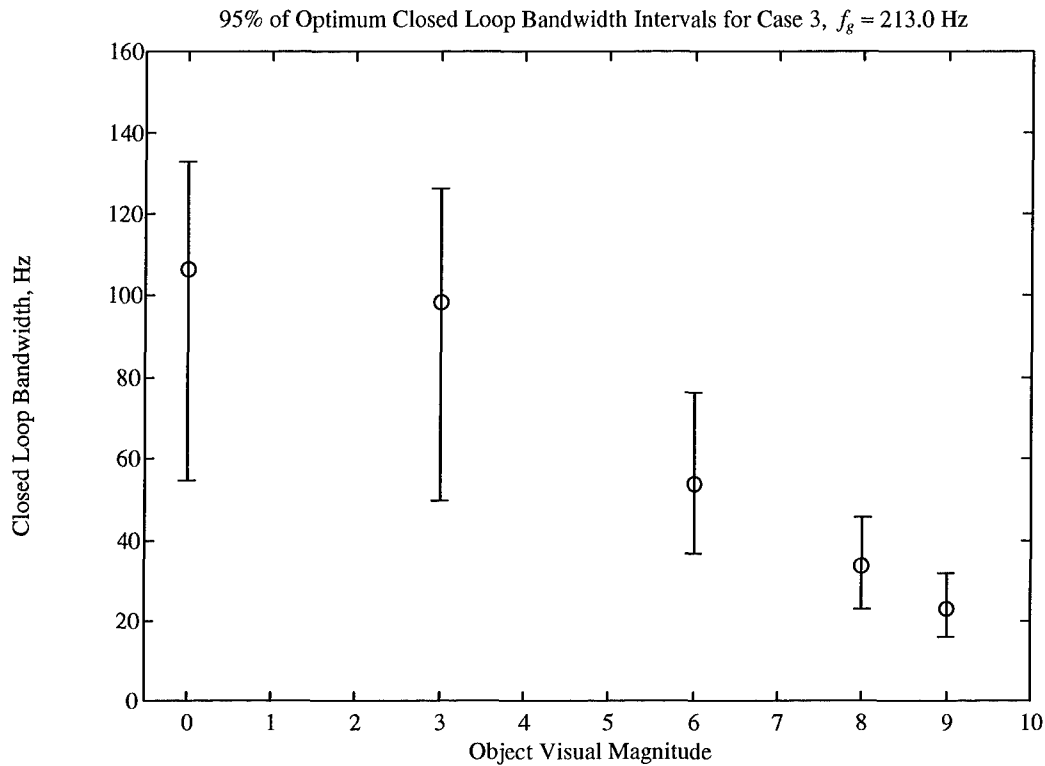
Figure 31. Optimal Closed Loop Bandwidth versus Object Visual Magnitudes for Case 3. The O's indicate optimal bandwidths that provide the maximum mean instantaneous Strehl ratio and the vertical bars indicate ranges of f_c settings that yield 95 percent of these maxima respectively.

worst case interval, Case 3, but just barely outside the best case interval, Case 2, so that performance would be slightly less than that of 95 percent of optimum under very good seeing conditions. However, when the seeing conditions are very good, the observed mean *ISR* value is still high so that this should not be a significant impact.

Figure 33 demonstrates the impact of properly choosing the closed loop bandwidth by considering f_c settings of 15 and 100 Hz and high and low light levels. Seeing conditions were for those given by Case 1, $r_0 = 0.05$ m and $f_g = 106.5$ Hz. The input object to be imaged was the simulated satellite shown earlier in Figure 5. In the first row, when light levels are high ($m_v = 0$), a closed loop bandwidth of 100 Hz is preferred over 15 Hz. However, if f_c is 15 Hz, the
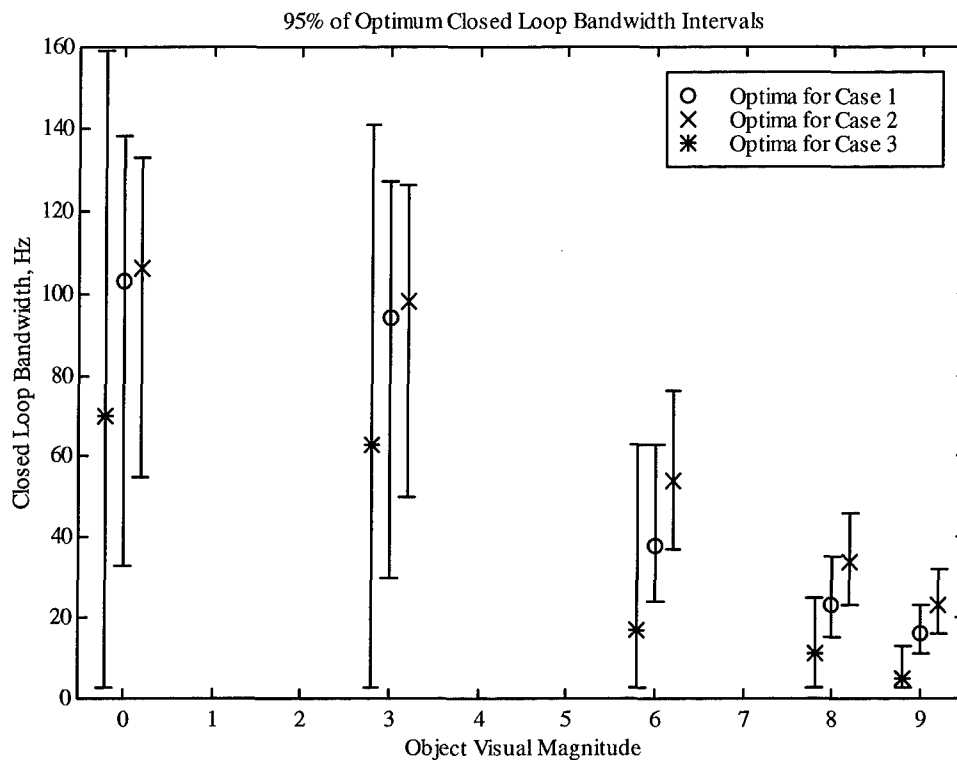
Figure 32. Optimal Closed Loop Bandwidth versus Object Visual Magnitudes for Cases 1, 2 and 3. The legend indicates the optimal bandwidth settings that provide the maximum mean instantaneous Strehl ratio for the different cases examined. The vertical bars indicate intervals of f_c settings that yield 95 percent of these maxima respectively.

the system performance based on the mean *ISR* metric is still only about 21 percent worse than if f_c is 100 Hz. In the second row, when light levels are low ($m_v = 9$), a closed loop bandwidth of 15 Hz provides better performance than if f_c is 100 Hz. In this case, if f_c is 100 Hz, the system performance based on the mean *ISR* metric is over 82 percent worse than if f_c is 15 Hz. Hence, for the specified seeing conditions under low light levels the optimum closed loop bandwidth is about seven times less than the Greenwood frequency. This result again shows the importance of properly choosing the closed loop bandwidth under low light levels.

(a) Mean ISR=0.584   (b) Mean ISR=0.743

(c) Mean ISR=0.414   (d) Mean ISR=0.074
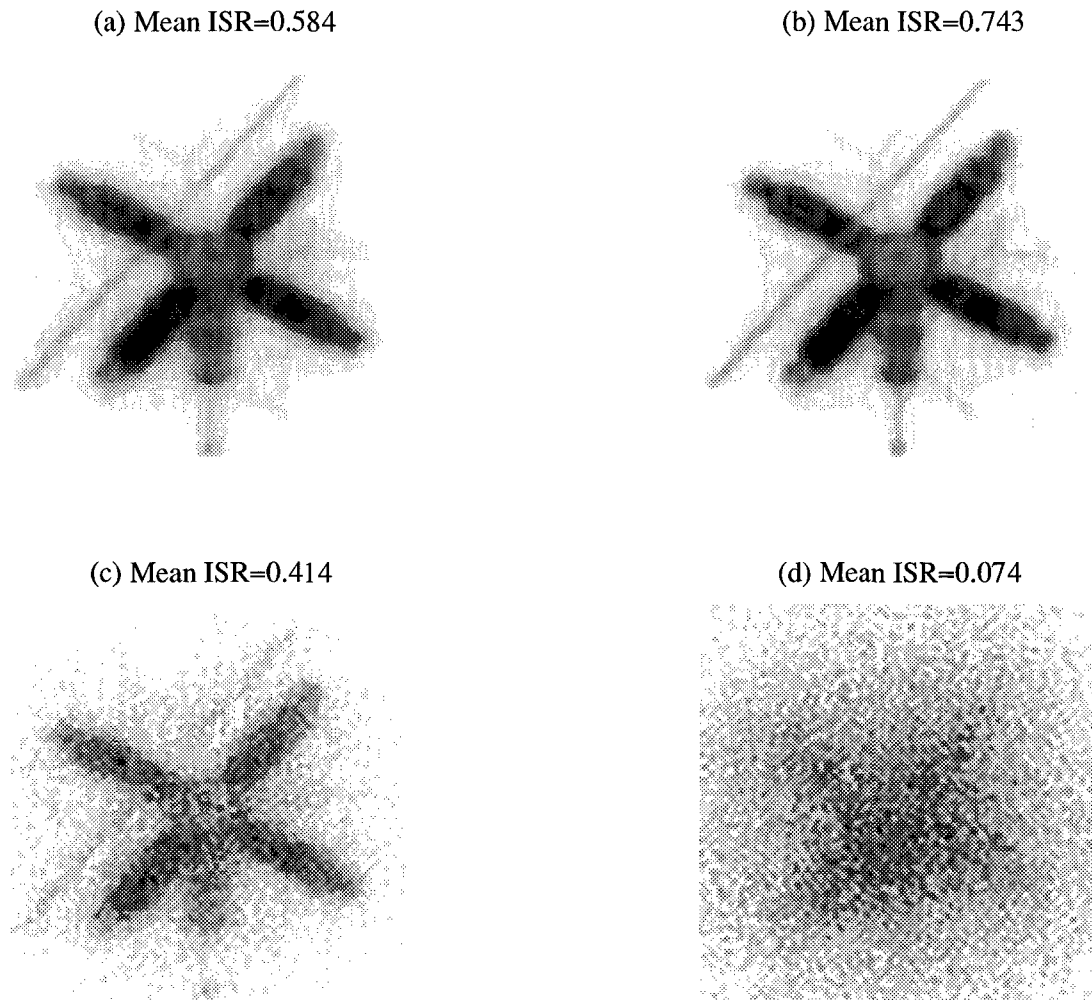
Figure 33. Sample Extended Object Output Images for Object Visual Magnitudes, m_v, of 0 and 9 and Indicated Closed Loop Bandwidths, f_c. Panel (a) has $m_v = 0$ and $f_c = 15$ Hz, Panel (b) has $m_v = 0$ and $f_c = 100$ Hz, Panel (c) has $m_v = 9$ and $f_c = 15$ Hz, and Panel (d) has $m_v = 9$ and $f_c = 100$ Hz. Seeing conditions were $r_0 = 0.05$ and $f_g = 106.5$ Hz.

Finally, Figure 34 shows the results of adding more light levels to the Case 2 scenario. Panel (a) simply shows the observed optimum closed loop bandwidth settings for light levels between $m_v = 0$ and 10. Panel (b) shows two piece-wise functions for the recommended closed loop bandwidth setting, f_c*. These function are based on fitting a linear regression model to the

(a) Observed optimums for Case 1
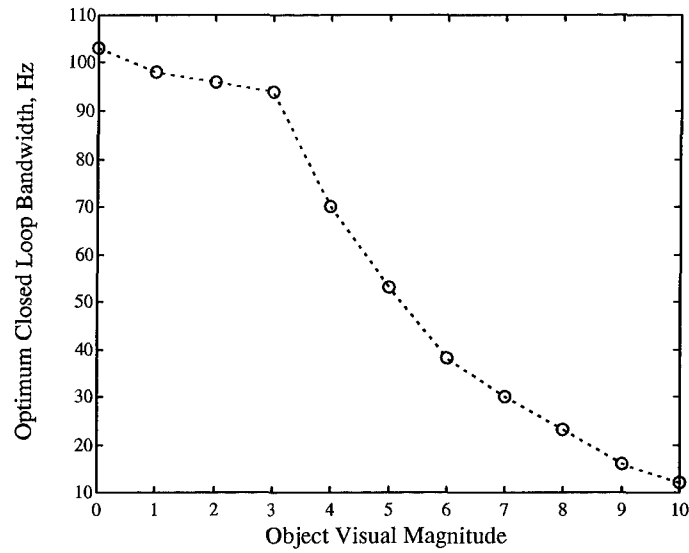
(b) Piece-wise functions fitted to observed optimums

$$f_c^* = -4.63\,m_v^{0.6} + 102.9$$

$$f_c^* = 356.9\,/\,m_v - 21.7$$
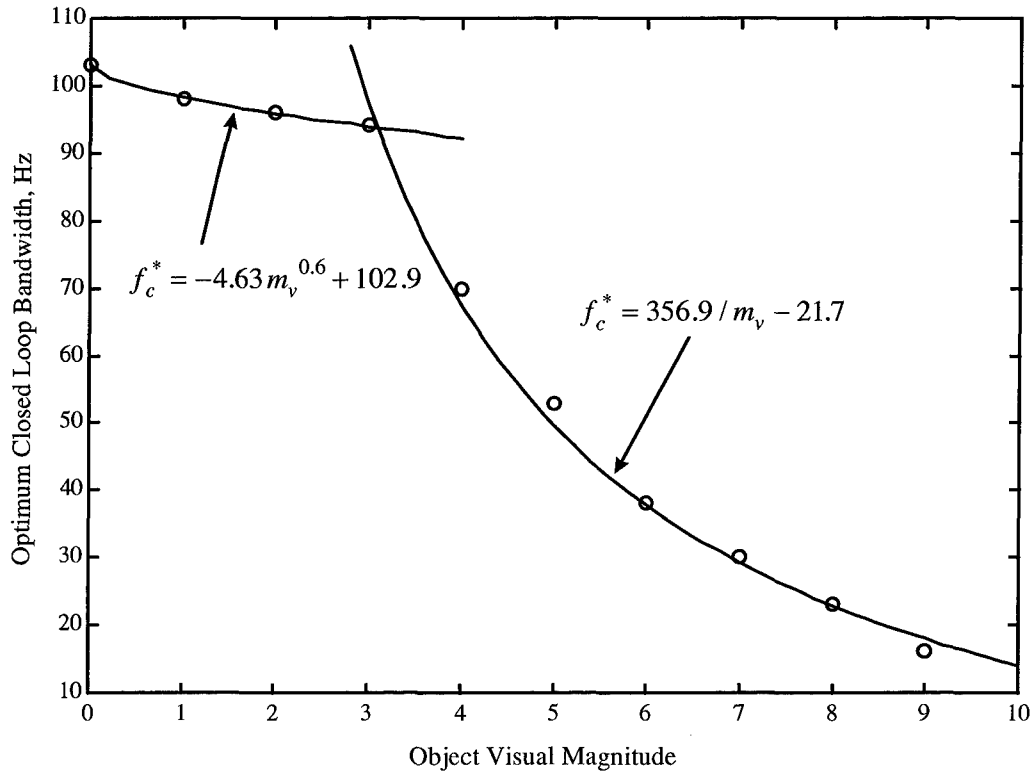
Figure 34. Recommended Closed Loop Bandwidth Settings, f_c^*, for the AMOS 1.6 Meter Adaptive Optics System. The object visual magnitude is given by m_v. Seeing conditions were r_0 = 0.05 m and f_g = 106.5 Hz (Case 1).

observed data for object visual magnitudes between 0 and 3 and 3 and 10 respectively. The correlation coefficients for these fitted function are both greater than 0.99

Conclusions. These results show that low closed loop bandwidths are optimal under low light level conditions, regardless of the Greenwood frequency. On the other hand, under high light levels, higher closed loop bandwidth settings are optimal and are close to the Greenwood frequency. Also, under a high light level scenario, to a point, low closed loop bandwidth settings will not significantly impair performance. This is especially true when seeing conditions are good so that the Greenwood or characteristic atmospheric frequency is low. When the light levels and Greenwood frequency are high, low closed loop bandwidth settings will begin to hamper performance because closed-loop control of the DM configuration will start to lag too far behind the changing atmosphere to provide good compensation.

It is believed that these results can primarily be explained by feedback control considerations. Specifically, high values of the system gain will be favored when the incident wavefront slopes can be measured accurately -- such as when there are high light levels to the WFS. On the other hand, when the measured slopes cannot be determined accurately due to insufficient signal-to-noise levels at the WFS, then high gain values may in turn lead to command signals to the DM that are also inaccurate. In this case lower gain settings will help to improve performance because wavefront measurement errors will tend to be averaged over more discrete time steps.

Finally, a strategy was presented for selecting the closed loop bandwidth to provide robust system performance for different target object light levels. Specifically, it is concluded that a prudent choice would be to set the closed loop bandwidth for a given light level to the optimum setting under poor seeing conditions. Based on maximum observed values of the mean *ISR* metric, it was shown that this strategy can provide results that are within 95 percent of optimal in

almost all cases. Figure 34 presented two functions that can be used to easily compute these recommended closed loop bandwidth settings. These functions require only a knowledge of the light level of the object to be imaged.

In summary, the most important conclusion of this research is that the predominant factor in determining optimal bandwidths is light level to the WFS, even under very poor seeing conditions as indicated by a very high Greenwood frequency. Hence, these results are important even for systems that incorporate artificial guide star capabilities. This is because such systems equipped with modest sized lasers will produce beacons with equivalent stellar visual magnitudes of about 8 to 10 [25, 36]. In this region, this research shows that the tolerance for selecting the proper closed loop bandwidth to achieve good imaging system performance is narrow.

V. Adaptive Optics System Trade-off Analysis

Introduction

This chapter presents a discussion of the trade-offs between fully compensated adaptive optics and hybrid imaging systems with partially compensated adaptive optics. The discussion keys in on the trade-offs to be made between decreasing spatial resolution, through an increase in system sample spacing, and the associated decrease in noise levels at the Wavefront Sensor (WFS). Specifically, a method to counteract the impact of low light levels on overall adaptive optics system performance without the use of laser beacons was examined. This method involves the notion of "clustering" smaller WFS subapertures into larger effective WFS subapertures via software instruction.

General Discussion

As mentioned in Chapter I, adaptive optics systems are quite complicated. Moreover, as the aperture of the primary mirror is made larger, such systems become increasingly sophisticated in terms of sampling requirements, control rates, and data transfer speeds, all of which translates to increased system cost [9:10-11]. Consequently, simpler adaptive optics systems are being explored. These simpler systems are known as partially compensated adaptive optics systems. They are characterized by fewer, larger WFS subapertures and fewer Deformable Mirror (DM) degrees-of-freedom than fully compensated systems [9:10].

Based on sample spacing considerations alone, the above characterization implies that a partially compensated AOS will not provide the same level of wavefront compensation as adaptive optics systems with more WFS subapertures and DM degrees-of-freedom. This is because finite sample spacing of the WFS and finite degrees-of-freedom of the DM result in the

same general type of system performance degradation -- the ability to sense and compensate for high spatial frequencies of the wavefront deformations [9:175]. In more precise terms, the cutoff frequency of the WFS is inversely proportional to the WFS sample spacing or sample interval. The sample interval of a WFS can be defined by how the system pupil is segmented into subapertures. Each of these subapertures can be defined by a side dimension, d, of a single WFS subaperture. As the number of WFS subapertures is increased across the pupil, d is decreased. Hence, as the sample interval is decreased, the cutoff frequency of the WFS will be increased [9:175].

In the same fashion, finite degrees-of-freedom of the DM or finite number of DM actuators also acts as a filter that limits high spatial frequency information in the reconstructed wavefront phase. As with the WFS, the cutoff frequency of the DM is inversely proportional to the spacing between actuators that are used to shape the DM surface. Therefore, as the spacing between actuators is decreased, the cutoff frequency of the DM will be increased [9:175].

Roggemann and Welsh provide a table that gives the aperture averaged mean square phase error after the removal of the first N number of Zernike modes from the incident degraded wavefront [9:94]. In short, these Zernike modes are derived from Zernike polynomials which are used to represent random wavefront degradations associated with a turbulent atmosphere. This table can be used to roughly estimate the residual phase fluctuation after wavefront compensation as a function of both the subaperture size and the Fried parameter, r_0. Figure 35 shows the aperture averaged mean square phase error after removal of the first N Zernike modes as a function of N, d, and r_0. Based on subaperture size considerations alone, this plot clearly shows that smaller subapertures will provide better atmospheric compensation regardless of the seeing conditions. For instance, when $r_0 = 0.05$ m and $d = 0.10$ m, removal of the first 6 Zernike modes from the incident wavefront is equivalent to the removal of all 21 modes when d is increased to

0.20 m. Similarly, when $r_0 = 0.40$ m and $d = 0.10$ m, removal of the first 3 Zernike modes from the incident wavefront is equivalent to the removal of the first 8 modes when d is increased to 0.20 m. This is because the cutoff frequency of the WFS is increased as d gets smaller.
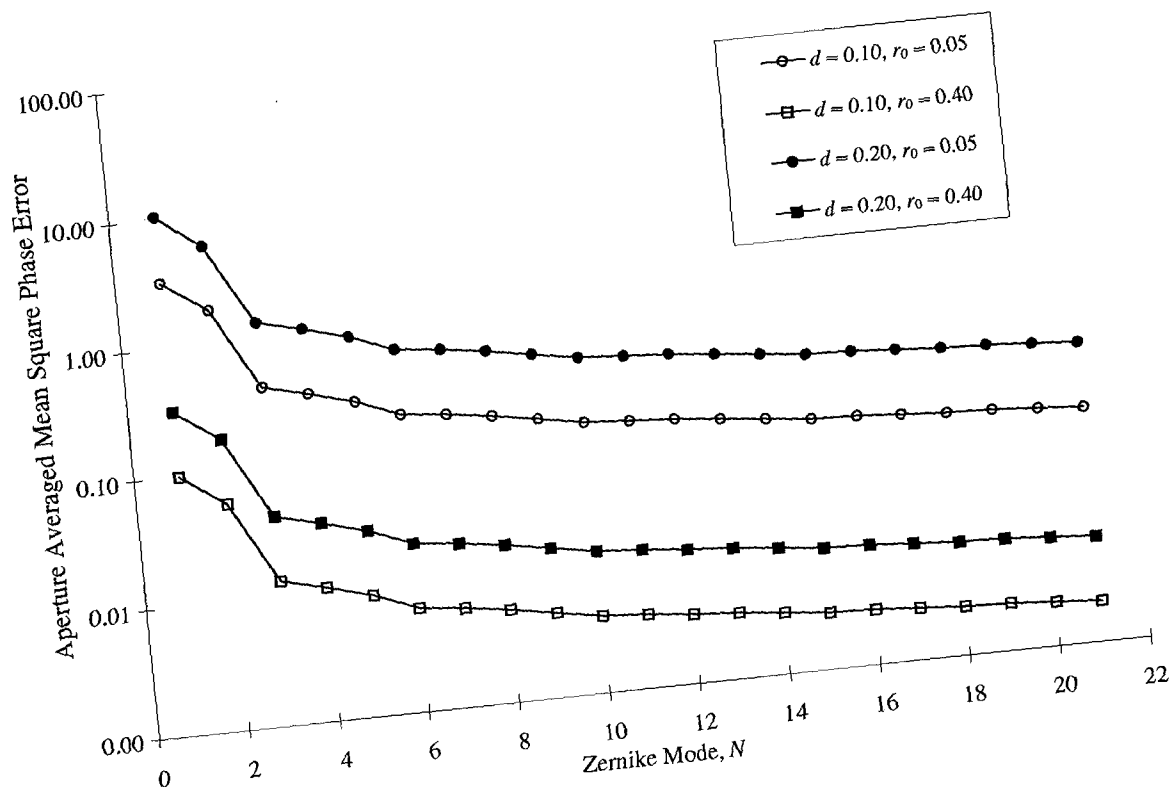
Figure 35. Aperture Averaged Mean Square Phase Error After Removal of the First N Zernike Modes. The WFS subaperture size is given by d and the Fried parameter is given by r_0.

Using the table discussed above, Roggemann and Welsh provide a general rule of thumb for "good" atmospheric compensation in regard to WFS and DM sample spacing that is very similar to the one given by Gavel *et al* in the previous chapter. That is, the AOS should be designed with a sample interval of the WFS and DM equal to the Fried parameter, r_0. Improving

96

performance limitations imposed by sample spacing calls for decreasing the WFS subapertures and DM actuator spacing. However, for typical observatories, r_0 can vary from 3 cm under poor seeing conditions to 20 or 30 cm when seeing is good [7:21]. Accordingly, to have a fully compensated system under the most adverse conditions requires many WFS subapertures and DM actuators. This increase in system complexity is directly translated to increased system cost. As such, AOS designs generally have WFS and DM sample spacing greater than r_0 [9:176].

Aside from cost and complexity issues, another reason for the implementation of a partially compensated AOS is to deal with finite light levels incident upon the WFS. As the number of WFS subapertures is increased, the total average number of photoevents seen by each subaperture is decreased because their surface area will also be decreased. Therefore, as degradations associated with finite sample spacing improve with decreasing subaperture size, the overall wavefront residual error may actually increase under low light levels. This is because the root mean square error of the slope measurements is inversely proportional to the total average detected photon count per subaperture measurement [9:186]. Hence, there is a fundamental trade-off between decreasing sample spacing to allow for finer wavefront sensing and compensation and increasing noise levels at the WFS.

Roggemann and Welsh provide a formula that gives the shot or photon noise at the WFS, which is directly related to the root mean square error of the slope measurements [9:187]. By considering the beacon to be a point source, the formula for the standard deviation of the shot noise, σ_{ns}, is

$$\sigma_{ns} = \frac{\sqrt{2}\pi}{d\,\overline{K}^{1/2}\int_{-1}^{1}\exp\left[-3.44\left(\frac{D|f_x|}{r_0}\right)^{5/3}\left(1-|f_x|\right)^{1/3}\right]\mathrm{d}f_x} \tag{30}$$

where $\overline{K}$ is the average photon count per subaperture for each WFS measurement, D is the pupil diameter, and f_x is the one-dimensional spatial frequency normalized to one at the optical cutoff frequency of the system. Using Equation (30), it is possible to plot the shot noise as a function of d, r_0, and light levels by numerical integration. Figure 36 shows such a plot assuming $D = 1.6$ m and a system sample rate of 1000 Hz, which is necessary for computing $\overline{K}$ under previous assumptions.
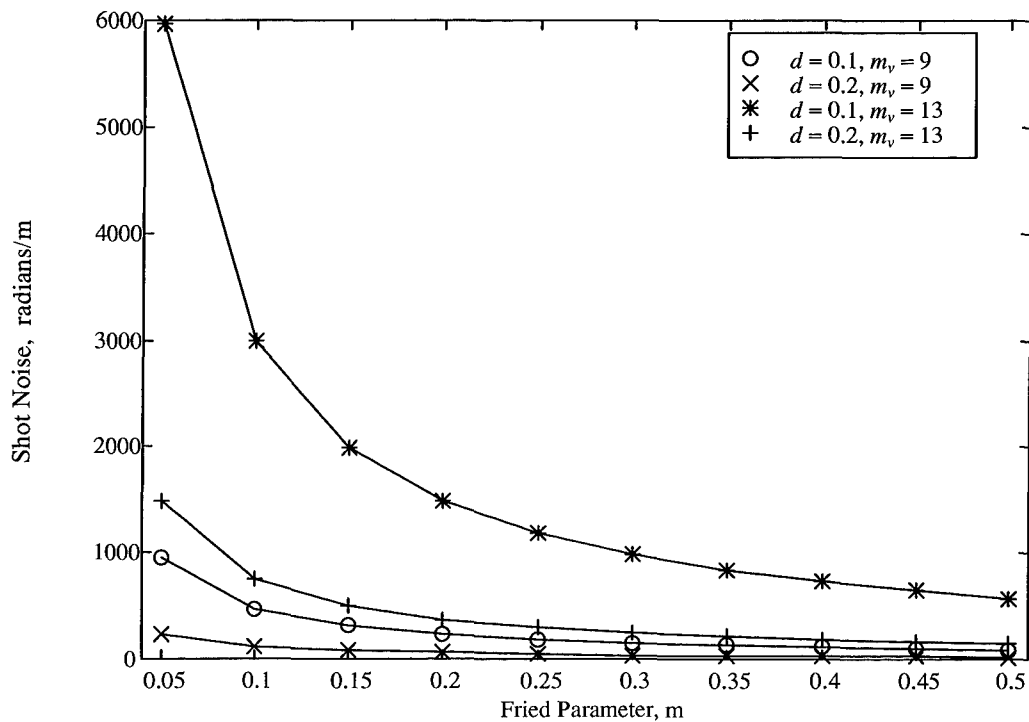
Figure 36. Shot Noise versus the Fried Parameter for Different Subaperture Sizes and Light Levels. The WFS subaperture size is given by d and the object visual magnitude is given by m_v.

Figure 36 shows that shot noise is substantially reduced as seeing conditions improve regardless of light levels. As expected from Equation (30), shot noise is also decreased when the

subaperture size is increased or the light level is increased. What is also noticeable in this plot is that as seeing conditions improve for the same light levels, the relative difference in shot noise between larger and smaller subapertures is decreased. Hence, there may exist a point of diminishing returns in reducing shot noise by increasing the subaperture size.

Clustering of Wavefront Sensor Subapertures

For an existing adaptive optics system such as the Air Force Maui Optical Station (AMOS) 1.6 meter system, it is not possible to reduce the size of the WFS subapertures or DM actuator spacing further. On the other hand, when seeing conditions are good, it may be beneficial to cluster together WFS subapertures in order to overcome high noise levels at the WFS under very low light levels. For instance, four subapertures with a side dimension of d could be clustered together via software control into one larger subaperture with a side dimension of $2d$ as shown in Figure 37. This method would quadruple the number of incident photons to each of the resulting clustered WFS subapertures. For feedback control consideration, the DM actuator spacing would also be doubled via software instruction [37:21]. When low SNR levels at the WFS are the dominant factor causing wavefront degradations, this technique might provide better overall performance than a more highly sampled system.

Of course this notion can be applied during the design phase as well. To the extent that budget constraints would allow, WFS subaperture size and DM actuator spacing could be made as small as possible to provide full compensation under the expected worst case operational scenario for r_0. If the target or WFS beacon is bright so that noise is relatively low, WFS subapertures could remain unclustered to compensate for high spatial frequencies of the wavefront deformations. As r_0 gets larger or the target gets dimmer, a clustering scheme could provide better overall performance by increasing the WFS signal-to-noise ratio.
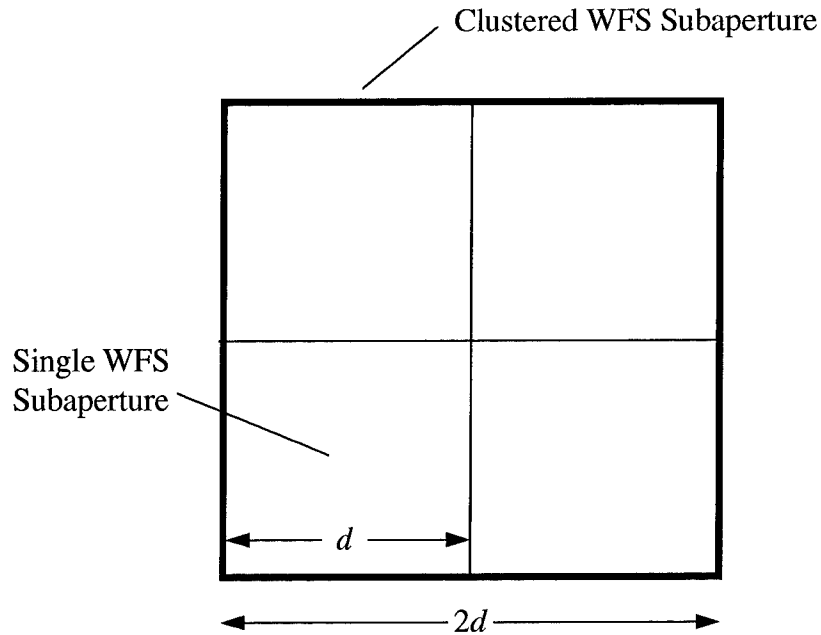
Figure 37. WFS Subaperture Clustering Scheme

The remainder of this chapter is geared at examining when clustering of WFS subapertures would benefit the performance of the AMOS 1.6 meter adaptive optics system. This work used the Instantaneous Strehl Ratio (*ISR*) defined in the previous chapter as a measure of system performance.

Experimental Setup

The work in this chapter again used the HYSIM adaptive optics simulation. As mentioned, this work modeled the AMOS 1.6 meter adaptive optics system. Figure 38 shows the input values used by HYSIM to model the AMOS system parameters. This figure also shows the values of other system controllable parameters and uncontrollable variables that were incorporated in the research in this chapter.

Controllable Parameters   **AMOS System Parameters**

- Number of frames: 200
- System sample rate: 1000 Hz
- System gain: 0.005 - 0.015
- Frame exposure time: 0.001 s
- Average long exposure time: 0.2 s

- Primary mirror diameter: 1.6 m
- Obscurration diameter: 0.335 m
- Imaging wavelength: 7.0E-07 m
- WFS wavelength: 6.0E-07 m
- Number of subapertures across pupil diameter: 8 and 16
- DM actuator separation in pupil plane: 0.22 and 0.11 m

System Variables

Inputs → AOS (Model by HYSIM) → Outputs

Uncontrollable Variables

- Average long exposure image
- Average long exposure OTF
- Vector of instantaneous Strehl ratio (*ISR*) values

- Extended object target distance: 6.5E+05 m
- Object and WFS visual magnitudes: 9 - 13
- Fried parameter: 0.05 and 0.40 m
- Average local wind velocity: 12.5 m/s
- Physical size of array holding input object: 40.0 m

Figure 38. Summary of Parameters Used to Model AMOS 1.6 Meter Adaptive Optics System and Variables Used in Clustering Study in Chapter V

Results

This section presents results of examining clustered versus unclustered WFS subapertures for the modeled AMOS 1.6 meter adaptive optics system. Based on the discussion presented in the previous sections, the potential operational scenario for invoking a clustering arrangement is under very low light levels. For the AMOS system, unclustered WFS subapertures will have a side dimension of 0.1 meters. Using the clustering scheme presented in Figure 37, the clustered WFS subapertures will have a side dimension of 0.2 meters. Hence, the system sample spacing is doubled by clustering. At the same time, clustering will boost SNR levels at the WFS by a factor of two. With this and the other parameters given in Figure 38 and summarized in Table 5, simulation runs were conducted for different object visual magnitudes assuming both unclustered and clustered WFS subapertures. It is noted that this work also made use of some of the results

101

from the previous chapter that indicated the desire for low system closed loop bandwidths under low light levels. Hence, Figure 39 compares the mean *ISR* metric presented in Chapter IV for unclustered and clustered WFS subapertures using a closed loop bandwidth of 15 Hz for both for Cases 1 and 2. Similarly, Figure 40 shows the mean ISR metric for unclustered and clustered WFS subapertures using closed loop bandwidths of 5, 10, and 15 Hz for Case 2.

Table 5. Summary of Parameters for Different Cases Studied in Chapter V

Case	r_0 (m)	v (m/s)	f_g (Hz)
1	0.05	12.5	106.5
2	0.40	12.5	13.3

Figures 39 and 40 show that for the modeled AMOS 1.6 meter adaptive optics system, there will never be a preference for clustering the WFS subapertures via software control, at least down to a visual magnitude of 13. At this and lower visual magnitudes, the system is performing so poorly that it may even be better to turn the adaptive optics off. This is because the low value of the mean *ISR* metric indicates that the system could start to perform worse than the AOS off case, when no compensation is provided.

Conclusions

This chapter presented a discussion on the trade-offs between decreasing the subaperture size to improve AOS performance limitations imposed by WFS and DM actuator sample spacing
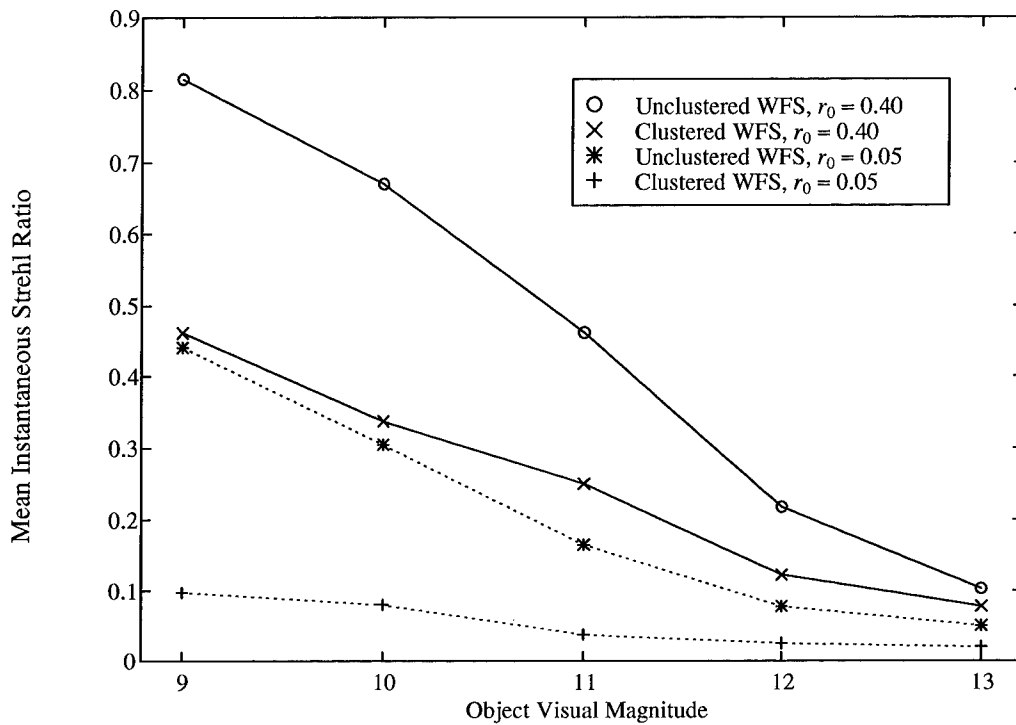
Figure 39. Mean Instantaneous Strehl Ratio versus Object Visual Magnitude for Clustered and Unclustered WFS Subapertures and Different Values of the Fried Parameter, r_0. The closed loop bandwidth was 15 Hz. Case 1 has $r_0 = 0.05$ meters and Case 2 has $r_0 = 0.40$ meters.

and increasing the subaperture size to reduce noise levels at the WFS. The results show that for the modeled AMOS 1.6 meter adaptive optics system, there will never be a preference for clustering the WFS subapertures via software control. These results are explained by the fact that when the subapertures were clustered, the amount of noise reduction at the WFS was not sufficient to overcome the increase in measurement errors associated with larger WFS subapertures at any of the light levels examined. It is concluded that this is a performance limitation of a system designed with relatively large subapertures such as the AMOS 1.6 meter system where $d = 0.10$ meters. This work again showed a preference for low closed loop bandwidths under low light levels.

103
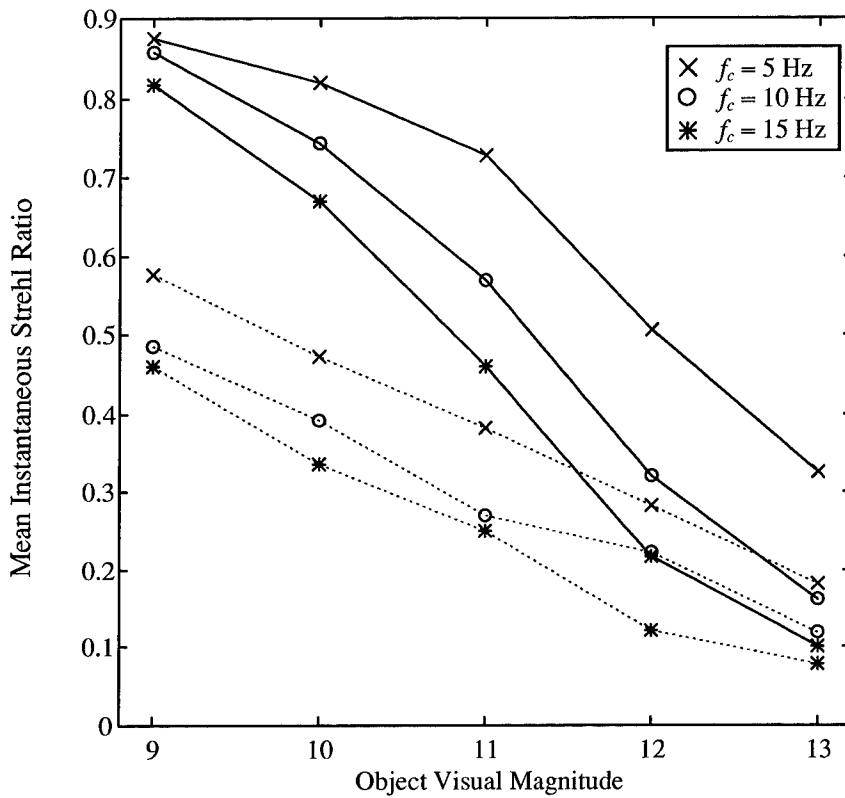
Figure 40. Mean Instantaneous Strehl Ratio versus Object Visual Magnitude for Different Closed Loop Bandwidths, f_c, and Clustered and Unclustered WFS Subapertures. The solid line shows unclustered subapertures and the dashed line shows clustered subapertures. Seeing conditions were $r_0 = 0.40$ meters and $f_g = 13.3$ Hz (Case 2).

VI. Summary and Recommendations

This chapter presents a summary of the research contributions provided by this dissertation effort and offers ideas for future related research topics.

Summary of Research Contributions

The following items are a summary of contributions obtained from this research:

1) Two novel image quality metrics were developed -- the Modified Strehl Ratio (MSR) and the Noise-Effective-Cutoff (NEC) frequency. It was shown that these two metrics, as well as the commonly used Strehl ratio, are highly correlated with the Root-Mean-Square Error (RMSE) between the detected image obtained with an AOS and the corresponding diffraction limited image. These results were shown to hold across a wide range of atmospheric seeing conditions, object light levels, and AOS settings. These results also hold for imaging both extended objects and point sources. These are important results because the MSR and NEC can be obtained and used in real field experiments to gauge AOS performance whereas the RMSE cannot. Appropriate decisions regarding system settings to maximize AOS performance can be made based on these metrics.

2) For a constant beacon light level, such as might be obtained with the use of a laser guide star, it was shown that the Strehl ratio cannot distinguish between different image qualities. On the other hand, the MSR retains a high degree of correlation with RMSE. This is because the MSR metric can account for both Wavefront Sensor (WFS) performance and measurement noise in the detected image, whereas the Strehl ratio only accounts for WFS performance.

3) As a subsidiary contribution, it was shown why the sample based average image spectrum Signal-to-Noise Ratio (SNR) approaches a limiting value of 0.886 as the normalized

spatial frequency approaches one. This is in contrast to the theoretical average image SNR which approaches a limiting value of zero as the normalized spatial frequency approaches one.

4) The most significant factor affecting AOS performance is the object light level that is used to drive the WFS. This is followed by the system gain, which is akin to the system closed loop bandwidth when the system sample rate is held constant. The next most important factor is the Fried parameter which is a measure of atmospheric seeing conditions. For a constant long exposure time, the least most important factor is the short exposure time. This is because about the same number of photoevents will be seen in the long exposure regardless of the short exposure time.

5) An empirical model of the AMOS 1.6 meter adaptive optics system performance based on the MSR was constructed. The fitted regression model can be used to accurately predict AMOS performance based on input values for the variables of system sample rate, system gain, object visual magnitude, Fried parameter, and wind speed. The coefficient of multiple correlation for this model was 0.991.

6) Based on the above results, it was decided to optimize the system closed loop bandwidth as a function of light levels and seeing conditions. Optimal closed loop bandwidth settings for the AMOS 1.6 meter system were determined. Moreover, it was shown that the proper choice of the closed loop bandwidth settings are much more critical under low light levels than under high light levels, regardless of the seeing conditions. Under low light levels, these results show that the optimal closed loop bandwidth settings are many times lower than the Greenwood frequency. This is an important result because the widespread rule of thumb specifies a closed loop bandwidth setting that is equal to the Greenwood frequency. Finally, as seeing conditions worsen, the optimal closed loop bandwidth settings shift to slightly higher values and visa versa and seeing conditions improve. This is as expected.

7) A method to produce robust system performance under widely varying atmospheric seeing conditions was developed. That is, a strategy was determined that led to the identification of the closed loop bandwidth setting for a particular light level that will yield performance that is almost always within 95 percent of the maximum, regardless of seeing conditions. This is an important result because the system operator need only know the predicted object visual magnitude that is to be imaged. The preferred system closed loop bandwidth can then be selected without having to worry about seeing conditions. Moreover, these results are important even for systems that incorporate artificial guide star capabilities. This is because such systems will produce beacons with equivalent stellar visual magnitudes of about 8 to 10. In this region, it is shown that selecting the proper closed loop bandwidth to achieve good imaging system performance becomes fairly delicate.

8) Trade-offs between fully compensated and partially compensated adaptive optics systems were identified. The most conspicuous trade-off is of course cost. In addition, fully compensated systems will be able to achieve a higher level of performance via finer wavefront sensing and compensation than partially compensated systems, when no regard is given to light levels. When noise dominates over limitations imposed by sample spacing, then partially compensated systems may actually perform better than fully compensated systems. This is because partially compensated systems can boost SNR levels at the WFS.

9) Based on the above assessment, the concept of clustering smaller WFS subapertures into larger effective subapertures was introduced. An experiment under this notion was conducted by modeling the AMOS 1.6 meter system. It was concluded that clustering of the AMOS subapertures will never benefit system performance, at least up to a very dim visual magnitude of 13. It is believed this is because the AMOS subapertures are already relatively

large, so that by clustering them the resulting wavefront errors due to sample spacing can never overcome the associated reduction in WFS noise.

10) In the early stages of this research, a spreadsheet was developed for the purpose of determining input values to HYSIM that would ensure that these values were what would actually be used in the simulation. That is, due to the discrete nature of the HYSIM code, rounding is often necessary to ensure proper accounting of dynamic variables as the simulation proceeds. For example, the phase screen that is used to model atmospheric turbulence is moved across the pupil of the telescope in discrete pixel amounts according to the value specified for the local wind speed. Hence, only wind speeds that correspond to discrete pixel shifts can actually be realized by HYSIM. If an input wind speed is specified that does not correspond to discrete pixel shifts, the nearest number of pixel shifts will be implemented. In this case, the actual modeled wind speed will be different from the one specified in the input data file. Using the mentioned spreadsheet allows for easily determining the values of the input data values that will in turn be the same as modeled by HYSIM after rounding. This spreadsheet was provided to personnel at the Starfire Optical Range (SOR) at Kirtland AFB for their use as well in running simulations with HYSIM.

11) Using the above spreadsheet, it was realized that for large telescope pupils, high wind speeds, or fine WFS subaperture spacing, the size of the random phase screen matrix was insufficient. That is, under any of these conditions the phase screen could quickly be exhausted as it is moved across the telescope pupil. After a particular phase screen is exhausted, a new random phase screen that is uncorrelated with the original phase screen is created and placed in the telescope pupil. Hence, the system will need to go through a transient period until it settles down again each time a new phase screen is created. This is not a big problem if the number of shifts per phase screen remaining after the system has settled down is large compared to the

number of shifts required to advance past the transient phase. On the other hand, if the converse is true, then the simulated results will show that the system is performing worse than it would if the phase screen was very large. Hence, after this problem was identified, the HYSIM code was modified to allow for the creation of larger phase screens.

12) Lastly, in Chapter IV some plots that showed the instantaneous Strehl ratio as a function of the short frame number indicated a problem with the phase screen shifting routine incorporated in HYSIM. Specifically, the system was taking much longer than expected to pass through the transient period. In some cases this period would exhaust a significant portion of the phase screen matrix. When dealing with phase screens of finite size, this means that the simulated results would again indicate that the system is performing worse than it should. This problem was tracked down and only a very minor modification to the code was needed to correctly implement shifting of the phase screen matrix. After this, plots of the instantaneous Strehl ratio as a function of the short frame number showed behavior as expected.

Recommendations and Suggestions for Future Research

The following items are a summary of recommendations for future research related to this dissertation.

1) Recall, Chapter I discussed the image quality metrics of ground resolution and Ground Resolved Distance (GRD). These metrics imply remote sensing systems looking toward the earth. It should be possible however to apply similar methods to systems looking skyward at extended space objects of known size/distance. Both of the above measures are a function of system attributes and target range and both methods rely on experienced photo interpreters to provide estimates of observed feature sizes. In this regard, Holst references the GRD based Imagery Interpretability Rating Scale developed for military applications and states that it can be

modified for environmental remote sensing. This inspires the notion for developing a similar rating scale for imaging space objects. HYSIM could be used to produce diverse image qualities of various different input objects as a function of seeing conditions, light levels, and target distances. The quality of the resulting images could be assessed by experienced photo interpreters and a rating scale could subsequently be derived therefrom.

2) Somewhat similar to the above notion, another idea was discussed in Chapter II that relates to image quality metrics. That is, the idea of using multiple observers to view and rank numerous images containing known amounts of degradations. From this, equations could be derived which relate these rankings to the amount of degradation [11:239]. Along these lines, a potential research scenario is to a) produce images under different amounts of atmospheric distortions and AOS control settings, b) have observers rank the quality of these images, c) train a neural network to classify the images, d) extract information from the neural network that can be used to predict image quality and gauge the saliency of controllable input parameters and uncontrollable factors that influence system performance, and e) develop an AOS image quality metric that can be used to optimize AOS performance.

3) Roggemann and Welsh indicate that minimizing the aperture averaged mean square residual phase is equivalent to maximizing the AOS Strehl ratio [9:195]. Developing an empirical relationship between the long exposure aperture averaged mean square residual phase and the long exposure Strehl ratio could be explored. Similarly, an empirical relationship between the instantaneous aperture averaged mean square residual phase and the instantaneous Strehl ratio would also be an interesting research topic. Alternatively, exploring relationships between residual phase error and other metrics examined in Chapter III, such as the root-mean-square error and the modified Strehl ratio may prove interesting.

4) In Chapter IV it was shown that by far the most significant factor affecting AOS performance is object light level. Throughout this research it was assumed that the object to be imaged was providing photons to both the WFS and imaging detectors. An RSM/DOE analysis could be developed to not only consider the factors discussed herein, but also to consider independent light levels to both the WFS and the imaging detector. Thus, performance assessments could be made of systems that are equipped with laser beacons. Since incorporating laser guide stars is a prevalent trend, such research would be very applicable to current and developing adaptive optics systems technology.

5) Chapter IV presented a subsection on the feedback control model used by HYSIM and which is commonly used by actual adaptive optics systems. This section inspires the notion of using neural networks to learn and then predict the incident wavefront phase as a function past wavefront measurements. These predictions could be used to compute command signals to the deformable mirror. Perhaps the system could respond more rapidly to the changing atmosphere under this method than the more conventional feedback control method which uses a weighted sum of past wavefront measurements to compute the next command signal. The main idea here is that under low light levels the need to significantly lower the closed loop bandwidth was shown. This implies that the system will respond more slowly to a changing atmosphere than if the closed loop bandwidth were higher. Under high Greenwood frequencies, this can start to significantly impair system performance. With a neural network, maybe the system can respond just as rapidly to the changing atmosphere regardless of light levels to the WFS.

6) Chapter V examined a relationship between the aperture averaged mean square phase error and subaperture side dimension, d. This chapter also examined a relationship between the WFS shot noise, subaperture size, light levels, and Fried parameter. A potential research topic would be to develop a relation between WFS shot noise and the aperture averaged

mean square phase error. With this, it might be possible to optimize the WFS subaperture side dimension as seeing conditions and light levels change. Additionally, closed loop bandwidth may also be included -- this recommendation begs the need for an RSM type analysis. In this case, the goal would be to develop an empirical model of the aperture averaged mean square phase error as a function of r_0, m_v, d, f_c.

7) Another potentially useful research topic would be to examine the affect of varying the wavelength and associated bandwidths used by the WFS and imaging detectors respectively. Longer wavelengths of light may be preferred over shorter ones since they are perturbed less by the atmosphere. Another consideration is choosing wavelengths that provide the most photons to the WFS and imaging detectors. In this research, the relative amount of light sent to these detectors was kept constant. Therefore, the number of photons incident upon each detector was only a function of the target brightness. However, one can also consider varying the relative amount of photons sent to each detector by virtue of an assumed beam splitting device. When a high level of compensation can be achieved, such as under high light levels or good seeing, more photons could be sent to the imaging detector to boost image spectrum SNR levels. When light levels are low, more photons could be diverted to the WFS for better atmospheric compensation.

Appendix A. Adaptive Optics System Simulation

This appendix presents a summary of the simulation that will be used in this research to model adaptive optics systems. This simulation is called the Hybrid Imaging Simulator (HYSIM). HYSIM is used to model and compute Optical Transfer Function (OTF) statistics, image spectrum statistics, and an average long exposure image for either a point source or an extended object under specified atmospheric conditions and AOS configurations.

Overview of HYSIM

Figure A1 provides an overview of the HYSIM flow process [37:6]. There are three major phases to the simulation: setup, main loop, and output. In the setup phase, AOS simulated control parameters are established. The majority of these parameters are read from an external input data file. The input data file also specifies information about the object to be imaged and the number of images (or frames) to be realized during the simulation. Next, statistics that establish atmospheric effects on the entrance pupil (i.e., the telescope aperture) are computed. Subsequent to this, the pupil template and adaptive optics grid structure (WFS and DM actuators) are calculated. Using this grid structure and a least squares reconstruction model, a DM control matrix is calculated that maps WFS measurements to DM actuator commands [37:5].

In the main loop phase, individual occurrences of each image frame are computed and statistics for each of these realizations are stored. At the beginning of each frame loop, a phase screen is created and is placed in the entrance pupil of the telescope using information from the setup phase. WFS and DM models are then used to simulate compensation of the aberrated wavefront. Next, the instantaneous OTF and short exposure image are computed. Appropriate
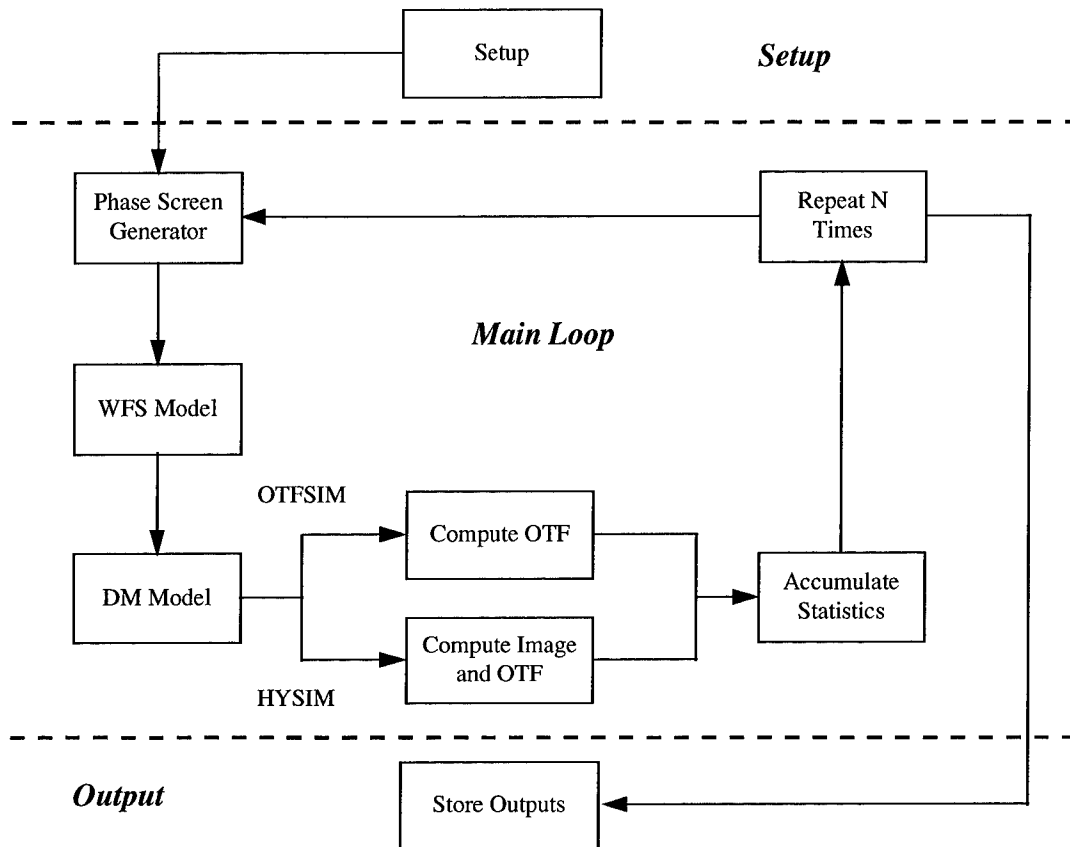
Figure A1. Flowchart for HYSIM

statistics are accumulated and the loop continues until the desired number of frames is reached [37:6-7].

Finally, the simulations enter the output phase where additional summary statistics such as the mean OTF and average image spectrum Signal to Noise Ratio (SNR) are computed. These output data along with the final averaged image are saved to files as specified in the input data file.

Details of HYSIM

HYSIM is a Fortran based adaptive optics simulation program developed by Michael C. Roggemann. Significant details of HYSIM pertinent to this dissertation are presented below.

114

HYSIM List File. The HYSIM list file, hysim5.dat, is an input data file read by the

executable HYSIM code. This input file allows for different seeing conditions, optical

configurations, and AOS parameters to be specified without having to alter the compiled HYSIM

code. Table A1 contains a sample hysim5.dat file. In this table, italicized characters are the

Table A1. Sample hysim5.dat Input Data File

100	*nframes* = number of frames
10	*iradius* = number of subapertures across mirror diameter, must be even
9393291	*iseed* = random seed, should be large and odd
1.0	*dmir* = primary mirror diameter, m
0.0	*lobs* = telescope obscuration diameter, m
7.0e-7	*wvl* = imaging wavelength, m
6.0e-7	*wfswvl* = wavelength for wavefront sensor, m
9.0e5	*ranget* = target distance, m
1000000.0	*phot* = average number of photons per image
0.25	*r0* = r_0 at 0.5 micron, m
2	*icase* = 1 - calibration, 2 - object imaging
1	*itilt* = 1 - correct tilt, 0 - don't correct tilt
1	*iao* = 1 - use adaptive optics, 0 - don't use adaptive optics
20.0	*nstep* = number of steps for integration on subaperture
0.12	*actsep* = actuator separation grid size in pupil, m
1000.0	*phcnt* = average # of photoevents per subaperture per integration time
200.0	*bw* = closed loop bandwidth, Hz
0.6	*loopgain* = loop gain
30.0	*wspd* = wind speed, m/s
0.002	*frtime* = image frame integration time, s
ocnr5	*name* = name of object input file array
10.0	*arrlen* = length of one side of object array, m
image	*avgnm* = average long exposure image file name
imagesnr	*snrnm* = average long exposure image spectrum SNR file
imageotf	*avgotf* = average long exposure OTF file name

variable names used in HYSIM while the left-hand column indicates the input value of these variables. It is assumed that all units of length are in meters. A more detailed description of these variables is also given below.

The variable *nframes* specifies how many loops of the program are to be made, each loop yielding a different short exposure image for a random occurrence of the atmosphere. *Iradius* specifies the number of subapertures across the telescope pupil. This variable also establishes the size of the subapertures for a given primary mirror diameter. *Iseed* is the variable name for the seed used by the random number generator in HYSIM. The primary mirror diameter of the telescope is specified by *dmir*. Most telescopes have a secondary mirror that partially obscures the primary mirror. The diameter of this secondary mirror can be specified by *Iobs*. The next two variables, *wvl* and *wfswvl*, specify the imaging wavelength and the WFS wavelength respectively. The distance from the telescope to the target object is given by the variable *ranget*. *Phot* specifies the desired number of photons to be captured per image and is related to object brightness.

The next variable, *r0* , is the atmospheric coherence diameter. Also known as the Fried parameter, this variable specifies the stability of the atmosphere. Per Fugate, r_0 may be thought of as the size of the typical turbulent "seeing cell" in the atmosphere [7:21]. The larger r_0, the more stable the atmosphere and the better the seeing. The converse applies for smaller values of r_0. Roggemann and Welsh state that the Fried parameter can be interpreted as the aperture size beyond which further increases in diameter result in no further increase in resolution [9:71]. The Fried parameter is a function of several variables including wavelength, altitude, and angle from the zenith to the target object. Having units of length, r_0 can vary from 5 cm or less under poor seeing conditions to 20 cm or more under good seeing conditions.

The next variable, *icase*, tells the simulation which type of object is to be imaged, 1 for a point source object or 2 for an extended object. *Itilt* specifies whether or not to compensate for tilt degradations in the incoming light caused by the atmosphere. Likewise, *iao* specifies whether or not to use adaptive optics (i.e., adaptive optic system on or off). *Nstep* is used to define how many integration steps to use in calculating the Jacobian that relates the sensitivity of WFS measurements to DM actuator commands. The total number of steps taken across each subaperture in calculating the Jacobian is 2 *nstep* + 1 [37:8]. The distance between adjacent DM actuators commands is given by *actsep*. This distance needs to be scaled proportionate to the diameter of the telescope aperture.

Phcnt specifies the desired number of photons to be seen by the WFS during each update and like *phot*, is related to the object brightness.. *Bw* is used to model the sample rate of the AOS in Hertz (Hz). In the simulation, *sample rate* = 10 *bw*. The variable *loopgain* is used to compute the closed loop bandwidth, f_c, of the AOS. Specifically, f_c = *gain* x sample rate where *gain* = *loopgain* / 2π. Using *bw*, HYSIM computes a variable called *deltatime* = 1/ (10 *bw*) in seconds, which is the shortest time used in the program. For each *deltatime* increment, the turbulent atmosphere is moved across the pupil of the telescope by an amount *deltatime* x *wspd* where *wspd* is the local wind speed in meters per second. *Frtime* is the variable used to specify the total exposure length of each frame realization. Hence, each frame will have seen *frtime* / *deltatime* movements of the atmosphere and the same number of corrections applied to the DM.

Name is the variable used to specify the file name of an extended object to be imaged. *Arrlen* is the size of an array used to hold the object. It is used to properly scale the array of the object to be imaged. The last three lines of the input data file, *avgnm, snrnm,* and *avgotf,* are variables used to specify output file names for the average long exposure image, the SNR of the average long exposure image spectrum, and the average long exposure OTF respectively.

117

Regular Space and Pixel Space. To allow for processing by computer, discrete computational forms are sometimes needed throughout the simulation. Nondiscrete measurements will be referred to as regular space and discrete forms will be called pixel space. As seen in the input data file, one specifies a primary mirror diameter and the number of WFS subapertures across this diameter. Internal to the HYSIM program, a constant that specifies the number of pixels (or grid points) across one of these subapertures is called *ispace*. With this and the former two variables *dmir* and *iradius*, the sampling of the entrance pupil is established.

For the numerical integration algorithm used (Simpson's 1/3 rule), *ispace* must be an odd number. In addition to affecting the obtained resolution of simulated images (i.e., increased resolution with higher values of *ispace*), the value of *ispace* is also related to the size of the array used to model atmospheric turbulence. That is, as *ispace* is increased, the distance between adjacent pixels is decreased. Therefore, for a particular wind speed, the phase array will move across more pixels with increasing values of ispace. Accordingly, as will be discussed, it may be necessary to adjust the size of the phase array to accommodate different values of *wspd*.

For contiguous (or touching) subapertures, the number of pixels across the entrance pupil is given by *ndiam* which can be computed as follows

$$ndiam = 1 + iradius \times (ispace - 1) \tag{A1}$$

This then allows for a conversion between regular space and pixel space. That is, the sample spacing in the entrance pupil, *dx*, is given by

$$dx = dmir / (ndiam - 1) \tag{A2}$$

For consistency with discrete Fourier transform algorithms (i.e., Fast Fourier Transform or FFT), the center location of an $n \times n$ array is given by pixel location $(n/2 + 1, n/2 + 1)$ using integer truncation. Figure A2 shows an example of four contiguous WFS subapertures with *ispace* = 5.
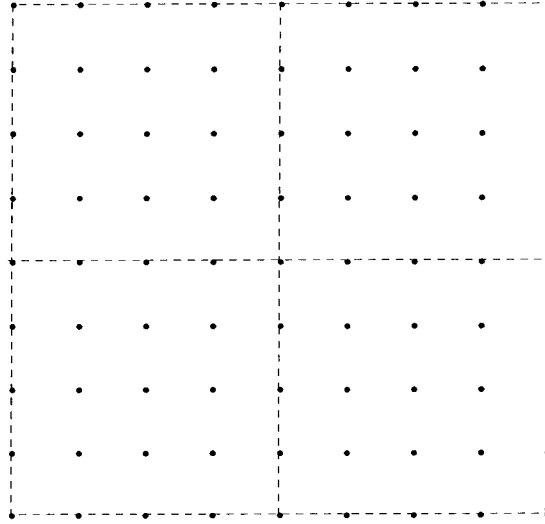
Figure A2. Example of Four Contiguous WFS Subapertures, Ispace = 5

Finally, the radius of the entrance pupil in pixels is computed as follows

$$radius = (ndiam - 1) / 2.0 \qquad (A3)$$

<u>HYSIM Setup Phase</u>. As previously mentioned, during the setup phase, input parameters are read from the input data file and the above pixel space computations are performed. Since the input data file specifies r_0 at a nominal wavelength of 0.5 μm, r_0 is calibrated to the imaging wavelength by the Kolmogorov relation between r_0 and λ [37:13]. If the calibrated value of r_0

divided by *dx* is less than 2.5 the program will abort because of inadequate sampling of the incoming light.

The array size for the phase screen is a function of several internal program constants. These are *nbig*, *nhuge*, and *n0* where $nhuge = 2^{nbig}$ and *n0*= 10 x *nhuge* + 1. As mentioned, the constant *nbig* must be chosen so that the resulting *n0* x *n0*array is large enough to prevent data near the edges from being clipped. With these parameters established, the maximum allowable shift of the phase screen is given by *zmaxshift* = 2 x [(*n0* / 2) - *radius*] x *dx*. If the shift specified via the input variables is greater than *zmaxshift* the program will abort, otherwise it continues and calls the subroutine <u>setup1</u>.

Setup1 is contained in the file setup.f. This subroutine computes an atmospheric turbulence Karhunen-Loeve correlation matrix which is used to form a random phase screen. Telescope pupil arrays are established next. The constants *msdim* and *madim* are used to manage the size of these arrays which are dependent on the number of WFS subapertures and the number of DM actuators. Next, subroutine <u>mktemp2</u> is called. This subroutine creates the pupil template array called *puptemp*. In short, this template is used to represent the telescope pupil in array form. Array elements have either a value of 1 or 0 depending upon whether or not array elements are in or outside the pupil respectively. That portion of the primary mirror obscured by the secondary mirror are also represented by array elements with a value of 0.

A WFS subaperture template is established by the subroutine <u>putsa</u> which is contained in the file ao4.f. This subroutine outputs four arrays that describe subaperture locations in pixel space. To be an active subaperture, the entire subaperture must be contained in the pupil. Putsa also provides the number of active subapertures (*numsa*). The subroutine <u>phdiffsd</u> computes the standard deviation (*stddev*) of the WFS measurement noise using a Hartmann WFS model [37:17-18].

A DM actuator template is created by the subroutine putact. This subroutine outputs two arrays that contain the *x* and *y* coordinates of each actuator in real space and the number of active actuators (*numac*). Active actuators are those that fall within the unobscured pupil.

Next, if the AOS is turned on, the optimal matrix *R* (*rji*) that maps WFS phase difference measurements *S* (*delph*) onto DM actuator commands *C* (*cj*) using a least squares method is calculated via the subroutine compr. Deformable mirror actuator commands are then computed as $C = R \cdot S$ [37:19-21]. Roggemann notes that number of subapertures must be the same or greater than the number of actuators or the matrix *R* will be nonexistent.

The last step prior to entering the main loop of the HYSIM program is to create an object array. If a point source is to be imaged, an object array that has all zeros except for a one at the center is constructed. If an extended object is to be imaged, it is read from the file listed in the input data file. The object array contains *lq* x *lq* real numbers where *lq* = 2 x *nsid* and *nsid* = 16 x *nhuge*. Scaling is then performed to adjust for the object size, its range, imaging wavelength, primary mirror diameter, number of pixels across the pupil, and array size of the object [37:22-24]. Scaling is carried out in the subroutine scaleobj.

HYSIM uses a linear system model of incoherent imaging represented by the following

$$i(x) = o(x) * h(x) \tag{A4}$$

$$I(u) = O(u) \times H(u) \tag{A5}$$

where *i(x)* represents the image, *o(x)* represents the object, *h(x)* represents the Point Spread Function (PSF) of the AOS combined with the atmosphere, and *x* represents spatial coordinates. The Fourier transform of each of these yields Equation (A5). *I(u)* is the image spectrum, *O(u)* is the object spectrum, *H(u)* is the OTF of the AOS combined with the atmosphere, and *u* represents frequency coordinates. As mentioned in the introduction, random variations in the

121

atmosphere cause the OTF to also be random. Object scaling is necessary for the FFT algorithm used in the program. This ensures that the sample spacing of the array containing the FFT of the object equals the sample spacing of the OTF array so that the arrays can be properly multiplied. Finally, the object spectrum is normalized so that the amplitude of the dc spatial frequency is one [37:25]. HYSIM then enters the main loop phase of the program.

HYSIM Main Loop Phase. In this phase of the program, random phase screens are created, their effects on OTF are determined, and appropriate statistics are collected.

At the start of each frame loop, a thin phase screen located in the entrance pupil of the telescope is created. This phase screen represents atmospheric turbulence and is created from a random draw to have a phase structure function defined by Kolmogorov turbulence [37:25-28]. Contained in the file scrn.f, this is carried out in the subroutine screenmkr. The output of screenmkr is a phase screen (*wf*) for r_0 containing phase values in radians. A scaled version of this (*sclwf*) is then computed for the WFS. The mean and variance of these phase screens are calculated and their means are subtracted from them via the subroutine stat. Next, if the AOS is turned on, the phase screens are used to provide an incident wavefront to the WFS and DM and the OTF is computed. If the AOS is turned off, the WFS and DM models compute the OTF without any correction for atmospheric turbulence.

The WFS model used by HYSIM is a Hartmann WFS model with touching subapertures. This model computes phase differences across the subapertures via the subroutine phdiff1 which uses Simpson's 1/3 rule for numerical integration. The phase difference for each subaperture is computed and stored in the array *delph*. These phase differences also include measurement noise calculated in the subroutine *phdiffsd*. The output *delph* of the WFS model is the input to the DM model which computes DM actuator commands. The commands sent to the DM attempt to form the mirror so that it has a conjugate shape to that of the distorted wavefront. After being

reflected by the DM, the residual wavefront error (*newphs*) is computed by the subroutine modpup. This corrected wavefront is then used to compute the OTF and short exposure image.

Prior to being presented to the WFS model, if specified in the input data file, tilt is removed from the phase screens. This is done by the subroutine tiltoff which divides the first moment of the wavefront array by the first moment of the array positions [37:30]. In effect this removes the pupil average tilt in the incoming distorted light. In practice, this is accomplished via a flat mirror placed in the light path that can be tilted in two axes under closed loop control. Then the scaled version of the phase screen (*sclwf*) is sent to the WFS model.

As stated, the WFS model calculates phase difference measurements and returns them to the array *delph*. The phase screen is then shifted via the subroutine shift. DM actuator commands (*cj*) are then calculated by matrix multiplication of *rji* and *delph*. The commands are then provided to the DM model (subroutine modpup). The subroutine returns the corrected wavefront *newphs*. If the AOS is turned off, the phase screen *wf* is copied to the array *newphs*. Afterwards, if tilt removal is on, any tilt introduced by the DM is again removed from the corrected wavefront. The compensated OTF (*ct4*) of the combined atmosphere and telescope is then computed via the subroutine compotf.

The compensated OTF is used to calculate the image spectrum, *fimg*, as *fimg = ct4 × fobj*. The inverse FFT of *fimg* then yields the compensated, unnormalized image, *img*. This array is then normalized so that there are *phot* (from the list file) number of photoevents in the image. Then a compensated, photon limited image (i.e., no other noise present other than the random arrival times of photons) is obtained by passing *img* to the subroutine poisson. The total number of photoevents in the image are counted (*zcnt1*) and the FFT of the image is stored in the array *ta1*. This is the end of the main loop. Prior to returning to the top of the loop, HYSIM compiles

sums of *ta1* and $|ta1|^2$ which are used to calculate the average long exposure image and the SNR of the image spectrum [37:33].

HYSIM Output Phase. As mentioned, the main loop is run through *nframes* number of times, each loop yielding a particular short exposure image for a random occurrence of the atmosphere. Running sums are accumulated after each loop. After *nframes* number of loops, HYSIM averages the compiled sums by dividing by *nframes*. The SNR of the image spectrum is calculated from the following equation

$$snr = \frac{\sqrt{nframes} \times \langle ta1 \rangle}{\sqrt{\text{var}\{ta1\}}}$$

(A6)

where $\langle \cdot \rangle$ is the expectation operator (average) and $\text{var}\{\cdot\} = \langle (\cdot)^2 \rangle - \langle \cdot \rangle^2$ [37:34]. The SNR of the image spectrum is stored in the list file specified in the input data file. Finally, the average long exposure image is obtained via the inverse FFT of the average image spectrum. The real part of this is stored in the file specified in the input data file. The resulting image array can then be displayed as desired by the user.

Compiling and Running HYSIM

To obtain an executable version of HYSIM, the following statement can be used

:~⟩ f77 hysim5.f ao4.f setup.f scrn.f matrix.f -o hysim5.e

where hysim5.f is the main HYSIM code and the other *.f files contain the various subroutines required for the program. This compilation step yields the executable file hysim5.e along with an object file, *.o, for each *.f file. If the input data file, labeled hysim5.dat, is located in the same directory as hysim5.e, the program can be executed by entering hysim5.e.

A more recent version of HYSIM, allows for larger phase screens and eliminates the setup.f file. This version consists then of the following elements -- hysim5z.f, ao4.f, zscrn.f, and matrix.f. To obtain executable code for this version of HYSIM, the following statement can be used

:~⟩ f77 hysim5z.f ao4.f zscrn.f matrix.f -o hysim5z.e

where as before hysim5z.f is the main HYSIM code and the other *.f files contain the various subroutines required for the program. This compilation step yields the executable file hysim5z.e along with an object file, *.o, for each *.f file. Again, if the input data file, labeled hysim5.dat, is located in the same directory as hysim5.e, the program can be executed by entering hysim5z.e.

Appendix B. Detailed Review of RSM and DOE

This appendix presents a more detailed review of the RSM and DOE methodologies that were introduced in Chapter III, including Taguchi's method and other methods that deal with dispersion as a system response.

Introduction to RSM and DOE Methods

In applying Response Surface Methodology (RSM) and Design of Experiments (DOE), it is presumed that a functional relationship exists between the system output, h, and the controllable and uncontrollable input variables. Say there are p controllable variables (i.e., $X_1, X_2,..., X_p$) and q uncontrollable variables (i.e., $Z_1, Z_2,..., Z_q$). The presumed functional relationship may be written as

$$h = f(X_1, X_2,..., X_p, Z_1, Z_2,..., Z_q). \qquad (B1)$$

As the values of the variables change, the response will also change so that Equation (B1) represents a $p+q+1$ dimensional response surface. When the system is operated with the $p+q$ variables set at definite levels, this is known as an experimental run [12:3]. If these levels are unaltered and the experiment is repeated, the measured response will most likely vary due to experimental error and inherent system variability. Hence, h represents the mean response for a particular set of input variables. An observed response, Y, will be distributed about this mean and the expected value of Y is h. For a particular run, the difference between Y and h is the error, e. Therefore, Equation (B1) can also be thought of as functional relationship which relates the observed system response to the controllable and uncontrollable input variables and error. In equation form [12:3]

$$Y = f(X_1, X_2, ..., X_p, Z_1, Z_2, ..., Z_q) + e. \qquad \text{(B2)}$$

Assuming the true underlying response surface is locally "smooth," the above functional

relationship can be locally approximated by an interpolation function $g(\xi, b)$, such as a

polynomial, where ξ is a vector that represents the input variables and b is a vector of

coefficients. This interpolation function provides a local empirical model of the system [12:11].

In short then, one approach to meeting the optimization objectives of this research effort is

to a) choose a suitable AOS performance metric, b) decide on appropriate ranges for the input

variables, c) perform experimental runs at set levels of the input variables, d) compute the

resulting metric (i.e., the system response) for each run, e) estimate interpolation function

coefficients by applying linear regression techniques to these data, f) determine the gradient of

steepest ascent/descent from the resulting empirical model, g) follow this gradient (by changing

the settings of the controllable input variables) until estimated AOS performance no longer

improves or the model is no longer valid, h) repeat the process as necessary. RSM and DOE

techniques establish fundamental procedures to carry out this process in a systematic fashion.

Box says [12:13]

> *The essentially iterative nature of response surface methodology (RSM) would ensure that, as the investigation proceeded, it would be possible to learn about (a) the amount of replication needed to achieve sufficient precision, (b) the location of the experimental regions of most interest, (c) appropriate scalings and transformations for the input and output variables, and (d) the degree of complexity of an approximating function, and hence of the design, needed at various states.*

Before the above process can be implemented, it is necessary to identify all controllable and

uncontrollable input variables. This may result in a very long list. One way to reduce this list to

a manageable size is to sit down with the system experts and determine what variables they believe are most important. Box also recommends a preliminary two-level fractional factorial *screening* design to identify variables "worthy of further study" [12:13]. An experiment that simultaneous varies the levels of input variables is called a factorial experiment [11:456]. An experiment that operates at all possible combinations of the levels of the input variables is called a full or complete factorial design. An experiment that incorporates a subset of all possible combinations is a fractional factorial design. And if there are only two levels of the input variables, this is a two-level fractional factorial. A design of this type can help in determining "critical process factors" and the direction of changes for these factors to improve the system response [11:456].

In RSM and DOE, the primary input variables are often referred to as factors. The effect of a factor, or <u>main effect</u>, is defined as the change in the system response that results from a change in the level of that factor [11:483]. Consider two factors, A and B, each with two levels, A_1 and A_2 and B_1 and B_2. The main effect of factor A is the difference between the average system response when $A = A_1$ and the average system response when $A = A_2$. The main effect of factor B can be determined in a similar fashion. When the difference in the system response between the levels of one factor is not the same at all levels of the other factors, then there is an <u>interaction</u> between the factors [11:484]. Identification of significant interactions is very important. For instance, one might conclude that a main effect is very minor when indeed it is very significant but it is masked by an interaction. Graphical displays can be very useful presenting such interactions.

<u>2^k Factorial Designs</u>. Experimental designs that incorporate k factors that each have two levels are known as 2^k designs [11:494]. This is because one complete replicate of all possible combinations of factors requires 2^k runs. For instance, an experiment involving $k = 3$ factors,

each at two levels, requires eight runs per replicate. In a design of this type, computation of main effects, interactions between factors, analysis of variance, and residual analysis is readily carried out via application of a multiple linear regression model. To make computations even more efficient, the levels of each factor are usually coded to a value of −1 for the low level of the factor and +1 for the high level. Consider three factors, A, B, and C, each with two levels. Table B1 represents a systematic method to code the levels of the factors and to carry out the eight runs for each replicate. In this table, the high level of a factor (+1) is represented by a + sign and the low level of a factor (−1) is represented by a − sign. The column with the I heading is the identity column. The signs for the main factor columns, A, B, and C, are shown in customary order. The signs for the remaining factor columns (interactions) are obtained by multiplying the respective main factor columns, row by row [11:503].

Table B1. A 2^k, $k = 3$ Design

Run	Factorial Effect							
	I	A	B	C	AB	AC	BC	ABC
1	+	-	-	-	+	+	+	-
2	+	+	-	-	-	-	+	+
3	+	-	+	-	-	+	-	+
4	+	+	+	-	+	-	-	-
5	+	-	-	+	+	-	-	+
6	+	+	-	+	-	+	-	-
7	+	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+	+

Table B1 has several important qualities. The most significant of which is that the sum of the products of signs in any two columns equals zero, meaning columns are orthogonal. Through

this design, an estimate of main effects and interactions can be obtained by multiplying the

system response for each run by the desired column sign, summing the result, and then dividing

this by one-half of the number of runs in the experiment (including replicates) [11:504]. The

Sum of Squares (SS) for any effect can be computed by squaring the effect and multiplying this

by $n2^k-2$, where n is the number of design replicates. This knowledge allows an analysis of

variance table to be constructed very efficiently. Table B2 shows the analysis of variance table

for the above three factor example. The mean square column is determined by dividing the sum

of squares by the appropriate degrees of freedom.

Table B2. Analysis of Variance Table for a 2^k, $k = 3$, Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F Statistic
A	SS_A	1	MS_A	MS_A / MS_E
B	SS_B	1	MS_B	MS_B / MS_E
C	SS_C	1	MS_C	MS_C / MS_E
AB	SS_{AB}	1	MS_{AB}	MS_{AB} / MS_E
AC	SS_{AC}	1	MS_{AC}	MS_{AC} / MS_E
BC	SS_{BC}	1	MS_{BC}	MS_{BC} / MS_E
ABC	SS_{ABC}	1	MS_{ABC}	MS_{ABC} / MS_E
Error	SS_E	$2^k(n-1)$	MS_E	
Total	SS_T	$2^k n-1$	MS_T	

From the analysis of variance of table one can conclude if main effects and interactions are

significant or not by comparing their computed F statistics to tabulated values of desired

confidence. Finally, it is noted that with a 2^k design coded as shown in Table B1, the coefficients

for the regression model are one-half of the corresponding effect estimate. This is because the regression coefficients measure the effect of a unit change in a factor on the mean of the system response and the effect corresponds to a two unit change in a factor (i.e., -1 to $+1$) [11:499]. The intercept of the regression model is simply the overall mean of all runs.

In addition to the analysis of variance table, there are two other methods that are useful in determining which effects are significant. First, one can compute the standard errors of the effects and compare these to the magnitude of the effects. Second, one can construct a normal probability plot [11:506].

Montgomery discusses the use of a pooled variance estimate as an overall variance estimate. This estimate is the same as the mean square error (MS_E) from the analysis of variance procedure. Using this, the estimated variance of an effect, V(effect), becomes [11:506]

$$V(\text{effect}) = MS_E / n2^{k-2} . \tag{B3}$$

The estimate of the standard error, s(effect), is the square root of V(effect). With this, one can compute approximate 95% confidence intervals for each effect as the effect $\pm 2s$(effect). If this interval does not include zero, then the effect can be considered significant.

A normal probability plot is also useful in assessing the significance of effects. If effects are not significant, then their estimates will behave like random samples drawn from a normal distribution with zero mean and when plotted, the effects will lie approximately along a straight line. Effects that do not fall on this line are significant [11:511]. To construct a normal probability plot, the effect of each factor are arranged in ascending order. Then one computes the cumulative normal density function (CDF) for these ordered effects and plots pairs of the

effect and the CDF on normal probability paper. The CDF for the ith ordered effect, $F(i)$, is computed by [12:129]

$$F(i) = (i - 0.5) / n \qquad \text{(B4)}$$

Alternatively, the inverse CDF can be computed and pairs of the effect and inverse CDF can be plotted. The inverse CDF for the ith ordered effect, $\Phi^{-1}[F(i)]$ is computed by [38]

$$\Phi^{-1}[F(i)] = [F(i)^{0.1349} - (1 - F(i))^{0.1349}] / 0.1975 . \qquad \text{(B5)}$$

One can also construct a normal probability plot of the residuals, the difference between an observed run and the estimate that would result from the fitted regression model. The residuals should fall approximately on a straight line as a check of normally distributed observations, which is an underlying assumption for the analysis of variance [11:465]. Two more assumptions are equal variances at each factor level and independently distributed observations. These can be checked by plotting the residuals against the factor levels and the mean fitted response at each factor level respectively [11:465-467]. Montgomery also suggests plotting residuals versus each variable as a way to determine if any of the process variables affect variability of the system response. He also presents a formal test procedure to ascertain system variability as a function of the process variables [11:526-528].

Sparsity of Effects Principle. As the number of factors included in an experiment increases, so does the number of effects. In a 2^k factorial design, if $k = 3$, there are a total of seven effects, when $k = 4$ there are 15 effects, and when $k = 6$ there are a total of 63 different effects. Fortunately, the sparsity of effects principle applies in most situations [11:508]. This principle

says the system is usually dominated by main effects and low order interactions. Hence, when k = 4 or 5, Montgomery says it is a common practice to run only a single replicate of the 2^k design and then pool the higher order interactions into the estimate of error [11:508].

Fractional 2^k Designs. When higher order interactions can be assumed to be insignificant, then a fractional design involving less than 2^k runs can be used to obtain estimates of the main effects and lower order interactions. A design that uses 2^{k-1} runs is called a one-half fraction of the 2^k design. Table B3 shows two possible setups for a half fraction of the 2^3 design. In the top half of Table B3, only those runs that have a + sign on the ABC effect are used. ABC is therefore known as the generator of this fraction. It is also noted that $I = ABC$ for the upper half of the table and this is called the defining relation for the design [12:515].

By examination of the top half of Table B3, one can see that $A = BC$, $B = AC$, and $C = AB$. In this case then, A and BC, B and AC, and C and AB are aliases. This is a consequence of using a fractional design. For this design then, by assuming the second order interactions are insignificant, the main effects are estimated using only four runs

Table B3. A 2^{k-1}, $k = 3$, Design

	Factorial Effect							
Run	I	A	B	C	AB	AC	BC	ABC
2	+	+	-	-	-	-	+	+
3	+	-	+	-	-	+	-	+
5	+	-	-	+	+	-	-	+
8	+	+	+	+	+	+	+	+
4	+	+	+	-	+	-	-	-
6	+	+	-	+	-	+	-	-
7	+	-	+	+	-	-	+	-
1	+	-	-	-	+	+	+	-

vice eight runs. Using this concept, it is possible to select a fractional design so main effects and lower order interactions are aliased with higher order interactions that are assumed to be insignificant [11:516]. Aliases can be determined by using the defining relation. Multiplying any effect by the defining relation yields the aliases for that effect. For example, the alias of B is found by

$$B = B \cdot I = B \cdot ABC = AB^2C = AC \tag{B6}$$

since any column multiplied by itself equals I, (e.g., $BB = B^2 = I$). If the bottom half of Table B3 was selected for the half fraction, the defining relation would be $I = -ABC$ and the aliases are $A = -BC$, $B = -AC$, and $C = -AB$.

Any easy way to construct a half fraction design is to write down the conventional signs for the main factors of a 2^{k-1} design and then equate the signs of the kth factor to the product of the signs of the preceding $k-1$ factors [11:517].

After conducting a half fraction experiment, perhaps one or more of the factors can be eliminated as not being important. If this is the case, the design will become a full factorial design. For example, say one constructed a half fraction design with $k = 4$, or a 2^{4-1} design, and determined one of the factors was not important. Then the design will "project" into a full factorial 2^3 design for any of the remaining three factors [11:519]. Hence, the suggestion by Box to use preliminary two-level fractional factorial <u>screening</u> design to identify important variables. Montgomery says that "this projection property is highly useful in factor screening as it allows negligible factors to be eliminated, resulting in a stronger experiment in the active factors that remain" [11:519].

Design Resolution. Design resolution is a term that describes the amount of aliasing present in a particular design. The most important of these are designs of resolution III, IV, and V. In resolution III designs, no main effects are aliased with any other effects, but main effects are aliased with second order interactions. In resolution IV designs, no main effects are aliased with any other main effects or second order interactions, but second order interactions are aliased with each other. In resolution V designs, no main effects or second order interactions are aliased with any other main effects or second order interactions, but second order interactions are aliased with third order interactions. By convention, Roman numeral subscripts are used to identify design resolution (e.g., 2_V^{5-1} is a half fraction of 2^5 with a design resolution of V) [11:520]. Designs of resolution III and IV are commonly used for screening experiments.

In addition to half fraction designs, designs involving even less runs are possible. Such designs are 2^{k-p} fractional factorial designs, where $1/2^p$ is the degree of fractionation. These designs are discussed in detail Montgomery [11:521]. Montgomery also discusses the selection of design generators to produce "more attractive alias structures" [11:521]. The idea is to produce the highest resolution possible for a given number of factors and runs. Tables are readily available that summarize these designs [11:522, 3:164].

Lack of Fit Analysis. The designs presented above are based on fitting first degree linear regression models to the data. Replication of runs was not required. Hence, a formal methodology for determining model adequacy is absent. However, by conducting more than one run at set values of the input variables, such a determination is possible.

In fitting a regression model, pure error sum of squares, SS_P, is always part of the error sum of squares, SS_E [12:72]. The remaining part being lack of fit sum of squares, SS_L. That is [12:72]

$$SS_E = SS_P + SS_L. \qquad (B7)$$

135

Box presents appropriate equations and replication requirements to conduct a formal lack of fit test. In summary, these techniques lead to the computation of an F ratio, $F = MS_L / MS_E$, where MS_L is equal to SS_L divided by its appropriate degrees of freedom. Hence, large values of F will discredit the fitted model [12:74]. In this case, Box suggests analyzing the residuals and taking remedial measures, such as transforming the input and/or output system response variables [12:268]. By transforming variables (e.g., $y = \log Y$), it is possible that a first degree model will still be adequate. If this does not prove successful, then the use of a higher order regression model (e.g., a second order or quadratic approximating function) may be necessary. Also, if the region of interest in an experiment involves "a turning point," then a model of at least second degree is required [12:305].

Central Composite Designs. A Central Composite Design (CCD) allows estimation of second order models. These designs incorporate two level factorial or fractional factorial designs that are augmented with additional points which allow estimation of purely quadratic effects [12:306]. CCDs can be subdivided into three portions -- a cube portion, a center portion, and a star portion. The cube portion consists of the high and low levels of all the factors as previously discussed (e.g., 2^k or 2^{k-p} designs). The center potion has a coded value of zero for all factors. And the star portion consists of $2k$ number of points where each of these points have a $\pm\alpha$ coded value for one factor and the other factors are set to a coded value of zero. [12:306]. Box presents a strategy for selecting the number of center points and the value of α so that the resulting CCD is both orthogonal and rotatable, which are both desirable properties for subsequent response surface characterization.

For a first order model, an orthogonal design ensures that the regression coefficients will be uncorrelated with each other [12:479]. In a second order design, orthogonality ensures that all

136

first and second order regression coefficient will be uncorrelated with each other [12:483]. A design is rotatable if the variance of the fitted response function at some design point is only a function of the distance from the center of the design to that point and not its direction from the center [12:440]. It is not always possible to achieve designs that are both orthogonal and rotatable. Common practice then is to select the number of center points to obtain an orthogonal design and then choose α to make the design as rotatable as possible [38].

As with first order models, formal test procedures exist to assess the adequacy of second order models. This comes down to further partitioning lack of fit sum of squares, SS_L, into sum of squares due to quadratic terms, SS_Q, and sum of squares due to other terms, SS_O [38]. We have

$$SS_L = SS_Q + SS_O. \qquad \text{(B8)}$$

One then computes of an F ratio, $F = MS_Q / MS_E$, where MS_Q is equal to SS_Q divided by its appropriate degrees of freedom. Hence, large values of F indicate second order terms are significant and should be included in the model.

If a second order model is appropriate, examination of the fitted response surface to identify optimal conditions is the next aspect of RSM. One approach is to use calculus based methods to identify stationary points, if they exist, via partial derivatives with respect to factors in the fitted model [38]. Another approach is canonical analysis. In this approach the coordinate axes are rotated to remove cross product terms (i.e., interaction factors) from the second order model. Also, when appropriate, the coordinate axes may also be translated to coincide with stationary points. The resulting model is then in <u>canonical form</u>. The canonical form allows identification of local optima and description/examination of the response surface [38]. RSM contains a host

of methodologies for investigating response surfaces, including the identification and elucidation of ridge systems [38].

Variance as a System Response Variable

In this section, RSM and DOE methods that treat the variance of some system response as the variable of interest are presented, including the Taguchi method.

Consider a hypothetical RSM problem where the system response is a function of two input variables that each have two levels coded at -1 and +1. Now, say we desire to maximize the mean system response as a function of the coded input variables, $x1$ and $x2$. Lets also say conduct three replicates of this design with results as shown in Figure B1. In this figure, we see that design point (a) has a high mean value and medium variance of the system response. Design point (b) has a slightly higher mean value than design point (a), but a high variance of the system response. Design point (c) has a low mean value and low variance of the system response. and design point (d) has a low mean value and medium variance of the system response. Hence, since we desire to maximize the mean system response as a function of the coded input variables, we might choose design point (b). However, since design point (a) has almost the same mean response as (b), but lower variance, this might be a better design point to operate at.

Now let us assume we want to simply operate our system so that the output has the lowest possible variance. Under this concept, design point (c) would be considered superior to all of the other design points shown in Figure B1. Of course, as was mentioned earlier, it might often be the case where we wish to maximize the mean response of the system while minimizing the variance of the response. This notion is explored further in the reviews that follow.

In the text Statistical Design and Analysis of Experiments Mason, Gunst, and Hess discuss statistical design techniques in regard to quality improvement studies [39]. In such studies the

138
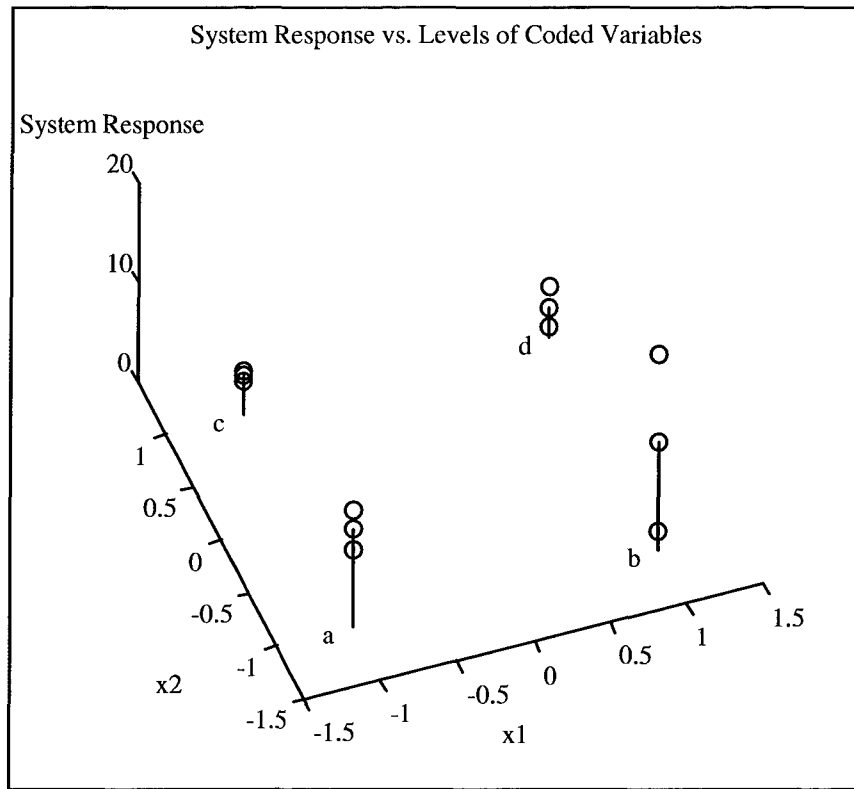
Figure B1. Sample of System Response at Different Levels of Coded Variables

response of interest includes both measures of product or process location (mean) and measures of dispersion (variance). Their text concentrates on analysis of location, but they state that measures of dispersion can also be used as response variables. In this case, several observations are made at the same levels of the factors and some measure of dispersion, such as the standard deviation of these responses is computed and used as the response variable for the particular factor level combinations of each design point. Mason et al state that when dispersion is of interest, that often the log of the standard deviation is used for each design. They also say that an equal number of replications, r, is required at each design point with $r > 4$ [39:109].

Sometimes both measures of location and dispersion are of interest together. Here they say that the response functions for the two measures can be overlaid graphically with contour plots to explore the tradeoffs in achieving optimal levels of the design factors.

Davidian and Carroll present a paper entitled *Variance Function Estimation* [40] in which they develop a general theory for variance function estimation. Their primary thesis is that the method used to obtain variance functions is important despite the fact that asymptotic theory implies otherwise. As a lead in, they state that "most variance function estimation procedures can be looked upon as regressions with 'responses' being transformations of absolute residuals from a preliminary fit or sample standard deviations from replicates at a design point." They feel that the second method is more efficient and propose that it is robust to "misspecification" of the fitted mean response model. This is because the second method does not use the mean function. They do point out the need for the number of replications to be at least two at each design point though for this method to be carried out.

In their paper entitled *Analyzing Dispersion Effects From Replicated Factorial Experiments*, Nair and Pregibon discuss structured modeling of variances and estimation of dispersion effects from two-level factorial experiments with replications [41]. From the context in which it is used, dispersion effects refer to the regression coefficients of the fitted variance model. The authors present a method to estimate dispersion effects. First they compute the sum of squared deviations for each design point as

$$Z_i = \sum_{j=1}^{J} (Y_{ij} - \mu_i)^2, \quad i = 1, \ldots, N \tag{B9}$$

where N represents each of the distinct points within the design, J is the number of replications, and μ_i will most often be the within-replication mean response. They then present a method

140

proposed by Bartlett and Kendall [42] that uses the log of the Z_i's to estimate the dispersion effects, γ_k, as

$$\hat{\gamma}_k = N^{-1} \left[\sum_{S(k+)} \log Z_i - \sum_{S(k-)} \log Z_i \right], \quad k = 2, \ldots, N \tag{B10}$$

where $S(k+)$ indicates +1 values in the respective factor column and $S(k-)$ indicates -1 values in the respective factor column. The authors say that the constant term γ_1 is a nuisance parameter so k is seen to go from 2 to N. They go to say that this form is the maximum likelihood estimator of the dispersion effects.

The authors also present another method to estimate dispersion effects by Box and Meyer [43]. The method can be used for unreplicated experiments assuming "effect sparsity" which says only a small proportion of the effects are important in relation to noise. This method is based on pooling squared deviations across high and low levels of each factor. Nair and Pregibon extend this methodology to the case with replications that results in the following form

$$\hat{\gamma}_k = \frac{1}{2} \left[\log \sum_{S(k+)} Z_i - \log \sum_{S(k-)} Z_i \right], \quad k = 2, \ldots, N \tag{B11}$$

where the Z_i's are the same as defined earlier. Nair and Pregibon also use this form in their example and compare results obtained with Equation (B11) to that obtained using Equation (B10). They state that the two methods are positively correlated with each other and agree. They state the second method is more efficient though when the stated assumptions hold.

141

Finally, in reference to their example, Nair and Pregibon state that "there is no apparent dependence of variability on mean level, so transformation of the data is not required. This also suggest that it is preferable to analyze the within-replication variances, Z_i's, rather than Taguchi's (1986) signal-to-noise (S-N) ratios" [44].

In their paper *Experimental Designs for Estimating Both Mean and Variance Functions*, Vining and Schaub discuss how statisticians have recently become concerned with both the expected value of the response of interest as well as its variability [45]. As such, they examine two experimental methods that will allow simultaneous estimation of both of these functions. The first method assumes that variance is constant over the region of interest. The second method involves a Bayesian approach that uses prior information about the variance function of interest.

The authors refer to the first method as the One-Step Approach. In this approach, they assume a second-order model for the response and a first-order model for the variance. They then state that an appropriate experimental strategy replicates the first-order portion (i.e., factorial points) of a second-order design, such as the CCD design that was discussed earlier. They present other modified forms of this that could also be used. For each of these schemes, they provide a comparison of the total number of runs required as a function of the number design replicates for three, four, and five factor designs. In the second approach, the authors present a Bayesian approach to generating designs that contain necessary replication for variance computations. They discuss the essentials of the theory needed to generate such designs.

Lucas presents a paper entitled *How to Achieve a Robust Process Using Response Surface Methodology* [46]. He proposes a new class of composite designs having mixed resolution as an alternative to Taguchi designs. He also provides an example of where it is <u>better not to obtain a robust process</u> in terms of customer satisfaction. Lucas starts off by reviewing RSM and

142

screening designs. He then states that Taguchi designs are "either direct applications of screening designs or are obtained by 'crossing' designs." He presents a three-level screening design with nine design points that can be used to examine up to four factors, a design that Taguchi calls L_9. Lucas also makes another point in regard to screening designs that is worth repeating here, he says

> *I am a believer in the 'main effects principle.' This is the empirical observation that linear main effects are more important than higher-order effects. This observation is supported by the large number of times that main effects dominate in a quadratic response surface experiment, even thought the fact that the quadratic response surface experiment was run indicates the experimenter felt curvature and interactions would be important. The main effects principle is one justification for using screening experiments.*

Lucas moves on to discuss quadratic response surface designs that are often used in "process optimization and improvement." He discusses how the results of such an analysis are presented as contour plots of the resulting response surface, providing some of the most important results from the experiment.

Next, Lucas discusses basic aspects of Taguchi designs. These will be discussed more later, but for now he points out that although this approach is simple, "it may not extract all the information in the data, and it may not find optimum parameter settings." He gives an example where he constructs a Taguchi design and then fits traditional regression models to the resulting data. He states that doing so enables one "to use all of the diagnostic procedures that have been developed for response surfaces."

In the next section, <u>Doing Better Than a Robust Process</u>, he discusses how Taguchi analysis ignores main effects of the outer array or noise variables. Referring back to the main effects principle, these factors might be as or more important than interactions between controllable

variables. With this knowledge, Lucas discusses how large effects of outer array variables can be used to supply products to customers who have different needs. This can be accomplished through product differentiation as a result of different system outputs under different occurrences of the values of the outer variables.

Finally, Lucas presents composite designs that can be used instead of Taguchi crossed arrays. These designs generally require less experimental runs that Taguchi designs, yet they still allow estimation of interaction effects between inner array variables. This is in contrast to Taguchi designs that only estimate interactions between inner and outer array variables.

In the paper *Dispersion Effects in Robust-Design Experiments with Noise Factors*, Steinberg and Bursztyn [47] concentrate on analyzing the data obtained from robust design experiments that include both design factors (controllable variables) and noise factors (uncontrollable variables). Their main purpose is to compare Taguchi type row analysis that calls for summarizing signal-to-noise ratios with that of direct modeling of noise factors as was discussed in the previous review. They state that the two methods can lead to "completely different conclusions about which factors affect variability."

In one example, as result of interactions, the authors show how the latter model can allow identification of settings of otherwise unimportant main effects to be judiciously chosen to control output variation. They also show how row analysis can lead to erroneous process conclusions. Steinberg and Bursztyn state that when noise factors are included in an experiment, it is important to model their effects explicitly in the subsequent analysis. Such full array analysis provides valuable information about which noise factors are really contributing to process variability.

As seen, many of the above sources offer more efficient experimental or are more accurate analysis strategies than Taguchi's. Nonetheless, many of his concepts are seen as very important

considerations in process improvement and optimization. Hence, a closer look at Taguchi's approach is worthwhile.

Taguchi Method. There are three major themes to Taguchi's philosophy. These are a) products and processes should be designed so that they are "robust" to external sources of variability, b) experimental design is a tool to help accomplish this objective, and c) operation on target is more important than conformance to specifications [11:533].

Whereas most engineering processes typically specify allowable tolerances for acceptable product outputs, Taguchi puts forth the notion that any deviation from the desired product target level results in a loss. He describes a loss function that relates the deviation from target to a loss incurred by society, which in most cases implies both the product producer and the consumer. Specifically, the loss function has a quadratic form

$$L(y) = k(y - T)^2 \qquad (B12)$$

where k is some constant with appropriate units (e.g., dollars per response unit), y is the system output or some measure of the product, and T is the target level. Hence, any deviation from the target implies a loss.

Following Taguchi's methodology, one experimental design would be established for the controllable factors and another design for the uncontrollable or noise factors. Taguchi has designed "orthogonal arrays," which means that factors can be evaluated independently of one another, that is the effect of one factor does not interfere with the estimation of the effect of another factor [48]. Typically, these orthogonal arrays call for the controllable factors to each be run at three levels (low, medium, and high) and uncontrollable factors to each be run at two levels (low and high). The arrays for the controllable and uncontrollable factors are called the

145

inner array and outer array respectively. The respective designs are then combined into a single parameter design layout where permutations of the inner array are run across all permutations of the outer array [11:534]. This rapidly increases the number of runs that need to be conducted. For instance, a typical experiment involving four controllable variables and three uncontrollable variables would require 72 runs!

After conducting these runs, the next step would be to compute the mean response and variability for each permutation of runs in the inner array. These data provide information on how the system performs under many different occurrences of noise. However, in Taguchi's methodology, the goal is not necessarily to choose the settings that yield the best mean response, $\bar{y}$. Rather, the desire is to choose settings that provide a satisfactory mean response with low variability. To do this, Taguchi suggests using a Signal-to-Noise Ratio (SNR) that is derived from the quadratic loss function [11:534]. Three types are commonly used. These are nominal is best (SNR_N), larger is better (SNR_L), and smaller is better (SNR_S). In equation form [11:534]:

$$SNR_N = 10 \log \left(\bar{y}/S^2 \right) \tag{B13}$$

$$SNR_L = -10 \log \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{y_i^2} \right) \tag{B14}$$

$$SNR_S = -10 \log \left(\frac{1}{n} \sum_{i=1}^{n} y_i^2 \right) \tag{B15}$$

where the y_i's are the different observations for each permutation of runs or row in the inner array and S^2 is the sample based variance of these observations. By inspection, these forms result in SNR values that are expressed in decibels.

From Montgomery, the factor levels that maximize the appropriate SNR are "optimal" [11:536]. In short, each permutation of factor levels from the inner array would be run for every combination of noise levels in the outer array. Then the system output metric would be computed for each run. Using these, the SNR would be calculated for each permutation of factor levels. The factor level permutation yielding the highest SNR would be selected as the best. In this manner, Taguchi proposes that these settings will be the least sensitive to uncontrollable factors that influence the system output.

Opponents of Taguchi's methodology disapprove of recommended design layouts because information regarding interactions is lost due to aliasing. This is due to the three level fractional factorials that have "very messy alias structures" [11:538]. As mentioned earlier, another issue is that the design structure can lead to a very large number of experimental runs. According to Montgomery, several alternative experimental designs would be superior to the inner and outer array setup of Taguchi [11:538]. As such, he proposes running all variables at two levels in a single design array at a design resolution of IV for example. This can cut the number of runs in half or more. He also suggests different assignments of controllable and uncontrollable variables that will allow different effects to be estimated without aliasing.

Next, picking the best level for each factor based on plots of average responses for that factor alone can lead to less than optimal conditions, especially if interactions are ignored. This begs the need for more formal response surface methods [11:541].

Finally, care must be given in using the SNR forms presented, the main problem being that the mean (location) and variability (dispersion) of the system response are confounded with each

other. For instance, perhaps the SNR might be driven up because the variability is decreasing

faster than the mean, yet the mean is approaching unacceptable values. A recommended

alternative is to isolate location and dispersion as separate system responses [11:543]. For

instance, the SNR_N form, Equation (B13), can be rewritten as follows

$$SNR_N = 10\log(\bar{y}) - 10\log(S^2) \qquad\qquad \text{(B16)}$$

Now, if the output is kept on target, we can maximize the SNR by minimizing the variance.

Whereas, using Equation (B13), it could be unclear if the SNR is increasing because the variance

is decreasing or the output is no longer on target.

Montgomery demonstrates how confounding is present in the other forms of SNR proposed

by Taguchi. Based on this, Montgomery recommends that these SNR forms not be used. Rather,

he proposes a better approach is to isolate location and dispersion effects and to analyze them as

separate response variables. This takes us back to many of the preceding articles. He mentions

that the log of the standard deviation, $\log(S)$, is a common response variable for analyzing

dispersion effects.

This appendix contains a more detailed documentation of the RSM/DOE analysis that was summarized in Chapter IV. The six initial factors that were examined in our model of the AMOS 1.6 meter adaptive optics system were shown in Figure 1 of Chapter IV. Table C1 below shows the factor settings for each of these factors and the corresponding design matrix. This table shows that the simulation runs were conducted in two blocks. It is pointed out that this was only done for convenience in organizing these runs and that there is no potential for confounding between the blocks since the simulated system output is purely dependent on factor settings and the input seed. In keeping with the methodology presented in Appendix B, the factors were coded to -1 for the low value of each factor and to +1 for the high value of each factor. Using the units provided earlier in Figure 17, the specific coding scheme for each factor is provided in Equations (C1) through (C6).

$$A = \frac{s - 750}{250}, \quad 500 \le s \le 1000 \tag{C1}$$

$$B = \frac{k_g - 0.471}{0.157}, \quad 0.314 \le k_g \le 0.628 \tag{C2}$$

$$C = \frac{t_s - 0.011}{0.009}, \quad 0.002 \le t_s \le 0.02 \tag{C3}$$

$$D = \frac{m_v - 7.5}{1.5}, \quad 6 \le m_v \le 9 \tag{C4}$$

$$E = \frac{r_0 - 0.15}{0.05}, \quad 0.1 \le r_0 \le 0.2 \tag{C5}$$

$$F = \frac{v - 18.75}{6.25}, \quad 12.5 \le v \le 25.0 \tag{C6}$$

Using the simulation data for the shown design matrix a normal probability plot of main, second order interactions, and third order interactions effects was constructed as shown in Figure C1. The most significant effects are labeled in this plot. Using this information, a regression model was fit to the data as summarized in Table C2. Table C3 provides a summary of the lack of fit analysis that was accomplished using the above model. The null hypothesis -- the fitted model is adequate -- requires the computed F statistic for lack of fit to be less than the critical value of $F_{1-\alpha,23,65}$. For $\alpha = 0.05$, the critical value is 1.70. Hence, the null hypothesis that the model is adequate is rejected, which implies the need for fitting a second order model.

Based on the above result, axial points were added to the original design matrix and additional simulation runs were made. This regression model is summarized in Table C4. Based on predictor variables having a P-value of less than 0.05, significant second order terms were BB and DD. The fitted regression model yields a coefficient of multiple correlation of 0.991. A large F value of about 668 also supports the regression model since its P-value is near zero.

Table C1. Factor Settings and Design Matrix for 2^{6-1} Resolution VI AMOS Model. These are the uncoded factor settings.

	SET4a: 2^{6-1}, Res VI Design - High Gain Block					
Run	A = sample rate (Hz)	B = gain	C = frame exposure time (s)	D = object visual magnitude	E = Fried parameter (m)	F = wind speed (m/s)
1	500	0.628	0.002	6.0	0.1	12.5
2	1000	0.628	0.002	6.0	0.1	25.0
3	500	0.628	0.02	6.0	0.1	25.0
4	1000	0.628	0.02	6.0	0.1	12.5
5	500	0.628	0.002	9.0	0.1	25.0
6	1000	0.628	0.002	9.0	0.1	12.5
7	500	0.628	0.02	9.0	0.1	12.5
8	1000	0.628	0.02	9.0	0.1	25.0
9	500	0.628	0.002	6.0	0.2	25.0
10	1000	0.628	0.002	6.0	0.2	12.5
11	500	0.628	0.02	6.0	0.2	12.5
12	1000	0.628	0.02	6.0	0.2	25.0
13	500	0.628	0.002	9.0	0.2	12.5
14	1000	0.628	0.002	9.0	0.2	25.0
15	500	0.628	0.02	9.0	0.2	25.0
16	1000	0.628	0.02	9.0	0.2	12.5

|	SET4b: 2^{6-1}, Res VI Design - Low Gain Block					
Run	A = sample rate (Hz)	B = gain	C = frame exposure time (s)	D = object visual magnitude	E = Fried parameter (m)	F = wind speed (m/s)
1	500	0.314	0.002	6.0	0.1	25.0
2	1000	0.314	0.002	6.0	0.1	12.5
3	500	0.314	0.02	6.0	0.1	12.5
4	1000	0.314	0.02	6.0	0.1	25.0
5	500	0.314	0.002	9.0	0.1	12.5
6	1000	0.314	0.002	9.0	0.1	25.0
7	500	0.314	0.02	9.0	0.1	25.0
8	1000	0.314	0.02	9.0	0.1	12.5
9	500	0.314	0.002	6.0	0.2	12.5
10	1000	0.314	0.002	6.0	0.2	25.0
11	500	0.314	0.02	6.0	0.2	25.0
12	1000	0.314	0.02	6.0	0.2	12.5
13	500	0.314	0.002	9.0	0.2	25.0
14	1000	0.314	0.002	9.0	0.2	12.5
15	500	0.314	0.02	9.0	0.2	12.5
16	1000	0.314	0.02	9.0	0.2	25.0
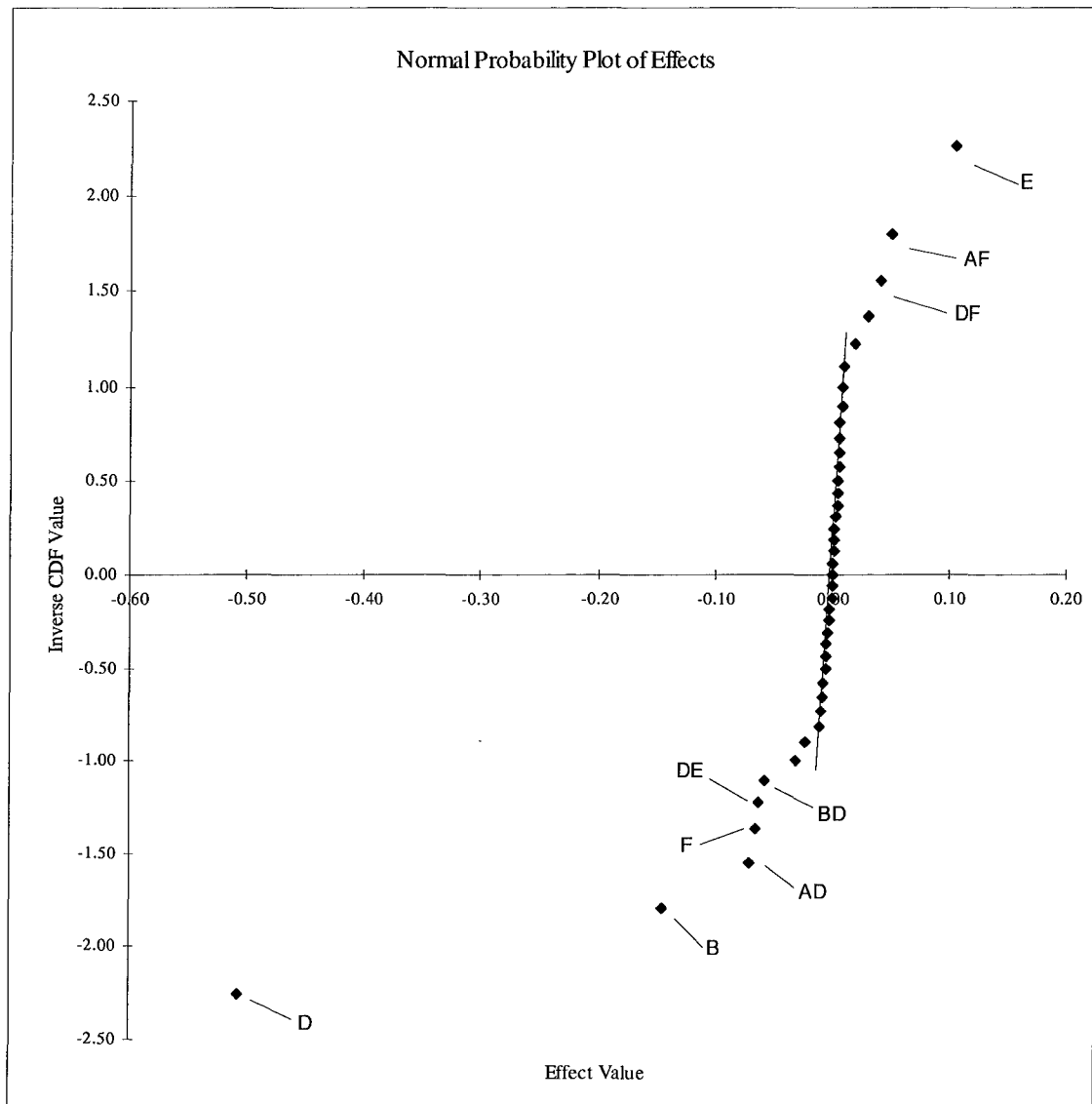
Figure C1. Normal Probability Plot of Effects of Factors Considered in RSM/DOE Analysis. This plot is based on the coded values of the factors.

Table C2. Linear Regression of *MSR*2 on Significant Effects with Two Center Point
Replications

```
┌─────────────────────────────────────────────────────────────────────────────┐
│           UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF MSR2                   │
│                                                                               │
│ PREDICTOR                                                                     │
│ VARIABLES      COEFFICIENT   STD ERROR     STUDENT'S T       P        VIF      │
│ ---------      -----------   ---------     -----------    ------     -----     │
│ CONSTANT         0.27951      0.00232        120.69       0.0000               │
│ D               -0.16875      0.00234        -72.12       0.0000      1.0      │
│ B               -0.04810      0.00234        -20.56       0.0000      1.0      │
│ AD              -0.02140      0.00234         -9.14       0.0000      1.0      │
│ BD              -0.01917      0.00234         -8.19       0.0000      1.0      │
│ F               -0.02352      0.00234        -10.05       0.0000      1.0      │
│ DE              -0.02065      0.00234         -8.82       0.0000      1.0      │
│ DF               0.01325      0.00234          5.66       0.0000      1.0      │
│ AF               0.01663      0.00234          7.11       0.0000      1.0      │
│ E                0.03462      0.00234         14.80       0.0000      1.0      │
│                                                                               │
│ R-SQUARED             0.9861     RESID. MEAN SQUARE (MSE)   5.256E-04          │
│ ADJUSTED R-SQUARED    0.9847     STANDARD DEVIATION             0.02293        │
│                                                                               │
│ SOURCE         DF      SS           MS           F         P                   │
│ ----------     ---   ----------   ----------   -----     ------                │
│ REGRESSION      9     3.28762       0.36529     695.00    0.0000               │
│ RESIDUAL        88    0.04625      5.256E-04                                   │
│ TOTAL           97    3.33387                                                  │
│                                                                               │
│ CASES INCLUDED 98    MISSING CASES 0                                          │
└─────────────────────────────────────────────────────────────────────────────┘
```

Table C3. Lack of Fit Analysis for First Order Model With Center Points Added

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F Statistic
Regression	3.28762	9	0.36529	695.04
Residual	0.04625	88	5.25568E-04	
Lack of Fit	0.03191	23	1.38739E-03	6.29
Pure Error	0.01434	65	2.20615E-04	
Total	3.33387	97		

153

Table C4. Linear Regression of *MSR*2 on Significant Effects, Including Second Order Terms BB and DD

```
                   UNWEIGHTED LEAST SQUARES LINEAR REGRESSION OF MSR2

PREDICTOR
VARIABLES     COEFFICIENT    STD ERROR      STUDENT'S T        P         VIF
---------     -----------    ---------      -----------      ------      -----
CONSTANT        0.32587      0.00369          88.20         0.0000
B              -0.04866      0.00219         -22.19         0.0000        1.0
D              -0.17016      0.00219         -77.62         0.0000        1.0
E               0.03463      0.00219          15.80         0.0000        1.0
F              -0.02405      0.00206         -11.66         0.0000        1.0
AD             -0.02140      0.00226          -9.47         0.0000        1.0
AF              0.01671      0.00220           7.61         0.0000        1.0
BD             -0.01917      0.00226          -8.48         0.0000        1.0
DE             -0.02065      0.00226          -9.14         0.0000        1.0
DF              0.01325      0.00226           5.86         0.0000        1.0
BB             -0.02364      0.00676          -3.50         0.0006        2.6
DD             -0.02381      0.00676          -3.52         0.0006        2.6

R-SQUARED             0.9827      RESID. MEAN SQUARE (MSE)   4.902E-04
ADJUSTED R-SQUARED    0.9813      STANDARD DEVIATION              0.02214

SOURCE         DF       SS           MS           F         P
----------     ---   ----------   ----------    -----    ------
REGRESSION      11    3.60128       0.32739     667.84    0.0000
RESIDUAL       129    0.06324      4.902E-04
TOTAL          140    3.66452

CASES INCLUDED 141    MISSING
```

154

Bibliography

1. King, Henry C. The History of the Telescope. New York: Dover Publications, Inc., 1955.

2. Gardner, Chester S., Byron M. Welsh, and Laird A. Thompson. "Design and Performance Analysis of Adaptive Optical Telescopes Using Laser Guide Stars," Proceedings of the IEEE, 78-11: 1721-1743 (November 1990).

3. Tyson, Robert K. Principles of Adaptive Optics. Boston: Academic Press, Inc., 1955.

4. About the NASA Kuiper Airborne Observatory, NASA Ames Fact Sheet, http://airsci-www.arc.nasa.gov/kao.

5. About the Air Force Maui Optical Station, Phillips Laboratory Fact Sheet, http://snickers.mhpcc.edu.

6. About the Starfire Optical Range, Phillips Laboratory Fact Sheet, http://plk.af.mil.

7. Wild, Walter J. and Robert Q. Fugate. "Untwinkling the Stars - Part II," Sky and Telescope: 20-27 (June 1994).

8. Fugate, Robert Q. and Walter J. Wild. "Untwinkling the Stars - Part I," Sky and Telescope: 24-31 (May 1994).

9. Roggemann, Michael C. and Byron M. Welsh. Imaging Through Turbulence. Boca Raton FL: CRC Press, 1996.

10. Roggemann, Michael C. and C.L. Matson. "Power Spectrum and Fourier Phase Spectrum Estimation by Using Fully and Partially Compensating Adaptive Optics and Bispectrum Postprocessing," Journal of the Optical Society of America, 9: 1525-1535 (1992).

11. Montgomery, Douglas C. Introduction to Statistical Quality Control (Second Edition). New York: John Wiley & Sons, 1991.

12. Box, George E. and Norman R. Draper. Empirical Model-Building and Response Surfaces. New York: John Wiley & Sons, 1987.

13. Holst, Gerald C. Electro-Optical Imaging System Performance. Bellingham WA: SPIE Optical Engineering Press, 1995.

14. Hecht, Eugene. Schaum's Outline Series - Theory and Problems of Optics. New York: McGraw-Hill, Inc., 1975.

15. Goodman, Joseph W. Introduction to Fourier Optics. New York: McGraw-Hill, Inc., 1968.

16. Muller, Richard A. and Andrew Buffington. "Real-time Correction of Atmospherically Degraded Telescope Images Through Image Sharpening," <u>Journal of the Optical Society of America, 64-9</u>: 1200-1210 (September 1974).

17. Al-Khafaji, Amir W. and John R. Tooley. <u>Numerical Methods in Engineering Practice</u>. New York: Holt, Rinehart, and Winston, Inc., 1986.

18. Lee, David J. <u>Objective Image Quality Metrics: Applications for Partially Compensated Images of Space Objects</u>. MS Thesis, AFIT/GSO/ENP/93D-03. School of Engineering, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, December 1993.

19. Lewis, Rhys. <u>Practical Digital Image Processing</u>. London: Ellis Horwood Limited, 1990.

20. Roggemann, Michael C., Craig A. Stoudt and Byron M. Welsh. "Image-spectrum Signal-to-Noise-Ratio Improvements by Statistical Frame Selection for Adaptive-Optics Imaging Through Atmospheric Turbulence," <u>Optical Engineering, 33-10</u>: 3254-3264 (October 1994).

21. Fried, D.L. "Probability of Getting a Lucky Short-Exposure Image Through Turbulence," <u>Journal of the Optical Society of America, 68</u>: 1651-1658 (1978).

22. Gardner, Chester S., Byron M. Welsh, and Laird A. Thompson. "Design and Performance Analysis of Adaptive Optical Telescopes Using Laser Guide Stars," <u>Proceedings of the IEEE, 78-11</u>: 1721-1743 (November 1990).

23. Ellerbroek, Brent L. "First-Order Performance Evaluation of Adaptive-Optics Systems for Atmospheric-Turbulence Compensation in Extended-field-of-View Astronomical Telescopes," <u>Journal of the Optical Society of America A, 11-2</u>: 783-805 (1994).

24. Gavel, D.T., J.R. Morris, and R.G. Vernon. "Systematic Design and Analysis of Laser-Guide-Star Adaptive-Optics Systems for Large Telescopes," <u>Journal of the Optical Society of America A, 11-2</u>: 914-924 (1994).

25. Max, Claire E. and others. "Issues in the Design and Optimization of Adaptive Optics and Laser Guide Stars for the Keck Telescopes," <u>SPIE Proceedings Vol. 2201</u>: 189-200 (1994).

26. Tyler, David W. and Janet S. Fender "Optimal Wavelength Selection for Adaptive Optics Telescopes," <u>SPIE Proceedings Vol. 2201</u>: 227-238 (1994).

27. Dayton, David C, Steven C. Sandven, and John D. Gonglewski. "Performance Simulation and Experimental Results of Low-Order Adaptive Optics Systems in Conjunction With Computer Postprocessing," <u>SPIE Proceedings Vol. 2201</u>: 260-264 (1994).

28. Idell, P.S. and A. Webster. "Resolution Limits for Coherent Optical Imaging: Signal-to-Noise Analysis in the Spatial-Frequency Domain, " <u>Journal of the Optical Society of America A, 9</u>: 43-56 (1992).

29. Goodman, Joseph W. <u>Statistical Optics</u>. New York: John Wiley & Sons, 1985.

30. Roggemann, Michael C. and C.L. Matson. "Power Spectrum and Fourier Phase Spectrum Estimation by Using Fully and Partially Compensating Adaptive Optics and Bispectrum Postprocessing," Journal of the Optical Society of America, 9: 1525-1535 (1992).

31. Neter, John, William Wasserman, Michael H. Kutner. Applied Linear Statistical Models, Third Edition. Homewood IL: Irwin, 1990.

32. Harrington, Patrick M. and Byron M. Welsh. "Frequency-domain Analysis of an Adaptive Optical System's Temporal Response," Optical Engineering, 33-7: 2336-2342 (July 1994).

33. Gaffard, J.P. and Corrine Boyer. "Adaptive Optics for Optimization of Image Resolution," Applied Optics, 26-18: 3772-3777 (September 1987).

34. Greenwood, Darryl P. "Bandwidth Specification for Adaptive Optics Systems," Journal of the Optical Society of America 67-3: 390-392 (1977).

35. Ellerbroek, Brent L. "Optimizing Closed-loop Adaptive-Optics Performance with Use of Multiple Control Bandwidths," Journal of the Optical Society of America A, 11-11: 2871-2886 (1994).

36. Lloyd-Hart, M. R. Angel, B. Jacobsen, D. Wittman, D. McCarthy. "Preliminary Closed-loop Results from an Adaptive Optics System Using a Sodium Resonance Guide Star," SPIE Proceedings Vol. 2201: 364-372 (1994).

37. Roggemann, Michael A. Documentation for Simulation Program: OTFSIM3 & HYSIM3. School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH.

38. Kenneth W. Bauer. Class handout, OPER 683, Response Surface Methodology. School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH, Summer 1994.

39. Mason, Robert L., Richard F. Gunst, and James L. Hess. Statistical Design and Analysis of Experiments. New York: John Wiley & Sons, 1989.

40. Davidian, M. and R.J. Carroll. "Variance Function Estimation," Journal of the American Statistical Association, Vol 82, No. 400: 10791091 (December 1987)

41. Nair, Vijayan N. and Daryl Pregibon. "Analyzing Dispersion Effects From Replicated Factorial Experiments," Technometrics, Vol 30, No. 3: 247-257 (August 1988).

42. Bartlett, M.S. and D.G. Kendall. "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation," Journal of the Royal Statistical Society, Ser B, 8: 128-138 (1946).

43. Box, G.E. and R.D. Meyer. "Dispersion Effects From Fractional Designs," Technometrics, Vol 28: 19-27 (1988).

44. Taguchi, G. Introduction to Quality Engineering. Tokyo: Asian Productivity Organization (1986).

45. Vining, G. Geoffrey and Diane Schaub. "Experimental Designs for Estimating Both Mean and Variance Functions," Journal of Quality Technology, Vol 28, No 2: 135-147 (April 1996).

46. Lucas, James M. "How to Achieve a Robust Process Using Response Surface Methodology," Journal of Quality Technology, Vol 26, No 4: 248-259 (October 1994).

47. Steinberg, David M. and Dizza Bursztyn. "Dispersion Effects in Robust-Design Experiments with Noise Factors," Journal of Quality Technology, Vol 26, No 1: 12-21 (January 1994).

48. Ross, Phillip J. Taguchi Techniques for Quality Engineering - Loss Function, Orthogonal Experiments, Parameter and Tolerance Design. New York: McGraw-Hill Book Company, 1988.

Vita

Robert T. Brigantic is a Major in the United States Air Force. ████████████ ████████████ He Graduated from Gresham Union High School in Oregon in 1979. After high school he attended Oregon State University in Corvallis, Oregon where he earned a Bachelor of Science Degree in Chemical Engineering in 1983. He then attended Air Force Officer Training School and was commissioned in September 1983.

Major Brigantic's first assignment was as a Chemical Engineer at the San Antonio Air Logistics Center, Kelly AFB, Texas. In 1986 he attended the Air Force Institute of Technology. He earned a Master of Science Degree in Space Operations and was recognized as a Distinguished Graduate in 1987. He was then assigned to the Consolidated Space Test Center, Onizuka AFB, California where he served as the Chief of the Engineering Division for Shuttle and IUS Operations. Major Brigantic also served as Operations Director for realtime support of operational Space Shuttle missions. Later he was selected to be the Executive Officer and the Squadron Section. In 1994 he entered the School of Engineering, Air Force Institute of Technology. Major Brigantic is a life member of the national engineering honor society - Tau Beta Pi, the Society for Imaging Science and Technology, the National Space Society, and the Institute for Operations Research and the Management Sciences.

████████████████████████████████

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1997	Doctoral Dissertation

4. TITLE AND SUBTITLE

OPTIMIZATION CONSIDERATIONS FOR
ADAPTIVE OPTICS DIGITAL IMAGERY SYSTEMS

5. FUNDING NUMBERS

6. AUTHOR(S)

Robert T. Brigantic, Major, USAF

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

Air Force Institute of Technology
2750 P Street
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT/DS/ENS/97-02

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

N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This dissertation had three objectives. The first objective was to develop image quality metrics that characterize Adaptive Optics System (AOS) performance. The second objective was to delineate input settings that maximize AOS performance. The third objective was to identify and characterize trade-offs between fully and partially compensated adaptive.

For the first objective, three candidate image quality metrics were considered: the Strehl ratio, a novel metric that modifies the Strehl ratio by integrating the modulus of the average system optical transfer function to a "noise-effective-cutoff" frequency at which some specified image spectrum signal-to-noise-ratio level is attained, and the noise-effective-cutoff frequency. It was shown that these metrics are correlated with the Root-Mean-Square Error between the detected image and the associated diffraction limited image and that they have traits that make them desirable for AOS performance metrics. For the second objective, optimum closed loop bandwidth settings were determined as a function of target object light levels and atmospheric seeing conditions. A strategy for selecting the closed loop bandwidth to provide robust system performance was also developed. For the third research objective, a qualitatively assessment of trade-offs between fully compensated and partially compensated adaptive optics systems was provided.

14. SUBJECT TERMS

Adaptive Optics System, Image Quality Metrics, Closed Loop Bandwidth, Strehl Ratio, Image Spectrum Signal-to-Noise Ratio, Noise-Effective-Cutoff Frequency, System Performance, Object Light Level, Atmospheric Seeing Conditions

15. NUMBER OF PAGES

171

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to *stay within the lines* to meet *optical scanning requirements*.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	-	Contract	**PR** - Project
G	-	Grant	**TA** - Task
PE	-	Program Element	**WU** - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

    DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
    DOE - See authorities.
    NASA - See Handbook NHB 2200.2.
    NTIS - Leave blank.

Block 12b. Distribution Code.

    DOD - Leave blank.
    DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
    NASA - Leave blank.
    NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.