UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Imaging the foveal cone mosaic with a MEMS-based adaptive optics scanning laser ophthalmoscope

Permalink
https://escholarship.org/uc/item/9th5r171

Author
Li, Yiang

Publication Date
2010

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Imaging the Foveal Cone Mosaic with a MEMS-based Adaptive Optics Scanning Laser Ophthalmoscope

By

Yiang Li

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Vision Science

in the

Graduate Divisions

of the

University of California, Berkeley

Committee in charge:

Professor Austin J. Roorda
Professor Masayoshi Tomizuka
Professor Martin S. Banks

Fall 2010

Imaging the Foveal Cone Mosaic with a MEMS-based Adaptive Optics Scanning Laser Ophthalmoscope

Copyright 2010

By

Yiang Li

University of California, Berkeley

Abstract

Imaging the foveal cone mosaic with a MEMS-based adaptive optics scanning laser ophthalmoscope

By

Yiang Li

Doctor of Philosophy in Vision Science

University of California, Berkeley

Professor Austin J. Roorda, Chair

Our knowledge of the structure of the human photoreceptor mosaic is mostly based on histological data. Imaging microscopic structure in intact eyes has traditionally been difficult due to structural imperfections in the eye's optics called aberrations. The introduction of adaptive optics (AO) into vision science has allowed us to access the living human retina at microscopic levels, opening up new possibilities for both basic and clinical research. This dissertation concerns the advancement of AO technology for retinal imaging while emphasizing its application to imaging the foveal cone photoreceptor mosaic in living human eyes. Foveal cones provide a fundamental challenge for today's AO systems due to their small size (2 μm diameter). As a result, much of my effort has been put towards improving AO system performance to resolve these small cells consistently. I have improved the wavefront correction capabilities of an adaptive optics scanning laser ophthalmoscope (AOSLO) using a single MEMS deformable mirror, so that the smallest foveal cones in some eyes can now be resolved. Specifically, many of the nonlinear characteristics of the particular MEMS device used have been negated in the new wavefront controller, and the wavefront reconstructor has been optimized by incorporating measurement noise and aberration (Kolmogorov) statistics. This contribution is significant because, prior to this research, the capability to image the entire foveal cone mosaic *in vivo* had never been demonstrated using this imaging modality.

Some basic scientific investigations were carried in parallel with the technical developments. Specifically, I used this MEMS-based AOSLO to investigate how foveal fixation is related to the cone density distribution and to determine the inter-subject variability of foveal cone density in relation to eye length. The foveae of 18 healthy eyes (18 subjects) with axial lengths from 22.86 mm to 28.31 mm were imaged and analyzed. The entire foveal cone mosaic was resolved in four eyes, but cones within 0.03 mm (≈ 0.1°) from the foveal center remained unresolved in most eyes. The preferred retinal

locus of fixation deviated significantly ($P < 0.001$) from the location of peak cone density for all but one eye. Retinal cone density decreased significantly ($P < 0.05$) with increasing axial length 0.30 mm away from the foveal center but not closer, so we can conclude that the axial myopia progression causes retinal stretch. However, how axial length affects cone density within the central fovea, or foveola, is swamped by other factors besides just cone density due to high levels of inter-subject variability observed there.

I dedicate this work to my family,
my three best friends,
and my teachers

“The one who doesn’t fall never stands up”
The Last Emperor

Table of contents

List of figures

List of tables

List of symbols

Units left blank mean either the quantity is unitless or varies best on the context

	Definition	Common units
j	square root of -1 unless specified otherwise	
n	Index of refraction	
λ	Light wavelength	nm or μm
c	Speed of light (3×10^8)	meters per second
k	Wavenumber ($2\pi/\lambda$)	μm^{-1}
ω	Natural frequency	s^{-1}
r	Radius of curvature	meters
F	Fresnel number	
f	Focal length	μm or mm
f_n	Spatial frequency along the n direction	m^{-1}
d	Diameter	μm or mm
A	Area	mm^2 or m^2
P	Perimeter	μm, mm or m
$T(x, y)$	Voronoi tile	
$d(x, y)$	Cone density	mm^{-2} or $degree^{-2}$
$s(x, y)$	Cone spacing	μm or mm
h	Height	μm or mm
I_j	Intensity count of the j^{th} pixel	
$\psi(x, y)$	Electric field	Vm^{-1}
$circ(x, y)$	Circle function	
$FT\{\bullet\}$	Fourier transformation	
s	Resolution limit, retinal feature size, cone spacing	μm or mm
$S(x, y)$	Irradiance	Watts per square meter
$<S>$	Strehl ratio	
$\delta(x, y)$	Phase constant	radians, wave or μm
$\phi(x, y)$	Continuous wavefront (either phase error or optical path difference)	radians, wave or μm
$\boldsymbol{\phi}$	Descretized wavefront vector	radians, wave or μm
ϕ^x	x-derivative of the wavefront	
ϕ^y	y-derivative of the wavefront	
Σ	Domain of an aperture or pupil	
Z	Arbitrary Zernike mode	radians, wave or μm
Z^x	x-derivative of a Zernike mode	radians, wave or μm
Z^y	y-derivative of a Zernike mode	radians, wave or μm
F_E	Electrostatic force	Newton (kg m)
V	Voltage	volts
C	Contrast	
δ	Phase constant	radians
r_0	Fried parameter	cm
$\mathbf{M}_{ij}$	i^{th} row and j^{th} column entry of matrix $\mathbf{M}$	depends on its entries
$\mathbf{I}$	Identity matrix	depends on its entries

$\mathbf{T}$	Interaction matrix	V^{-2}
Λ	covariance matrix	depends on its entries
Φ	Power spectral density	depends on the spectrum
$\mathbf{e}$	Fitting error	depends on its entries
$\mathbf{y}$	Vector of wavefront slope measurements	
$\mathbf{c}$	Vector of Zernike coefficients	μm
z^{-1}	One step delay	
α	Regularization factor	
κ	Integrator gain	
m	Integrator leak	

List of abbreviations

ACD	Anterior chamber depth
ANSI	American National Standards Institute
AO	Adaptive optics
AOSLO	Adaptive optics scanning laser ophthalmoscope
BMC	Boston Micromachines Corporation
CCD	Charge-coupled device
DM	Deformable mirror
FFT	Fast Fourier transform
FIR	Finite impulse response
Hz	Hertz
IFFT	Inverse fast Fourier transform
ILC	Iterative learning control
IPT	Matlab Image Processing Toolbox
LQ	Linear quadratic
LQR	Linear quadratic regulator
MEMS	Micro-electro mechanical systems
OCT	Optical coherence tomography
OPD	Optical path difference
PID	Proportional-integral-derivative
PRL	Preferred retinal locus
PSD	Power spectral density
PSF	Point spread function
PSI	Phase shifting interferometer
RMF	Retinal magnification factor
RMS	Root mean square
RPE	Retinal pigment epithelium
SLD	Superluminescent diode
SLO	Scanning laser ophthalmoscope
SVD	Singular value decomposition

Acknowledgements

Looking back at the years spent at Berkeley, I could not have imagined how things, both good and not so good, have turned out. What I have accomplish and learned here could not have been done anywhere else, and it is the sole reason why I was able to land multiple job offers before even completing my degree. For that, I have many people to thank. The following statements are true and, besides the first few of which are obvious, are in no particular order.

- To no surprise, I have to first acknowledge my dissertation chair, Professor Austin Roorda. Even though we rarely saw eye-to-eye in terms of how to carry out a project, he for the most part has allowed me to work how I saw fit which I think was the right decision. Due to my personality for stressing fundamental understanding in every aspect of a project, scientific progress on my part was often slow, but very solid. I would often take the engineering part very seriously even if it meant taking courses in a completely different field just to gain fundamental understanding. Austin was always very patient in allowing me to do so, and in return, I have matured in my technical ability which the lab has directly benefited greatly from (although often not recognizing it).
- I would like to thank my best friend Crystal who always looked through all this bullshit and liked me for me and always encouraged me to be the person I wanted to be. Although we can no longer talk as often, her phone calls every week were often all that was good in my dark world allowing me to persevere through the first two years of graduate school.
- My friends Matt and Aaron, whom I have been so horrible in keeping in touch with, were the only reasons why I left high school unscathed. I guess you could say that they are the reasons why I appear “normal” to most people. Thank for keeping me in line and up-to-speed in life outside academia.
- Thanks to Professor Geunyoung Yoon for hiring me to work in his lab at the end of my junior year knowing fully that I was completely useless and had a good chance of remaining that way. I remember being extremely frustrated at that time after several failed attempts at research and was quite desperate to find just a little bit of success. Geunyoung not only introduced to the interesting yet practical field of physiological optics, he also taught me how to learn and be independent in this sometimes ruthless profession.
- Thanks to Professor James Zavislan for playing a similar role as Geunyoung in my career. Although I did not enjoy his classes at all at that time, the only undergraduate notes I revisited while writing this thesis were those from his classes. Ironically, notes from the class I most enjoyed were never revisited.
- I was tracing back to how I got into vision science in the first place which ultimately ended at Professor Walter Makous. Unlike all the other professors mentioned, I met Walt when I was a freshman at the University of Rochester where he left me a written note in Burton Hall which was where I resided. Apparently members of my church had contacted him and he responded positively to drive me to church if I wanted. I knew not of his brilliance and his importance in the vision science community. If all professors were like him, the academic community would be a much more pleasant.

- I would like to thank Professor Mona El-Sherif for her guidance during my thesis writing. The year 2010 is a World Cup year, so I wasn't able to start writing my dissertation until July 6th. With her strict schedule, rules and uncompromising naggings, I was able to complete my dissertation in time, so I will no longer have to worry about it when I start work overseas.
- Thanks to the machinists in the optometry school, particularly Nick Lines, for putting up with my never ending requests, poorly executed drawing and ridiculous deadlines. In return, I have bought them beer on occasion.
- Thanks to my coaches and teammates during my time on CAL Lightweight Crew. They were right, it was probably the best year of my life in many ways, and I will forever be proud to be a part of that program. Never in my life have I experienced this level of unity, commitment and friendship. What I have learned there will undoubtedly shape how I live the rest of my life.
- Thanks to Mark Jellison and the many friends I work out with at TAC, particularly Stef, Mona, Virginie, Peter, Johnny among others. You have given me something to look forward to while I am in lab on Mondays, Wednesdays and Fridays.
- Thanks to the many friends I have made while in Berkeley, many of whom left my life as they completed their respective programs and made me experience much melancholy. One era that I had a hard time getting over when it ended was a two to three year span involving Anna, Paul and Jen among several others whom I ran together several times a week. We also ran many races together and hung out socially which negated a lot of the pain that had accumulated over consecutive evenings of working alone in the lab. Looking back at all this, I realize that we must all move on eventually. I am glad that they are all doing so well these days.
- I would like to thank Professor Martin Banks for the opportunity to work with such a distinguished vision scientist and for accepting to be on my dissertation committee a few weeks before I needed to file.
- I thank Professor Masayoshi Tomizuka and other Faculty members from the Department of Mechanical Engineering for their excellent instruction on control systems. They were extremely difficult but important courses, and I do not regret a single moment from the countless evening I have spent doing problem sets and studying for exams. They have shaped how I now perceive technically challenging research problems.
- Finally, in some convoluted way, I would like to thank two professors in the Vision Science group for many personal insults and disparaging comments to me. Their ability to methodically break me down and make me feel worthless was flawless. In return, I discouraged all new Vision Science grad students from working in their labs succeeding for the most part. Nevertheless, these two individuals have made me a mentally tough and emotionless when it came to research matters. This was beneficial for my graduate career as it made me very productive. It also taught me exactly how not to treat my current and future students. Ironically, I do not think I like the person I am turning into, and that is part of the reason why I am leaving academia.

Curriculum Vitae

The author was born in Shanghai, the People's Republic of China in 1984 and immigrated with his family to the United States in 1991. Upon obtaining citizenship in 2001, he pursued degrees in mathematics and optics at the University of Rochester and graduated Cum Laude and with high honors respectively in addition to a minor in French. While completing his undergraduate coursework, he conducted honors research in a project titled "tear-film wavefront sensing", which was presented at the Association for Research in Vision and Ophthalmology Annual Meeting in 2005 and was eventually published in *Optics Express*. However, his keen interest in the underlying mathematical framework of the physical sciences soon led him to the mathematically rich field of adaptive optics. In 2005, he began his graduate studies at the University of California, Berkeley dividing his time between the School of Optometry and the Department of Mechanical Engineering. He has carried out his doctoral dissertation research in the areas of digital image processing, control theory and optics all applied toward advancing adaptive optics for retinal imaging under the supervision of Professor Austin Roorda as well as committee members Martin S. Banks and Masayoshi Tomizuka. Recently, he has accepted the position of project engineer at Abbott Medical Optics and will commence employment there on September 1, 2010.

1. Introduction

Adaptive optics (AO) refers to technology that reduces the amount of aberrations in an optical system in real-time. Originally developed to compensate for the wave aberrations caused by fluctuations in the atmosphere, AO has received considerable attention for vision science applications since it was first applied to the eye in 1997. It was that year when Liang and colleagues presented the first ever images of single cone photoreceptors in a living human eye.[1] Analogous to how the earth's atmosphere degrades the image quality of ground-based telescopes, aberrations due to the eye's optical imperfections degrade retinal image quality making it difficult for clinicians and scientists to observe the microscopic structures of the retina. Adaptive optics (AO) aims to remove these degradations by correcting the eye's wavefront aberrations in real-time and typically does so in a closed-loop fashion. How this done is illustrated in Figure 1.1, which is a schematic of the original flood-illuminated AO system at the University of Rochester. Like most complex systems that operate in closed loop, there is a sensor, the Shack-Hartmann wavefront sensor; there are actuators, the ones that push and pull a reflective membrane which we call the deformable mirror (DM); and there is a control computer which is responsible for taking sensor data and calculating the appropriate voltage signals to send to the actuators. Our ability to bypass the eye's optics and observe the living human retina at the microscopic level depends greatly on how well the aberrations are corrected. Therefore, it is important to both clinicians and basic researchers that we continue to push the performance of AO systems so that routine imaging of both healthy and diseased eyes can eventually become standard.

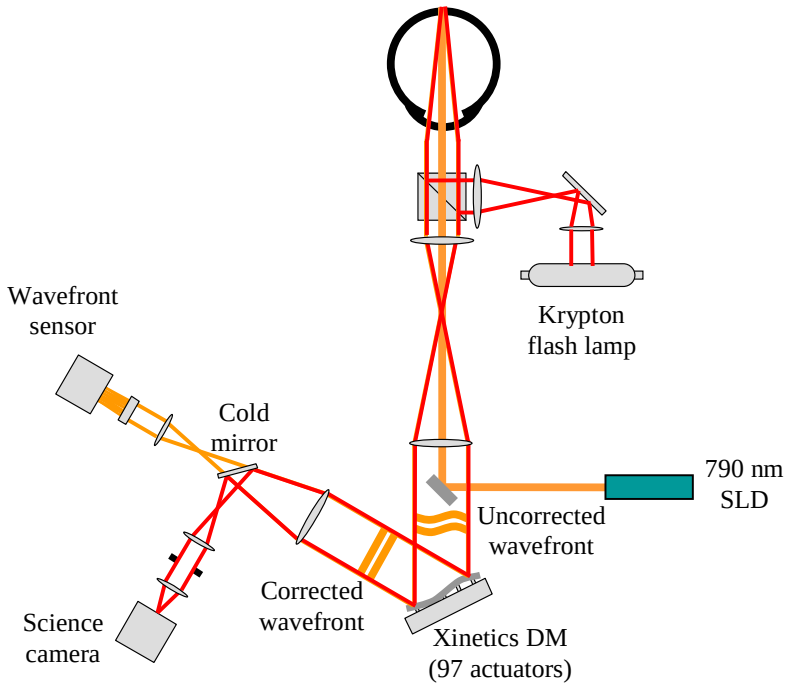


Figure 1.1: Schematic diagram of the AO flood-illuminated ophthalmoscope at the University of Rochester.

Prior to the work presented in this dissertation, the Rochester flood-illuminated AO ophthalmoscope was the only system that has demonstrated the capability to resolve the entire foveal cone mosaic. There are not very many examples though. To my knowledge, there are only four such foveal cone images in the literature[2, 3]. A cropped version of one of these images acquired from a healthy eye is given in Figure 1.2 displaying that every single cone was resolved, and they were resolved with enough contrast that they could probably all be identified with minimal guesswork. Over the years, AO has been successfully integrated into scanning laser ophthalmscopes (SLO)[4-7], optical coherence tomography (OCT)[8-12] and other flood-illuminated systems[13, 14]. Despite all these advances, none of these new systems has shown complete images of the healthy foveal cone mosaic. This could be due to a number of factors such as the DM used, imaging modality, optical system alignment, control system performance, etc. For example, the Rochester flood-illuminated AO system employs traditional 97 channel piezo-actuated DM (Xinetics) which has a proven track record in astronomical AO but is not used in any other AO imager currently online. It could also be due to differences in imaging modality, and that illuminating the retina with a completely incoherent light source results in a better quality image. Whatever the actual reason may be, we know that the 97 channel Xinetics offers similar spatial resolution as the Boston Micromachines Corporation (BMC) Multi-DM MEMS device (140 channels over a square grid minus the four corners) but has a superior finish to the reflective surface. Figure 1.3 is a microscopic image of one of our BMC devices illustrating this. These surface imperfections will contribute to the quality of the reflected beam, and the errors will propagate to the wavefront measurements and ultimately affect final wavefront correction if the problem is not addressed at other stages of the AO control loop.

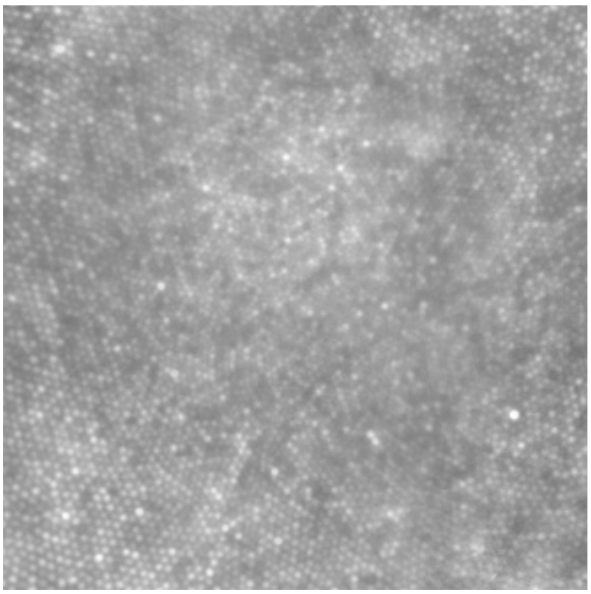


Figure 1.2: A 1° by 1° image of the cone mosaic centered about the approximate foveal center acquired by University of Rochester's AO flood-illuminated ophthalmoscope[2, 3]

Addressing the performance and robustness of current AO retinal imagers is challenging because doing so in an unbiased manner requires images of the same retina acquired under nearly identical conditions at different times. Pertaining to the correction of the eye's wave aberrations, there have been several studies that addressed the properties and performance of different DMs as well as bandwidth requirements of a real-time correction[15-18]. However, the conclusions have been primarily based on modeling and wavefront error calculated from Shack-Hartmann data, which I will show often overestimates the quality of the actual wavefront correction. In this dissertation, I will describe some of the improvements I made to the control system of an AO scanning laser ophthalmoscope (AOSLO)[7]; in particular,r the refinement of many required calculations that exist throughout the AO control loop. The instrument I worked on was the Berkeley MEMS-based AOSLO[6], which I will refer to as AOSLOII. These improvements have allowed us to image the smallest cone photoreceptors near and sometimes at the very center of the fovea, which led to some new scientific/clinical findings concerning cone density and axial myopia.

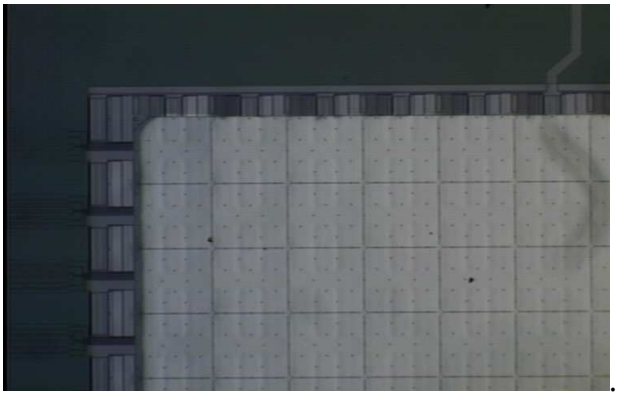


Figure 1.3: Image taken with a 2× microscope objective of a corner of one of our BMC MEMS devices (the same model used in AOSLOII) showing undesirable print-through due how it is manufactured. Actuator pitch is about 400 µm.

1.1 Challenges

Figure 1.4 is a 1° by 1° image of a healthy foveal cone mosaic acquired using AOSLOII in 2008 illustrating exactly the problems we still face when imaging the fovea with an AOSLO. I had already made some improvements to the AOSLOII control system, but the performance was in general not as robust as it currently is. The nearly uniform hexagonal packing structure of cone photoreceptors is clearly visible immediately outside of the central fovea (about 0.25° in this case but varies across individuals) but gradually fades toward the foveal center forming a darker region with a speckle like appearance. We can rule out these light spots in the center as cone photoreceptors because they are noticeably larger than those just outside the central fovea, and the overall appearance of the mosaic is not in agreement with the retinal images acquired with the Rochester flood-illuminated AO system (Figure 1.2). There are several reasons why most AOSLO images have this type of appearance in the foveal region. Lack of sufficient lateral resolution is most likely responsible for why the smallest foveal cones remain unresolved, and this could be only due to differences in the DM. The speckle-like appearance of the center of the fovea may

be a consequence of using a light source with a low coherence length and may also have to do with anatomical differences of foveal cones as many textbooks have depicted them to be much thinner and longer than the ones outside the fovea. The paragraphs below list some of the issues that I have considered concerning foveal imaging with an AOSLO. In any case, improving AO system performance in any way will only improve how well we can observe the foveal cone mosaic or any other retinal feature for that matter.

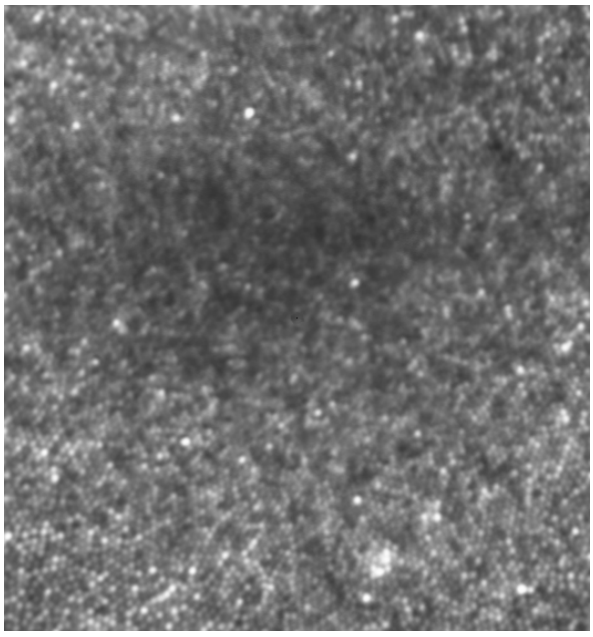


Figure 1.4: A 1° by 1° image of the cone mosaic centered about the approximate foveal center acquired with AOSLOII. The central foveal cones are clearly unresolved.

1.1.1 Visual and adaptive optics

To first approximation, the human eye is a relatively simple optical system made up of only three primary components: the cornea, the iris and the crystalline lens. The cornea and lens provide the optical power needed to focus incident light onto the retina with the iris acting as the aperture stop of the system regulating the amounting of light that makes it to the retina. However, this optical system is not perfect; it suffers from optical aberrations. Aberrations blur the image formed on the retina which reduces visual acuity and contrast sensitivity while also making it more challenging for clinicians and vision scientists to look inside the eye. Our understanding of the eye's optical properties has greatly increased due to fairly recent developments in ocular aberrometry and in particular the Shack-Hartmann wavefront sensor. Ocular aberrations cause the measured wavefront to depart from a plane. This departure is often referred to as the wavefront error or simply the wavefront. Since the first Shack-Hartmann sensor for the eye was demonstrated[19], the representation of the aberration structure of the eye has almost been exclusively based on Zernike polynomials which, as I will go into detail in Chapter 2, has both merits and limitations. When the wavefront is expressed as a linear combination of Zernike polynomials, it becomes evident that the low order modes explain most of the wavefront error in the eye. As it is now widely known, the low order Zernike modes that cause blur are the defocus and astigmatism terms, and most of that can be corrected by

conventional means (i.e. spectacles and contact lenses). High order modes cannot be corrected by conventional means, but they also do not affect visual quality significantly when the pupil is less than 3 mm in diameter[20]. However, if we were to image the retina through a smaller diameter pupil, diffraction significantly limits lateral resolution. Therefore, imaging through a large pupil is desirable in terms of reducing diffraction but doing so is only beneficial if the high order aberrations can be corrected.

Simple eye models are useful for assessing the lateral resolution achievable with AO. The following analysis, where the Gullstrand model eye with four refracting surfaces (Table 1.1) is used, is oversimplified, unrealistic and completely ignores the temporal fluctuations present in ocular aberrations, but it is nevertheless relevant for this discussion. Given these basic optical parameters, the cardinal points of the system can be found by doing a first-order ray trace. They are shown schematically in Figure 1.5. The eye model is quite simple except for perhaps that the nodal points, N_1 and N_2, are displaced from the principal planes, H_1 and H_2, because the refractive indices of the image space (vitreous) and object space (air) being different. The Rayleigh criterion for the resolution limit is given by:

$$s = \frac{1.22\lambda f}{d} \tag{1.1}$$

where λ is the illumination wavelength in the vitreous, f is the focal length of the eye defined by the distance H_2 to the retina (approximately 22.32 mm) and d is the diameter of the pupil. AOSLOII is set up to image over a 6 mm diameter pupil, so based on the Rayleigh criterion, the smallest resolvable feature on the retina with 840 nm light is about 2.8 µm. The Rochester flood-illuminated AO system can theoretically do a little better because it uses 790 nm light ($s \approx 2.6$ µm). These calculations apply to a diffraction limited eye which can only occur if the eye's wave aberrations are fully corrected with AO. If any uncorrected residual aberrations remain, which is the case for all real AO systems, the resolution will suffer. It is quite common for authors to state that their AO equipped retinal imager can achieve correction levels at which the root-mean-square (RMS) of the residual wavefront error of about 100 nm or less over a dilated (> 6 mm diameter) pupil[6, 7, 17, 21]. Assuming those are really their achievable residual wavefront errors, then the following expression for the Rayleigh criterion applies[22]:

$$s = \frac{1.22\lambda f}{d\sqrt{<S>}} \tag{1.2}$$

where $<S>$ is the Strehl ratio which is related to the RMS wavefront error by:

$$<S> \approx e^{-\left(\frac{2\pi}{\lambda}RMS\right)} \tag{1.3}$$

as long as the wavefront error is "small" (i.e. RMS being less than a quarter of a wave). Equations 1.2 and 1.3 are extremely useful because they allow us to assess how the resolution limit is related to the focal length of the wavefront compensated eye, the imaging wavelength and wavefront error in a very simple manner. AOSLOII has two wavelength options, 680 nm (red) and 840 nm (near infrared), but for convenience (primarily due to subject comfort since the retina is less sensitive to infrared), the preferred imaging wavelength is 840nm. We do not have any control over the size of the eye being imaged, which can range from 22 mm to over 28 mm long. Assuming that the eye's focal length is about 2 mm shorter than the total length of the eyeball, it is possible to graphically observe how the achievable resolution limit will differ for different people

(Figure 1.6). This simple exercise demonstrates that the theoretical resolution limit for retinal imaging is worse for longer eyes than for shorter eyes. So if AO retinal imagers today are indeed operating near the diffraction limit, we would expect that a short eye would yield higher quality images of foveal cone mosaics.

Table 1.1: Gullstrand model eye

	Front corneal surface	Back corneal surface	Front lens surface	Back lens surface
Radius of curvature (mm)	7.8	6.9	10.2	-6.0
Refractive index	1.38	1.3374	1.42	1.336
Thickness (mm)	0.535	3.1	4.0	24.18

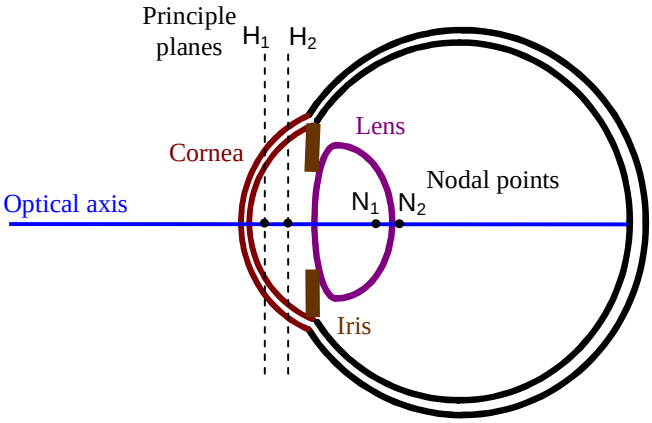


Figure 1.5: Schematic diagram of a four-surface eye model illustrating that the principle planes do not coincide with the nodal points.

So far, I have not mentioned the temporal component of the eye's wave aberrations. Since so many quality images of cone mosaics have been acquired with basically a static AO correction[1, 6, 7, 17], it may seem that a static correction would be sufficient. Although a tenfold improvement can be obtained with a good static correction, it will still only be about halfway to diffraction limited performance based on the Strehl ratio[23]. Based on the assumption that the aberration structure of the eye can be completely described by a linear combination of Zernike polynomials up to 5^{th} order, Hofer *et al.*[23] claim that a closed loop bandwidth of 1 to 2 Hz is all that is needed to achieve diffraction limited imaging performance. It should be clarified that closed loop bandwidth is not the same as the system sampling rate. Roughly speaking, a 2 Hz bandwidth requirement would require a sampling rate beyond 30 Hz. It is now known that Zernike polynomials up to 10^{th} order still fail to capture the entire ocular wavefront structure[24, 25], so the actual bandwidth requirement for diffraction-limited performance is likely to be even greater. Diaz-Santana *et al.* also examined bandwidth requirements for ocular aberration compensation but did so with a much faster AO system albeit without retinal imaging capabilities. They found that there is a lot of power in the fluctuation of the eye's wave aberrations above 30 Hz, so we can further improve the Strehl ratio by increasing the bandwidth. However, they did not specify how they reconstructed the wavefront from

Shack-Hartmann sensor data, and they did not present any evidence based on actual images of the retina.

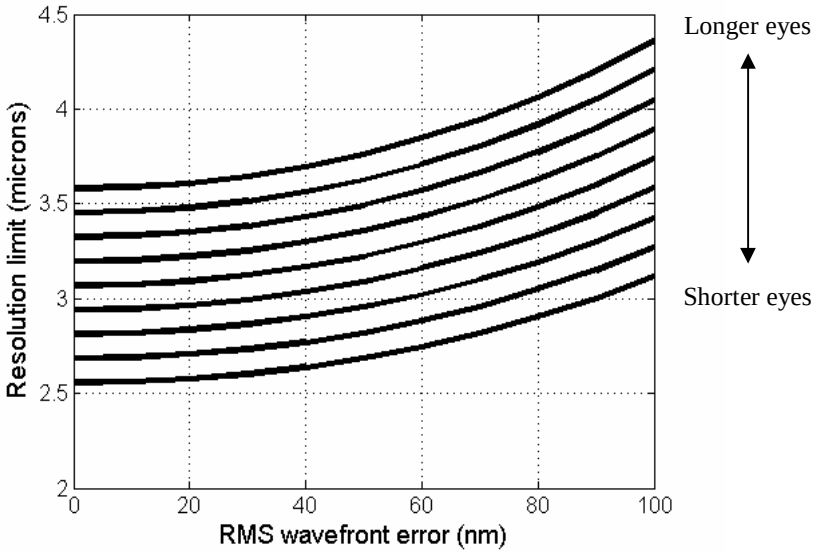


Figure 1.6: Resolution limit based on the Rayleigh criterion as a function of the RMS wavefront error for various eye lengths (22 to 28 mm). Focal length was taken to be 2 mm shorter than the total eye length

1.1.3 Foveal cone photoreceptors

The word fovea means *pit* in Latin. The fovea of the eye, called the fovea centralis, is the part of the retina responsible for fine spatial vision. The anatomical pit that characterizes the fovea centralis, which I will refer to as the fovea from now on, is about 1.5 mm wide and contains the highest density of cone photoreceptors[26]. According to the Gullstrand eye model, 1.5 mm converts to about 5.1 degrees of visual angle which is more than five times wider than the AOSLO image in Figure 1.4. Clearly, we can resolve most of the cone photoreceptors inside the fovea. The center of the fovea is called the foveola. The size of the foveola is somewhat arbitrarily defined, much like the fovea, and has been reported to be 200 to 400 µm in diameter in various textbooks. Perhaps it is more sensible to define the foveola as the region containing no rod photoreceptors as Yuodelis and Hendrickson[27] did. They reported a diameter of 683 µm. It is in the foveola that the individual cone photoreceptors have continued to challenge the resolution limit of today's AO systems.

Since *en vivo* imaging of foveolar cones is still very challenging, many investigators still use cone density data from histological and/or psychophysical studies. The most comprehensive histological investigation of human cone photoreceptor distribution was undoubtedly conducted by Curcio *et al.*[28] Figure 1.7a is an image of the foveal cone mosaic from their classic paper. According to their work, foveal cone density exhibited a great deal of inter-subject variability especially at the fovea center. With the exception of the two very high density foveas, which were mentioned by the investigators to have been possibly affected by tissue shrinkage, the smallest foveal cones were at least 2 µm in diameter which should be a good approximation for minimum cone spacing in the rod-free fovea[27-30]. Based on the Rayleigh criterion (Figure 1.6), we would not be able

to resolve the smallest foveal cones in certain individuals, but there were other individuals in Curcio's study[28] with minimum cone spacing greater 3.5 µm which should quite easily be resolved.

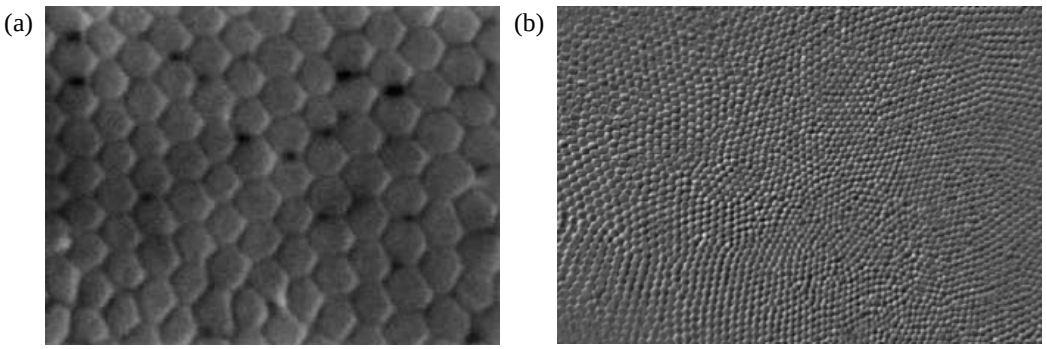


Figure 1.7: In vitro images of foveal cone mosaics acquired using differential interference contrast microscopy. (a) An image, about 30 µm across, from Figure 1 in Curcio *et al.*[28] and (b) an image, about 200 µm across, from Figure 1 in Wojitas *et al.*[31]

It is also possible that anatomical features of foveolar cones besides their size make them difficult to resolve with an AOSLO. As mentioned earlier, cone photoreceptors have been depicted to be both thinner and longer as they are closer to the foveal center. The fact that they become longer can be seen in a high quality OCT scan (Figure 1.8) where the distance between the two highly reflective interfaces of cones is greater in the foveal center. Consider the simple optical model of a cone of the left side of Figure 1.8, the first reflection r_1 occurs at the junction between the inner and outer segments, and the second reflection r_2 occurs at the interface of the outer segment and the retinal pigment epithelium (RPE). Assuming that absorption is negligible, a field incident on the retina, ψ, results in two main fields, ψ_1 and ψ_2, reflected from the cone photoreceptors. The expressions that describe this process are given by:

$$\begin{cases} \psi_1 = r_1\psi \\ \psi_2 = (1+r_1)\,r_2\psi \end{cases} \tag{1.4}$$

where the $1 + r_1$ term for the ψ_2 is the amplitude (not irradiance) transmittance of the first interface. Immediately outside the very center of the fovea, the two reflections are about the same in magnitude and they nearly do not interfere (electric fields have zero correlation) because the reflective interfaces are separated by a distance longer than the coherence length of the imaging light source (10 µm according to manufacturer Superlum Ltd.). Therefore, it will be impossible for the total reflected light from one cone to completely interfere with that from an adjacent cone reducing interference artifacts (speckle noise) in the final image. But from the OCT image in Figure 1.8, we can readily observe that the top reflection (r_1) gets dimmer toward the fovea center which is quite prominent considering that the image is on a log scale. If r_1 goes to nearly zero, we are effectively only left with:

$$\psi_2 = r_2\psi \tag{1.5}$$

which will be free to interfere (correlate) with the ψ_2 of its neighbors raising the speckle noise in the retinal image and swamping the reflectance signals representing the actual

cone photoreceptor locations. Whether or not this model is of practical concern to AOSLO imaging is completely based the lateral resolution of the system. If we can focus a smaller spot of light onto the retina with a better AO correction, then there will be less opportunity for interference to occur between adjacent cones.

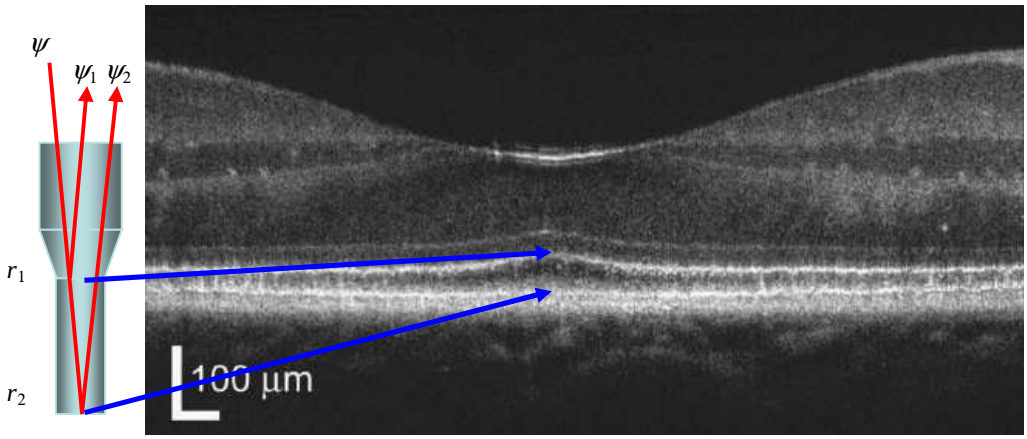


Figure 1.8: A cross section through the foveal center acquired using an AO OCT system from Figure 3 in Bigelow *et al.*[32] On the left is a model of a foveal cone photoreceptor with the arrows indicating approximately where the two lines of high reflectivity takes occur (see text for model description).

1.2 Adaptive optics scanning laser ophthalmoscope

Roorda *et al.* (2002)[7] incorporated AO into an SLO for the first AOSLO (AOSLOI), which successfully resolved individual cone photoreceptors in the living human retina but only as close as 0.5° (≈ 150 µm) away from the center of the fovea. The AOSLO initially had the advantage over the conventional flood-illuminated AO ophthalmoscope in being able to perform optical slicing of different tissue layers of the retina and the ability to record retinal videos as opposed to a single snapshot. Over the years, these advantages have proved to be valuable tools in our lab as well as others for assessing retinal blood flow[33-35], Ganglion cells in the macaque[36, 37], and RPE cells[38, 39].

AOSLOII was designed differently from AOSLOI because the aim was to make the system as small and compact as possible for the goal of deploying it in a clinical setting[40]. This was made possible by employing a tiny MEMS DM whose clear aperture is about the same size as the pupil, making the system much smaller. The AO control system operates over the optical path shown in Figure 1.9a. The near-infrared beam is provided by an 840 nm superluminescent diode (SLD) (Superlum Ltd., Russia) and a photomultiplier tube (Hamamatsu, Japan) is used for light detection. The DM (BMC, USA) has a 12 by 12 actuator array minus the corner actuators providing a total of 140 degrees of freedom. The current state of AOSLOII is more or less the same as how it was originally described by Zhang *et al.*[6] with the exception that I swapped out the old 8-bit DM driver for a higher step resolution (14-bit) driver that currently comes standard with any Multi-DM purchase and also upgraded to a 5.5 µm advertised stroke DM.

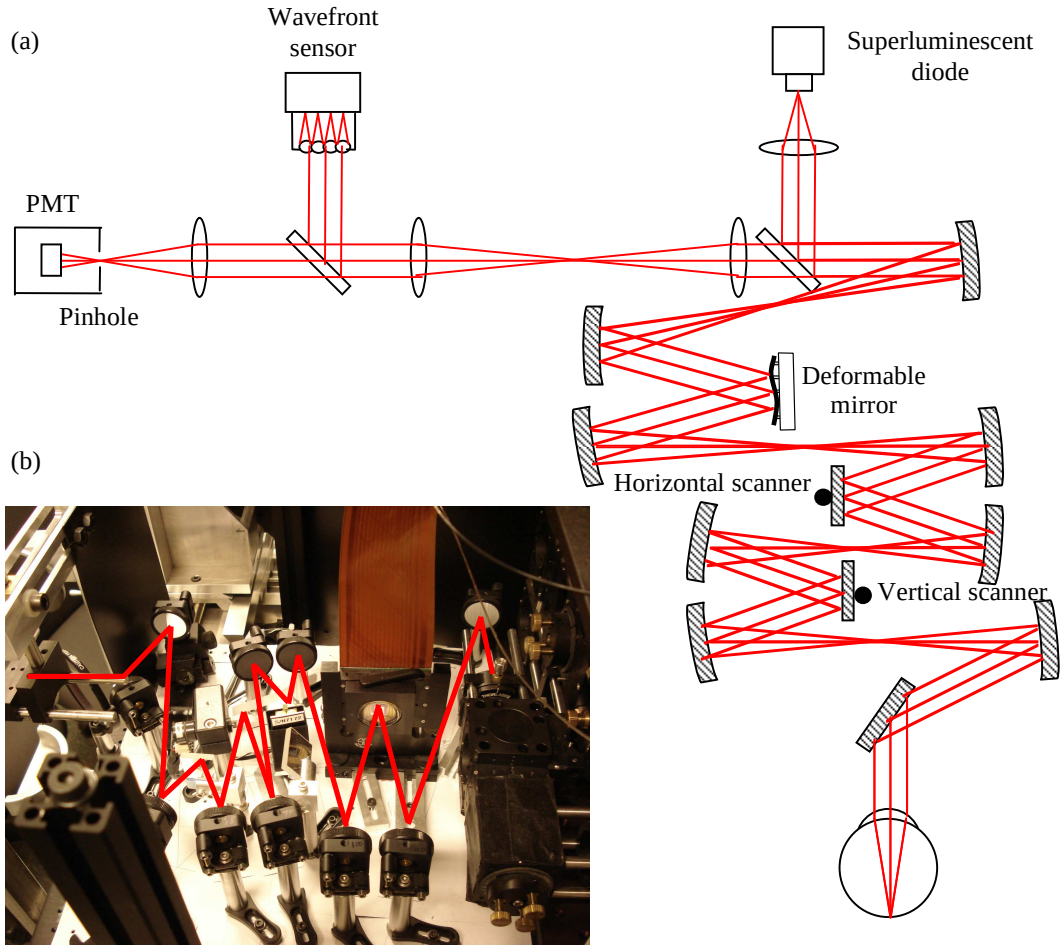


Figure 1.9: (a) Schematic diagram of AOSLOII and (b) a top view of the system with the optical path overlaid in red.

1.3 Purpose and structure of dissertation

The goal for this dissertation is to provide the mathematical framework for wavefront estimation and control specific to AO for vision science and demonstrating how we have benefited specifically from improving AO system performance in this way. Since this dissertation is for a Vision Science degree, some basic science research on the topics of foveal fixation and eye growth has also been conducted. The work carried out spans the disciplines of optics, controls and vision science. The mathematics that complements these fields, especially Fourier analysis and linear algebra, will often appear. Complex numbers will also appear due to having to work with electromagnetic waves. A lot of effort has been put towards the implementation of these ideas in a real AO system, particularly AOSLOII. And it was primarily through these efforts that I was able to demonstrate that the smallest foveal cone photoreceptors can indeed be resolved using an AOSLO. From these experiences, I hope to convey to other scientists, engineers and clinicians the simple necessity of a more rigorous treatment of wavefront sensing and control. Even though the work here specifically addresses an AO system for correcting ocular wave aberration, the MEMS device I used is just as new to the applications in

astronomy, so I hope AO developers from other disciplines can benefit from this research as well.

In Chapter 2, a detailed explanation of Shack-Hartmann wavefront sensing is given beginning with a discussion on how an electromagnetic wave propagates through the eye's optics and leading to each of the steps in the Shack-Hartmann sensing process currently carried out on AOSLOII. I will then develop the wavefront reconstruction (estimation) problem with minimal assumptions. Since Zernike polynomials have become almost synonymous with ocular wavefront sensing, the properties of Zernike polynomials and how they are affected by the measurement and estimation procedure will be addressed analytically and numerically. It will become apparent that the standard wavefront estimation procedure is valid for certain applications, but not others. Zonal and Fourier methods for wavefront reconstruction are presented and analyzed as well.

Chapter 3 discusses the second half of the problem which is to actively compensate ocular wavefront error. The wavefront control problem is presented from mathematics to real-time software implementation. Specifically, four wavefront reconstruction (control) algorithms were designed and implemented on AOSLOII. Three of the four algorithms are of the zonal type and one is based on Zernike polynomials. The details on each of the AO control strategies will be followed by a discussion on stability and comparisons based on both residual wavefront error and retinal image quality. Finally, I will describe how I have implemented the entire AO control system as five separate tasks with the detailed function for each task described by a finite state machine.

While Chapters 2 and 3 focused on improving AO system performance to push the lateral resolution limit of the AOSLO, I report on some of the first investigations on the structure of the foveal cone mosaic in living human retinas in Chapter 4. Using AOSLOII, I have imaged the foveae of more than 20 eyes over the course of this work. However, to avoid potential bias, only one eye per subject (18 eyes, 18 subjects) was used in the investigations described in Chapter 4. I have written custom software to identify the cone photoreceptors in an AOSLO image and to generate topographic maps representing cone density from the identified cone locations. Together with ocular biometry, I looked at how foveal cone density varied across individuals with different sized eyes which typically corresponded to different levels of refractive error. The results of this study have important implications regarding whether or not visual acuity is fundamentally limited by the sampling rate of foveal cones. Furthermore, it is often assumed that we fixate with the part of the retina with the highest cone density. But is this assumption valid at the microscopic level? With most if not all of the foveal cones resolved for each retina, we can precisely locate the point of peak cone density for each retina and compare it against the individual's fixation points.

Acknowledgements

Michael Helmbrecht took the photograph in Figure 1.2 at the Berkeley Sensor and Actuator Center. I acknowledge Yuhua Zhang for many discussions regarding the resolution of AOSLOII. Pavan Tiruveedhula assisted with the installation of the DM driver electronics.

2. Measuring monochromatic ocular aberrations

2.1 Introduction

A Shack-Hartmann type sensor is used for wavefront measurement in AOSLOII[6]. To my knowledge, there are no AO equipped retinal imagers online using a different wavefront sensing modality. Accurate wavefront measurement is a critical component for effective aberration compensation and robust closed-loop AO operation. In this chapter, I start by presenting the principles of Shack-Hartmann wavefront sensing for ocular aberrometry including the basic components making up a typical Shack-Hartmann sensor. The discussion will then be focused on my investigations into wavefront reconstruction algorithms for estimating the wavefront (as opposed to those for controlling the wavefront which will be discussed in Chapter 3).

Shack-Hartmann wavefront sensors as standalone systems have become the norm for measuring and analyzing ocular aberrations. These devices are now ubiquitous in both research and clinical settings to a point where one can easily be mislead into believing that most of the important fundamental principles associated with Shack-Hartmann wavefront sensing are well understood. Although this is very much true in general, a close inspection of the literature would reveal this not to be the case for measuring ocular aberrations. In vision science and associated clinical disciplines (optometry and ophthalmology), it is especially important to understand the calculations required in the wavefront measurement process. Consider wavefront guided LASIK for example. Improper interpretation of wavefront data may lead to potentially disastrous results.

The design of ocular Shack-Hartmann sensors today has not deviated significantly from the very first designed given in Liang *et al.*[19] The basic components that makes up the sensor (minus the light delivery components) is not much more than an afocal telescope formed by two lenses, a lenslet array and a detector (typically a charged-coupled device (CCD)) as depicted in Figure 2.1. The basic principle is that the eye's wave aberrations produces spatial variations in the phase component of the field located at the pupil plane when observed from the outside, and it is of our interest to recover this quantity. At first, a telescope may not seem necessary as only a single lens is needed to image the field at the pupil onto the lenslet array. However in order to get the maximal signal level at the detector, the focal plane of the lenslet array should also be conjugate to the retina. Therefore, the telescope serves two purposes: 1) relaying the field at the pupil plane onto the lenslet array, and 2) allowing the illuminated spot on the retina to be reasonably in focus on the CCD by the lenslet array assuming that the retina coincides with the back focal plane of the eye (emmetropic eye). The raw data outputted by the Shack-Hartmann sensor is simply a digital image of a spot pattern produced by each of the micro-lenses (subapertures) focusing the light onto the CCD. If the incident wavefront has local slope over a particular subaperture, it will induce a shift in the position of the focused spot proportional to this slope. The set of local wavefront slopes within the pupil is all that is needed to recover the ocular wavefront error. Details on obtaining wavefront slopes and methods for reconstructing the wavefront from the slopes will be rigorously addressed in this chapter in section 2.2.

Light delivery into the eye is typically accomplished by some means of collimation if the source is originally uncollimated (i.e. a diode) before it is introduced

into the main optical path via a beamsplitter. Liang's original design employed a 532.8 nm He-Ne laser and a relatively low density lenslet array by today's aberrometry standards (7 subapertures across a 5.4 mm diameter pupil), but he and other colleagues would immediate develop an improved device with a better quality and denser lenslet array as well as attempting to harness the polarization of the bean to suppress corneal reflections[20]. Image quality, as depicted in these earlier papers, were rather poor again by today's standards and only static wavefront measurements could be made most likely due to the technology available at the time. Nevertheless, their work established the fact that the Shack-Hartman wavefront sensor as an objective yet simple technique to measure ocular wave aberrations.

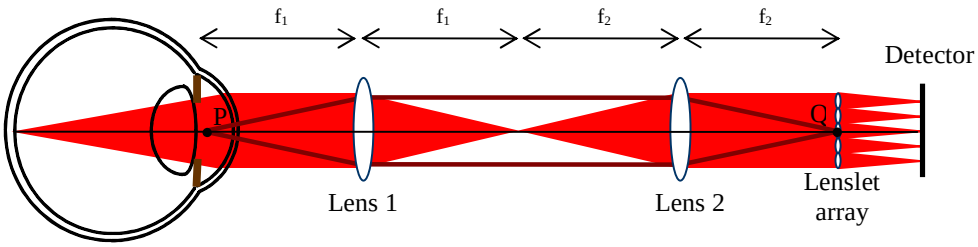


Figure 2.1: Schematic diagram of a Shack-Hartmann wavefront sensor. Lenses 1 and 2 are separated by their focal lengths (f_1 and f_2). Focal length selection is based on the desired lateral magnification of the eye's entrance pupil, located at P when imaged onto the lenslet array. For the AOSLO, this magnification is 0.889.

2.1.1 Shack-Hartmann sensor design

This section is intended to establish some general background for ocular Shack-Hartmann wavefront sensing. Hardware improvements to the earlier designs mentioned above have included changes in the light source, resolution (spatial and temporal) and dynamic range. How each of these components affects the final output image of Shack-Hartmann spot pattern is important for guiding algorithm development down the line. For example, the use of a polarizing beam-splitter to first bring light into the eye and then separate reflections from the retina from that of the cornea was not particularly effective. Sticking to this method would require one to rely on the image post-processing methods for separating which would add unnecessary computational overhead to the wavefront sensing process.

Off-axis illumination, where the entry beam is displaced from the pupil center (Figure 2.2a), is a simple and widely used solution to this problem[41]. The principle behind the formation of a sharp image in a confocal imager such as the AOSLO requires both the entry exit beams to be pre-corrected, so off-axis illumination cannot be applied[22, 42]. Instead in AOSLOII, a manually adjustable aperture is placed in a retinal conjugate plane, as shown in Figure 2.2b, with the reason being that light reflected from the retina will be better focused at this plane than light reflected from the cornea. So most of the light reflected off the retina passes through the aperture while that from the cornea is mostly blocked. The critical assumption here is that the eye's wave aberrations are relatively low. For example, if we had an infinitely small aperture, the field at the aperture plane becomes a point source. In this case, a plane wave will be created by lens 2 due to the point source at its primary focal point no matter how much aberrations are

present in the eye (i.e. the wavefront sensor will not detect any refractive error so no control signals will be prescribed to the DM). The size of the aperture must be adjusted with the tradeoff between corneal reflection and measurement accuracy in mind. It is often the case when imaging individuals with relatively high refractive error that a trial lens must be placed in front of the eye to correct for most of the refractive error before the wavefront can even be properly measured.

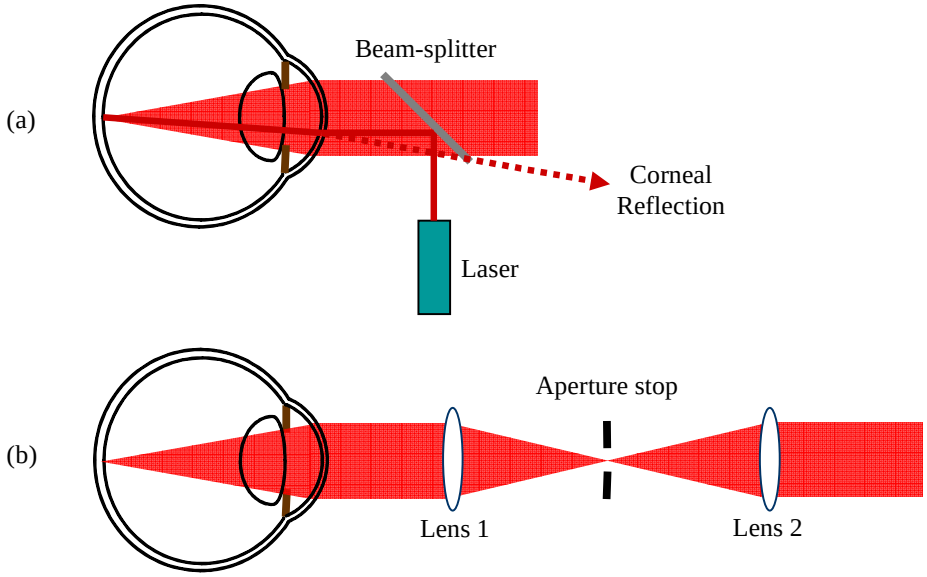


Figure 2.2: Two methods for countering reflections from the corneal front surface. (a) Off-axis illumination forces corneal reflections off to an angle, and (b) an aperture place at a retinal conjugate blocks most of the light reflected from the cornea because they are not at a sharp focus at that plane.

Laser speckle adds undesirable correlated noise to the Shack-Hartmann image that the accuracy to which the location of each focused spot can be identified may be compromised. The coherence length of the He-Ne lasers used in earlier wavefront sensors were very long (> 1 m) which contaminated the spot array image with this type of correlated noise. This issue was addressed by several investigators more recently (2001) by switching to the less coherent superluminescent diode (SLD) which is the type of light source employed in AOSLOII[23, 43, 44]. Hofer *et al.*[23] also placed a high speed scanning mirror at a plane conjugate to the pupil, which effectively smoothes the Shack-Hartmann image by rapidly scanning the laser spot across a small patch of retina but also makes the spots larger[23]. SLD technology has steadily improved both in terms of spectral band-width and cost, so it is to no surprise that it became rapidly adopted for wavefront sensing. A scanning system, which is obviously more complicated and costly, may no longer be necessary for most applications especially with the application of more sophisticated image processing algorithms to Shack-Hartmann images

Finally, there are other components that have found their way into research and/or commercially available Shack-Hartmann sensors that may or may not be beneficial for an AOSLO. A well-known problem that arises when designing a Shack-Hartmann sensor is the tug-of-war between dynamic range and sensitivity. More often then not, the solution

was influence mainly by the applications. Pantanelli *et al.*[45] constructed a wavefront sensor with quadrature masking system where basically 25 percent of the subapertures (even spaced out) are analyzed at a time allowing for extremely high dynamic range without sacrificing sensitivity. This advancement however does not really benefit AO retinal imaging as having one will decrease the system bandwidth by a factor of four. Another popular addition to the generic Shack-Hartmann sensor design would be a badal system[23, 46, 47]. A badal system allows the distance between lens 1 and 2, as labeled in Figures 2.1 and 2.2, to be adjustable which allows for defocus correction. A badal system has been integrated into an AOSLO as well[5] relieving some of the stroke requirements of the DM. Perhaps the most popular or profitable upgrade to make is to opt for a denser lenslet array. Several research groups have taken advantage of the fine sampling capabilities offered by denser arrays to measure irregularities due to tear film[24, 48, 49]. To no surprise, in addition to these laboratory-based developments, many commercially available Shack-Hartmann sensors, such as Abbott Medical Optics' COAS-HD and Imagine Eyes' HASO 32, are now equipped with very high density lenslet arrays.

2.1.2 Pupil function

Prior to discussing wavefront reconstruction, it is necessary to establish the terminology and the mathematical background associated with the optical wavefront including how it affects retinal image quality. This section reviews the necessary fundamentals of wave optics in order to derive an expression for the eye's pupil function because the phase component of this function is what we are try to measure using the Shack-Hartmann sensor. Furthermore, I have come across some misnomers and/or inconsistencies in the literature in which I would like to address during in my treatment of this topic. In these discussions, scalar diffraction theory is valid (optical elements involved are much larger than the imaging wavelength (840 nm)) and will be used exclusively, while the polarization state of the laser, although important in certain cases, will be ignored.

Although somewhat trivial, it should be first clarified to avoid confusion that the terms exit and entrance pupil depend on the direction of light propagation. When describing image formation on the retina, the entrance pupil is just the pupil of the eye. But in retinal imaging, the object of interest becomes the retina, so the definitions must be swapped. For a perfect eye, this scenario is illustrated in Figure 2.3 where a point on the retina emanates a spherical wave which is perfectly countered by the eye's optical power so that plane waves leave the eye.

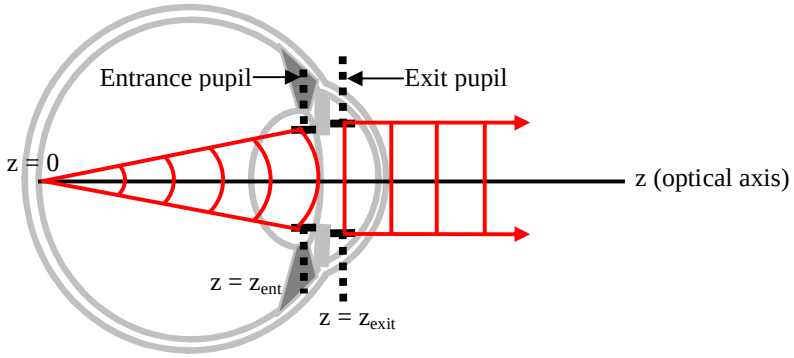


Figure 2.3: Schematic of a diffraction limited (perfect) eye illustrating that a point source at the retinal plane (z = 0) creates a diverging spherical wave that is exactly canceled by the eye's optical power, and as a result, perfect plane waves leave the exit pupil.

The general expression for the electric field component of an electromagnetic wave is given by:

$$\psi(\mathbf{x},t) = E(\mathbf{x},t)e^{jF(\mathbf{x})} \tag{2.1}$$

where **x =** (*x*, y, *z*) are pupil coordinates and optical axis and $j = (-1)^{1/2}$. Units of the electric field are electric potential (volts) divided by distance which is completely defined by the amplitude term (*E*) of equation 2.1 because the phase component (*F*), which contains the wavefront, is required to be unitless. The spherical wave incident at the entrance pupil plane can be described by:

$$\psi_{ent}(r,t) = \frac{E_0}{r}e^{-j\omega t}e^{jkr} \tag{2.2}$$

where *r* is the radius of curvature of the wavefront, *k* is the wavenumber and ω is natural frequency. It should be noted that since the field, $\psi_{ent}(r,t)$, is in the vitreous chamber, the wavenumber ($k = 2\pi/\lambda$) is defined such that the wavelength (λ) is that of the illumination (840 nm) divided by the refractive index of the vitreous (≈ 1.33). Paraxial approximation of the radius of curvature via first order binomial expansion is:

$$r \approx z + \frac{x^2 + y^2}{2z} \tag{2.3}$$

indicating that a sphere is well approximated by a parabola near the optical axis (paraxial region). Substituting Equation 2.3 into Equation 2.2 and evaluating it at $z = z_{ent}$, we obtain the desired approximation for a spherical wave in Cartesian coordinates:

$$\psi_{ent}(\mathbf{x},t) = \frac{E_0}{z_{ent}}e^{-j\omega t}e^{jk\left(z_{ent}+\frac{x^2+y^2}{2z_{ent}}\right)} \tag{2.4}$$

The quadratic term is omitted in the denominator of the amplitude since it is small with respect to z_{ent}. This cannot be done in phase term because the quadratic term is not small in comparison to the wavelength term that gets divided into it (recall $k = 2\pi/\lambda$). In simple examples such as purely spherical and plane waves, the behavior of the amplitude term is tractable and can be expressed analytically ($\frac{E_0}{r}e^{-j\omega t}$ and $E_0e^{-j\omega t}$ for spherical and plane waves respectively), but this would not be possible for more complicated wavefronts. For

this reason, I will proceed by representing the amplitude with just a generalized function as in equation 2.1. This is clearly an over-simplified model of the actual physics involved and is designed to hide certain mathematical complications. However, it is nevertheless valid for what I am trying to convey. Doing so changes the expression for the paraxial approximation of a spherical wave into a more canonical appearance:

$$\psi_{ent}(\mathbf{x},t) = E_{ent}(\mathbf{x},t)e^{jk\left(z_{ent}+\frac{x^2+y^2}{2z_{ent}}\right)} \tag{2.5}$$

The textbook definition for wavefront is a surface of constant phase, which in the case for the spherical wave would be any surface proportional to the quadratic phase component of equation 2.5:

$$F(\mathbf{x}) = k\frac{x^2+y^2}{2z_{ent}} = \frac{2\pi}{\lambda}OPD(x,y) \tag{2.6}$$

where OPD stands for the on optical path difference from the plane wave term in Equation 2.5 ($e^{jkz_{ent}}$) which has become somewhat standard in recent literature on ocular aberration when specifying or plotting the wavefront error across the pupil. The choice as to whether wavefront is defined strictly as optical phase shift (radians) or some form of OPD (microns or number of waves) makes no fundamental difference and should be obvious given the context.

The Shack-Hartmann sensor measures the wavefront at the exit pupil, not the entrance pupil (pupils defined as in Figure 2.3). Since the exit pupil is the image of the iris formed by the cornea, and the entrance pupil is the image of the iris formed by the crystalline lens, the pupils are images of each other. Under the assumption of linearity, this allows for the field at the entrance pupil to be related to that at the exit pupil by a complex multiplicative factor called the system transfer function. This transfer function, which is completely characterized by the eye's optical components (cornea, iris and crystalline lens), can be expressed in the familiar amplitude and phase form:

$$T(x,y) = T_0(x,y)e^{jU(x,y)} \tag{2.7}$$

Unlike expressions for the field, the system transfer function is unitless, and it is specified by the shape of the aperture (a circle in most cases), any transmission variation caused by scatter and absorption throughout the system, and phase effects of the optical system (focus power and aberrations). Since the pupils are images of each other, we have:

$$\psi_{exit}(\mathbf{x}',t) = T(x,y)\psi_{ent}(\mathbf{x},t) \tag{2.8}$$

where $\mathbf{x}' = (m_p x, m_p y, z)$ with m_p being pupil magnification since the exit pupil is about 10 percent larger than the entrance pupil.

The spatial profile of the amplitude term in the transfer function determines the amount of light reaching each subaperture of the wavefront sensor since irradiance is just the square amplitude. For a perfect eye (Figure 2.3), all light is transmitted inside the entrance pupil and no light is transmitted outside of the pupil, so the amplitude transmittance term is simply:

$$T_0(x,y) = \begin{cases} 1 & (x,y) \in \Sigma \\ 0 & (x,y) \notin \Sigma \end{cases} \tag{2.9}$$

where Σ is the spatial domain defined by the entrance pupil. This trivial transmittance model has become standard when performing retinal image quality related calculation such as the eye's point spread function (PSF). However, even a healthy real eye will have

some degree of absorption and scatter of light throughout its optical system. This observation is illustrated in Figure 2.4 which gives grayscale topographic map representations of the amplitude (magnitude) distribution across the pupil from Shack-Hartmann data for a healthy emmetropic eye and that from a model eye. The grayscale values were normalized for comparison, and it can be readily observed that the irradiance distribution is much more uniform in the model eye which is expected as it is only representative of the light source. The irradiance distribution for the real eye tapers off toward the edges of the pupil indicating that the trivial transmittance model may not be a particularly accurate choice. Interestingly, it has recently been verified that this is primarily due to the directional waveguide properties of photoreceptors although other optical factors may play significant roles as well[50]. Similarly, the phase term cannot be expressed accurately in analytical form due to imperfections in the eye's optics. However for a perfect eye, the phase effect should be equal and opposite of the phase of the field in the entrance pupil. The transfer function for an optical system with plus power can be approximated by[51]:

$$T(x,y) = T_0(x,y)e^{-j\frac{k}{2f}(x^2+y^2)} \tag{2.10}$$

where f is the focal length (positive) of the system. For a perfect eye, the focal length must match the radius of curvature of the field at the entrance pupil ($f = r$ in general or $f = z_{ent}$ with paraxial approximation), so an expression for the field immediately to the right of the exit pupil can be obtained by substituting Equation 2.10 into Equation 2.8.

$$\begin{aligned}\psi_{exit}(\mathbf{x}',t) &= T_0(x,y)E_{ent}(\mathbf{x},t)e^{jkz_{exit}} \\ &= E_{exit}(\mathbf{x}',t)e^{jkz_{exit}}\end{aligned} \tag{2.11}$$

Equation 2.11 is a plane wave which is exactly what is expected to emerge from the exit pupil of a perfect eye. When the eye's optics is not perfect, an additional phase term must be present in the system transfer function:

$$T(x,y) = T_0(x,y)e^{-j\frac{k}{2f}(x^2+y^2)}e^{j\phi(x,y)} \tag{2.12}$$

where I will define the function, $\phi(x, y)$, as the wavefront error which encompasses all phase effects caused by the optical system not including the focusing power that it is supposed to have. As before, we can substitute Equation 2.12 into Equation 2.8 to obtain an expression for the field at the exit pupil.

$$\begin{aligned}\psi_{exit}(\mathbf{x}',t) &= T_0(x,y)e^{j\phi(x,y)}E_{ent}(\mathbf{x},t)e^{jkz_{exit}} \\ &= E_{exit}(\mathbf{x}',t)e^{j(kz_{exit}+\phi(x,y))}\end{aligned} \tag{2.13}$$

The pupil function is defined as:

$$P(x,y) = T_0(x,y)e^{j\phi(x,y)} \tag{2.14}$$

which is simply the system transfer function but with the phase component containing only the effects due to wavefront error. Therefore, the system transfer function acting on the diverging spherical wave in the entrance pupil is mathematically equivalent to the pupil function applied to a plane wave. There have been instances in the literature where the pupil function has been referred to as the electromagnetic wave (field) at the pupil[52, 53]. This is incorrect because the pupil function is a unitless quantity, and the actual field is a function of both time and the direction of propagation (z).

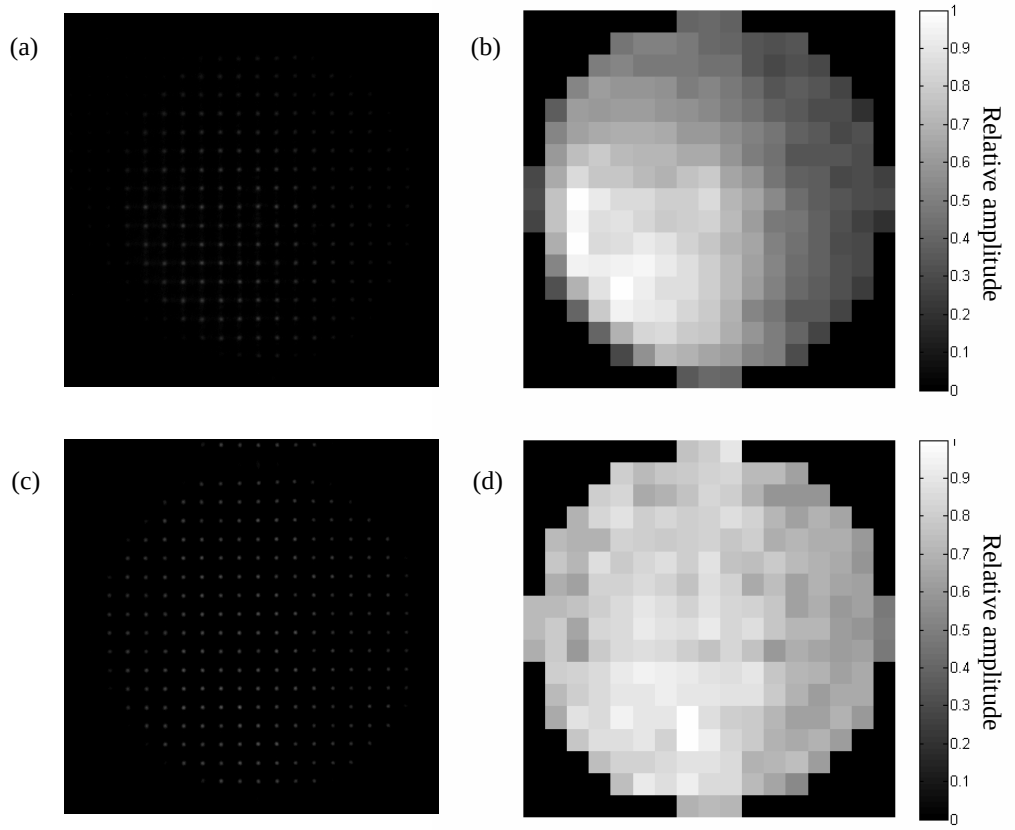


Figure 2.4: (a) Shack-Hartmann image acquired using AOSLOII for a healthy emmetropic eye. (b) Grayscale surface map of the distribution of amplitude (magnitude) at each subaperture across the pupil based on the square root of the mean intensity count of the spot. (c) - (d) Shack-Hartmann image and irradiance distribution for a model eye.

2.1.3 Point spread function

We continue the discussion on wave optics with the propagation of light through the eye's optical system in order to determine the field distribution on the retina. Consider first the simplest case where the eye's back focal plane coincides with the retina as illustrated in Figure 2.5. Light from a point object at infinity generates a plane wave at the pupil plane P. Given the pupil function, how do we estimate the field distribution, and more importantly the irradiance distribution, on the retina? A complete treatment of matter would require a rigorous discussion on diffraction theory which is beyond what I have covered in my dissertation research. However, it is nevertheless fundamentally relevant for retinal imaging so it will be formally reviewed.

The Fresnel number for the eye is the number of π phase shifts that occur inside the exit pupil (defined as in Figure 2.5) as observed from the retina and is given by:

$$F = \frac{a^2}{\lambda f} \tag{2.15}$$

where *a* is the exit pupil radius and *f* is the distance between the exit pupil and the retina, which is the wave optics (as opposed to geometrical optics) definition of focal length.

Clearly the Fresnel number will differ between individuals and between the two eyes of the same individual, but we can still make a reasonable estimate on the range of values using simple models such as the popular Gullstrand-Le Grand schematic eye. AOSLO imaging is typically done (limited to until recently) over a 6 mm diameter entrance pupil using near-infrared (840 nm) light which converts to about 630 nm in the vitreous chamber (n = 1.336). Furthermore, the length of the eye ranges from 20 to 29 mm, and with the exit pupil located about 4 mm behind the corneal apex, a reasonable range for f would be 16 to 25 mm[54]. Assuming that the exit pupil is 10 percent smaller than the entrance pupil, the Fresnel number for a human eye should fall between 590 and 740. According to scalar diffraction theory[51], the Fraunhofer approximation is only valid when the Fresnel number for the system is much less than one, so at the very least, a Fresnel approximation should be used. Letting (x, y) and (ξ, η) be Cartesian coordinates in the entrance pupil and across the retina respectively and dropping the time-dependence term for now, the integral for Fresnel propagation uses to determine the field at the retina (amplitude PSF) is

$$\psi_{retina}(\xi,\eta,z)=\frac{1}{j\lambda z}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_{exit}(\frac{x}{m_p},\frac{y}{m_p},z)e^{\frac{jk}{2z}\left[(\xi-x)^2+(\eta-y)^2\right]}d\left(\frac{x}{m_p}\right)d\left(\frac{y}{m_p}\right) \qquad (2.16)$$

where ψ_{exit} is the field in the exit pupil and recall from above that m_p is pupil magnification (≈ 1.1) which is often overlooked when performing these calculations. The field in the exit pupil can be expressed in terms of the system transfer function (Equation 2.10 and 2.12) and the field in the entrance pupil

$$\psi_{exit}(\frac{x}{m_p},\frac{y}{m_p},z)=T(x,y)E_{ent}(x,y)e^{jkz} \qquad (2.17)$$

We can simplify the Fresnel diffraction integral by assuming that the incident plane wave has unit amplitude across the entire pupil. Substituting Equation 2.17 into Equation 2.16 with some rearranging of terms, we get:

$$\psi_{retina}(\xi,\eta,z)=-\frac{je^{jkz}}{\lambda z m_p^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}T(x,y)e^{\frac{jk}{2z}\left[(\xi-x)^2+(\eta-y)^2\right]}dxdy \qquad (2.18)$$

By substituting the general form (includes phase error ϕ) of the system transfer function (Equation 2.12) into Equation 2.18 and expanding the quadratic term in the argument of the exponent, we get.

$$\psi_{retina}(\xi,\eta,z)=-\frac{je^{jkz}}{\lambda z m_p^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}T_0(x,y)e^{-\frac{jk}{2f}(x^2+y^2)}e^{j\phi(x,y)}e^{\frac{jk}{2z}(\xi^2+\eta^2)}e^{-\frac{jk}{z}(x\xi+y\eta)}e^{\frac{jk}{2z}(x^2+y^2)}dxdy \qquad (2.19)$$

This expression could be greatly simplified by noticing that $z = f$ at the retinal plane. It should be clarified that f is the distance from the eye's exit pupil to the retinal plane, not the paraxial focal length of the eye's optical system. In other words, we would ideally want the back focal and retinal planes to coincide. If they do not, the difference must be accounted for by adding the appropriate amount of defocus on top of any other aberrations represented in the phase error term ϕ. Evaluating Equation 2.19 at $z = f$, we get the following for the amplitude PSF:

$$\psi_{retina}(\xi,\eta) = -\frac{je^{jkf}}{\lambda f m_p^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T_0(x,y)e^{j\phi(x,y)}e^{\frac{jk}{2z}(\xi^2+\eta^2)}e^{-\frac{jk}{z}(x\xi+y\eta)}dxdy$$
$$= -\frac{je^{jkf}e^{\frac{jk}{2f}(\xi^2+\eta^2)}}{\lambda f m_p^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T_0(x,y)e^{j\phi(x,y)}e^{-\frac{jk}{f}(x\xi+y\eta)}dxdy \quad (2.20)$$

which is proportional to the far-field (Fraunhofer) if the eye's optical transfer function did not have a focusing term. We can also write Equation 2.20 in terms of the pupil function (Equation 2.14) where it can be observed that the structure of the PSF is completely characterized by the pupil function.

$$\psi_{retina}(\xi,\eta) = -\frac{je^{jkf}e^{\frac{jk}{2f}(\xi^2+\eta^2)}}{\lambda f m_p^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y)e^{-\frac{jk}{f}(x\xi+y\eta)}dxdy \quad (2.21)$$

When considering retinal image quality, the irradiance PSF is actually more relevant as it represents the two-dimensional spatial impulse response of any imaging system. In SI unit (watts/m^2), irradiance is related to the field by:

$$S = \frac{1}{2}nc_0\varepsilon_0\left\langle|\psi|^2\right\rangle_t \quad (2.22)$$

where the refractive index of the vitreous $n \approx 1.336$, c_0 is the speed of light in vacuum and ε_0 is the permittivity of free space. However, the physical principle does not change if we choose not to include these extra constants in front of the final expression for the irradiance PSF.

$$S(\xi,\eta) = \frac{1}{\lambda^2 f^2 m_p^4}\left|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y)e^{-\frac{jk}{f}(x\xi+y\eta)}dxdy\right|^2 \quad (2.23)$$

Consider the special with a perfect eye with a circular symmetric pupil function. The pupil function in this ideal case will be completely real:

$$P(x,y) = circ(x,y) = \begin{cases}1 & \sqrt{x^2-y^2} < a \\ 0 & otherwise\end{cases} \quad (2.24)$$

Substituting Equation 2.24 into Equation 2.23 and converting into polar coordinates, the integral can be solved analytically, and the resultant irradiance PSF is the Airy pattern:

$$S_{Airy}(\rho) = \frac{4J_1(k\rho a/f)^2}{m_p^4(k\rho a/f)^2} \quad (2.25)$$

where $J_1(\bullet)$ is the Bessel function of the first kind, $\rho = \sqrt{\xi^2+\eta^2}$ and a is the radius of the exit pupil. Figure 2.5 shows the PSF specified by an Airy pattern.

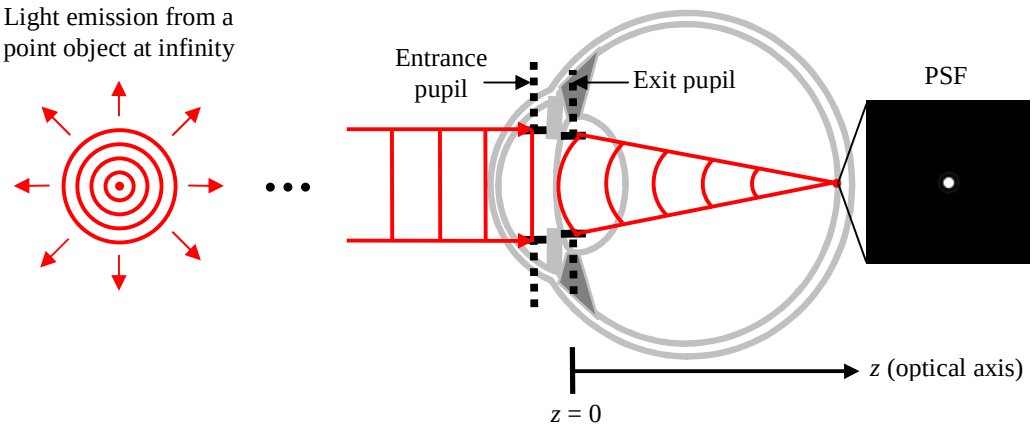


Figure 2.5: Formation of an image of a point source located at infinity for a perfect (diffraction limited) eye

When aberrations are present, the phase component of the pupil function is not zero, and the Equations for Fresnel diffraction (2.20 and 2.21) cannot be solved analytically and therefore must be evaluated numerically. If phase is accurately estimated from the wavefront sensor measurements and the amplitude transmittance is either known or modeled, the point spread function of the eye can than be computed via a Fourier transformation by considering the following substitutions into Equation 2.21:

$$\begin{cases} u = \dfrac{\xi}{\lambda f} \\ v = \dfrac{\eta}{\lambda f} \end{cases} \tag{2.26}$$

where u and v are horizontal and vertical spatial frequency coordinates defined such that if ξ and η were specified in mm, then u and v will be in cycles per mm. The integral for Fresnel propagation becomes

$$\psi_{retina}(\xi,\eta) = -\frac{je^{jkf}e^{j\pi\lambda f(u^2+v^2)}}{\lambda f m_p^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y)e^{-j2\pi(xu+yv)}dxdy \tag{2.27}$$

which is proportional to the standard expression for the two-dimensional Fourier transform:

$$H(u,v) = FT\{h(x,y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)e^{-j2\pi(xu+yv)}dxdy \tag{2.28}$$

where the function being transformed for our application is the pupil function. Initially, the terms outside of the integrals in Equation 2.27 may appear to be problematic. Rest assured that they are not and can be ignored since the numerator becomes 1 after taking the modulus square of amplitude PSF, and what is left is just the scalar term $\left(\lambda f m_p^2\right)^{-2}$. These calculations can be carried out using the following Matlab (The MathWorks, Inc., Natick, MA) commands:

```
psi = (1/(lambda*f*mp^2))*fftshift(fft2(fftshift(pupil_function)))      % Field at the exit pupil plane
no_pixels = sum(sum(pupil_function(pupil_function~=0)))                 % Number of pixels
```

```
S = psi.*conj(psi)./no_pixels;                    % Normalized irradiance PSF
```

The "fftshift" function needs to be applied to the pupil function prior to taking the Fourier transform because the origin in Matlab is in the top-left corner rather than the center. Another "fftshift" must be applied after taking the Fourier-transform so that the PSF can be plotted correctly. The normalization step makes it so that the highest value in the calculated irradiance PSF is the Strehl ratio.

Implementing the PSF calculations requires a two-dimensional array of numbers representing the wavefront departure from a perfect plane. Processing the wavefront sensor image to obtain the wavefront gradient and reconstructing the wavefront from its gradient is discussed in the following two sections of this chapter. For estimating the PSF, the reconstruction algorithms currently used for assessing AOSLOII performance are based on Zernike polynomials because they provide an analytical representation of the wavefront error (Figure 2.6a). The advantage of having an analytical expression is that the wavefront can be evaluated at as many points as necessary which is desirable to bring out the details in the PSF (Figure 2.6b). Basic zonal reconstruction algorithms, which I will argue to be more accurate in the sections below, evaluates the wavefront error at each subaperture (17 points across the pupil in AOSLOII) although sensible interpolation methods have been implemented outside of vision science[55]. Currently, the PSF is only used to monitor the Strehl ratio in real-time, so proper modeling aimed at determining the physical size of the computed PSF has taken lower priority. More often then not, a static value of 16.67 mm is used as the effective focal length of the eye and used to estimate the size of the PSF. As the need for more sophisticated image post-processing techniques (i.e deconvolution[56]) increases, we can expect more accurate models for reconstructing and scaling the PSF to be developed for AO retinal imaging.

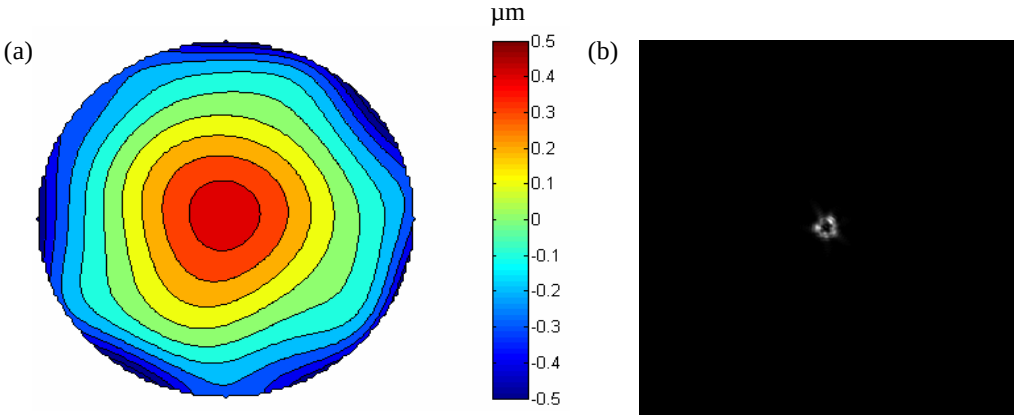


Figure 2.6: (a) Contour map representation of the wavefront error for a healthy emmetropic eye and (b) the corresponding irradiance PSF. The number of pixels across the diameter of the wavefront map was chosen somewhat arbitrarily to be 129.

2.2 Wavefront sensing

Now that the physical principles behind the ocular wavefront have been established, the next step is to produce meaningful data, namely the wavefront gradient, from a raw digital Shack-Hartmann image. Sharing similar history with most technical developments in AO (i.e. control systems, computationally efficient algorithms, turbulence modeling,

etc.), nearly all the literature on algorithm for estimating wavefront slopes for Shack-Hartmann sensors are written in the context of astronomical AO systems. Therefore, a thorough review of the literature on this topic for the purpose of identifying and testing concepts that may apply to both sky and ocular wavefront sensing would involve enough research effort to warrant an entirely separate dissertation. However, based on what has been described about AO retinal imagers that are currently online, there seems be surprisingly few advancements in the process of obtaining wavefront gradient measurements from the raw image beyond what was first described by Liang *et al.*[19] and some practical improvements can be made without much effort. This is a pity considering that the final wavefront measurement and control signal are computed from the wavefront gradient dictating the accuracy and precision of the wavefront correction.

As illustrated earlier in Figure 2.1, if the eye is placed along the optical axis such that the pupil plane coincides axially with the primarily focal point of first telescope lens, the field at the pupil will be relayed perfectly onto the lenslet array. Positioning of the eye in AOSLOII is carried out with a unit that contains a chin rest and temple mounts that is attached to a three axis positioning stage. Clearly, eye and head movements prevent the actual pupil from being positioned exactly on-axis and at the theoretical pupil plane, but the system appears to be insensitive to small (within a few mm) lateral and axial misalignments. At the other end of the telescope, the lenslet array samples the incoming field which produces the two-dimensional grid of spots on the CCD as shown in Figure 2.4. How these spot locations are related to the wavefront as well as the algorithm currently used in AOSLOII to estimate spot locations are discussed below. Since most of the work done in ocular wavefront sensing involve Zernike polynomials, I will begin with a discussion on their basic properties before covering the details on what I have implemented on AOSLOII for the purpose of improving the robustness and accuracy of the wavefront measurement process.

2.2.1 Zernike Polynomials

Describing the eye's wave aberrations with Zernike polynomials have become so ubiquitous that a naïve individual may be lead to believe that they are intrinsically tied in to the wavefront measurement process. This false impression is likely due to the fact that most ophthalmic Shack-Hartmann sensors are not being used in closed-loop AO systems but are stand alone systems that do little more than just recover and plot the wavefront error. Since the measurement accuracy of these devices cannot be evaluate by any absolute means such as the sharpness of an image of the foveal cone mosaic, making objective conclusions regarding their performance will be extremely difficult if not impossible. I will begin with an introduction regarding this popular set of basis functions and will continue the discussion on how they are incorporated into wavefront measurement and control in Sections 2.3 and 3.3.

An aberrated wavefront represented as a linear combination of Zernike polynomials is described mathematically by:

$$\phi(x,y) = \sum_{j=0}^{N-1} c_j Z_j(x,y) \tag{2.29}$$

where N is the total number of Zernike modes (the first N modes in this case) and j is called the mode number. The set of coefficients $\{c_j\}_{j=0}^{N}$ determines the weight of each

particular Zernike mode in terms of how much it contributes to the structure of the wavefront. Different authors have indexed these polynomials in different ways, so to avoid confusion the standard Optical Society of America indexing scheme[57] has been adopted. Zernike polynomials up to 4th order (first 14 modes not including piston, the 0th mode) are plotted in Figure 2.7 where each Zernike mode Z_n^m is indexed with *m* being the angular frequency and *n* being the radial order (or simply "order"). The order is defined as the highest degree in the set of polynomials (the power of *x* or *y*), so fundamentally it determines the smoothness of the final reconstructed wavefront. The choice for the highest radial order to fit can be somewhat subjective and application dependent but is ultimately limited by sampling density of the lenslet array. For example, if only refraction (sphere and cylinder) is of interest, only a second order representation is necessary. On the other hand, if we were interested in how the tear film affects the wavefront profile, including Zernike modes up to 10th order may not even suffice[24, 25]. Once the highest radial order has been decided, the total number of modes can be found by:

$$N = \frac{(n+1)(n+2)}{2} \tag{2.30}$$

However, we must ignore the modes piston (not plotted in Figure 2.7), tilt $\left(Z_1^{-1}\right)$ and tip $\left(Z_1^{1}\right)$ because they cannot actually be measured using a Shack Hartmann sensor, so the actual number of modes estimated is *N* - 3. This is acceptable because these modes do not affect the final retinal image quality. The indices *m* and *n* can be combined into a single index called the mode index which I used in Equation 2.29:

$$j = \frac{n(n+2)+m}{2} \tag{2.31}$$

Especially in the computations steps involved in wavefront reconstruction, having multiple indices is only redundant when the purpose is only to track the order of a set functions or vectors.

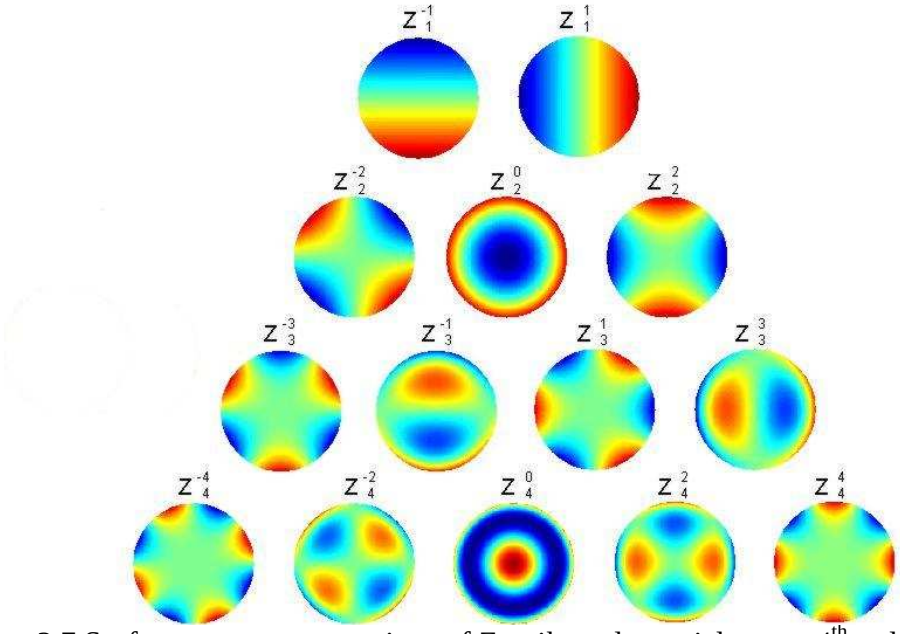


Figure 2.7 Surface map representations of Zernike polynomials up to 4th order using the OSA indexing scheme[57].

Zernike polynomials come in different forms (typically distinguishable only by the type of normalization used). The vision community has more or less settled on using the form recommended by ANSI which are normalized in such a way that the set of functions form a complete orthonormal (as opposed to only orthogonal) basis over a unit circle. Since the area of a unit circle is π, the inner product is defined as:

$$\left(Z_i, Z_j\right) \triangleq \frac{1}{\pi}\int_{\Sigma} Z_i(x,y) Z_j(x,y)\,dxdy = \delta_{ij} \tag{2.32}$$

where Σ is the pupil domain and δ_{ij} is the Kronecker delta:

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{2.33}$$

The pupil coordinates (x, y) clearly must be normalized by the pupil radius (i.e. $\rho = \sqrt{x^2 + y^2} \leq 1$) prior to evaluating Equations 2.29 and 2.32. Table 2.1 gives analytical expressions of Zernike polynomials up to 4th order in both the more common polar coordinates as well as in Cartesian coordinates. These equations can be automatically generated in closed form using somewhat complex formulae as explained in Noll's classic paper on Zernike polynomials and atmospheric turbulence[58]. But in practical computer implementations, it will be more computationally efficient to simply hard code all the equations up to the maximum order desired. This way, each equation will not have to be regenerated every time they are needed.

Table 2.1 Orthonormal Zernike polynomials up to 4th order

j	Polar coordinates	Cartesian coordinates
0	1	1
1	$2\rho\cos\theta$	$2y$
2	$2\rho\sin\theta$	$2x$
3	$\sqrt{6}\rho^2\sin 2\theta$	$2\sqrt{6}xy$
4	$\sqrt{3}\left(2\rho^2-1\right)$	$\sqrt{3}\left(x^2+y^2-1\right)$
5	$\sqrt{6}\rho^2\cos 2\theta$	$\sqrt{6}\left(x^2-y^2\right)$
6	$\sqrt{8}\rho^3\sin 3\theta$	$2\sqrt{2}y\left(3x^2-y^2\right)$
7	$\sqrt{8}\left(3\rho^3-2\rho\right)\sin\theta$	$2\sqrt{2}y\left(3x^2+4y^2-2\right)$
8	$\sqrt{8}\left(3\rho^3-2\rho\right)\cos\theta$	$2\sqrt{2}x\left(3x^2+4y^2-2\right)$
9	$\sqrt{8}\rho^3\cos 3\theta$	$2\sqrt{2}x\left(x^2-3y^2\right)$
10	$\sqrt{10}\rho^4\sin 4\theta$	$4\sqrt{10}xy\left(x^2-y^2\right)$
11	$\sqrt{10}\left(4\rho^4-3\rho^2\right)\sin 2\theta$	$2\sqrt{10}xy\left(4y^2+4x^2-3\right)$
12	$\sqrt{5}\left(6\rho^4-6\rho^2+1\right)$	$\sqrt{5}\left(6(x^2+y^2)^2-6(x^2+y^2)+1\right)$
13	$\sqrt{10}\left(4\rho^4-3\rho^2\right)\cos 2\theta$	$\sqrt{10}\left(x^2-y^2\right)\left(4(x^2+y^2)-3\right)$
14	$\sqrt{10}\rho^4\cos 4\theta$	$\sqrt{10}\left(x^4-6x^2y^2+y^4\right)$

2.2.2 Wavefront gradient

It is more reasonable to define the wavefront gradient, rather than the reconstructed wavefront, as the variable measured by a Shack-Hartmann sensor because information is lost during wavefront reconstruction (estimation). Nevertheless, a significant amount of computations are required just to get to that stage. The wavefront gradient is as:

$$\nabla\phi(x,y)=\left(\frac{\partial\phi(x,y)}{\partial x},\frac{\partial\phi(x,y)}{\partial y}\right) \tag{2.34}$$

For notation convenience throughout the rest of this dissertation, I will denote the first derivation of the wavefront by:

$$\begin{cases}\phi^x(x,y)\triangleq\dfrac{\partial\phi(x,y)}{\partial x}\\ \phi^y(x,y)\triangleq\dfrac{\partial\phi(x,y)}{\partial y}\end{cases} \tag{2.35}$$

Depending on the sensor model selected, most wavefront reconstructor designs make the assumption of Equation 2.29 sampled over the pupil at the subaperture centers even though the average gradient over each subaperture would be a much more accurate model[59-61]. How is the wavefront gradient related to the focused spot locations in the Shack-Hartmann image? The answer is entirely trigonometric.

Table 2.2: AOSLOII wavefront sensor properties. The frame-rate is dependent on the user's chosen exposure time which general depend on the reflectivity of the retina imaged.

Subaperture diameter	328 µm
Subaperture focal length	24000 µm
CCD pixel pitch	12.9 µm
Pupil magnification	0.889
Frame-rate	< 25 Hz
Image dimensions	512 × 512 pixels (after 2 × 2 binning)

Figure 2.8 illustrates how the spot pattern image is formed. A plane wave incident on an array of micro-lenses will form a regular grid of spots, which I will call the reference spot pattern. The locations of these focused spot images need not be measured as they can be generated mathematically if pupil magnification and subaperture diameter are known (Table 2.2).When the incident field involves an aberrated wavefront (Figure 2.8b), a particular spot location may depart from its corresponding reference point due to the wavefront having non-zero slope (on average) across that particular subaperture. Based on similar triangles, the relationship between each of these departures and the corresponding average wavefront slopes is

$$\begin{cases} \left\langle \frac{\partial \phi(x,y)}{dx} \right\rangle = \frac{1}{area(\Sigma)} \int_{\Sigma} \frac{\partial \phi(x,y)}{dx} dxdy = \frac{m}{f} \Delta x \\ \left\langle \frac{\partial \phi(x,y)}{dy} \right\rangle = \frac{1}{area(\Sigma)} \int_{\Sigma} \frac{\partial \phi(x,y)}{dy} dxdy = \frac{m}{f} \Delta y \end{cases} \tag{2.36}$$

where Δx and Δy are departures from the reference spot pattern, m is the pupil magnification, f is subaperture focal length, and Σ is the domain for a particular subaperture. These parameters for AOSLOII are given in Table 2.1. Although not strictly required, it is advisable to store wavefront gradient data in physical units (as opposed to pixels) as wavefront calculation down the line will be much more intuitive. Due to the gradient field containing both x and y components, there will be twice as many measurements as there are samples of the wavefront. In AOSLOII, the field across a 6 mm diameter pupil is sampled by 213 subapertures so the "output" of the wavefront sensor is a 426 element vector.

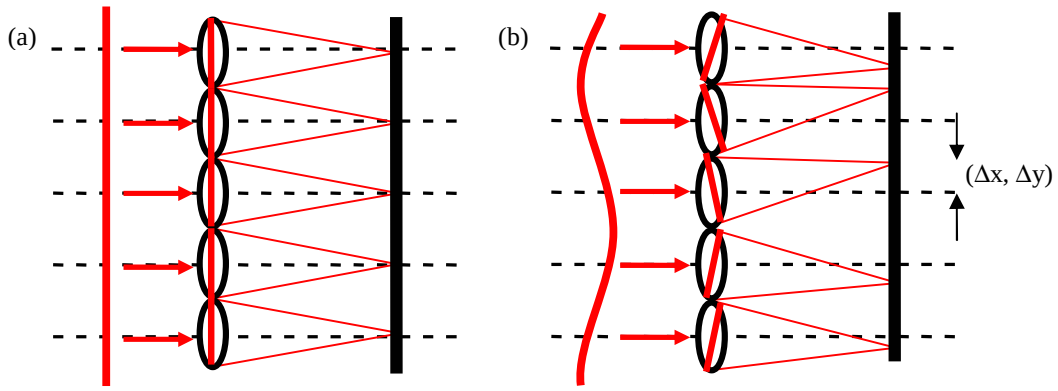


Figure 2.8: (a) Plane wave produces zero local wavefront slope across each subaperture so the resultant pattern of focused spots fall on a perfect grid. The average slope of the wavefront over a particular subaperture when the incident beam is aberrated (b) is generally non-zero and the corresponding focused spot image will displace from the reference grid.

In the case where the wavefront is expressed as a linear combination of Zernike polynomials, the wavefront gradient must be express in Zernike terms as well. By taking the gradient of Equation 2.29, we get:

$$\nabla\phi(x,y)=\sum_{j=0}^{N-1}c_j\nabla Z_j(x,y) \tag{2.37}$$

This leads to separate expressions for x and y components of the wavefront gradient as indicated in Equation 2.34. The x and y wavefront slopes can be expressed as:

$$\begin{cases}\phi^x(x,y)=\sum_{j=0}^{N-1}c_j\dfrac{\partial Z_j(x,y)}{\partial x}\\ \phi^y(x,y)=\sum_{j=0}^{N-1}c_j\dfrac{\partial Z_j(x,y)}{\partial y}\end{cases} \tag{2.38}$$

In Table 2.1, we saw that Zernike polynomials can be expressed analytically in Cartesian coordinates. This allows us express the partial derivatives in Equation 2.38 analytically as well. The x and y derivatives of individual Zernike modes up to 4th order are given in Table 2.3. Again, for the purpose of making the notation more convenient, I will denote Zernike polynomial derivatives with:

$$\begin{cases}Z_j^x(x,y)\triangleq\dfrac{\partial Z_j(x,y)}{\partial x}\\ Z_j^y(x,y)\triangleq\dfrac{\partial Z_j(x,y)}{\partial y}\end{cases} \tag{2.39}$$

Table 2.3: Analytically generated Zernike polynomial derivatives up to 4th order

j	Z_j^x	Z_j^y
1	0	2
2	2	0
3	$2\sqrt{6}y$	$2\sqrt{6}x$
4	$4\sqrt{3}x$	$4\sqrt{3}y$
5	$2\sqrt{6}x$	$-2\sqrt{6}y$
6	$6\sqrt{8}xy$	$\sqrt{8}y\left(3x^2-3y^2\right)$
7	$6\sqrt{8}xy$	$\sqrt{8}y\left(3x^2+9y^2-2\right)$
8	$\sqrt{8}\left(9x^2+3y^2-2\right)$	$6\sqrt{8}xy$
9	$\sqrt{8}\left(3x^2-3y^2\right)$	$-6\sqrt{8}xy$
10	$\sqrt{10}\left(12x^2y-4y^3\right)$	$\sqrt{10}\left(4x^3-12xy^2\right)$
11	$\sqrt{10}\left(24x^2y+8y^3-6y\right)$	$\sqrt{10}\left(8x^3+24xy^2-6x\right)$
12	$\sqrt{5}\left(24x^3+24xy^2-12x\right)$	$\sqrt{5}\left(24x^2y+24y^3-12y\right)$
13	$\sqrt{10}\left(16x^3-6x\right)$	$\sqrt{10}\left(6y-16y^3\right)$
14	$\sqrt{10}\left(4x^3-12xy^2\right)$	$\sqrt{10}\left(4y^3-12x^2y\right)$

2.2.3 Shack-Hartmann image analysis

Raw Shack-Hartmann sensor data is a digital image of a spot array pattern with their saliency depending on a number of factors including the quality of the light source, reflectivity of the retina, and the aberrations of the eye. Since how the position each spot in a Shack-Hartmann image displaces from its reference position is proportional to the local wavefront slope, the accuracy to which these spots can be located is extremely important as any error will affect all calculation that will follow. The estimated spot locations are called centroids. Many different types of centroid finding algorithms, some of which can be quite complex[52, 62], have been proposed, so the definition of centroid used here is more of an umbrella term that generalizes many different ways for calculating spot locations rather than the formal mathematical definition, which is just the geometric center of a set of points. Singer[63] provides a good description on AO software for the flood-illuminated AO system at the University of Rochester where he breaks down the major computations into three step that I will parallel: (1) setting regions of interest, (2) preparing the (Shack-Hartmann) image and (3) calculating centroids. These steps together currently encompass more than half of the computation overhead for real-time control of AOSLOII taking typically more than 4.5 milliseconds (Figure 2.10).

Local regions of interest, or search boxes, are first generated from the reference grid (where the spots would theoretically be located in an aberration-free system). Search boxes are initialized to be as large as possible as long as adjacent boxes do not overlap. This corresponds to be about 23 by 23 pixels (296 by 296 μm). The search boxes are also drawn on top of the Shack-Hartmann image for the user to see as shown in Figure 2.9b.

This step has become a popular feature due to practical reasons as often times with real eyes, a spot may fall partially or completely outside of its corresponding search box. Being able to visualize the location and size of the search boxes quickly prompts the operator to adjust the pupil location (regenerating the reference grid), adjust the subject's head position, and/or change the spectacle correction if the individual being imaged has a large amount of low order aberrations that cannot be corrected with the limited stroke of the MEMS DM. The latter operations could potentially not be necessary for imagers that employ badal optometers and high-stroke DMs for low order wavefront correction[5, 64].

Preparing the Shack-Hartmann image prior to executing more sensitive calculations is desirable as a clean high-contrast spot pattern, free from spurious reflections and noise, insures that a spot location will be found accurately. However as it intended to push AO equipped retinal imagers to clinical deployment, it is undesirable for imaging requirements to be so stringent that the room must be pitch dark and that only "good subjects" (low aberrations, good fixation ability, highly reflective retina, etc) can be reliably imaged. With this in mind, we have included additional calculation steps to make the wavefront slope measurements more robust to many unforeseen circumstances (spurious reflections, dim retinas, blurry spots, etc.). These include CCD exposure time that can be adjusted through the user interface, background image subtraction and an iterative type centroid finding algorithm[23, 47, 65]. These features were most likely not coded into the software that operated the earliest systems[1, 19, 20] since they are only secondary when it comes to demonstrating the functionality of a Shack-Hartmann sensor and that some of the computations may have been too burdensome for the computers available then. As mentioned by Singer[63] regarding the software for the flood-illuminated system in Rochester that it was more preferable at that time to set up excellent imaging condition rather than add more image processing calculations that may potentially lower the system bandwidth.

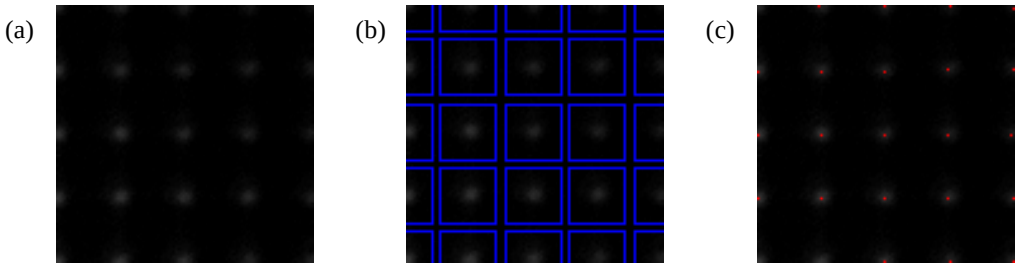


Figure 2.9: Steps in analyzing a Shack-Hartmann image (a) Close up view of the Shack-Hartmann image from Figure 2.4a. (b) Same image with the search boxes displayed for the user to see. (c) Estimated centroids which were rounded to the nearest pixel for display purposes.

The last step is the actual centroid calculation which is done on the now processed Shack-Hartmann image. The algorithm used by Liang *et al.*[19] was the center of mass on the intensity counts in each search box:

$$\begin{cases} x_c(k) = \dfrac{\sum_j x_j I_j}{\sum_j I_j} \\ y_c(k) = \dfrac{\sum_j y_j I_j}{\sum_j I_j} \end{cases} \tag{2.40}$$

where I is intensity count which is summed over the support of the search box of interest and k is iteration number which was originally just one. The accuracy of the centroid finding algorithm can be improved by doing the center of mass calculation iteratively.[23, 47, 63] This process involves shrinking the search box and then re-centering it around the previously computed centroid location before computing the center of mass again and doing this over and over again. If a spot lied very close to an edge of the search box, then one side of the spot intensity distribution may fall outside of the search box biasing the initial center of mass estimate. In this situation, an iterative center of mass algorithm will clearly result in a more accurate estimate since the search box will gradually become better centered over the spot. However, it may take a lot of iterations before the search box becomes nearly centered and accurate estimate is produced. A practical fix that we have adopted is to first find the location of the pixel with the highest intensity count, which we will refer to as the max-in-box operation:

$$\left(x_c(1), y_c(1)\right) = \left\{x, y \middle| I(x,y) = \max\left(I_j\right), \forall j \in \Sigma\right\} \tag{2.41}$$

where j is used to index the pixels in search box Σ. The max-in-box operation only replaces the first center of mass calculation since it is a lot less sensitive to misaligned search boxes than the center of mass, but it is also less accurate when the search box is well-centered as it completely ignores the irradiance distribution of the spot and is only sensitive to the nearest pixel. With this max-in-box and iterative center of mass combination, I have found empirically that the repeatability of the centroid measurement does not really improve beyond three iterations (four total if including the first max-in-box operation) as compared to the six or more iterations described in Hofer *et al.*[23]. Each newer, smaller box is formed by reducing all four center-to-side distances by two pixels. Hofer stated that centroid detection was most accurate and repeatable when the final search box was about the same width as the diffraction pattern for a single subaperture[23]. By inspection of our Shack-Hartmann images, the irradiance distribution of the individual spots seems to resemble more of a Gaussian distribution than that of a far-field pattern of a square aperture (having lenslets over the subapertures effectively brings the far-field pattern to the near field). In any case, the diameter of the diffraction pattern of a single lenslet is about 120 μm which spans across roughly 10 pixels, and the size of our last search box is 11 by 11 pixels. Lastly, it is worth mentioning that we also employ a global threshold condition whose value is made accessible through the user interface. The purpose of this threshold value is to roughly determine whether pixel values in a particular subaperture are due to noise or not. If not enough pixels (default is less than five pixels) meet the selected threshold criteria, then the centroid is considered not to be found and the master list of stored centroid locations in the program will not get updated for that particular entry. What this does is to ensure that when a centroid cannot be

computed, whatever measured value that was from the previous frame actually gets used. I credit much of the current level of robustness of the control loop for AOSLOII to this modification as it is almost completely immune to poor measurements caused by actions such as blinking.

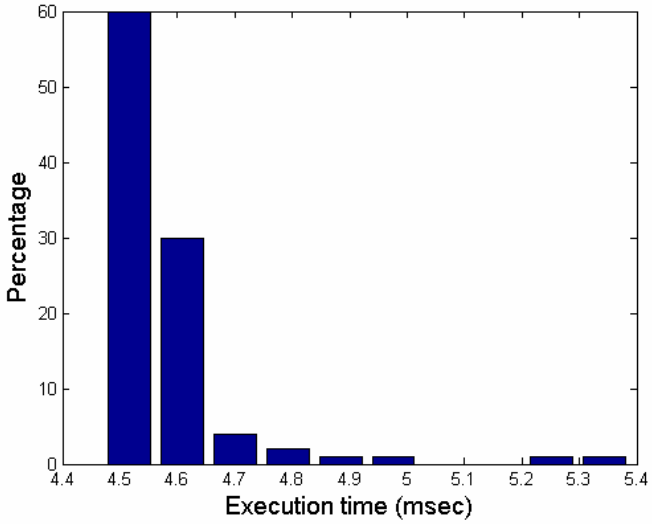


Figure 2.10: Amount of time taken to measure the wavefront gradient (subtract background image, execute centroid finding algorithm, and calculate wavefront slopes) and all so plot the centroid locations on the user interface (Figure 2.9c).

2.3 Wavefront reconstruction (estimation)

In modern texts, wavefront reconstruction can refer to either estimating the wavefront profile from sensor measurement (wavefront gradient) or calculating the commands for reproducing the wavefront profile with an active optical element such as a DM. Strictly speaking, the former case is nothing more than an integration problem albeit over a circular domain (pupil) and a finite sample of gradient measurements. In some system designs, the two cases may differ only by a scaling factor, so they become the same problem (i.e. AO systems with matching subaperture and DM actuator arrays along with zero crosstalk between actuators). As for others, such as all the AO retinal imagers currently online which have many more subapertures than actuators, the two cases become quite distinct and therefore must be treated separately. This section is dedicated to only the wavefront estimation type of reconstructor and their application in vision science. Wavefront reconstruction for AO control is covered in Chapter 3.

The majority of the literature on wavefront reconstruction was not written with applications in visual optics in mind. In fact, wavefront reconstructors in general have received scant attention since the first ocular Shack-Hartmann sensor was developed by Liang *et al.*[19] giving the impression that the first algorithm applied to ocular wavefront sensing is also the standard. However, Liang himself, according to his classic paper[19], recognized that there were many different approaches to the wavefront reconstruction problem, and that he simply chose to use the modal estimation method developed by Cubalchini[61]. The lack of rigorous work on wavefront reconstruction associated with the study of the eye's optics provided the motivation to undertake this research. The rest of this chapter lays out the mathematical foundations for three wavefront reconstruction

(estimation) algorithms, loosely categorized as being based on either Zernike, zonal and/or Fourier, which were implemented as part of this dissertation. The discussions below will begin with an analytical description of the problem before moving on to constructing the equivalent discrete models that can be implemented on a computer and solved numerically.

2.3.1 The classic Cubalchini method

Noll is usually credited with the first rigorous description of wavefront correction and Zernike polynomials[58], but Cubalchini is very likely to have been the first person to analyze the numerical problem on calculating Zernike coefficients from Shack-Hartmann type measurements[61]. For this reason, I will refer to the wavefront reconstructor he described as the Cubalchini method, which is a modal reconstructor in the sense that the quantities estimated are the individual Zernike coefficients. Once the coefficients are determined, the estimated wavefront can then be generated via Equation 2.29 using the analytical expressions of the Zernike polynomials (Table 2.1). The Cubalchini method was the reconstructor of choice for the very first ocular Shack-Hartmann sensor developed by Liang *et al.*[19], and is to my knowledge still the most widely, although not always appropriately, used algorithm in vision science as well as in applications in related disciplines.

Some basic properties of Zernike polynomials were stated earlier in Section 2.2.1, and two different indexing schemes (a double and a single) were introduced. For estimation purposes, only a single index is necessary so only the Zernike mode number will be used. If the wavefront can be measured directly, fitting Zernike polynomials to a wavefront is a standard two-dimensional curve fitting problem. Given a function representing the wavefront (ϕ), the i^{th} Zernike coefficient is calculated by evaluating the inner product (Equation 2.32) of the wavefront with the i^{th} Zernike polynomial:

$$\left(\phi, Z_i\right) = \frac{1}{\pi}\int_{\Sigma}\phi(x,y)Z_j(x,y)dxdy \tag{2.42}$$

Substituting Equation 2.29 for function ϕ, we get:

$$\left(\phi, Z_i\right) = \frac{1}{\pi}\sum_{j=0}^{N-1}c_j\int_{\Sigma}Z_j(x,y)Z_i(x,y)dxdy \tag{2.43}$$

We can immediately see that the integral in Equation 2.43 vanishes due to the orthonormality of Zernike polynomials except for when $i = j$. Therefore,

$$c_i = \left(\phi, Z_i\right) \tag{2.44}$$

which can be generalized to calculating any number of Zernike coefficients by:

$$\mathbf{c} = \begin{bmatrix} c_3 \\ c_4 \\ \vdots \\ c_{N-1} \end{bmatrix} = \begin{bmatrix} \left(\phi, Z_3\right) \\ \left(\phi, Z_4\right) \\ \vdots \\ \left(\phi, Z_{N-1}\right) \end{bmatrix} \tag{2.45}$$

As mentioned earlier, the first three Zernike modes (piston, tilt and tip) cannot be properly measured with a Shack-Hartmann sensor and do not affect retinal image quality, so they are always omitted during wavefront reconstruction.

Unfortunately, we cannot apply Equation 2.44 directly to wavefront reconstruction because the Shack-Hartmann sensor does not directly measure the

wavefront but rather the wavefront gradient. The result obtained in Equation 2.45 is also a special case for the solution to a linear least squares problem where the cost functional minimized is:

$$J(\mathbf{c}) = \frac{1}{2}\left\| \phi(x,y) - \sum_{j=0}^{N-1} c_j Z_j(x,y) \right\|^2 \tag{2.46}$$

$$= \frac{1}{2\pi}\int_{\Sigma}\left(\phi(x,y) - \sum_{j=0}^{N-1} c_j Z_j(x,y) \right)^2 dxdy$$

$$= \frac{1}{2}(\phi,\phi) - \sum_{j=0}^{N-1} c_j (\phi, Z_j) + \frac{1}{2}\sum_{j=0}^{N-1} c_j^2$$

which is proportional to the variance of the wavefront fitting error. The argument of the cost functional (**c**) is the vector of Zernike coefficients. Consider the estimation of an arbitrary Zernike coefficient c_i, simply differentiate the cost function with respect to c_i and set the resultant expression equal to zero:

$$\frac{\partial J(\mathbf{c})}{\partial c_i} = -(\phi, Z_i) + c_i = 0 \tag{2.47}$$

which is the same as Equation 2.44. The same logic applies when dealing with the gradient of both the wavefront and the set of Zernike polynomials. The cost functional can still be defined as the fitting error, but in terms of the wavefront gradient:

$$J(\mathbf{c}) = \frac{1}{2}\left\| \begin{bmatrix} \phi^x(x,y) \\ \phi^y(x,y) \end{bmatrix} - \sum_{j=0}^{N-1} c_j \begin{bmatrix} Z_j^x(x,y) \\ Z_j^y(x,y) \end{bmatrix} \right\|^2 \tag{2.48}$$

$$= \frac{1}{2\pi}\int_{\Sigma}\left(\phi^x(x,y) - \sum_{j=0}^{N-1} c_j Z_j^x(x,y) \right)^2 dxdy + \frac{1}{2\pi}\int_{\Sigma}\left(\phi^y(x,y) - \sum_{j=0}^{N-1} c_j Z_j^y(x,y) \right)^2 dxdy$$

$$= \frac{1}{2}\left[(\phi^x,\phi^x) + (\phi^y,\phi^y) \right] - \sum_{j=0}^{N-1} c_j \left[(\phi^x, Z_j^x) + (\phi^y, Z_j^y) \right] + \frac{1}{2}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1} c_j c_k \left[(Z_j^y, Z_k^y) + (Z_j^x, Z_k^x) \right]$$

Clearly the residual wavefront gradient version of the cost functional is less appealing than the just residual wavefront version. This is primarily due to the fact that we now need to minimize the fitting error for both the x and y components of the wavefront gradient. This task is further complicated by the fact that higher order ($n > 2$) Zernike modes lose their orthonormality if the derivative is taken[61, 66].

$$\begin{cases} (Z_i^x, Z_j^x) \neq \delta_{ij} & i > 5 \text{ or } j > 5 \\ (Z_i^y, Z_j^y) \neq \delta_{ij} & i > 5 \text{ or } j > 5 \end{cases} \tag{2.49}$$

Nevertheless, we can still derive an expression for an arbitrary coefficient c_i in more or less the same manner by differentiating Equation 2.48 with respect to c_i and setting the resultant equation equal to zero.

$$\frac{\partial J(\mathbf{c})}{\partial c_i} = -(\phi^x, Z_i^x) - (\phi^y, Z_i^y) + \sum_{k=0}^{N-1} c_i \left[(Z_k^x, Z_i^x) + (Z_k^y, Z_i^y) \right] = 0 \tag{2.50}$$

This result is a system of linear equations that can be expressed in matrix form:

$$\begin{bmatrix} \left(Z_3^x, Z_3^x\right)+\left(Z_3^y, Z_3^y\right) & \left(Z_3^x, Z_4^x\right)+\left(Z_3^y, Z_4^y\right) & \cdots & \left(Z_3^x, Z_{N-1}^x\right)+\left(Z_3^y, Z_{N-1}^y\right) \\ \left(Z_4^x, Z_3^x\right)+\left(Z_4^y, Z_3^y\right) & \left(Z_4^x, Z_4^x\right)+\left(Z_4^y, Z_4^y\right) & & \vdots \\ \vdots & & \ddots & \vdots \\ \left(Z_{N-1}^x, Z_3^x\right)+\left(Z_{N-1}^y, Z_3^y\right) & \left(Z_{N-1}^x, Z_4^x\right)+\left(Z_{N-1}^y, Z_4^y\right) & \cdots & \left(Z_{N-1}^x, Z_{N-1}^x\right)+\left(Z_{N-1}^y, Z_{N-1}^y\right) \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{N-1} \end{bmatrix} = \begin{bmatrix} \left(\phi^x, Z_3^x\right)+\left(\phi^y, Z_3^y\right) \\ \left(\phi^x, Z_4^x\right)+\left(\phi^y, Z_4^y\right) \\ \vdots \\ \left(\phi^x, Z_{N-1}^x\right)+\left(\phi^y, Z_{N-1}^y\right) \end{bmatrix} \tag{2.51}$$

which I will represent generically by $\mathbf{Bc} = \mathbf{d}$ where $\mathbf{B} \in \mathbb{R}^{N \times N}$, $\mathbf{c} \in \mathbb{R}^{N}$ and $\mathbf{d} \in \mathbb{R}^{2N_s}$ where N_s is the number of subapertures (N_s = 213 for 6 mm diameter pupil). Each entry of matrix **B** can be determined analytically[58] but the entries of vector **d** must be computed numerically since ϕ^x *and* ϕ^y are measured quantities. Once Equation 2.52 is properly set up, obtaining the best estimate of vector **c** is just a matrix inversion:

$$\hat{\mathbf{c}} = \mathbf{B}^{-1}\mathbf{d} \tag{2.52}$$

The derivations thus far have been primarily analytical. The Cubalchini method bypasses these tedious analytical steps needed to arrive at Equation 2.51 by setting up the problem in discrete space from the beginning. When working in the discrete domain, functions, such as the Zernike polynomials, are discretized into vectors by evaluating each function at the sampled locations in the pupil and vectorizing the result. When that is done, we can express the sampled wavefront gradient fitting error in the form:

$$\mathbf{e} = \mathbf{y} - \mathbf{Z}'\mathbf{c} \tag{2.53}$$

where the measurement vector **y** can be defined as the measured x- and y-slopes vectorized and stacked one on top of the other:

$$\mathbf{y} \triangleq \begin{bmatrix} \phi^x \\ \phi^y \end{bmatrix} \in \mathbb{R}^{2N_s} \tag{2.54}$$

The columns of matrix **Z'** are the vectorized Zernike polynomial gradients with x-derivatives stacked on top of the y-derivatives (the order of stacking must be consistent with the measurement vector **y**):

$$\mathbf{Z}' \triangleq \begin{bmatrix} Z_3^x & Z_4^x & \cdots & Z_{N-1}^x \\ Z_3^y & Z_4^y & \cdots & Z_{N-1}^y \end{bmatrix} \in \mathbb{R}^{2s \times N-3} \tag{2.55}$$

The linear least squares problem can now be constructed by minimizing the square of the wavefront gradient fitting error:

$$J(\mathbf{c}) = \frac{1}{2}\mathbf{e}^{\mathbf{T}}\mathbf{e} \tag{2.56}$$

The result of from differentiating Equation 2.56 with respect to vector **c** and setting it to zero is the best estimate of the Zernike coefficient vector **c** in the sense of least squares.

$$\hat{\mathbf{c}} = \left(\mathbf{Z}'^{\mathbf{T}}\mathbf{Z}'\right)^{-1}\mathbf{Z}'^{\mathbf{T}}\mathbf{y} \tag{2.57}$$

where the columns of $\left(\mathbf{Z}'^{\mathbf{T}}\mathbf{Z}'\right)$ are linearly independent, making the matrix invertible, as long as the number of estimated Zernike coefficients does not exceed the number of subapertures[67]. By inspection, we can see that Equation 2.58 is a convenient approximation of Equation 2.52:

$$\begin{cases} \mathbf{B} \approx \left(\mathbf{Z}'^{\mathbf{T}}\mathbf{Z}'\right) \\ \mathbf{d} \approx \mathbf{Z}'^{\mathbf{T}}\mathbf{y} \end{cases} \tag{2.58}$$

The matrix $\left(\mathbf{Z}'^{\mathbf{T}}\mathbf{Z}'\right)^{-1}\mathbf{Z}'^{\mathbf{T}}$ is the wavefront reconstructor which Cubalchini refers to as the least squares estimator. Cubalchini used slightly different notation in all his

derivations, but the mathematics are identical. The notation we use is more straightforward to implement in modern numerical programming environments such as Matlab. For most systems, the reconstructor only needs to be computed once and then stored. The number of visual optics papers that explicitly state that they calculate Zernike polynomial coefficients using Cubalchini's method is surprisingly few[19, 47, 68-75], while the majority of visual optics studies only refer to Liang *et al.*[19] without elaborating further on their choice of reconstruction method. Since no work in vision to my knowledge has explicitly stated that the method implemented for calculating Zernike coefficients differed from Liang's original work, it is reasonable to assume that Zernike coefficients to date have been calculated using Cubalchini's method.

2.3.2 Zonal reconstruction

Zernike polynomials have been discussed in great detail so far, but looking back at the original problem, what we have is a model stating that the measured quantity is equal to the wavefront gradient plus noise, which can be modeled as:

$$\left(y_x(x,y), y_y(x,y)\right) = \nabla\phi(x,y) + \left(v_x(x,y), v_y(x,y)\right) \tag{2.59}$$

where v_x and v_y are measurement noise. Formally speaking, the boundary conditions imposed on Equation 2.59 are of the Neumann type over the circular boundary defined by the pupil. Given this description, are Zernike polynomials, or any other set of polynomials for that matter, required to solve Equation 2.59? Zonal wavefront reconstruction describes a set of algorithms where the wavefront is computed directly from wavefront gradient measurements without assistance for any particular set of analytical basis functions. Equation 2.59 cannot be discretized and solved directly because the gradient operator is not self-adjoint. The only property of interest is that self-adjoint operators, when expressed in matrix form, are symmetric, so a rigorous discussion on this subject is unnecessary. However, the Laplace operator is self-adjoint. Taking the divergence of both sides of Equation 2.59 allows us to assess the same problem but involving a Laplacian instead of the gradient operator:

$$\begin{aligned} \nabla^2\phi(x,y) &= f(x,y) \\ &= \nabla\bullet\left(y_x(x,y), y_y(x,y)\right) - \nabla\bullet\left(v_x(x,y), v_y(x,y)\right) \\ &= \left(\frac{\partial y_x(x,y)}{\partial x} + \frac{\partial y_y(x,y)}{\partial y}\right) - \left(\frac{\partial v_x(x,y)}{\partial x} + \frac{\partial v_y(x,y)}{\partial y}\right) \end{aligned} \tag{2.60}$$

which is a Poisson's equation which can be solved by approximation via finite difference methods and writing the result as a system of linear equations:

$$\mathbf{B}\phi = \mathbf{d} \tag{2.61}$$

Matrix $\mathbf{B} \in \mathbb{R}^{s\times s}$ implements a discrete Laplacian, $\phi \in \mathbb{R}^s$ is the vector containing the desired wavefront values (one value per subaperture) stacked column by column across the pupil, and $\mathbf{d} \in \mathbb{R}^s$ is obtained by taking the divergence of the measurement vector. Setting up Equation 2.61 before we can even considering how to solve it ($\mathbf{B}$ is ill-conditioned) is not a trivial task because the eye's pupil is round. Classic papers[59, 60] on zonal reconstruction and Off-the-shelf matrix manipulation tools (i.e. relevant resources include the Matlab function *convmtx2.m*, a recent image deblurring book with Matlab examples[76], and a Matlab implementation for zonal wavefront reconstruction is given in a recent book by Dai[77]) both assume the data to be of rectangular support. Since real eyes

involve data with circular support, we cannot directly use any of the available software. Our task is complicated further by the fact that the pupil can vary in size requiring the software to include functions that appropriately trim or augment the matrices and vectors involved in the calculations based pupil size. These challenges are what attracted my interest in the zonal wavefront reconstruction problem both at the fundamental level and software implementation.

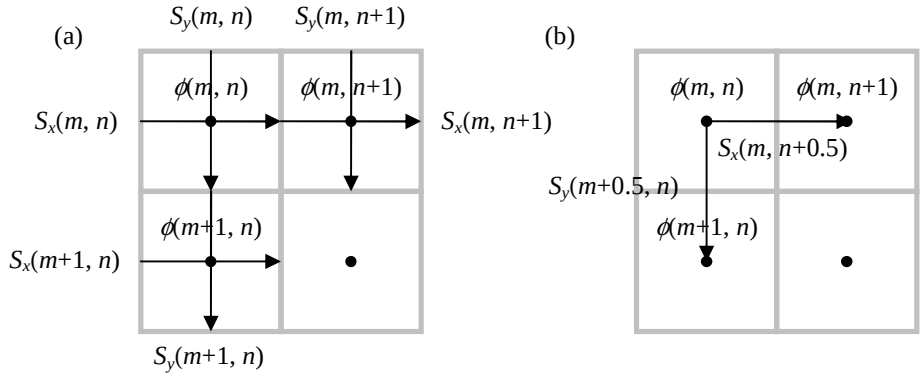


Figure 2.11: A patch of 2 by 2 subapertures (outlined in light gray) illustrating how wavefront slopes, S_x and S_y, are related to the sampled wavefront, ϕ, for the (a) Southwell and (b) Hudgin sensor configurations

In order to perform zonal wavefront reconstruction, operators must be discretized and modified to match the configuration of the wavefront slope measurements. For Shack-Hartmann sensors, the default geometry is called the Southwell configuration[60] which is characterized by the wavefront samples coinciding with the wavefront slope measurements (Figure 2.11a). I consider this configuration to be most intuitive because it is designed to estimate the wavefront error at each subaperture. However, an indirect computational route is required for the wavefront slopes to be related to the desired wavefront values located at the subaperture centers. Wavefront slope data in the Southwell configuration must be first converted to the Hudgin configuration[78] as illustrated for a patch of four subapertures in Figure 2.11b. Locally, the calculation is just a simple average of adjacent slopes:

$$\begin{cases} S_x(m, n+0.5) = \dfrac{S_x(m, n+1) + S_x(m, n)}{2} \\ S_y(m+0.5, n) = \dfrac{S_y(m+1, n) + S_y(m, n)}{2} \end{cases} \tag{2.62}$$

where S_x and S_y are scalars representing the local wavefront slopes and m and n indexes the subapertures. With some abuse of notation, the half index step in Equation 2.62 and Figure 2.11b is used to represent slopes that fall in between adjacent subapertures in the Hudgin configuration. Under the Hudgin configuration, the slopes can be related directly to the wavefront samples:

$$\begin{cases} S_x(m,n+0.5) = \dfrac{\phi(m,n+1)-\phi(m,n)}{d} \\ S_y(m+0.5,n) = \dfrac{\phi(m+1,n)+\phi(m,n)}{d} \end{cases} \tag{2.63}$$

which becomes the standard definition of the derivative if we the limit of *d* as it approaches zero. To estimate the entire wavefront, the calculations must be applied across the entire pupil. The original formulation of this problem is in matrix form[59]:

$$\Gamma\phi = \mathbf{Dy} \tag{2.64}$$

where matrices **D** and Γ implement Equations 2.62 and 2.63 respectively across the entire pupil and the measurement vector **y** is defined as before in Equation 2.54. Since matrix Γ computes both the x- and y-derivatives, if implemented sequentially, this matrix takes the form:

$$\Gamma = \begin{bmatrix} \Gamma_x \\ \Gamma_y \end{bmatrix} \tag{2.65}$$

where the interesting continuous-to-discrete space comparison can be made:

$$\begin{cases} \dfrac{\partial\phi(x,y)}{\partial x} \leftrightarrow \Gamma_x\phi \\ \dfrac{\partial\phi(x,y)}{\partial y} \leftrightarrow \Gamma_y\phi \end{cases} \tag{2.66}$$

This also means that matrix Γ has almost twice many rows as it does columns (392 by 213 for a 6 mm pupil in AOSLOII), so like the Zernike modal reconstructor from the last section, we have a least squares problem in our hands. To estimate the wavefront, multiply both size of Equation 2.64 by $\Gamma^{\mathbf{T}}$:

$$\Gamma^{\mathbf{T}}\Gamma\phi = \Gamma^{\mathbf{T}}\mathbf{Dy} \tag{2.67}$$

This is the same as Equation 2.61 by observing that:

$$\begin{cases} \mathbf{B} = \left(\Gamma^{\mathbf{T}}\Gamma\right) \\ \mathbf{d} = \Gamma^{\mathbf{T}}\mathbf{Dy} \end{cases} \tag{2.68}$$

Equation 2.64 (or 2.67) is not trivial to solve because $\Gamma^{\mathbf{T}}\Gamma$ is singular. If we choose to ignore why a singularity exists, the best solution can be obtained by taking the singular value decomposition (SVD) of Γ:

$$\Gamma = \mathbf{UDV}^{\mathbf{T}} \tag{2.69}$$

where the columns of the unitary matrices **U** and **V** for orthnormal basis sets for the wavefront gradient and the wavefront respectively and matrix **D** is diagonal whose entries are the singular values of this model. If the setup of Equation 2.64 is correct, then the smallest singular value would correspond to piston. Zeroing the smallest singular value after inverting **D**, the least squares estimate of the wavefront becomes:

$$\begin{aligned} \hat{\phi} &= \Gamma^{+}\mathbf{Dy} \\ \Gamma^{+} &= \mathbf{VD}^{+}\mathbf{U}^{\mathbf{T}} \end{aligned} \tag{2.70}$$

where the diagonal matrix $\mathbf{D}^{+}$ is obtained by taking the reciprocal of each non-zero entry in **D** while leaving the "near-zero" entries alone. Notice so far that the slopes from the Hudgin configuration (Equations 2.62 and 2.63) are only used to formulate the problem

and are actually never calculated. Equation 2.70 is valid, but a more sensible approach would be to recognize that certain wavefront profiles (i.e. piston) are in the nullspace of Γ and preconditioning can be applied to solving Equation 2.67. The set of wavefront profiles that span the nullspace of Γ are called unobservable modes with the definition being very much analogous to that in control theory. Currently, the two modes we define to be unobservable are piston and global waffling (Figure 2.12), since any wavefront formed by a linear combination of piston and waffle is theoretically invisible to the Shack-Hartmann sensor. Using preconditioning, the least squares solution takes the following form:

$$\hat{\phi} = \left(\Gamma^{\mathbf{T}}\Gamma + \mathbf{V}\mathbf{V}^{\mathbf{T}}\right)^{-1}\Gamma^{\mathbf{T}}\mathbf{D}\mathbf{y} \tag{2.71}$$

where the columns of matrix **V** (different from the one in Equations 2.69 and 2.70) are the vectorized unobservable modes. It is worth noting that Equation 2.71 is the solution to minimizing the linear quadratic (LQ) cost functional:

$$J(\phi) = \frac{1}{2}\left\|\mathbf{D}\mathbf{y} - \Gamma\phi\right\|_2^2 + \frac{1}{2}\phi^{\mathbf{T}}\left(\mathbf{V}\mathbf{V}^{\mathbf{T}}\right)\phi \tag{2.72}$$

which is analogous to the LQ problem in optimal control where instead of designing a matrix to penalize the input, the **V** matrix in zonal reconstruction penalizes the wavefront making sure that improbable wavefront shapes do not appear in the final estimate. We currently run Equation 2.71 in real-time on AOSLOII, and to my knowledge, no other AO retinal imager runs this type of reconstructor.

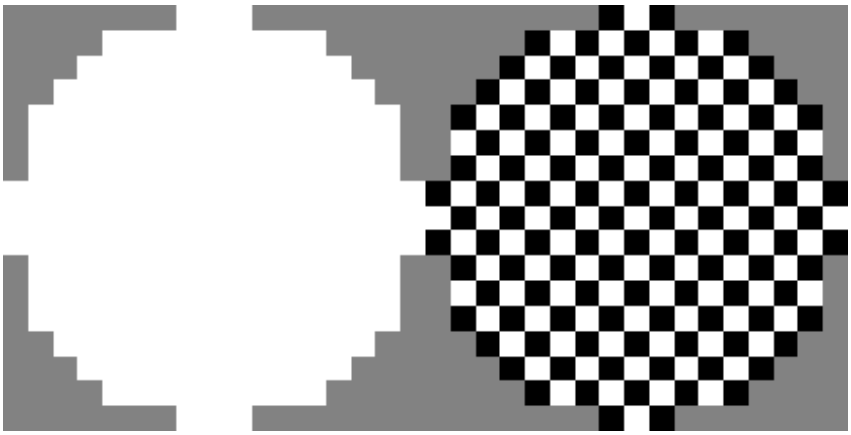


Figure 2.12: Piston (left) and global waffle (right) wavefront modes at subaperture resolution for 6 mm diameter pupil

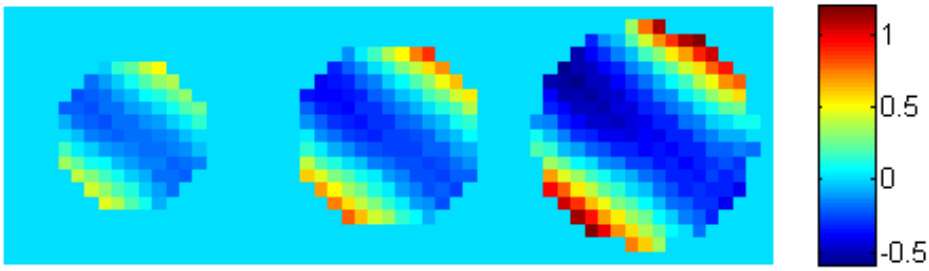


Figure 2.13: From left to right, reconstructed wavefronts for a 4, 5 and 6 mm diameter pupils from the same Shack-Hartmann image (model eye with a cylindrical trial lens)

In ocular wavefront sensing applications where high frequency features in the wavefront may be of interest (i.e tear film irregularities, corneal scaring, refractive surgery outcome assessment, etc.), fitting Zernike polynomials to the wavefront may result in undesirable levels of smoothing (Figure 2.13)[24, 25, 48, 79, 80]. Zonal reconstruction would be the more appropriate choice for these applications. The use of Zernike polynomials have been criticized as being responsible for poorer visual outcomes after LASIK[81], and newer wavefront sensors used to guide refractive surgery such as AMO's WaveScan Wavefront^{TM} System no longer rely on Zernike polynomials. In many of these studies, the algorithm used is actually of the Fourier transform type, but as discussed below, zonal and Fourier-based reconstructors are fundamentally the same[24, 49, 80]. Perhaps the reason why zonal reconstruction is not currently more widely used, at least in basic research, is due to implementation difficulties. In particularly, trimming or augmenting matrices Γ and **D** automatically in order to accommodate measurements over a circular domain is anything but trivial task. We have developed zonal reconstruction code for variable circular pupils (Figure 2.13) in both Matlab and C with the Matlab version provided in the Appendix B.1.

2.3.3 Fourier methods

Although considered to be a modal reconstructor, I consider Fourier methods to be zonal because they essentially solve the same problem but just in the frequency domain. The resultant reconstructed wavefronts using the two methods are generally difficult to differentiate, but the two are easily distinguishable from Zernike polynomial based reconstruction (Figure 2.13). As one might imagine, the purpose of developing Fourier methods is so that the wavefront resconstruction problem can be solved faster[82, 83]. And I am quite certain this is why Fourier methods along with other time saving modalities (i.e. the conjugate gradient) have received more serious attention in astronomy (as opposed to vision science) where future systems are designed to support thousands of degrees of freedom (subapertures and actuators). Current vision AO systems and wavefront sensors in vision science will obviously not benefit to that extent because the difference in speed between matrix and Fourier methods differ less and less with decreasing degrees of freedom. However, the future may still be bright for Fourier methods in vision science as they are well suited for dealing with arbitrary pupil sizes and shapes.

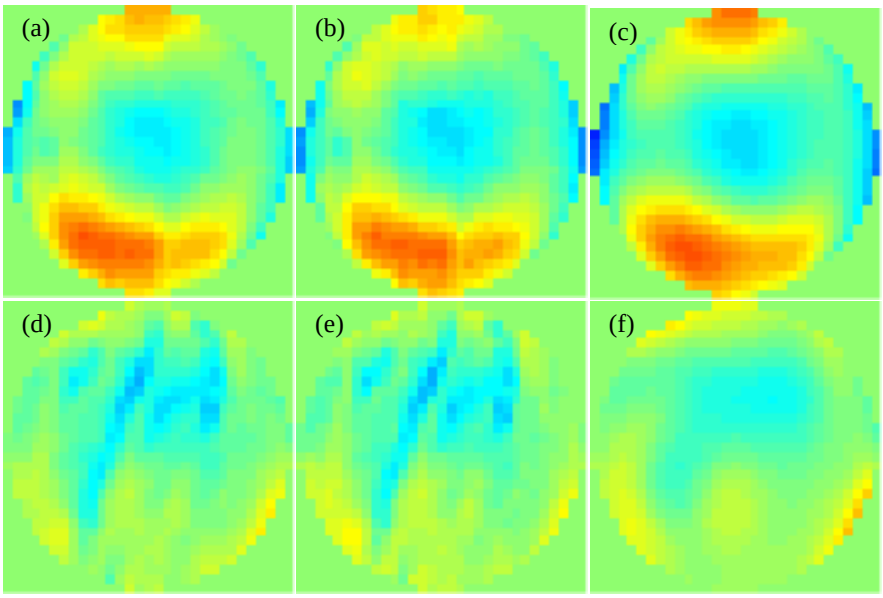


Figure 2.14: Wavefront maps of a normal eye with no tear breakup reconstructed using (a) zonal, (b) Fourier and (c) Zernike methods. Wavefront maps of a different normal eye after tear breakup reconstructed using (d) zonal, (e) Fourier and (f) Zernike methods. Shack-Hartmann images from a very high density wavefront sensor used in an earlier study is more appropriate for making qualitative comparisons between zonal and Zernike polynomial based methods since evaluating Zernike polynomials over a denser sampling grid better represents there use in ocular wavefront sensing

My treatment of wavefront reconstruction methods based on the Fourier transform will be very cursory as many of the subtle details that have been developed over the years were neither thoroughly investigated nor implemented as part of this dissertation. Therefore, a detail analysis on the underlying principles of Fourier methods is currently beyond my expertise. However, the idea behind Fourier transform methods is very simple: differential equations become algebraic equations in the frequency domain. Recall ϕ^x and ϕ^y being the x- and y- derivatives of the wavefront, the relationship between spatial and frequency domain representations of the wavefront gradient is:

$$\nabla\phi(x,y) \leftrightarrow \begin{cases} FT\left\{\phi^x\right\} = uFT\left\{\phi(x,y)\right\} \\ FT\left\{\phi^y\right\} = vFT\left\{\phi(x,y)\right\} \end{cases} \tag{2.73}$$

where *FT* denotes the two-dimensional Fourier transform and *u* and *v*, like earlier, are the spatial frequency coordinates. Like the zonal reconstructor, we then take the divergence of Equation 2.73 to obtain the Laplacian of the wavefront:

$$\nabla^2\phi(x,y) \leftrightarrow uFT\left\{\phi^x\right\} + vFT\left\{\phi^y\right\} = \left(u^2+v^2\right)FT\left\{\phi(x,y)\right\} \tag{2.74}$$

We can now isolate the wavefront, by dividing both sides by $u^2 + v^2$ and take the inverse Fourier transform of the result:

$$\hat{\phi}(x,y) = FT^{-1}\left\{\frac{uFT\left\{\phi^x\right\} + vFT\left\{\phi^y\right\}}{\left(u^2+v^2\right)}\right\} \tag{2.75}$$

Unlike zonal reconstruction, there are plenty of off the shelf tools that can be directly applied to the study of Fourier methods. In Matlab for example, *fft2.m* and *ifft2*.m implement the Fourier and inverse Fourier transforms, and the frequency coordinate space (u, v) is automatically generated by calling *freqspace.m*. However, I have found it less "conversion friendly" when it comes to the more important real-time implementation in a high level language such as C.

Equation 2.75 may give the impression that we are done, but in fact the reconstructed wavefront will likely contain large errors. The original proposed Fourier method was designed for rectangular pupils since the discrete Fourier transform (DFT) operates on a rectangular grid[83]. Since most optical systems are circular, one is generally forced to zero-pad the sampled wavefront gradient in order to form a square grid. The new square pupil contains data within the pupil but zeros outside meaning that the gradient across the pupil boundary will be incorrect. This assumption will undoubtedly lead to significant errors when evaluating Equation 2.75. There are several algorithms that have been developed to solve this problem. One way is to estimate the slopes across the boundary before taking the Fourier transform[82]. The other method is based on the Gershberg algorithm which solves this boundary condition iteratively by recalculating the wavefront gradient from the wavefront reconstructed via Equation 2.75, replace the calculated gradient with the measured gradient which can be done only inside the pupil while leave everything be outside the pupil and repeat Equation 2.75[84]. This Gershberg based Fourier reconstructor was used to generate Figures 2.13b and 2.13e. Gershberg type algorithms are not as computationally efficient because the reconstruction must be done several times. Nevertheless, they are still faster than matrix inversion methods. Recently, a unity frequency response reconstructor based on the Gershberg algorithm has been developed[85]. Without considering sensitivity to measurement error (i.e. sensor noise), this would be the most accurate wavefront reconstructor to date and is worth exploring for ocular wavefront sensing applications.

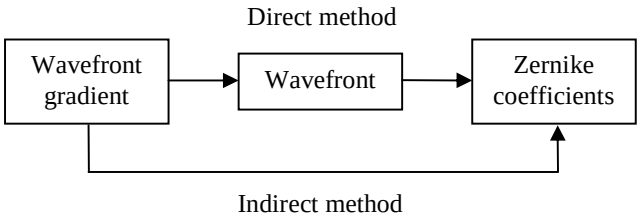


Figure 2.15: Direct and indirect routes for computing Zernike polynomial coefficients

2.4 Discussion

2.4.1 Direct and indirect routes to fitting Zernike polynomials

I would like to discuss whether or not minimizing the following two cost functionals (Equations 2.46 and 2.48) produces the same set of Zernike polynomial coefficients:

1. $$J(\mathbf{c}) = \frac{1}{2}\left\| \phi(x,y) - \sum_{j=0}^{N-1} c_j Z_j(x,y) \right\|^2$$

2. $$J(\mathbf{c}) = \frac{1}{2}\left\| \begin{bmatrix} \phi^x(x,y) \\ \phi^y(x,y) \end{bmatrix} - \sum_{j=0}^{N-1} c_j \begin{bmatrix} Z_j^x(x,y) \\ Z_j^y(x,y) \end{bmatrix} \right\|^2$$

The first is the case of fitting Zernike polynomials to a wavefront, which, as indicated in Figure 2.15, is the direct method because we are fitting Zernike polynomials directly to the wavefront. In the second case (the basis the Cubalchini method), Zernike polynomials are indirectly fitted to the wavefront by fitting the first derivatives of Zernike polynomials to the wavefront gradient. If these two processes are equivalent, then they should produce the same set of Zernike coefficients. Piston, tilt and tip will be left inside the expression for the sake of discussion. It is helpful to first state some important mathematical properties of Zernike polynomials[86, 87]:

Property 1: The 0^{th} mode is piston (DC component of the wavefront) and is defined to be unity across the entire pupil.

Property 2: All modes except for piston have zero mean.

Property 3: If piston is removed from the wavefront, the square root of the sum of squares of the coefficients yields the RMS wavefront error.

Property 4: The value of each Zernike coefficient does not depend on the number of modes in the expansion.

Property 1 is by definition, and Property 2 is a consequence of Property 1. This can be observed by taking an inner product of piston with another mode:

$$\begin{aligned}\left(Z_0, Z_j\right) &= \frac{1}{\pi}\int_{\Sigma} Z_0(x,y) Z_j(x,y)\,dxdy \\ &= \frac{1}{\pi}\int_{\Sigma} Z_j(x,y)\,dxdy \\ &= 0, \quad j \neq 0\end{aligned} \tag{2.76}$$

where the second line is just the mean of the j^{th} Zernike mode, which vanishes due to orthonormality. Property 3 can be directly derived from standard formula for computing the rms of an arbitrary waveform:

$$\phi_{rms} = \sqrt{\frac{1}{\pi}\int_{\Sigma} \phi^2(x,y)\,dxdy} \tag{2.77}$$

where π is the area of the normalized pupil. If we substitute Equation 2.29 in for $\phi(x, y)$, the RMS becomes:

$$\phi_{rms} = \sqrt{\frac{1}{\pi}\int_{\Sigma} \left(\sum_{j=0}^{N-1} c_j Z_j(x,y)\right)^2 dxdy} \tag{2.78}$$

But since Zernike polynomials are mutually orthonormal, we can remove all the cross terms and obtain a much simpler expression for the wavefront rms:

$$\phi_{rms} = \sqrt{\sum_{j=0}^{N-1} c_j^2} \tag{2.79}$$

This is an important result because the RMS value of a continuous function can be computed via a finite summation, but this result is valid only if the chosen basis functions are orthonormal. Property 4 was already derived earlier (Equation 2.44) which was:

$$c_j = \left(\phi, Z_j\right) = \frac{1}{\pi}\int_{\Sigma} \phi(x,y) Z_j(x,y)\,dxdy$$

Notice that three of the four properties are completely consequences of orthonormality, the exception being the Property 1 which is a definition. Therefore, they are all if and only if statements.

In order to answer whether or not the direct and indirect routes for calculating Zernike coefficients are the same, Property 4 is of particular interest because it has much more implications than appearing to be just a minor computational convenience. As stated earlier, ocular Shack-Hartmann sensors are ubiquitous and it is more likely than not for different devices to use different numbers of modes in their expansions. According to Property 4, suppose a particular Zernike coefficient, call it c_i, was measured for an individual using several different devices, the results from each measurement should be directly comparable assuming all other parameters (i.e. noise level, sampling density, pupil size, etc.) are reasonably similar. However, this frequently assumed property does not hold if the indirect route (Cubalchini method) is taken. In Section 2.3.1, a linear system of equations for computing the Zernike coefficients from the wavefront gradient was analytically derived resulting in Equation 2.52 which we denoted by **Bc** = **d**. The solution (Equation 2.53) was simply $\hat{\mathbf{c}} = \mathbf{B}^{-1}\mathbf{d}$ which exists because the set of Zernike gradient polynomials form a linearly independent but non-orthogonal basis[61, 66]. Using this notation, the estimated value for c_i becomes:

$$c_i = \sum_{j=0}^{N-1} \left[\mathbf{B}^{-1}\right]_{ij} \mathbf{d}_j \tag{2.80}$$

$$= \sum_{j=0}^{N-1} \left[\mathbf{B}^{-1}\right]_{ij} \left[\left(\phi^x, Z_j^x\right) + \left(\phi^y, Z_j^y\right)\right]$$

where $\left[\mathbf{B}^{-1}\right]_{ij}$ is the i^{th} row and j^{th} column entry of matrix $\left[\mathbf{B}^{-1}\right]$. By inspection, in order for the value of c_i to be dependent only on the i^{th} Zernike gradient polynomial and thus for Property 4 to hold, the entire i^{th} row of matrix $\left[\mathbf{B}^{-1}\right]$ must be zero except for its ij^{th} entry. This would require $\left[\mathbf{B}^{-1}\right]$ to be diagonal which is not the case because the Zernike gradient polynomial are non-orthonormal[61, 66]. Therefore, if we choose to indirectly estimate the Zernike coefficients, we are in fact fitting some non-orthonormal set of polynomials to the wavefront proving that the direct and indirect routes for estimating Zernike coefficients are indeed fundamentally different. An immediate consequence of this observation is that the convenient expression for calculating the RMS wavefront error (Equation 2.79) is no longer valid, so perhaps it is to no surprise that Cubalchini actually did not included this expression in his classic paper[61].

In a study conducted by Prieto *et al.*[47], the authors fitted different numbers of Zernike polynomials to a single data set using the Cubalchini (indirect) method and showed that the number of modes estimated does have an effect on the resultant Zernike coefficient estimates. The same type of behavior was observed from a 4^{th}, 6^{th}, and 10^{th} order Zernike fit (indirect) to wavefront slope data from AOSLOII (Table 2, left half). These results provide empirical evidence against the validity of Property 4 when Zernike coefficients are estimated indirectly. Consider the simple, but direct, alternative: first estimate the wavefront using zonal construction, and once that is done, we can fit Zernike polynomials directly to the wavefront through standard curve-fitting. Recall from Section 2.3.2 that if the proposed least squares zonal reconstructor (Equation 2.71) is used, the result is simply a vector of local wavefront estimates (one per subaperture).

$$\hat{\phi} = \left(\Gamma^{\mathbf{T}}\Gamma + \mathbf{V}\mathbf{V}^{\mathbf{T}}\right)^{-1} \Gamma^{\mathbf{T}}\mathbf{D}\mathbf{y}$$

The reconstructed wavefront expressed as a linear combination of Zernike polynomials can be express in matrix form as well by:

$$\hat{\phi} = \mathbf{Z}\mathbf{c} \tag{2.81}$$

where if we choose to fit the first *N* modes to the wavefront, matrix **Z** will then have *N* columns with each column being the appropriate (vectorized) Zernike polynomial sampled at each subaperture location:

$$\mathbf{Z} \triangleq \begin{bmatrix} \mathbf{z}_0 & \mathbf{z}_2 & \cdots & \mathbf{z}_{N-1} \end{bmatrix} \tag{2.82}$$

The least squares solution is simply:

$$\hat{\mathbf{c}} = \left(\mathbf{Z}^{\mathbf{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathbf{T}}\hat{\phi} \tag{2.81}$$

At this point, it is interesting to compare inner product definitions in the continuous and discrete spatial domains. As one can see, the inner product expression is much simpler in the discrete spatial domain:

$$\begin{aligned} \frac{1}{\pi}\int_{\Sigma} Z_j(x,y)Z_j(x,y)dxdy &\leftrightarrow \frac{1}{N}\mathbf{z}_i^{\mathbf{T}}\mathbf{z}_j \\ \frac{1}{\pi}\int_{\Sigma} \phi(x,y)Z_j(x,y)dxdy &\leftrightarrow \frac{1}{N}\mathbf{z}_j^{\mathbf{T}}\phi \end{aligned} \tag{2.82}$$

The discrete case can be derived from the continuous case by replacing the integral with a summation, recognizing that $dxdy \approx \pi/N$, and rewriting the result in vector form. Discretely sampled Zernike polynomials are only approximately orthonormal:

$$\frac{1}{N}\mathbf{z}_i^{\mathbf{T}}\mathbf{z}_j \approx \delta_{ij} \tag{2.83}$$

and so the condition will only become weaker as we attempt to estimate higher and higher order terms due to aliasing. However, if the Zernike modes of interest were appropriate for the sampling density of the wavefront sensor, then the approximation made in Equation 2.83 is valid and the least square solution given in Equation 2.81 becomes:

$$\begin{aligned} \hat{\mathbf{c}} &= \left(N\mathbf{I}\right)^{-1}\mathbf{Z}^{\mathbf{T}}\hat{\phi} = \frac{1}{N}\mathbf{Z}^{\mathbf{T}}\hat{\phi} \\ \Rightarrow c_i &= \frac{1}{N}\mathbf{z}_i^{\mathbf{T}}\hat{\phi} \end{aligned} \tag{2.84}$$

which is analogous to the continuous spatial domain case presented earlier (Equation 2.44) assuring that the value of each Zernike coefficient does not depend on the number of modes used in the expansion (Property 4). Although somewhat trivial at this point, the results from using the direct method are given in the right half of Table 2.4 showing that the coefficients are indeed independent of the order selected for the expansion. The estimated coefficients gradually become less and less accurate as the order of the mode of interest gets higher because Equation 2.83 will not be a good approximation if we sample a very high order mode too coarsely. Therefore, there is not a sharp cutoff where beyond a specific mode we can start disregarding the rest of the estimates, but the general vicinity to where this would occur will clearly be based on the sampling density of the wavefront sensor. As a guideline, I recommend up sampling about four times the Nyquist limit which is somewhat conservative. So for a 17 by 17 subaperture sensor, the estimated Zernike coefficients up to 4^{th} and 5^{th} order should be valid for all practical purposes.

Table 2.4: Zernike polynomial coefficients computed using both indirect (left half) and direct (right half) modal reconstruction methods. Defocus estimates (bold) seem to be the only mode that remained reasonably unaffected by the number of terms used in the expansion.

j	Indirect (Cubalchini method)			Direct		
	4^{th} order	6^{th} order	10^{th} order	4^{th} order	6^{th} order	10^{th} order
3	0.0001	-0.0005	-0.0013	0.0055	0.0055	0.0055
4	**-0.2215**	**-0.2213**	**-0.2214**	**-0.2178**	**-0.2178**	**-0.2178**
5	0.0057	0.0052	0.0046	-0.0025	-0.0025	-0.0025
6	-0.0002	-0.0121	-0.0112	0.0098	0.0098	0.0098
7	0.0033	-0.0027	-0.0031	0.0018	0.0018	0.0018
8	-0.0116	0.0017	0.0014	0.0025	0.0025	0.0025
9	0.0113	0.0115	0.0110	0.0137	0.0137	0.0137
10	0.0045	-0.0061	-0.0061	-0.0029	-0.0029	-0.0029
11	-0.0071	-0.0095	-0.0099	0.0102	0.0102	0.0102
12	0.0225	0.0235	0.0234	0.0115	0.0115	0.0115
13	-0.0039	0.0120	0.0127	0.0052	0.0052	0.0052
14	0.0169	0.0046	0.0059	-0.0009	-0.0009	-0.0009

2.4.2 Zonal reconstruction noise propagation

According to paper on wavefront reconstructors by Southwell[60], modal estimation was shown to be superior to zonal estimation. This statement may have caused some misconceptions regarding zonal reconstruction, particularly for ocular wavefront sensing, as some authors have gone as far as stating explicitly that modal reconstruction is better than zonal reconstruction[68, 88, 89]. It should be mentioned that Southwell's work was based on Legendre polynomials over a rectangular grid, and more importantly, his conclusion was based on noise propagation. More specifically, Southwell showed that a Legendre modal reconstruction tended to have a lower noise propagation coefficient than the proposed zonal method[60]. This should not be surprising because the total number of modal coefficients estimated is generally much lower than the number of local wavefront values estimated via a zonal approach. For example in AOSLOII, we estimate Zernike coefficients up to 10^{th} order which is only 63 numbers as opposed to the 213 subapertures where zonal estimates are made (6 mm diameter pupil).

Characterizing the performance of different reconstruction algorithms based on sensitivity to noise is important, but it does not in any way help determine the accuracy of the reconstructed wavefront. Since zonal reconstruction directly seeks an optimal wavefront profile rather than a linear combination of some predetermined set of functions, it is fundamentally more accurate than modal methods when measurement noise is negligible. In fact, it is only possible to achieve band-limited performance using zonal reconstruction[85]. If measurement noise cannot be neglected, and it is either impossible or impractical to further optimize the Centroid finding algorithm, then modal reconstruction may be more appropriate. However, that is not to say that the "standard" zonal reconstructor cannot be improved in terms of noise propagation. Consider first the zonal reconstructor whose noise propagation properties were assessed by Southwell:

$$\mathbf{E} = \left(\Gamma_e^{\mathbf{T}}\Gamma_e\right)^{-1}\Gamma^{\mathbf{T}} \tag{2.85}$$

$$\Gamma_e = \begin{bmatrix} \Gamma \\ \mathbf{p}^{\mathbf{T}} \end{bmatrix}$$

The **p** term augmented to the bottom our original design matrix Γ is just a vector of ones, so a full row of ones has added to Γ. The noise propagation C_{np} of the reconstructor **E**, derived from taking the SVD of Γ_e, is the sum of its inverse singular values squared (σ_j):

$$C_{np} = \sum_{j=1}^{s} \sigma_j^{-2} \tag{2.86}$$

where *s* is the number of subapertures (number of columns in Γ). This makes sense because if Γ_e is poorly conditioned, then there will be at least one very small singular value which is all it takes to produce very large noise propagation according to Equation 2.86. Since the square of the singular values of Γ_e are just the eigenvalues of matrix $\left(\Gamma_e^{\mathbf{T}}\Gamma_e\right)$, the noise propagation is also the sum of the inverse eigenvalues of $\left(\Gamma_e^{\mathbf{T}}\Gamma_e\right)$. Furthermore, notice that:

$$\left(\Gamma_e^{\mathbf{T}}\Gamma_e\right) = \begin{bmatrix} \Gamma^{\mathbf{T}} & \mathbf{p} \end{bmatrix} \begin{bmatrix} \Gamma \\ \mathbf{p}^{\mathbf{T}} \end{bmatrix} = \Gamma^{\mathbf{T}}\Gamma + \mathbf{p}\mathbf{p}^{\mathbf{T}}$$

so the zonal reconstructor addressed in Southwell[60] is nearly identical to that given by Equation 2.71 with the exception that only piston (**p**) is penalized. Since both piston and waffle modes are explicitly penalized in Equation 2.71, the zonal reconstructor currently in use on AOSLOII is better conditioned than the one considered in Southwell[60]. The noise propagation is simply the sum of the inverse eigenvalues of matrix $\Gamma^{\mathbf{T}}\Gamma + \mathbf{V}\mathbf{V}^{\mathbf{T}}$. If lower noise propagation is desired, regularization techniques can be used:

$$\mathbf{E} = \left(\Gamma^{\mathbf{T}}\Gamma + \mathbf{V}\mathbf{V}^{\mathbf{T}} + \alpha\mathbf{I}\right)^{-1}\Gamma^{\mathbf{T}}\mathbf{D} \tag{2.87}$$

$$\alpha \triangleq \frac{\sigma_{\mathbf{v}}^2}{\sigma_{\phi}^2}$$

where **I** is the identity matrix and α is ideally the noise-to-signal ratio. Equation 2.87 is a Wiener filter solution. Since the noise variance is not known exactly, α must be empirically tuned until the desired noise propagation level is achieved. If α is too small, we end up with the original reconstructor, while setting it to be too large will cause the reconstructor to converge closer toward a zero matrix which has zero noise propagation but will also be useless. Noise propagation does not seem to be an issue for wavefront reconstruction on AOSLOII, so no regularization is used ($\alpha = 0$).

2.5 Conclusions

To conclude, I have reviewed some of the physical optics involved in the image formation of the eye and verified that the standard procedure for calculating the eye's PSF is based on Fresnel approximation. Opting for a full Rayleigh-Sommerfield treatment of diffraction would probably not make any practical difference. However, the parameters used for the PSF calculation should be examined more closely. For example, the location and size of the eye's exit pupil can be more systematically estimated if certain biometry data is available. Also, a Zernike polynomial representation of the wavefront may be an over-smoothed version of the actual wavefront which will directly

affect the PSF making it appear much better than it really is. A wavefront reconstructed using a zonal reconstructor may aid in generating a more realistic PSF.

Estimating Zernike coefficients indirectly from the wavefront gradient (i.e. Cubalchini method[61]) has been analytically and numerical shown to be different from fitting Zernike polynomials directly to the wavefront. Two practical consequences result from this observation: (1) the RMS wavefront error is no longer simply the square root of the sum of squares of the estimated coefficients and (2) individual coefficients will be different depending on the number of modes used in the expansion making comparisons between different studies more difficult.

Finally, when assessing wavefront reconstruction methods, reconstruction accuracy and sensitivity to measurement error are separate properties. In terms of accuracy, zonal reconstructors (Fourier methods included), if set up properly, are optimal. If the errors in the wavefront slope measurements are substantial enough to warrant consideration of noise propagation properties of the wavefront reconstructor, a modal reconstruction will most likely be more accurate than a zonal estimate that has not been regularized. However, a simple Wiener filter can always be included in zonal reconstruction to match or exceed the noise propagation properties of a modal reconstruction at the expense of accuracy.

Acknowledgements

Austin Roorda provided the Matlab code for analytical expressions of the Zernike polynomial derivatives. I acknowledge Sandipan Mishra for linear algebra assistance in the early stages of developing the zonal reconstructor for circular pupils. Kamran Ahmad and Geunyoung Yoon provided the Matlab code for automatically generating zonal reconstruction matrices Γ and $\mathbf{D}$ for arbitrary circular pupil sizes. James Zavislan's undergraduate physical optics notes were extensively used in the derivations at the beginning of the chapter. Charlie Campbell and Geunyoung Yoon provided the original Matlab code for the implementation of the Fourier reconstructor using the Gershberg algorithm. Finally, Curt Vogel provided much guidance to my understanding of the mathematics behind much of the topics covered in this chapter.

3. Controlling monochromatic ocular aberrations

3.1 Introduction

Controlling wavefront aberrations in real-time is a difficult task. It is probably the most complicated component of an AO system as it involves many disciplines. In order to implement an effective closed loop design, one needs to consider factors regarding anatomy, optics, mechanics, electronics, real-time computing, and most of all control theory. Not properly addressing the significance of one particular topic will not lead to any danger in the sense of injury or severe property damage (although "snapped" DM actuators do inevitably occur and are expensive and time consuming to replace), but they should nevertheless be respected if we are to avoid further delays in bringing vision AO systems into the next generation and closer toward clinical deployment. As the number of AO applications in vision grows, we can expect increases in the number of users not familiar with AO and in patients with more challenging optics (i.e. post-LASIK, dry eyes, etc.). This reason alone motivates the need to make improvements to current AO system performance and robustness so that clinical and scientific throughput (better quality images from a larger pool of patients) can be improved. Some of the content presented in this chapter are taken out of my conference paper[90] with only trivial modifications.

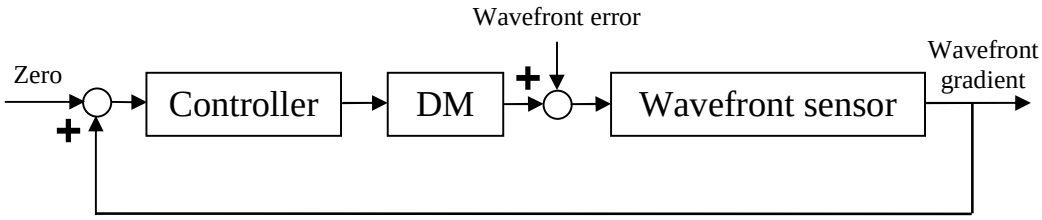


Figure 3.1: Descriptive block diagram representing the AO control loop. The reference input is zero because the desired flat wavefront has a first derivative of zero.

Figure 3.1 is a typical block diagram for a closed loop AO system in any discipline. AO control is a multivariate problem where the states are some kind description of the wavefront error. Whether these modes are in terms of basis functions (i.e. Zernike polynomial coefficients) or local wavefront heights seems to be based on the preference of the engineer. AO control for retinal imaging can be extremely simple from a traditional control systems point of view and extremely difficult where principles beyond the traditional control theory, particularly wave optics, are required to implement a functional system. It is easy because the plant, which is primarily the DM, is quasi-static which means that its response is effective instantaneous. This is due to most CCD cameras used for ocular wavefront sensing having frame rates between 10 and 100 Hz, but the bandwidth of the DM is on the order of kilohertz[91]. Therefore, the only dynamics that the controller must deal with are those of the eye (i.e. eye movement, tear file, aqueous, etc.) which are intrinsically stable as the worst thing that can happen is the subject blinks during feedback control. In any case, these dynamics are never modeled in practice but rather treated as a disturbance input which for the most part can be rejected with an integrator. The AO control problem is difficult because optical phase (wavefront) is an infinite dimensional spatial process, so the state, which I will denote with ϕ, is

theoretically an infinite dimensional vector. However, there are only a finite number of subapertures and actuators to sample and control the state, so a serious component of the AO control problem involves isolating only the states that affect retinal image quality the most and that are both observable and controllable. Another reason why AO control is difficult is because different systems use different DMs (Figure 3.2). There is not a single control algorithm that can be applied to all DMs as they can deform through very different mechanisms. For example, controllers designed to operate continuous facesheet DMs would not be able to drive segmented piston, tip and tilt DMs because the actuators are uncoupled and have three degrees of freedom in the later case. Therefore, most AO controllers tend to be at least partially system specific making streamlining efforts to be somewhat impractical.

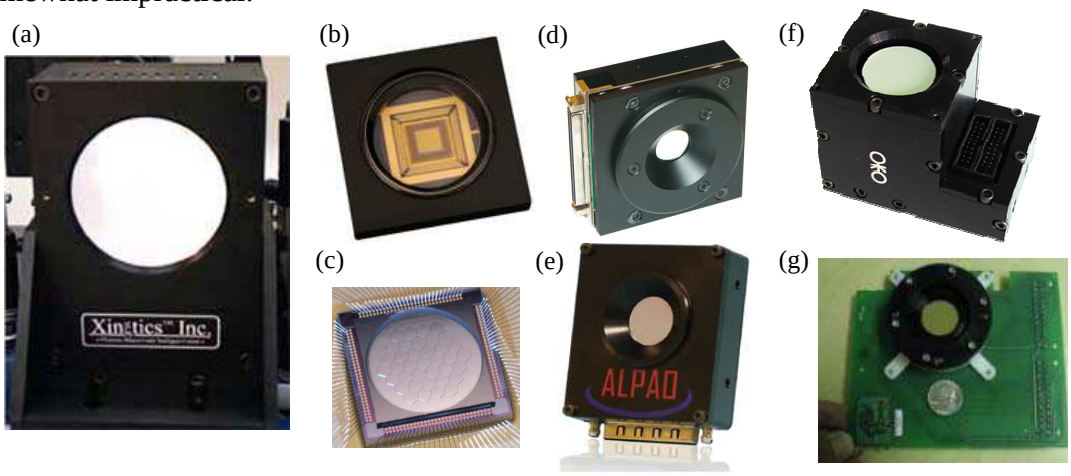


Figure 3.2: Commercially available DMs that have been applied to AO retinal imaging. (a) Piezo DM by Xinetics (both 37 and 97 actuator models have been used) used in the earliest systems. MEMS devices include the (b) BMC Multi-DM (used in AOSLOII) and the (c) PTT 111 by IRIS AO. High stroke magnetic deformable mirrors include the (d) mirao^{TM} from Imagine Eyes and (e) several different models available from ALPAO. Cheaper, lower density piezo-based devices (37 actuators) from (f) OKO Technologies and (g) AOptix have also been used in retinal imaging[10, 92].

The purpose of this chapter is to describe AO controller upgrades for AOSLOII that have allowed us to resolve the smallest foveal cones for the first time with an AOSLO and a MEMS-based DM for that matter (no complete foveal cone mosaic images exist from the Rochester flood illuminated AO system when an alternative DM (including the BMC MEMS device) was used[93]). As one can imagine, the process of developing and implementing an AO system where a clinician or researcher can simply place a patient in AOSLOII and begin acquiring retinal images with only a few key strokes extends beyond just the mathematical equations for AO control. Therefore, it is imperative to address all practical aspects of the project especially the interface between the optics (see Chapter 2) and the control system as well as the real-time requirements. AOSLOII employs the BMC Multi-DM (Figure 3.1b) which is a continuous facesheet DM. As schematized in Figure 3.2, these devices are arrays of capacitors that compress due to electrostatic force (F_E) when voltage is applied. The relationship between F_E and the applied voltage (V) is given by[94]:

$$F_E = \frac{\varepsilon_0 A V^2}{2(g - w)} \tag{3.1}$$

where A is the area of an electrostatic actuator (0.16 mm^2) and g is the nominal size of the gap between the facesheet (w) and the top surface of the actuators. This relationship is nonlinear, but as long as $g \approx g - w$ (small deflections), electrostatic force behaves linearly with the voltage square. Another subtlety exhibited by these devices is that the actuators can only be pulled down, so in order to achieve bi-directional actuation, the DM must operate at a bias. These turned out to be minor inconveniences, but the modeling of the restoring force imposed by the deformed facesheet and the resultant nonlinear cross-coupling between neighboring actuators for the BMC MEMS devices is a very complicated problem and is an area of active reseach[94-98]. Although I believe these modeling efforts are important, the effectiveness of currently available models are very difficult to judge because they have only been tested on simulations or highly constrained situations (i.e. correcting a phase plate or producing a lower order Zernike term such as defocus). In other words, this research can greatly benefit from implementation and testing on-site for either astronomical or vision science AO systems. No AO retinal imaging system currently online is running a type of controller based on an actual physical model of the DM, but the reason for this may be just due to the lack of a concerted effort among researchers with proper backgrounds (i.e. MEMS, control systems, optics and vision science).

The complications that have arisen by using the BMC Multi-DM seem to have encouraged many research groups to use more expensive alternative correctors. Even though several instruments have been developed with the Multi-DM as its sole wavefront corrector[6, 21, 32, 36], AOSLOII may be the only system remaining that is currently online as others have swapped for a different DM or have been rebuilt to incorporate multiple correctors. For example, several systems are using the BMC device together with a second higher stroke DM, typically one of the magnetic DMs, in a woofer-tweeter configuration[4, 64, 99, 100]. This is a pity because only MEMS-based devices have any reasonable chance of achieving clinically accessible costs. But have these additions lead to higher lateral resolution? Based on our results, this does not seem to be the case. Details on how we have updated the controllers for AOSLOII to improve lateral resolution are given below. Since the eye contains no isolated point sources, quantifying the actual AO system performance can be problematic. A somewhat qualitative but clinically relevant benchmark for any AO retinal imager is whether or not the smallest foveal cone photoreceptors can be resolved. Examples of resolved foveal cone mosaics are provided.

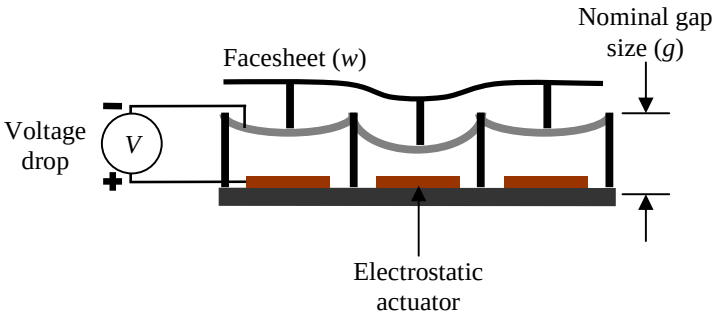


Figure 3.3: Cross-section schematic across three actuators for the BMC MEMS device. The voltage applied to the center actuator is higher than those for the two on each side.

3.1.1 Background

Many articles and dissertations have been written about AO controllers in other disciplines (i.e. astronomy). This is not the case for vision science. Improving AO performance in vision science has typically involved adding and upgrading components. For example, the DM used in the Rochester flood illuminated system was switched from a 37 channel to a 97 channel Xinetics device along with a faster (25 Hz) CCD camera for wavefront sensing[2, 17, 63, 101]. Furthermore, some AO retinal imagers that have come online more recently have adopted a woofer-tweeter configuration involving two DMs[64, 102, 103], and another group demonstrated multi-conjugate AO retinal imaging which involved multiple DMs as well as wavefront sensor beacons[14]. The refinement of AO system performance, particularly at the control system level, has rarely been addressed[104-106]. Since the history of AO control in vision science is nearly nonexistent, the task of bringing controllers currently used in vision AO systems to comparable levels of sophistication as those used in astronomical telescopes will surely be a challenge. The work presented in this chapter marks the first step to improving the resolution of AO retinal imagers by using more advanced controllers; an important design component that is often overlooked in vision science applications.

The first AO retinal imager, reported in 1997 by Liang *et al.*[107], established that the living human retina can be imaged at the microscopic level. As expected for any new instrument developed for a completely new application, the methodology to which the first AO retinal imager was realized was probably not carried out as systematically as its counterparts in astronomy. The first AO controller used in retinal imaging was poorly designed and resembled more of an active rather than adaptive optics system by definition. The Zernike coefficients were first calculated using the Cubalchini method[61] and a wavefront map was generated by evaluating Equation 2.29 (given again here) at many points across the pupil:

$$\phi(x,y) = \sum_{j=0}^{N-1} c_j Z_j(x,y)$$

Voltages required to produce the height of the wavefront at the actuator center locations are then computed, typically from deflection versus voltage curves provided by the manufacturer, and prescribed to the DM. In closed loop, the residual wavefront after the previous frame is measured and the corresponding height samples are added to those from the previous frame, the result behaves like an integral action but a poorly designed

one because spatial coupling between adjacent actuators were not modeled. Together with the assumption that sampling the Zernike-based wavefront profile only at the discrete actuator points is sufficiently accurate proved to have many consequences. One of which is that the DM surface can oscillate without eventually converging to a correction and actuator clipping (exceeding stroke range in one direction only) may occur after a certain number of loop iterations. In dealing with these effects, the sampled residual wavefront heights are multiplied by a small gain (around 0.3 to 0.4) before converting to voltages, and the system user had to manually stop the loop and hold the last voltage sequence when the system was at a state where the observed retinal image quality was qualitatively optimal[1]. This was possible because convergence was on the order of minutes. Since this type of system does not actually track the changes in the ocular wavefront, it is by definition an active, rather than adaptive, optics system. Nevertheless, this type of control strategy was the starting point for many later AO retinal imagers[6, 7] including AOSLOII.

Having to manually stop the AO control loop to avoid saturation and gradual reduction in retinal image quality is clearly undesirable. Eventually, better closed loop performance was achieved via control algorithms resembling more closely to how AO systems in vision science are operated today. The merit behind using influence functions was soon recognized but was implemented in a way that involved Zernike polynomials[43, 108]. The basic idea was that the response of each DM actuator must first be measured, and these measurements are used to construct a matrix model for the plant (DM plus wavefront sensor) which is called the interaction matrix. Using the Cubalchini method described in Chapter 2, each measurement results in a vector of Zernike coefficients which we will denote by $\mathbf{c}_i$ for the case of the i^{th} actuator response. The interaction matrix **B**, which relates the DM input vector **u** to the "measured" Zernike coefficients **c**, can then be generated column by column using these measurements:

$$\mathbf{u} = \mathbf{B}\mathbf{c} \tag{3.2}$$
$$\mathbf{B} \triangleq \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_{N_a} \end{bmatrix}$$

where N_a is the number of DM actuators. To make this into a closed loop control algorithm, the matrix **B** is inverted using some type of pseudoinverse (**B** is not square) and an integrator can be used to iteratively update the DM:

$$\mathbf{u}(k) = \mathbf{u}(k-1) - \mathbf{B}^{+}\mathbf{c}(k) \tag{3.3}$$

where k indexes the control iteration and $\mathbf{B}^{+}$ is the pseudoinverse of matrix **B**. These earlier studies may have had difficulties obtaining an accurate interaction matrix model or the wavefront sensor measurements may have been too noisy because the update term on the right side of Equation 3.3 needed to be multiplied by a small gain in order to achieve a more accurate convergence but only at the expense of speed. Both these procedures still deviate from the current standard basis for closed loop AO control in both astronomy and vision science, which involves modeling the plant with a static interaction matrix directly from the wavefront gradient measurements[109, 110]. The control algorithm is essentially an integrator in series with some type of inverse of the interaction matrix. To my knowledge, all AO retinal imagers that use a continuous facesheet DM besides ours use this approach, albeit the simplest version (either Moore-Penrose pseudoinverse or Tikhonov regularization [4, 5, 32, 36]). I expanded upon the standard AO controller design by: 1) incorporating the static nonlinear actuator response into an input linearization step and 2) implementing three less trivial control algorithms for AOSLOII. Each algorithm

optimizes a particular cost functional and uses a standard integrator law but with some minor tunable features. A comparison of the control algorithms was undertaken and the results discussed at the end of this chapter.

3.1.2 AO control loop

The objective of a general AO control system is best illustrated with a simple schematic (Figure 3.4a) where the aberrated wavefronts emanating from either the turbulent atmosphere or the eye falls on the surface of a DM, and the reflected beam is ideally free of aberrations. Mathematically, the aberrated wavefront is added to the wavefront induced by the DM surface, which is just twice the magnitude of the surface profile, and in the ideal scenario, the reflected beam would have a planar wavefront. As indicated in the corresponding block diagram (Figure 3.4b), the reflected beam is measured by the wavefront sensor and fed back to the controller so the DM commands for the next cycle can be computed. The expression for the measured error is given by:

$$\mathbf{e}(k) = \mathbf{H}(z^{-1})\mathbf{G}\mathbf{u}(k) + \mathbf{H}(z^{-1})\phi(k) \tag{3.4}$$

where the descriptions for the filters and signals are provided in Figure 3.4b. The measured error signal $\mathbf{e}(k)$ is actually the measured wavefront gradient equal to Equation 2.54 if no correction is applied by the DM. The best correction is achieved when the residual wavefront is flat or the wavefront gradient is zero. However, as mentioned in Chapter 2, this is not always the case due to the presence of unobservable wavefront modes (i.e. waffle). Since the DM is quasi-static, it is modeled to have an instantaneous reaction upon input, so the only plant dynamics is a pure delay element due to the CCD exposure time of the wavefront sensor.

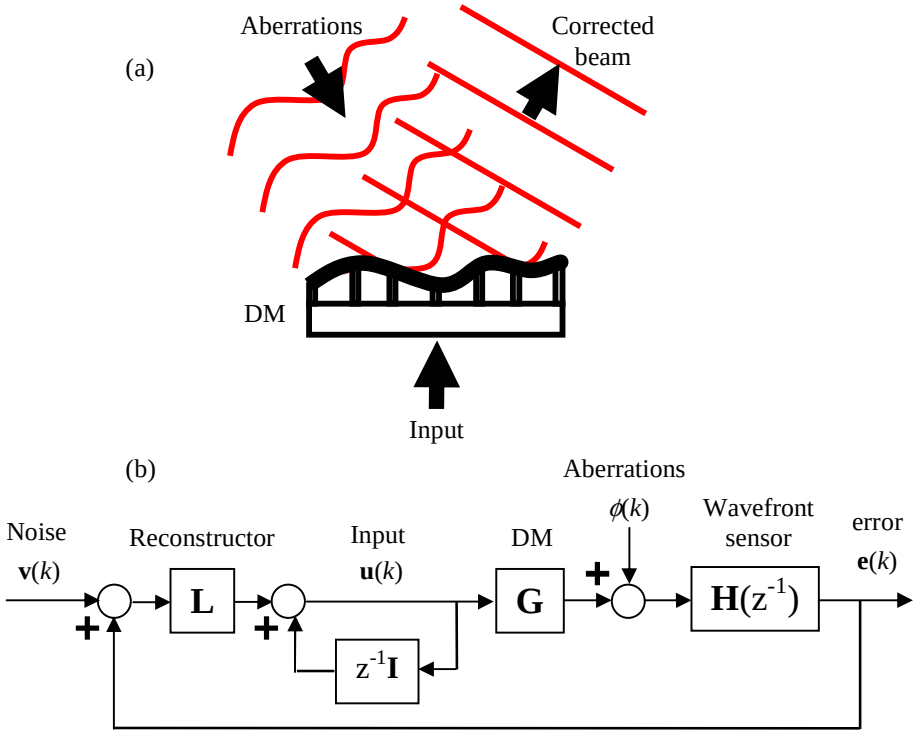


Figure 3.4: (a) Schematic representation of wavefront correction illustrating that the incoming wave aberrations are added to the DM surface to produce the resultant corrected beam and (b) the corresponding closed loop block diagram. Matrix **I** is identity whose dimensions depend on the length of the input $\mathbf{u}(k)$ and z^{-1} is delay.

Based on the standard interaction matrix model, the DM and wavefront sensor combined are modeled by the matrix **T**, which is clearly a gross simplification but can nevertheless be very effective if the wavefront reconstructor is carefully designed. Details on generating matrix **T** are provided in sections 3.1.3 and 3.1.4. It should be clarified at this point that wavefront reconstruction in this chapter refers to the estimation of the DM commands rather than the actual wavefront. This is often a source of confusion in the field as the two problems can often appear to be similar mathematically. Based on the interaction matrix model, the measured wavefront gradient produced by the actuated DM surface when the incident beam is a plane wave ($\mathbf{y} = 0$) simplifies Equation 3.4 to:

$$\mathbf{e}(k) = \mathbf{T}\mathbf{u}(k-1) \tag{3.5}$$

where the one step delay is due to the exposure and readout time of the wavefront sensor CCD camera. When the input beam is aberrated, Equation 3.5 becomes:

$$\mathbf{e}(k) = \mathbf{T}\mathbf{u}(k-1) + \mathbf{y}(k) \tag{3.6}$$

where $\mathbf{y}(k)$, defined by Equation 2.54, is the uncorrected wavefront gradient measured at cycle k. In fact, $\mathbf{y}(k)$ can never explicitly determined when the optics is set up in the standard configuration (Figure 1.1) where the wavefront sensor is placed after the DM. Therefore, only the residual wavefront is seen by the Shack-Hartmann sensor, and the residual wavefront gradient, $\mathbf{e}(k)$, is measured.

The standard integrator law employed by most AO systems is given by:

$$\mathbf{u}(k) = \mathbf{u}(k-1) + \mathbf{L}[\mathbf{e}(k) + \mathbf{v}(k)] \tag{3.7}$$

where matrix **L** is the wavefront reconstructor and **v**(*k*) includes both electronic and photonic noise. One may wonder why simple higher order controllers, such as a PID type control, are rarely used to control AO systems. For systems where the wavefront sensor CCD frame rate is fast enough to capture the dynamic behavior of the DM, one can most definitely benefit from having a higher order controller. In AOSLOII, the response of the MEMS DM is orders of magnitude faster than the wavefront sensor's frame rate, so even if there are undesirable properties such as a slow rise time or high oscillations in the DM step response, the wavefront sensor will not see it so adding a proportional or derivative term will be useless. It is worth mentioning that many authors have referred to Equation 3.7 as a proportional control law[5, 6, 17, 111]. This is incorrect and should be clarified to avoid future confusion on the subject. The actual integrator implemented on AOSLOII includes a scalar gain and a forgetting factor:

$$\mathbf{u}(k) = (1 - m\kappa)\mathbf{u}(k-1) + \kappa\mathbf{L}[\mathbf{e}(k) + \mathbf{v}(k)] \tag{3.8}$$

where the user can tune the gain κ to any value between 0 and 1 and the leak *m* to any value between 0 and 0.0l. I think this a sensible way for defining the forgetting factor ($1 - m\kappa$) because for example, if the gain (κ) were set to zero, the forgetting factor would not be allowed to go below one. Assuming all aberrations are both observable and controllable, $0 < \kappa \leq 1$ is all that is necessary for asymptotic stability and a perfect correction can be achieved by setting $\mathbf{L} = -\mathbf{T}^{-1}$, $\kappa = 1$ and $m = 0$. But as we will see from a stability analysis, the actual stability will depend more on how the wavefront reconstructor **L** deals with unobservable and uncontrollable modes as well as noise. In any case, combining Equations 3.6 and 3.8 gives us an expression for how the error updates at each cycle:

$$\mathbf{e}(k+1) = \left[(1 - m\kappa)\mathbf{I} + \kappa\mathbf{TL}\right]\mathbf{e}(k) + \kappa\mathbf{TL}\mathbf{v}(k) \tag{3.9}$$

The objective is to design the reconstructor matrix **L** so that the error (residual wavefront gradient) is eliminated in as few control cycles as possible while remaining stable.

When the noise **v**(*k*) appears to be dominating the wavefront measurement, stability becomes an important issue. The exposure time of the CCD camera can be heuristically adjusted through the user interface increasing the light level and subsequently the signal to noise ratio. The tradeoff for doing this is a reduction in temporal bandwidth which currently appears to be less critical than the accuracy in estimating **e**(*k*).

3.2 Calibration and modeling

The use of an interaction matrix (**T**) to model the plant implies linearity, so proper linearization of the input signal is needed before we can even generate an interaction matrix. So far, I have only discussed what the input vector **u**(*k*) means physically. In the control algorithm design, I conveniently assumed that the plant input is linearly related to the wavefront sensor output (residual wavefront gradient **e**(*k*)). However, the physical inputs to the DM are voltages while the DM driver has its own data number system. To begin exploring the possibilities, we needed a direct way to measure actuator influence functions preferably represented by an accurate map of the surface profile of the DM. An influence function is essentially the impulse response of the DM with the constraint that the actuators are defined over a discrete grid but the DM surface is for all practical

purposes continuous. A Shack-Hartmann sensor can technically be used to visualize DM influence functions, but the low sampling densities of a few subapertures per actuator is just not ideal for certain measurements (i.e. height, distribution width, linearity, etc.). Therefore, we resorted to interferometry to carry out this task.

3.2.1 Interferometer

I constructed an interferometer for characterizing the BMC MEMS device. The system is a Twyman-Green modification of the Michelson interferometer[112] where the reference arm mirror is cemented onto the front of piezo-actuator so that phase shifting can be accomplished. Figures 3.5a and 3.5b are a photograph of the interferometer with optical path overlaid and a schematic diagram of the system. The light source is a 635 nm diode laser (OZ optics) where light is brought into the optical path via a fiber connector. The field emitted from the end of the fiber is diverging and is collimated by lens 1. The collimated beam is then split into reference and objective arms via a cube beam splitter. The objective arm beam reflects off the BMC MEMS device, which is our test object, and the reference arm beam reflects off a flat mirror placed roughly the same distance away from the beam splitter as the DM. A convenience of the Twyman-Green setup is that the OPD between the objective and reference arms could be huge, and we would still get high contrast irradiance interference fringes (interferogram) by recombining the two beams. This is due to the coherence length of the laser being on the order of a meter. What this means mathematically is that the fields for the two beams (ψ_{ref} and ψ_{obj}) are correlated and the resultant irradiance, which is the interferogram, will have an oscillatory term. Recall from Chapter 2 that irradiance is simply the modulus square of the field (Equation 2.2 without the constants):

$$S = \left\langle |\psi|^2 \right\rangle_t$$

where the electric field is a function of both space and time, but the *x*, *y*, and *t* arguments were dropped for notational convenience. The brackets denote the time average:

$$\langle \psi \rangle_t = \lim_{T\to\infty} \frac{1}{T} \int_0^T \psi(t)\,dt$$

When we have the sum of two coherent fields, the resultant irradiance distribution specifying an interferogram is given by:

$$\begin{aligned} S &= \left\langle \left(\psi_{ref} + \psi_{obj}\right)\left(\psi_{ref}^* + \psi_{obj}^*\right) \right\rangle_t \\ &= S_{ref} + S_{obj} + 2\,\mathrm{Re}\left\langle \psi_{ref}\psi_{obj}^* \right\rangle_t \end{aligned} \tag{3.10}$$

The wavefront is encoded in the cross-correlation term in Equation 3.10 so its visibility is especially important. Light detection was achieved using a relatively old 8-bit CCD camera (480 by 640 pixel device from The Imaging Source) and digitized with a USB framegrabber (Zarbeco, LLC, Randolph, NJ). Notice that in the schematic diagram (Figure 3.5b), the CCD is not placed at the focal point of lens 2. Whether the beam is in focus on the CCD array or not does not affect the visibility of the interferogram. However, if it was at focus focal point, the interferogram would be shrunk to nearly a point on the CCD making the acquired image useless. Instead, I placed the CCD conjugate to the DM surface. This is obviously not a requirement, but the benefit of doing

so is that the resultant image is the DM with interference fringes overlaid on top (Figure 3.6a)

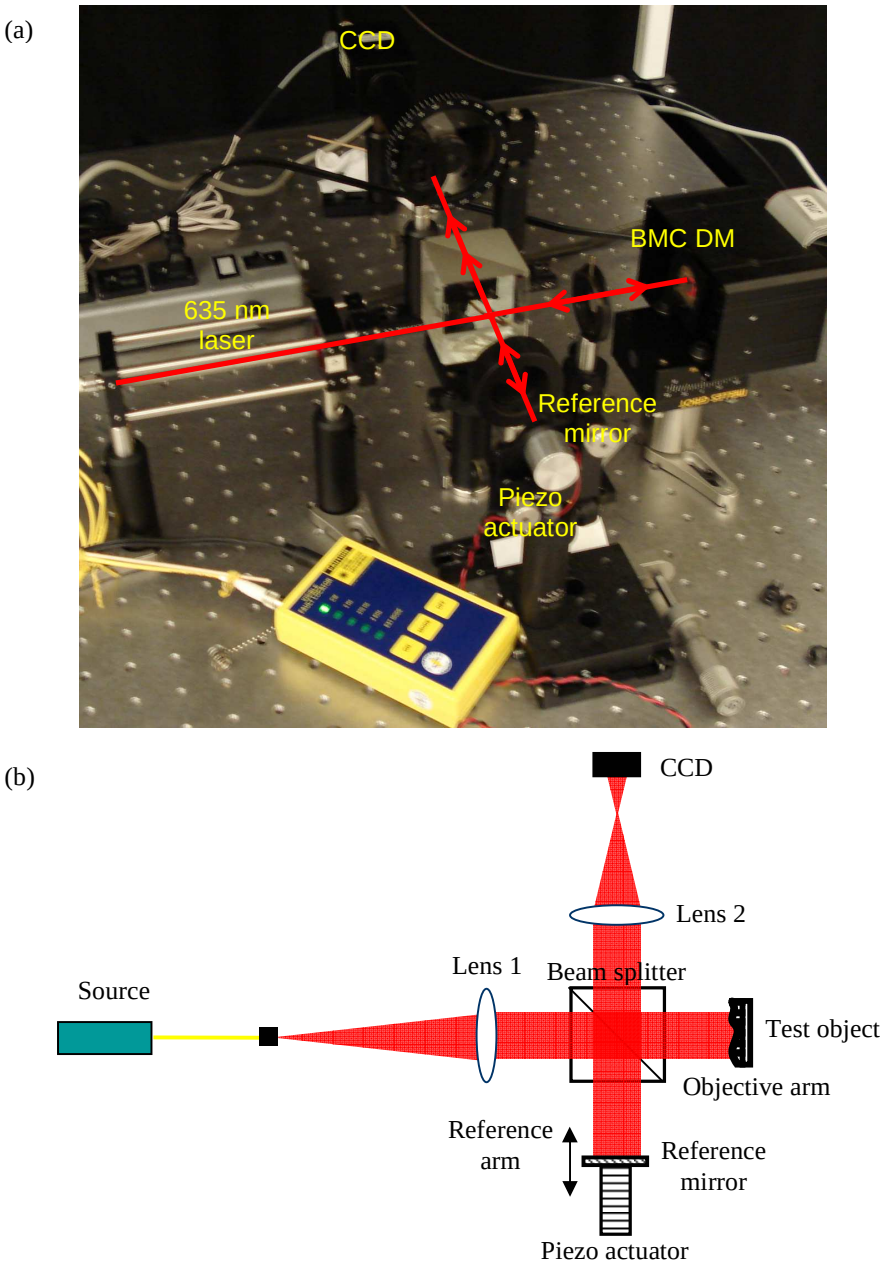


Figure 3.5: (a) Top view of the PSI that I constructed with the optical path overlaid in red and (b) the corresponding schematic diagram illustrate the beam behavior.

An interferogram with a pulled actuator and the corresponding surface profile is given in Figure 3.6. A single interferogram, such as the one in Figure 3.6, encodes the wavefront in waves as a repetitive contour map where the height difference between a dark contour and the next dark contour is an OPD of one wave (635 nm). The OPD between adjacent pixels will clearly be much less than a wave. Since these differences

can be converted to the wavefront gradient by knowing the laser wavelength and the pixel size, the familiar zonal wavefront reconstruction (estimation) problem for Shack-Hartmann sensor data comes to mind. In fact, an interferometer is just an extremely high resolution wavefront sensor that samples the wavefront gradient in the Hudgin configuration (Figure 2.11b). However, many practical issues arise if we want to unwrap a raw interferogram into a wavefront or surface height map such as the one in Figure 3.6b. Examples include:

1. A bright pixel next to a dark pixel can mean either a phase lead or a phase lag
2. Noise level and fringe contrast
3. The 8-bit discretization of the interferogram
4. Computation time

The last issue is due to the large number of pixels compared to subapertures in a Shack-Hartmann sensor, the computational overhead involved in inverting a matrix equivalent to the matrix Γ (Equation 2.65) for zonal reconstruction (estimation) will be orders of magnitude greater. Volumes have been written in dealing with these issues, but a lot of them can be avoided by implementing what is known as phase shifting interferometry (PSI). PSI exploits the wave nature of light to extract the phase (wavefront) from an interferogram. The process is most easily explained by referring back to Equation 3.10. Without any loss of generality, it can be assumed that the reference and objective fields are completely coherent and the initial phase to be zero. This way, an expression for the detected irradiance can be very simple:

$$S(x,y) = a(x,y) + b(x,y)\cos\left(\phi(x,y)\right) \tag{3.11}$$

where a is the DC term ($S_{ref} + S_{obj}$) and the cross-correlation term has been replaced with a cosine. For notation simplicity, it will just be implied from now on that all the terms have a spatial dependence as well (i.e. $a = a(x, y)$). Equation 3.11 is consistent to what we would expect for the irradiance detected an any particular pixel in that if there were no phase error ($\phi = 0$) or a phase error that is an integer multiple of 2π, the irradiance would be $a + b$ making the pixel bright (constructive interference), and if the phase error were an odd multiple of π, the irradiance would be $a - b$ making the pixel dark (destructive interference). Again, the phase error can easily be converted into OPD by noting that 2π is one wave (530 nm). The fringe contrast of an interferogram is defined by the expression:

$$C = \frac{\max\{S\} - \min\{S\}}{\max\{S\} + \min\{S\}} \tag{3.12}$$

which by inspection of Equation 3.11 can be simplified to:

$$C = \frac{a + b - (a - b)}{a + b + (a - b)} = \frac{b}{a} \tag{3.13}$$

We generally operate at a contrast level of about 70 percent which appears to be sufficient for the required calculation for obtaining the phase.

It may appear that obtaining ϕ from Equation 3.11 is trivial but is actually impossible because we do not know the values of a and b, which effectively gives us three unknowns but only one equation. A good estimate of the contrast only gives us how a and b are related to each other but not their actual values. The task is made even more

difficult by the fact that the irradiance is discretized into pixel values between 0 and 255 due to the 8-bit CCD used (the actual range is less as it would depend on the fringe contrast). To generate the necessary minimum of two more equations, a phase shifting mechanism is required, and that is the role of the piezo actuator (Low-Voltage Actuator, Thorlabs) placed behind the reference mirror. By applying a voltage to the piezo, the entire interferogram will experience a constant phase shift δ. This is expressed mathematically as:

$$S = a + b\cos(\phi + \delta) \tag{3.14}$$

In order to convert δ into OPD, simply divide it by 2π and multiply by the wavelength. The distance that the reference mirror had to travel is actually half the OPD as light travels this distance on the way to the mirror and travels the same distance again coming away from the mirror. In principle, only two phase shifts (three equations) are needed, but this bare minimum option can lead to significant errors in the recovered phase if the actual distance shifted deviated from what was prescribed[113]. For example, the 120° three-step algorithm[113] requires an interferogram taken at $\delta = 60°$, 180° and 300°. A 60° shift corresponds to only about 53 nm which is an extremely short distance to push a piezo actuator accurately especially in open loop. I do not believe the current PSI is capable of achieving this level of accuracy. Even though many improved three-step algorithms have been developed, I opted for a slightly more forgiving four-step algorithm for practical reasons that will become apparent. The standard four-step algorithm requires four interferograms acquired sequentially with quarter-wave phase shifts[114]:

$$\begin{cases} S_1 = a + b\cos(\phi) \\ S_2 = a + b\cos(\phi + 0.5\pi) \\ S_3 = a + b\cos(\phi + \pi) \\ S_4 = a + b\cos(\phi + 1.5\pi) \end{cases} \tag{3.15}$$

By applying some simple trigonometric identities, the phase can be extracted from the four interferograms by the simple expression:

$$\phi = \tan^{-1}\left(\frac{S_4 - S_2}{S_1 - S_3}\right) + \delta \tag{3.16}$$

where the added phase constant δ is to make the range of the function be from 0 to 2π (see Table 3.1 for definitions). Equation 3.16 cannot be evaluated for certain conditions because the argument can be singular or ambiguous at certain pixels. We must have a sensible solution for dealing with situations when we are forced to divide by zero or to evaluate zero divided by zero. Table 3.1 provides the required logic applied at each pixel before Equation 3.16 can be evaluated. After this is done, the resultant phase map must be unwrapped since all the values are modulo 2π. Phase unwrapping over a rectangular grid is a fairly straightforward process involving basically just an integrator equation[115]:

$$\begin{aligned} &\phi(x+1, y) = \phi(x, y) + V\left[\phi_w(x+1, y) - \phi(x, y)\right] \\ &V(x) \triangleq \left[x - 2\pi\,\mathrm{int}(x/\pi)\right] \end{aligned} \tag{3.17}$$

where the wrapped phase is now denoted by ϕ_w and the initial condition was simply selected to be the value of the top-left corner pixel. Equation 3.17 must be applied one

row at a time where since the first row is scanned left to right, the second row must then be scanned right to left to preserve phase continuity along the edges. This zigzagging scan pattern continues until the entire phase is unwrapped. It should be noted that due to the setup of the Twyman-Green interferometer, there will inevitably be some tilt fringes in the raw interferograms, and tilt and tip modes will exist in the unwrapped phase. The magnitude of these modes were estimated via linear regressions in both the x- and y-directions and subtracted from the unwrapped phase. The resultant phase (wavefront) can finally be converted to a high resolution topographic map of the DM surface (Figure 3.6b) by calculating the corresponding OPD.

Table 3.1: Logic implied in Equation 3.16

Condition	Action	δ
$S_4 = S_2$ and $S_1 > S_3$	$\phi = 0$	0
$S_4 = S_2$ and $S_1 = S_3$	$\phi = 0.5\pi$	0
$S_4 > S_2$ and $S_1 < S_3$	Evaluate $\tan^{-1}(\cdot)$	π
$S_4 = S_2$ and $S_1 < S_3$	$\phi = \pi$	0
$S_4 < S_2$ and $S_1 < S_3$	Evaluate $\tan^{-1}(\cdot)$	π
$S_4 < S_2$ and $S_1 = S_3$	$\phi = 1.5\pi$	0
$S_4 < S_2$ and $S_1 > S_3$	Evaluate $\tan^{-1}(\cdot)$	2π
$S_4 = S_2$ and $S_1 = S_3$	0	0
Otherwise	Evaluate $\tan^{-1}(\cdot)$	0

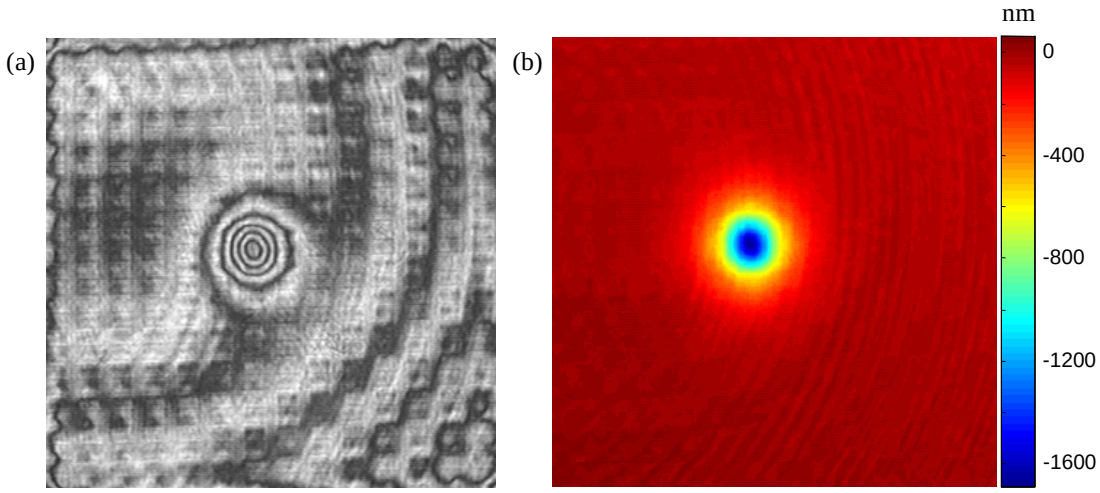


Figure 3.6: (a) Raw interferogram produced by the interference of the reference and objective beams. A differential voltage was applied to a single DM actuator pulling the facesheet down. (b) The corresponding topographic of the DM surface obtained by phase shifting, phase unwrapping, and proper scaling to wavelength of the beams (details given throughout section 3.2.1).

3.2.2 Input linearization

We would like to define the control input $\mathbf{u}(k)$ in such a way so that its entries can take on both positive and negative values and that the transfer function from the control input to the wavefront gradient output is approximately linear. Since actuation is achieved by

applying an electrostatic force, we can rule out the actual voltage applied to the DM as a potential candidate. Many researchers have measured peak deflections as a function of the applied voltage and fitted the measurements with a second-order polynomial. They then perform a linearization step (usually not described in detail but must involve solving for the roots of the fitted polynomial), so that they can treat their DM and wavefront sensor combination as a linear system[91, 99, 116-118]. All but one of these studies used an interferometer for the DM characterization process. For that one study, measurements were made using a Shack-Hartmann sensor, so it should not be surprising that their results were somewhat different from the others as they used a third-order polynomial model.[99] It is difficult to gauge the effectiveness of these linearization attempts based on closed loop performance as they all required many cycles (> 5) to bring the RMS wavefront error to about the optimal correction. The fact that these experiments were from either correcting an artificial model eye or reproducing a particular Zernike mode on the DM is particularly surprising as I would expect that correcting purely static aberrations should only take one or two cycles if the system was properly linearized.

A likely explanation for this observation is given as follows. Recall that a bias voltage needs be applied to the entire DM in order to actuate in both directions. It seemed very likely that the deflection versus voltage curve could be different if the DM characterization was done with respect to the bias position. This turned out to be indeed the case, and with the resultant second-order polynomial fits were used to linearize the input, the AO correction converged much faster even for real eyes[90]. But how is the bias voltage determined? The biased position should accommodate a maximum positive and negative single actuator deflection of equal magnitude. The corresponding voltage input for the ideal biased position was verified with our PSI to be about 190 volts. How this was done was through a series of trial and error experiments involving the following steps:

1. Pick a bias voltage and apply it to all the actuators
2. Release the actuator of interest to 0 volts and take a deflection measurement.
3. Apply the maximum voltage to the same actuator and take that deflection measurement.
4. If the absolute values of these two measurements are nearly equal, then the tested bias voltage is optimal. Otherwise, increase or decrease the bias voltage guess and start over.

If we had selected our bias voltage based on the original deflection versus voltage curves (i.e. driving one actuator with all the other actuators at zero voltage), we would have grossly underestimated the ideal bias voltage.

Through these exercises, it became apparent that a much more intuitive strategy existed. The linearization strategies mentioned so far, including that described in my paper[90], requires us to define the control input $\mathbf{u}(k)$ in units of deflection (typically in μm) which is not intuitive. More importantly, it creates an inconvenience because no two BMC MEMS devices have the same response, even if they are of the same model. Thus, each new DM must go through the same calibration process of measuring deflection versus voltage curves and updating the fitted polynomial coefficients in the control software. Figure 3.7 plots deflection against squared voltage, and it is clear that the

relationship is almost linear. This is consistent with Equation 3.1 in that electrostatic force is proportional to the squared voltage. Furthermore, we can conveniently derive the ideal bias voltage by looking at where the line crosses a deflection of zero (DM bias position) in Figure 3.7. The voltage range for driving the current MEMS DM in AOSLOII is from 0 to 265 volts, so half way in this range in terms of the squared voltage:

$$V_{bias} = \sqrt{\frac{V_{\max}^{\ 2}}{2}} \tag{3.18}$$

results to about 187.4 volts which is very close to our empirically determined result of 190 volts. So the squared voltage is clearly a practical candidate for the linearized input, but in order to allow for bi-directional actuation, the squared voltage must be offset by the square of the bias voltage. Suppose $u_i(k)$ is the i^{th} entry of the linearized input vector $\mathbf{u}(k)$, its value is related to the actual voltage applied to i^{th} actuator $V_i(k)$ by:

$$u_i(k) = V_i^2(k) - V_{bias}^2 \tag{3.19}$$

Linearizing the input in this manner is much simpler than the methods mentioned above, according to convergence rates that we typically achieve, it seems to also be superior (see sections 3.2.2 and 3.3.3 for experimental evidence).

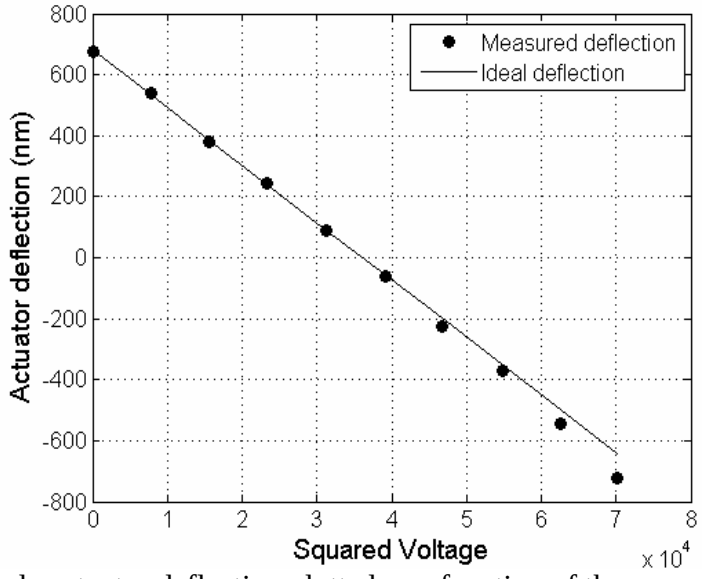


Figure 3.7: Single actuator deflection plotted as a function of the squared voltage. The DM measured is the 5.5 µm Multi-DM from BMC. Notice that the peak-valley dynamic range is more like 1.5 µm.

3.2.3 Interaction matrix

Identification of the interaction matrix **T** is done experimentally by introducing a flat wavefront into the system with a model eye and then measuring the static response of all the actuators one by one[109, 110]. The DM is initially fixed at the biased position, and then starting from the first actuator, the following actions are executed:

1. Completely release the current actuator (0 volts) while keeping all the others at

the bias voltage

2. Measure and record the wavefront gradient. Call the resultant vector **a**
3. Pull the current actuator completely (265 volts) while keeping all the others at the bias voltage
4. Measure and record the wavefront gradient. Call the resultant vector **b**
5. Compute the appropriate column of the interaction matrix. The i^{th} actuator corresponds to the i^{th} column of the interaction matrix which is defined by:

$$\mathbf{t_i} = \frac{1}{u_{\max} - u_{\min}}\left[\mathbf{a} - \mathbf{b}\right] \tag{3.20}$$

where u_{max} and u_{min} are the minimum and maximum values that the entries of the linearized input vector u(k) can take:

$$\begin{cases} u_{\max} = V_{\max}^2 - V_{\max}^2 = 355112.5 \\ u_{\min} = V_{\min}^2 - V_{\min}^2 = -355112.5 \end{cases} \tag{3.21}$$

This procedure is repeated until the response of each actuator has been measured. The interaction matrix can then generated by:

$$\mathbf{T} \triangleq \left[\mathbf{t_1} \quad \mathbf{t_2} \quad \ldots \quad \mathbf{t}_{N_a}\right] \tag{3.22}$$

where N_a is the number of actuators which is 140 for the DM used in AOSLOII. The actuator indexing scheme does not really matter as long as it is kept consistent at every calculation to follow that involves the input vector **u**(*k*). For a 6 mm diameter pupil, the dimensions of the interaction matrix are 426 by 140. In practice, I found that generating several interaction matrices (≈ 10) and then taking the average as the final interaction matrix benefits system performance. The Shack-Hartmann sensor measurement will always contain some electronic and photon noise which gets inserted into the interaction matrix **T**. Averaging several interaction matrices increases the signal to noise ratio or its entries just like how frame-average an images sequence of a scene results in a better signal-to-noise ratio in the final image.

3.3 Wavefront reconstruction (control)

Wavefront reconstruction for control differs from estimation (Chapter 2) in that we are interested in calculating the control signal, which in our case is the linearized input vector **u**(*k*). In other words, we are interested in physically reconstructing the wavefront from wavefront gradient measurements using a DM. The wavefront reconstructor was denoted by the matrix **L** in the figures and equations above, and this section discusses several designs for this matrix. According to the only error budget conducted for a vision AO system, most of the uncorrected wavefront error was attributed to poor wavefront reconstructor design, particularly aliasing, and the temporal bandwidth which was limited by the 16 Hz sampling rate of their Shack-Hartmann sensor[105]. The design of the wavefront reconstructor has taken relatively low priority in the development of vision AO systems so far. It should be clear from the evidence provided in this chapter that wavefront reconstruction and AO control in general can greatly benefit retinal image quality and should probably be higher on the priority list in future projects.

The AO control problem differs substantially from most applications given in textbooks on control systems. One exception is iterative learning control (ILC) which to

my surprise is nearly identical to the AO control problem in many respects. A general ILC update law takes the form[119]:

$$\mathbf{u}_{k+1} = \mathbf{Q}\mathbf{u}_k + \mathbf{L}\mathbf{e}_k \tag{3.23}$$

where the distinction from the AO control problem is the fact that the entries of vectors **u** and **e** are comprised of input and error values over time rather than space. Matrix **L** is also never square in AO control as it is generally assumed to be for ILC. In ILC, the index *k* is for tracking the number of iterations that a specific task is run rather than the number of control loop cycles. The ILC **Q** and **L** matrices are referred to as filters. Specifically, matrix **L** is called the learning filter, which plays a similar role as the wavefront reconstructor in AO control. Matrix **Q** is most often the identity matrix but is sometimes a low pass filter aimed at make the learning process more robust. Both these filters are temporal and generally one dimensional. In the AO control case, matrix **Q** has almost always been either identity or a simple forgetting factor (Equation 3.8) that can be adjusted heuristically. If it were designed as a low pass filter, it would be a two-dimensional spatial filter rather than a one-dimensional temporal filter as in the ILC case. The one exception is the AO controller at the W.M. Keck Observatory whose integrator plus Smith-predictor compensator can be rearranged into the following form:

$$\begin{aligned} \mathbf{u}(k) &= \mathbf{Q}(z^{-1})\mathbf{u}(k-1) + \kappa\mathbf{L}\mathbf{e}(k) \\ \mathbf{Q}(z^{-1}) &= \left((l-w) + lwz^{-1}\right)\mathbf{I} \end{aligned} \tag{3.24}$$

where *w* is set to 0.25 and *l* is the forgetting factor which they set to 0.999 for bright stars and 0.99 for not so bright stars[120]. So **Q** is also a two tap temporal low pass filter which would be equivalent to a filter in the iteration domain for the ILC case. This added temporal filter pushes the AO system's bandwidth at the expense of having more error in rejecting the static and low frequency aberrations. Presumably because most of the eye's wave aberrations are static while the rest are more slowly varying compared to their atmospheric counterparts, I have found that adding this temporal filter more often than not actually reduces retinal image quality, so I do not recommend applying it to vision AO systems.

The question remains on how to best design the wavefront recontructor. The four wavefront reconstructor designs implemented and tested on AOSLOII, which span from basic plant inversion to more advanced quadratically optimal designs as well as a modal reconstructor (Zernike polynomials), are covered in the four subsections below. The name I have given to each reconstructor design is by no means official as different people have used different names to refer to more or less the same thing. Just to keep the equations as compact as possible, I will assume the gain κ and leak *m* to be 1 and 0 respectively as they do not contribute to the purpose of the discussion.

3.3.1 Pseudoinverse

This is the naïve solution to the AO control problem where the reconstructor is designed to be the inverse of the interaction matrix **T**. Since **T** is not square, we have to use the pseudoinverse and the control law becomes:

$$\mathbf{u}(k) = \mathbf{u}(k-1) - \mathbf{T}^{+}\mathbf{e}(k) \tag{3.25}$$

where the negative sign before the pseudoinverse of matrix **T** is to the fact that our closed loop system matrix is **I + TL** and we want to make that as close to the zero matrix as possible. Vision scientists refer to this algorithm as "direct-slope control" which is not

really a term recognized in other AO communities. This has been the standard control strategy (minus the input linearization steps) used in AO systems for vision science since Hofer *et al.*[17] first reported using it in 2001. Although being somewhat novel to the vision science community, according to Wallner's classic paper in 1983[121], most astronomical AO systems were already using this control method well before 1980. Just to put this matter in perspective, there are probably over a hundred publications that describe more sophisticated AO control algorithms, and at least a few of them would be applicable to controlling AO systems for vision science.

Equation 3.25, implemented as it is, has been verified to be unstable, and the reason is because the interaction matrix **T** is poorly conditioned. A widely used technique to invert **T** is by taking the SVD and then ignoring the smallest singular values during the inversion. The number of singular modes to ignore is found experimentally by determining the minimum number of modes that must be dropped while preserving stability. This is typically done using a model eye through trial and error. The truncated SVD scheme turned out to be quite unreliable in practice in the sense that it often failed to correct what seemed to be a perfectly healthy eye when all the other reconstructor designs considered did fine. I believe this is mainly due to the relatively high amount of mismodeling error associated with using an interaction matrix to model a MEMS DM making the system more sensitive to high frequency mode buildup like localized waffling[82]. Thus, in order to include this algorithm in a performance comparison, we would be limiting ourselves to testing only on individuals whose wave aberrations can be effectively compensated via the SVD algorithm. This limitation is inconvenient and could potentially bias our results. As an alternative, I considered a more stable version of the pseudoinverse which comes from the limiting process:

$$\mathbf{T}^{+} = \lim_{\alpha \to 0} \left(\mathbf{T}^{\mathbf{T}} \mathbf{T} + \alpha \mathbf{I} \right)^{-1} \mathbf{T}^{\mathbf{T}} \tag{3.26}$$

By making the regularization factor α easily accessible from the user interface, we can heuristically obtain a working reconstructor of the form:

$$\mathbf{L} = -\left(\mathbf{T}^{\mathbf{T}} \mathbf{T} + \alpha \mathbf{I} \right)^{-1} \mathbf{T}^{\mathbf{T}} \tag{3.27}$$

where α, like that of the zonal estimator described in Chapter 2, should ideally be the noise to signal ratio of the system. The difference between regularization and the truncated SVD is that instead of ignoring the smallest singular values, a constant α is added to every singular value prior to taking the reciprocal. So even though we did not explicitly solve the high frequency amplification problem, at least the inversion is more stable resulting in a more reliable AO controller.

3.3.2 Statistical weighting

The type of statistically weighted wavefront reconstructor I describe here is more or less what is also used at W.M. Keck Observatory.[120] It is currently the default reconstructor used in AOSLOII. Atmospheric turbulence is often modeled to follow Kolmogorov statistics, but there is evidence that the spatial power spectrum of the eye's wave aberrations also follow the Kolmogorov model[122]. Even though evidence for this is limited to just one short paper, the relevance of the work warrants for some exploratory efforts into how we may exploit ocular aberration statistics for AO control systems in vision science. Going back to our linear model of the AO system (Equation 3.6):

$$\mathbf{e}(k) = \mathbf{T}\mathbf{u}(k-1) + \mathbf{y}(k)$$

where $\mathbf{y}(k)$ contains the uncompensated wavefront gradient that is not actually measured during closed loop operation. We want to determine the best estimate of the input $\mathbf{u}(k)$ given measurement $\mathbf{e}(k)$ under the assumption that the eye's wave aberration are either static or slowly varying. This assumption allows us to make the critical approximation:

$$\mathbf{y}(k) \approx \mathbf{y} + \mathbf{v}(k) \tag{3.28}$$

where $\mathbf{v}(k)$ is the measurement noise which is assumed to be Gaussian and white. The best estimate for $\mathbf{u}(k)$ in the least squares sense is given by the conditional expectation:

$$\hat{\mathbf{u}}(k) = -E\left[\mathbf{u}(k)\middle|\mathbf{e}(k)\right] \tag{3.29}$$

Assuming that $\mathbf{e}(k)$ and $\mathbf{u}(k)$ are stationary Gaussian random processes, the conditional expectation can be express in the convenient form:

$$\begin{aligned} E\left[\mathbf{u}(k)\middle|\mathbf{e}(k)\right] &= m_{\mathbf{u}} + \Lambda_{\mathbf{ue}}\Lambda_{\mathbf{ee}}^{-1}\left[\mathbf{e}(k) - m_{\mathbf{e}}\right] \\ &= m_{\mathbf{u}} + \Lambda_{\mathbf{ue}}\Lambda_{\mathbf{ee}}^{-1}\mathbf{e}(k) \end{aligned} \tag{3.30}$$

where $m_{\mathbf{u}}$ and $m_{\mathbf{e}}$ are the means of vectors $\mathbf{u}(k)$ and $\mathbf{e}(k)$ respectively, and $\Lambda_{\mathbf{ue}}$ and $\Lambda_{\mathbf{ee}}$ are covariance matrices involving the appropriate vectors:

$$\Lambda_{\mathbf{ue}} \triangleq \left\langle \mathbf{u}\mathbf{e}^{\mathbf{T}} \right\rangle$$

$$\Lambda_{\mathbf{ee}} \triangleq \left\langle \mathbf{e}\mathbf{e}^{\mathbf{T}} \right\rangle$$

The last equality in Equation 3.30 holds because $\mathbf{e}(k)$ is the measured wavefront gradient with tilt and tip already removed so $m_{\mathbf{e}}$ is zero. By expanding the covariance matrices under the assumption that $\mathbf{u}(k)$ and $\mathbf{e}(k)$ are uncorrelated with each other, we may rewrite Equation 3.40 in the matrix form:

$$E\left[\mathbf{u}(k)\middle|\mathbf{e}(k)\right] = m_{\mathbf{u}} + \Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\left[\mathbf{T}\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}} + \Lambda_{\mathbf{vv}}\right]^{-1}\mathbf{e}(k) \tag{3.31}$$

Assuming zero piston ($m_{\mathbf{u}} = 0$) and substituting Equation 3.31 into Equation 3.29, we arrive at the optimal estimate of the input vector $\mathbf{u}(k)$ in the open loop case:

$$\hat{\mathbf{u}}(k) = -\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\left[\mathbf{T}\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}} + \Lambda_{\mathbf{vv}}\right]^{-1}\mathbf{e}(k) \tag{3.32}$$

where the wavefront reconstructor is given by:

$$\mathbf{L} = -\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\left[\mathbf{T}\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}} + \Lambda_{\mathbf{vv}}\right]^{-1} \tag{3.33}$$

Law and Lane[123] showed that this reconstructor is equivalent to the "optimal estimator" (open loop) described in Wallner's classic paper[121]. The same reconstructor has also been referred to as Bayesian[121], maximum *a posteriori*[124], and a special case of the minimum-variance design[125]. To avoid confusion, I will refer to this design as the statistically weighted reconstructor because it is nothing more than the solution to a statistically weighted least squares problem.

Wallner's derivation was for an optimal open loop AO control scheme for a Linear DM and wavefront sensor. According to his simulations, the statistically weighted reconstructor in open loop performed similarly to the pseudoinverse reconstructor in closed loop under good (low noise) conditions but was superior at low light (noisy) conditions. The key assumption was the linear DM. Even though we linearized the control input, that by no means accounted for all the nonlinearities present in these BMC MEMS devices. We still must rely on closed loop control to deal with mismodeling errors and iteratively reduce the residual wavefront error because we cannot obtain the

best possible correction in just a single step. I implemented the statistically weighted reconstructor (Equation 3.3) with the pure integrator:

$$\mathbf{u}(k) = \mathbf{u}(k-1) + \mathbf{L}\mathbf{e}(k)$$

and its performance for correcting a model eye based on the RMS wavefront error is given in Figure 3.8 (the trend labeled as "Penalizing Kolmogorov statistics only"). What we can immediately observe is that the AO system is not closed-loop stable as the RMS wavefront error started to increase after just a few cycles. This simple test also demonstrates why closed loop control is still required. The wavefront error present in a model eye is completely static, but it required an additional measurement and correction cycle to bring the RMS value below 100 nm. For open loop control, the residual wavefront gradient measurements would not be available to iteratively refine the control signal, so the RMS wavefront error will remain near where it was after the first cycle (≈ 250 nm for the example given in Figure 3.8). But as I will show, the statistically weighted wavefront reconstructor is an optimal design and provides a superior correction in comparison to the standard pseudoinverse reconstructor but only if we can come up with a stable version.

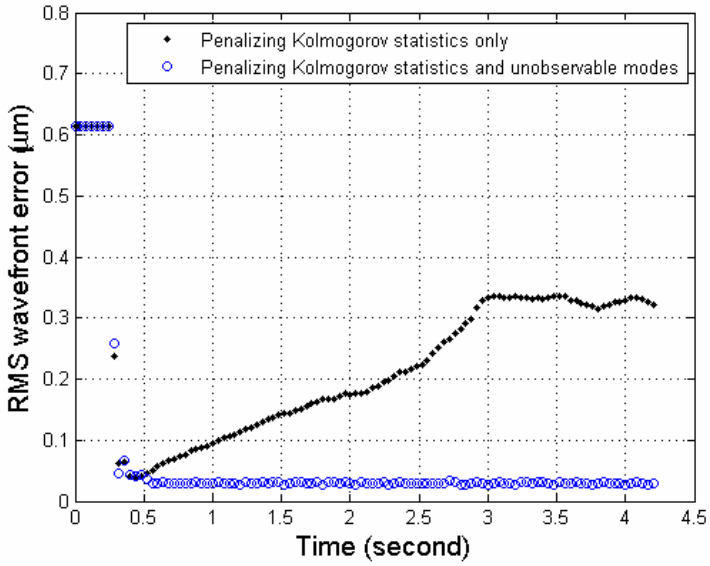


Figure 3.8: Performance of the statistically weighted reconstructor under integral control based on the calculated RMS wavefront error for correcting a static aberration profile (model eye). Closed loop stability was only observed when unobservable modes were explicitly penalized.

Recall that the piston term ($m_{\mathbf{u}}$) was ignored in the derivation of the statistically weighted reconstructor. This would not affect open loop because piston, or any other unobservable DM mode (i.e. tilt and tip), cannot accumulate in open loop. In close loop control, noise and mismodeling errors will inadvertently produce unobservable modes on the DM which will accumulate. In other words, if we keep adding shapes on the DM that are invisible to the wavefront sensor, we will eventually run out of stroke and the correction will actually start getting worse (in the case for piston buildup, the DM will eventually be in a state where the actuators can only move in one direction). This is

exactly what we see in Figure 3.8 for the case where unobservable modes were not penalized (black dots). To explicitly penalize unobservable modes, consider the following identity pertaining to the statistically weighted reconstructor:

$$-\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\left[\mathbf{T}\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}+\Lambda_{\mathbf{vv}}\right]^{-1}=-\left[\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}\mathbf{T}+\Lambda_{\mathbf{uu}}^{-1}\right]^{-1}\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1} \tag{3.34}$$

A proof of this is given in Appendix A. Notice that the right side of Equation 3.34 is the solution to the LQ problem where the cost functional:

$$J(\mathbf{u}(k))=\frac{1}{2}\mathbf{e}^{\mathbf{T}}(k)\Lambda_{\mathbf{vv}}^{-1}\mathbf{e}(k)+\frac{1}{2}\mathbf{u}^{\mathbf{T}}(k)\Lambda_{\mathbf{uu}}^{-1}\mathbf{u}(k) \tag{3.35}$$

is minimized with respect to input $\mathbf{u}(k)$. The idea, as we will see, is very much analogous to the frequency shaped LQR (linear quadratic regulator) design except that we are dealing with two-dimensional spatial frequencies and a system that is quasi-static.

Until this point, I have not mentioned how to generate the covariance matrices. Since we assumed that our measurement noise is white, the simpler of the two is the noise covariance $\Lambda_{\mathbf{vv}}$ which is diagonal. Determining the actual variance values that is the diagonal entries of $\Lambda_{\mathbf{vv}}$ is very difficult if not impossible because 1) they are subject dependent since different retinas have different reflectivity and 2) the variance of the measured wavefront gradient is due to both noise and actual fluctuation in the ocular wavefront. A simple, but reasonable, estimate for the diagonal entries would be values that are inversely proportional to the irradiance distribution across the pupil as discussed in Chapter 2 (Figure 2.4). This way, darker subapertures get weighted less than the brighter subapertures in the cost functional. Generating the inverse of matrix $\Lambda_{\mathbf{uu}}$ is much more difficult. Assuming that the wavefront is approximately proportional to the controller input, a sparse approximation for the inverse wavefront covariance matrix corresponding to the Kolmogorov spectrum has been proposed[126]. According to Noll[58], the Kolmogorov power spectral density (PSD) is given by:

$$\Phi(f_r)=\left(0.023/r_0^{5/3}\right)f_r^{-11/3} \tag{3.36}$$

where f_r is spatial frequency and r_0 is the Fried parameter which is used to describe astronomical seeing. The constant, $0.023/r_0^{5/3}$, is derived from atmospheric turbulence and has nothing to do with the structure of ocular aberrations, so the absolute magnitude of the PSD is not of interest. Given that the PSD is proportional $f_r^{-11/3}$ and then making the not-so-subtle approximation:

$$f_r^{-11/3}\approx f_r^{-4} \tag{3.37}$$

we can derive a very convenient functional relationship through a series of analytical manipulations[55] that I will not repeat here. The functional relationship is given by:

$$\iint u(f_r)\Phi^{-1}(f_r)v(f_r)df_r=\iint\left[\nabla^2u(r)\right]\left[\nabla^2v(r)\right]dr \tag{3.38}$$

where u(r) and v(r) are general square integrable functions. When applied to our problem, the following continuous and discrete spatial domain analogy can be made:

$$\iint\left[\nabla^2u(r)\right]\left[\nabla^2u(r)\right]dr\leftrightarrow\mathbf{u}^{\mathbf{T}}\Lambda_{\mathbf{uu}}^{-1}\mathbf{u} \tag{3.39}$$

and it becomes evident that penalizing the input with a weighting matrix based on Kolmogorov statistics is approximately equivalent to penalizing the Laplacian (curvature) of the input. In matrix form, this approximation can be written as:

$$\Lambda_{\mathbf{uu}}^{-1}\approx\alpha\mathbf{C}^{\mathbf{T}}\mathbf{C} \tag{3.40}$$

where α is again made to be a tunable regularization factor and matrix **C** is the convolution matrix form of the finite impulse response (FIR) filter approximation of the two-dimentional Laplacian operator. As specified in my paper[90], I first implemented the more common Laplacian filter which is specified by the convolution mask:

$$\frac{1}{8}\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{3.41}$$

A better alternative can be obtained using concepts from image processing. Consider the standard one-dimensional Laplacian filter:

$$H(z) = \frac{1}{4}\left(z^{-1} - 2 + z\right) \tag{3.42}$$

where the z-transform variable applies to the spatial domain here. Next, apply the McClellan transformation to this one-dimensional design in order to get the two-dimension equivalent. The resultant filter is specified by the convolution mask:

$$\begin{bmatrix} 0.0625 & 0.125 & 0.0625 \\ 0.125 & -0.75 & 0.125 \\ 0.0625 & 0.125 & 0.0625 \end{bmatrix} \tag{3.43}$$

The advantages afforded by this filter over the more common design (Equation 3.41) are best illustrated by comparing the shape of their frequency responses (Figure 3.9). The alternative filter design clearly has much more circular uniform frequency response which is a desirable property if we want the same filter applied in all directions: an assumption that is implicit in the "turbulence" model (Equation 3.36) assumes.

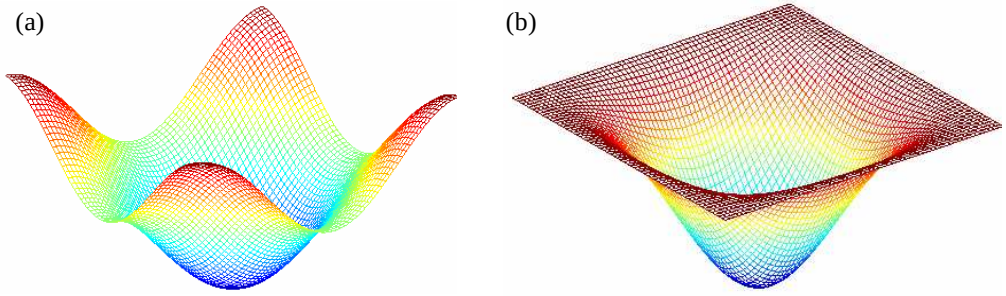


Figure 3.9: Frequency response of the (a) common two-dimensional Laplacian filter design (Equation 3.41) and (b) an alternative design obtained via a McClellan transformation of the standard one-dimensional Laplacian filter (Equation 3.43)

It is now more evident as to why the statistically weighted wavefront reconstructor, in its current form, is not closed loop stable. Analogous to the LQR design, while the weighting matrix for the state, **e**(*k*), is only required to be positive semi-definite, the weighting matrix on the input is required to be strictly positive-definite to keep the input bounded. Unfortunately, the weighting matrix $\mathbf{C}^{\mathbf{T}}\mathbf{C}$ does not satisfy this requirement because the nullspace of the Laplacian operator **C** is spanned by piston, tilt and tip (Figure 3.10). This should be quite obvious because differentiating any one of

these three modes twice will result in zero. Therefore, these three unobservable modes must be penalized explicitly by modifying our cost functional:

$$J(\mathbf{u}(k)) = \frac{1}{2}\mathbf{e}^{\mathbf{T}}(k)\Lambda_{\mathbf{vv}}^{-1}\mathbf{e}(k) + \frac{1}{2}\mathbf{u}^{\mathbf{T}}(k)\left[\Lambda_{\mathbf{uu}}^{-1} + \mathbf{V}\mathbf{V}^{\mathbf{T}}\right]\mathbf{u}(k) \tag{3.44}$$

where piston, tile and tip make up the columns of matrix **V**. By minimizing this new cost function with respect to **u**(k), we finally arrive at an implementation for the statistically weighted reconstructor that is also closed loop stable:

$$\mathbf{L} = -\left[\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}\mathbf{T} + \Lambda_{\mathbf{uu}}^{-1} + \mathbf{V}\mathbf{V}^{\mathbf{T}}\right]^{-1}\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1} \tag{3.45}$$

Figure 3.8 clearly shows closed loop stability when using this reconstructor which penalizes both Kolmogorov statistics and the three unobservable modes.

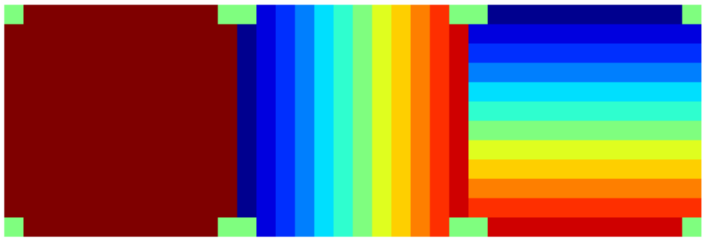


Figure 3.10: From left to right: piston, tilt and tip modes in the control input domain. Note the corner actuators are not included in the data

3.3.3 Local waffle penalty

In Chapter 2, I talked about global waffle which is a single unobservable mode (Figure 2.12). Waffling patterns can also be of the local type, and according to reports across a variety of AO disciplines, their presence in AO systems is common producing undesirable artifacts in the image[82, 127-130]. Waffle modes are created by driving adjacent actuators in opposite directions producing a voltage map resembling a checkerboard pattern. Patches of this pattern are often observed when the pseudoinverse reconstructor, particularly the truncated SVD method, is used. Since they are not well sensed by the wavefront sensor, they can slowly build up in the control loop degrading retinal image quality in the process. Some examples of local waffle modes are given in Figure 3.11.

Penalizing local waffle modes is almost identical to statistical weighting[82]. In the original design, there is no weighting on the measurement vector **e**(k) and the weighting on the input **u**(k) is the convolution matrix form of the FIR filter[82]:

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{3.46}$$

which is proportional to the first derivative operator. The cost functional I used to obtain the local waffle penalty reconstructor is:

$$J(\mathbf{u}(k)) = \frac{1}{2}\left\|\mathbf{e}(k)\right\|_2^2 + \frac{1}{2}\mathbf{u}^{\mathbf{T}}(k)\left[\alpha\mathbf{F}^{\mathbf{T}}\mathbf{F} + \mathbf{V}\mathbf{V}^{\mathbf{T}}\right]\mathbf{u}(k) \tag{3.47}$$

where **F** is the convolution matrix form of local waffle filter (Equation 3.46) and α is a tunable regularization factor. The corresponding reconstructor is given by:

$$\mathbf{L} = -\left[\mathbf{T}^{\mathbf{T}}\mathbf{T} + \alpha\mathbf{F}^{\mathbf{T}}\mathbf{F} + \mathbf{V}\mathbf{V}^{\mathbf{T}}\right]^{-1}\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1} \tag{3.48}$$

Matrix **V** is the same as that used in the statistically weighted reconstructor. Strictly speaking, since local waffle is proportional to the first derivative, piston is the only mode that needs to be included in **V**, but empirical tests have revealed that tilt and tip modes (Figure 3.10) still tend to accumulate if they are not penalized. Therefore, they are included in the current local waffle penalty reconstructor design.

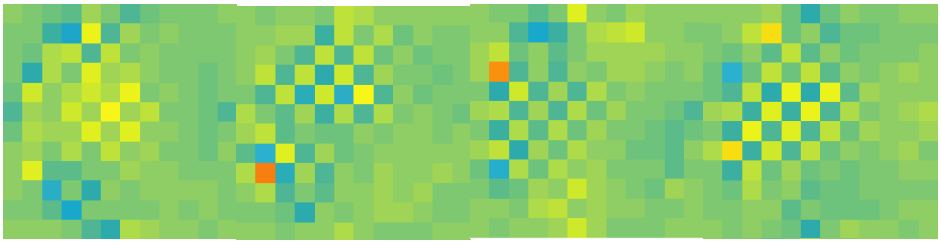


Figure 3.11: Localized waffling structure in the input domain

3.3.4 Zernike polynomials

Lastly, if a finite number of Zernike polynomials can accurately represent the eye's wave aberrations, projecting the input vector **u**(k) onto a Zernike spanned subspace could potentially improve system robustness during less-than-ideal experimental conditions. Furthermore, this approach, which is often referred to as modal reconstruction, provides a convenient platform for designing custom reconstructors for correcting only a particular set of Zernike modes (i.e. defocus, astigmatism, etc.) and leaving all other modes intact. This added flexibility would not benefit retinal imaging but could prove to be a useful tool in vision testing applications.

Recall the Zernike polynomial pyramid given in Chapter 2 (Figure 2.7).An input vector defined by the first N Zernike modes (minus piston, tilt and tip) can be written as:

$$\mathbf{u}(k) = \sum_{j=3}^{N-1} c_i(k)\mathbf{z}_j = \mathbf{Z}\mathbf{c}(k) \tag{3.49}$$

where c_j and $\mathbf{z}_j$ are the j^{th} Zernike coefficient and mode if vectorized form respectively:

$$\begin{cases} \mathbf{Z} = \begin{bmatrix} \mathbf{z}_3 & \mathbf{z}_4 & \dots & \mathbf{z}_{N-1} \end{bmatrix} \in \mathbb{R}^{140\times N} \\ \mathbf{c} = \begin{bmatrix} c_3 & c_4 & \dots & c_{N-1} \end{bmatrix}^{\mathbf{T}} \in \mathbb{R}^{140} \end{cases} \tag{3.50}$$

It is worth noting that since we are currently working in the input space, Equation 3.49 is fundamentally different from Equation 2.29:

$$\phi(x,y) = \sum_{j=0}^{N-1} c_j Z_j(x,y)$$

By substituting Equation 3.49 into Equation 3.6:

$$\mathbf{e}(k) = \mathbf{T}\mathbf{u}(k-1) + \mathbf{y}(k) \tag{3.51}$$

the resultant expression can be interpreted as an equivalent AO system but with the input vector being made up of Zernike coefficients:

$$\mathbf{e}(k) = \mathbf{T}\mathbf{Z}\mathbf{c}(k-1) + \mathbf{y}(k) \tag{3.52}$$

Under the same assumptions as the previous reconstructor designs, the Zernike polynomial reconstructor is found by minimizing the following cost functional with respect to the Zernike coefficient vector **c**(k):

$$J(\mathbf{c}(k)) = \left\|\mathbf{e}(k)\right\|_2^2 + \alpha^2 \left\|\mathbf{c}(k)\right\|_2^2 \tag{3.53}$$

The optimal open loop estimate of the Zernike coefficient vector is given by:

$$\hat{\mathbf{c}}(k) = -\left[(\mathbf{TZ})^{\mathbf{T}}(\mathbf{TZ})\right]^{+}(\mathbf{TZ})^{\mathbf{T}}\mathbf{e}(k) \tag{3.54}$$

Equation 3.54 is not particularly useful because we are not interested in the Zernike coefficients. In order to obtain the final reconstructor, we need to left-multiply both sides of Equation 3.54 by matrix **Z** since:

$$\hat{\mathbf{u}}(k) = \mathbf{Z}\hat{\mathbf{c}}(k) \tag{3.55}$$

Therefore, the Zernike polynomial reconstructor is:

$$\mathbf{L} = -\mathbf{Z}\left[(\mathbf{TZ})^{\mathbf{T}}(\mathbf{TZ})\right]^{+}(\mathbf{TZ})^{\mathbf{T}} \tag{3.56}$$

It was stated in Chapter 2 regarding curve fitting Zernike polynomials to a sampled wavefront that the Zernike polynomials lose their orthonormality when evaluated over a discrete grid. The same applies to this Zernike polynomial reconstructor. Since the DM grid is even coarser than the wavefront sensor lenslet array, matrix **Z** here is even more ill-conditioned. I have experimentally verified that the Zernike polynomial reconstructor in Equation 3.56 is not closed-loop stable in AOSLOII no matter how high the regularization factor is set. But when the Gram-Schmidt orthogonalization procedure was applied to the columns of matrix **Z** prior to evaluating Equation 3.56, the resultant reconstructor was verified to be closed-loop stable.

3.4 Discussion

I have covered both the theory and implementation of the different AO control algorithms implemented on AOSLOII. In this section, I discuss some of the specific concerns and limitations I have encountered in carrying out this work. A particular topic of interest is how different AO controllers compare with each other in terms of performance, so I will also discuss the stability and performance of the four wavefront reconstruction algorithms in the paragraphs below. The data presented here for the algorithm performance comparison are from my paper[90].

3.4.1 Interferometry concerns

Many important concepts and subtleties concerning interferometry, especially PSI techniques, are very much beyond my expertise, so I will not be able to address my concerns on this topic in detail like the others. Using the current interaction matrix model of the AO system, the interferometric measurements were not particularly critical besides verifying the ideal bias position/voltage for the DM. However, if we were to start exploring the practical application of more advanced models[94, 97, 131] of these BMC MEMS devices, a more accurate and precise PSI would most definitely be required.

As stated earlier, the standard four step phase detection algorithm was used because there is a convenient way to test whether or not the phase shift induced by the piezo is a quarter-wave (≈ 159 nm). Since we already have one measurement without performing a phase shift, only three quarter-wave shifts are needed to obtain the four required interferograms. Without much additional effort, a fourth and final shift can be added to the routine and the fifth interferogram should ideally be:

$$S_5 = a + b\cos(\phi + 2\pi) \tag{3.57}$$

So S_5 should equal S_1 since $\cos(\phi + 2\pi) = \cos(\phi)$, but in practice, two such measurements would never be exactly equal due to noise and errors in the phase shift which leads to what is referred to as detuning error. Since the current PSI setup has no position feedback from the piezo actuator, I am particularly concerned with this type of error. My attempt at managing detuning error are as follows:

1. Acquire five interferograms using the estimated nominal voltage applied to the piezo actuator for phase shifting.
2. Use the following formula to calculate the average phase shift corresponding to the nominal voltages applied:

$$\delta = \cos^{-1}\left(\frac{1}{2}\frac{S_5 - S_1}{S_4 - S_2}\right) \tag{3.58}$$

3. If the angle is between 80° and 100°, execute phase detection and unwrapping calculations.

The criterion for calculating the phase in step 3 may appear to be too tolerant, but the success rate for achieving even that kind of precision in open loop is less than 50 percent. Recall the deflection versus squared voltage plot given in Figure 3.7, generating the dataset for such an analysis extremely difficult with this level of success rate. According to Malacara *et al.*[114], the standard four step algorithm is quite sensitive to detuning error. A slightly different algorithm, known simply as "four steps in X"[114], is not as sensitive and may be worth testing, but it is not as straightforward to implement. In any case, I would recommend fitting a strain gauge to the piezo actuator for position feedback before investigating more advanced DM modeling techniques.

3.4.2 AO loop stability

The following discussion is mostly from my paper with some minor modifictions[90]. Stability for a closed loop AO system is primarily numerical because other than a single time-delay element, the plant model has no temporal dynamics. As mentioned earlier, this is because the DM is quasi-static with respect to CCD frame-rate of the wavefront sensor. The expression for the residual wavefront gradient (error) dynamics is given by Equation 3.9 (assuming default values for the integral gain and leak ($\kappa = 1$ and $m = 0$)):

$$\mathbf{e}(k+1) = \left[\mathbf{I} + \mathbf{TL}\right]\mathbf{e}(k) + \mathbf{TL}\mathbf{v}(k)$$

The stability for the AO system depends on the eigenvalues of matrix **I + TL** assuming that it is static. The eigenvalues of **I + TL** are less than or equal to one with at least one eigenvalue being exactly one which corresponds to the piston mode. In practice, high spatial frequency aberration modes that could not be sensed or controlled will also behave as if the corresponding closed-loop eigenvalues are one. How the eigenvalues are distributed depends of the reconstructor design. In order to investigate convergence and steady state error, consider the Lyapunov function candidate:

$$V(\mathbf{e}(k)) = \mathbf{e}^{\mathbf{T}}(k)\mathbf{e}(k) \tag{3.59}$$

Assuming that the error and noise terms are uncorrelated, substituting the error dynamics expression above into Equation 3.59, we get:

$$V(\mathbf{e}(k+1)) = \mathbf{e}^{\mathbf{T}}(k)\left[\mathbf{I} + \mathbf{TL}\right]^{\mathbf{T}}\left[\mathbf{I} + \mathbf{TL}\right]\mathbf{e}(k) + \mathbf{v}^{\mathbf{T}}(k)\left[\mathbf{TL}\right]^{\mathbf{T}}\mathbf{TL}\mathbf{v}(k) \tag{3.60}$$

Letting **W** be the weighting matrix on the input **u**(k), it can be shown that:

$$V(\mathbf{e}(k+1)) = \mathbf{e}^{\mathbf{T}}(k)\mathbf{e}(k) - \mathbf{e}(k)\mathbf{L}^{\mathbf{T}}\left[\mathbf{T}^{\mathbf{T}}\mathbf{T} + 2\mathbf{W}\right]\mathbf{L}\mathbf{e}(k) + \mathbf{v}^{\mathbf{T}}(k)\left[\mathbf{TL}\right]^{\mathbf{T}}\mathbf{TL}\mathbf{v}(k) \quad (3.61)$$

Stability would require the difference along the trajectory to be non-positive, that is:

$$\Delta V(\mathbf{e}(k)) = V(\mathbf{e}(k+1)) - V(\mathbf{e}(k)) \le 0 \quad (3.62)$$

By taking the difference between Equations 3.61 and 3.59, we arrive at the desired stability criterion:

$$\Delta V(\mathbf{e}(k)) = \mathbf{v}^{\mathbf{T}}(k)\left[\mathbf{TL}\right]^{\mathbf{T}}\mathbf{TL}\mathbf{v}(k) - \mathbf{e}(k)\mathbf{L}^{\mathbf{T}}\left[\mathbf{T}^{\mathbf{T}}\mathbf{T} + 2\mathbf{W}\right]\mathbf{L}\mathbf{e}(k) \le 0 \quad (3.63)$$

which implies that:

$$\mathbf{v}^{\mathbf{T}}(k)\left[\mathbf{TL}\right]^{\mathbf{T}}\mathbf{TL}\mathbf{v}(k) \le \mathbf{e}(k)\mathbf{L}^{\mathbf{T}}\left[\mathbf{T}^{\mathbf{T}}\mathbf{T} + 2\mathbf{W}\right]\mathbf{L}\mathbf{e}(k) \quad (3.64)$$

Equation 3.64 is important because it completely summarizes why closed loop AO systems, at least how they currently operate, are actually never guaranteed to bring the error to zero. The inequality in Equation 3.64 implies that for $V(\mathbf{e}(k))$ to decrease if $\mathbf{T}^{\mathbf{T}}\mathbf{T} + 2\mathbf{W}$ is positive definite and that the quadratic form of $\mathbf{e}(k)$ on the right of Equation 3.63 is dominating the quadratic form of $\mathbf{v}(k)$ on the left of Equation 3.63. In other words, Equation 3.64 specifies a region where $V(\mathbf{e}(k))$ is assured to decrease. But we can do ourselves a favor by designing a wavefront reconstructor (**L**) that does not 1) amplify noise and 2) produce shapes on the DM surface that are difficult to sense. Therefore, in less than ideal conditions (i.e. real eyes), we would always expect the statistically weighted reconstructor to outperform the other designs considered.

If noise is "small", the stability criterion simplifies to:

$$\mathbf{e}(k)\mathbf{L}^{\mathbf{T}}\left[\mathbf{T}^{\mathbf{T}}\mathbf{T} + 2\mathbf{W}\right]\mathbf{L}\mathbf{e}(k) \ge 0$$

and the analysis becomes much more definitive. Matrix **W** should be strictly positive definite to keep the input bounded and avoid actuator clipping. For our pseudoinverse and Zernike polynomial reconstructor designs, **W** was proportional to the identity matrix (Tikhonov regularization) so positive definiteness is trivial. For the other two reconstructor designs, we manually identified specific modes that needed to be explicitly penalized in order to establish positive definiteness. It follows immediately that:

$$\left[\mathbf{T}^{\mathbf{T}}\mathbf{T} + 2\mathbf{W}\right] \succ 0 \quad (3.65)$$

which may lead us to believe that $\Delta V(\mathbf{e}(k)) < 0$ guaranteeing asymptotic stability in the absence of noise $\mathbf{v}(k)$. However, the wavefront reconstructor **L** is a generalized inverse of a non-square matrix **T** that has more rows than columns, so it cannot have full column rank. Therefore, we can only guarantee that:

$$\begin{aligned} &\mathbf{L}^{\mathbf{T}}\left[\mathbf{T}^{\mathbf{T}}\mathbf{T} + 2\mathbf{W}\right]\mathbf{L} \succeq 0 \\ &\Rightarrow \Delta V(\mathbf{e}(k)) \le 0 \end{aligned} \quad (3.66)$$

so the system is only stable in the sense of Lyapunov. For the Zernike polynomial reconstructor, simply replace the input vector $\mathbf{u}(k)$ with the Zernike coefficient vector $\mathbf{c}(k)$ and the same stability analysis applies.

A more thorough stability analysis would require accurate modeling of electrostatic actuation coupled with the membrane deformation properties of the DM. A mathematical model of the type of MEMS device used in this study has been assessed[97]. It was not adopted for this study because the model's predicted membrane response did not closely match the actual membrane response of our MEMS device.

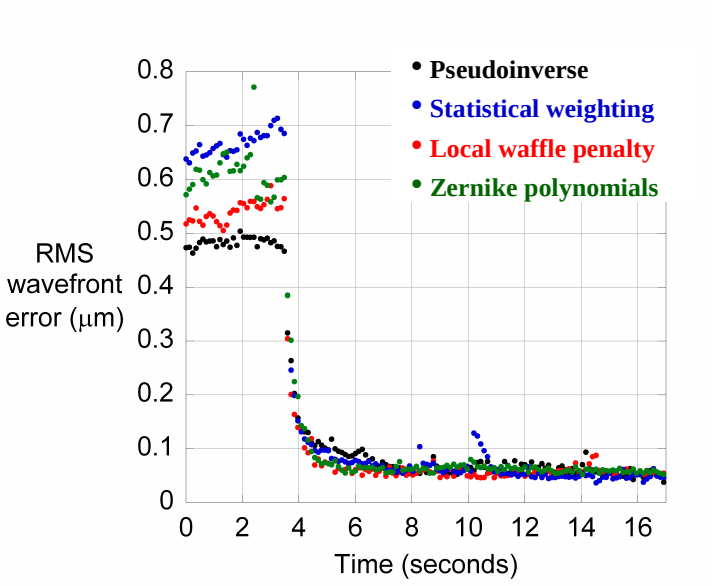


Figure 3.12: Performance based on the estimated RMS wavefront error

3.4.2 Performance comparison

Performance evaluations based on actual retinal images are carried out by repeatedly imaging the same eye but using a different algorithm in each session. This must all be done in a single setting to ensure that the imaging conditions between sessions are nearly identical. I kept individual imaging sessions short (~20 seconds) and administered them only minutes apart to minimize subject fatigue which may bias the comparisons. The center of the raster scan was placed approximately 0.4 degrees from the subjects' preferred retinal locus (PRL), and the scan angle was 1°. An obvious image quality metric for assessing performance would simply be the RMS wavefront error. I used the zonal reconstructor (Equation 2.71) for estimating the residual wavefront for reasons given in Chapter 2. Figure 5 shows typical RMS wavefront error trends for correcting a healthy eye. What we can immediately observe is that all four trends converges quickly to near the best corrected state and more or less remained there until the end of each imaging session. Does this mean that all four wavefront reconstructors performed equally well? The acquired images of the retina (Figure 3.13) do not seem to support this claim. Figure 3.13b is noticeably brighter and sharper than Figure 3.13a, but according to the RMS wavefront error, they were acquired under more or less the same optical conditions. Although not directly verifiable, I attribute this observation to the presence of certain aberration profiles from either the DM or the eye can lie beyond the sampling capabilities of the Shack-Hartmann sensor.

(a) (b)

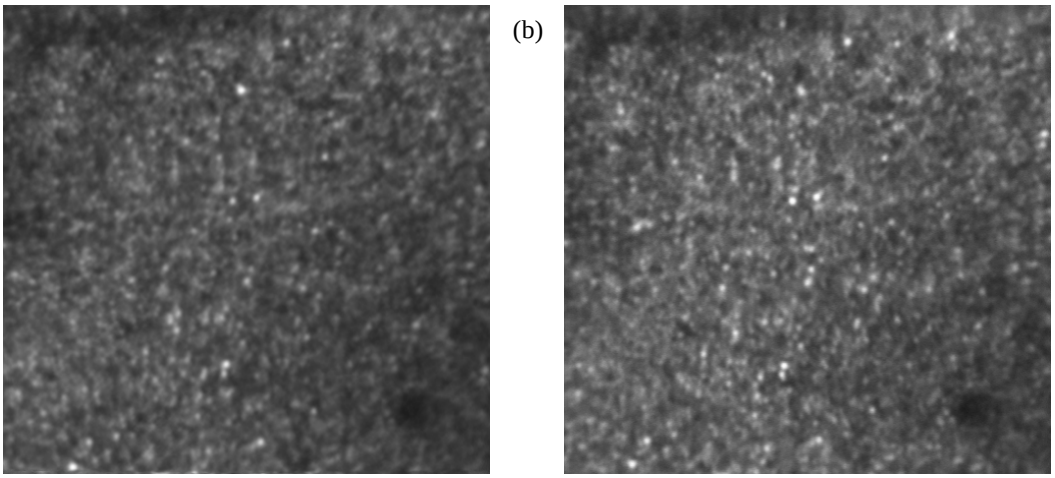


Figure 3.13: AOSLOII images for a healthy eye acquired at the exact same retinal location using (a) the Zernike polynomial reconstructor and (b) the statistically weighted reconstructor. Images are about 0.8° by 0.8° (0.25 mm by 0.25 mm) with the foveal center at approximately the bottom right corners.

The pixel values in an AOSLO image are more direct indicators of retinal image quality for reasons tied to the image formation process of a confocal system including SLO. In an AOSLO, a spot of light is raster scanned on the retina and the reflected light gets descanned on the way back, since light travels much faster than the speed of the scanning mirrors, and eventually reaching a point detector (the PMT in Figure 1.9). So only a small area on the retina is illuminated and detected at a particular moment in time. The final retinal image is actually constructed pixel by pixel from light detected at different time points. Ignoring the uncommon path errors, the irradiance PSF of this double-pass process evaluated at the plane of the confocal pinhole is given by the autocorrelation of the single-pass irradiance PSF[22, 42]:

$$S_{double-pass}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(x',y')S(x'+x,y'+y)dx'dy' \tag{3.67}$$

The PMT detects the total irradiance transmitted by the pinhole, so a particular pixel value of a retinal image is always proportional to the transmitted irradiance corresponding point in the raster scan[42]. Mathematically, the value of the j^{th} pixel in an image is given by:

$$I_j \propto \left\langle \int_{\Sigma} S_{double-pass}(x,y)dxdy \right\rangle_t \tag{3.68}$$

where the actual temporal integration time of the PMT can be estimated from the support of the retinal image (512 by 512) and the imaging frame-rate (30 Hz). Minimizing the residual wavefront error using AO condenses the spatial spread of the double-pass PSF, so more irradiance is transmitted through the confocal pinhole and increasing the pixel value. In other words, a brighter image is a better image and that was clearly illustrated in Figure 3.13. Obviously, looking at individual pixel values would be mostly meaningless as the eye is always moving, but the mean pixel value of the image should be a robust indicator of relative image brightness and quality. Figure 3.14 plots the mean pixel values as functions of time for the same eye whose RMS wavefront error and retinal image were given in Figures 3.12 and 3.13. Based on this image quality metric, it becomes clear that

the statistically weighted reconstructor provides superior image quality. This result has been verified to be consistent for two other eyes (two subjects) tested.

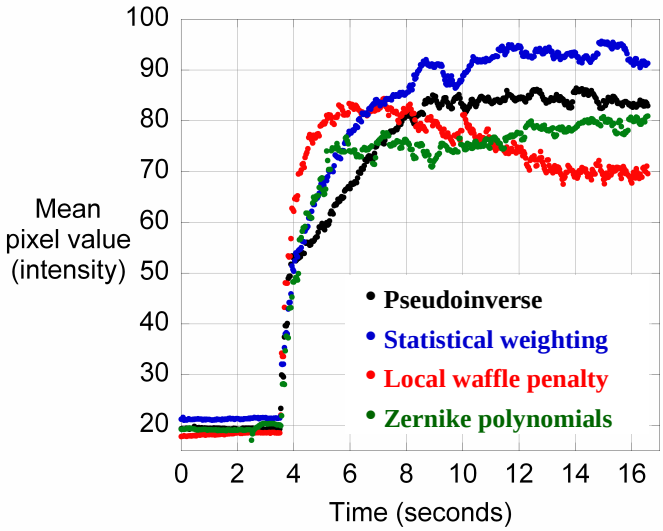


Figure 3.14: Performance based on the mean pixel values of the retinal images

I should mention that we have upgraded both software and hardware components of the AO controller for AOSLOII since acquiring the data presented in this subsection. As stated in the introduction, we are now running a higher stroke MEMS DM with higher step resolution driver electronics. These upgrades should only improve AO performance independent of the control algorithm, so there is little reason to believe that the relative performance among the different reconstructor designs should change. Since the statistically weighted reconstructor should theoretically calculate a better correction, its performance with respect to other reconstructor designs should only improve with these system upgrades. Furthermore, the original implementation of the statistically weighted reconstructor is, in principle, inferior to the version currently implemented as the inverse noise covariance matrix was assumed to be proportional to identity and the more common Laplacian filter mask (Equation 3.41) was used to approximate the inverse input convariance matrix. Based on the comparisons already conducted and the increased number of high quality images of the foveal cone mosaic that we have been able to acquire, I believe it is not really necessary to continue testing whether or not the statistically weighted reconstructor design is superior to the other reconstructors considered in this study. In my opinion, more effort should probably be spent on improving optimal reconstructor designs and their flexibility (i.e. changes in pupil size and shape) as well as developing more sophisticated DM models.

3.5 Conclusions

Improving the AO control system has yielded quantifiable improvements in retinal image quality. The (near) linearity observed between the DM surface deflection and squared voltage should be exploited for controlling BMC MEMS devices. Stability analysis of the standard integrator AO control law revealed that under the presence of noise, the

wavefront reconstructor dictates system behavior. Since, the statistically weighted reconstructor is designed to respond less to noisier measurements while penalizes high frequency DM modes, it proved to be superior to the standard pseudoinverse reconstructor ("direct-slope control") and the two other reconstructor designs tested (local waffle penalty and Zernike polynomials).

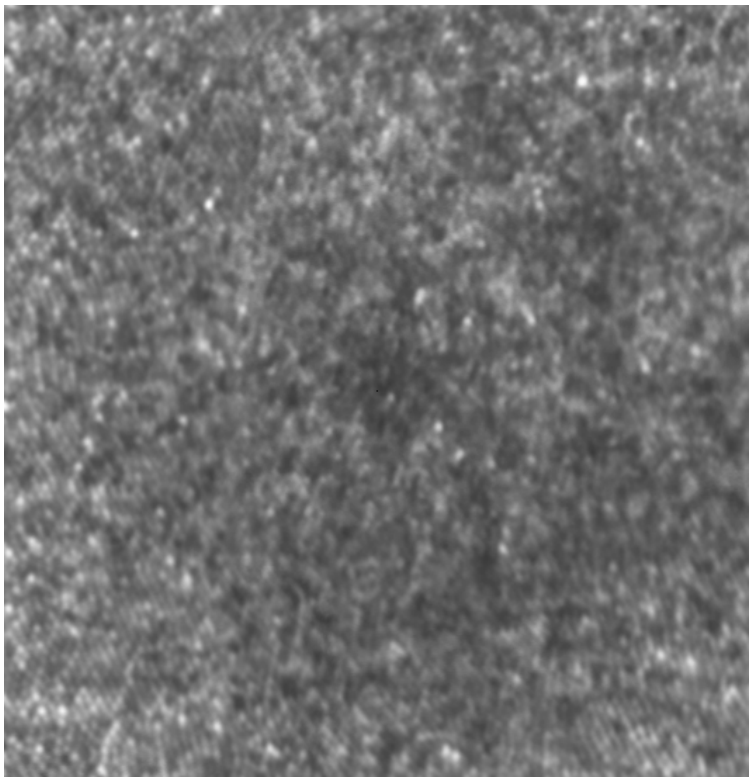
Two quantitative image quality metrics were used to evaluate the performance of the control algorithms: 1) RMS residual wavefront error and the 2) mean pixel value of the acquired retinal image. Even though the four reconstructors under integral control performed similarly according to the computed RMS wavefront error, they did not all produce retinal images of similar quality. The mean pixel value is a more sensitive indicator of retinal image quality because it is directly related to the double-pass PSF. This important limitation regarding the use of the computed RMS wavefront error to characterize AO system performance should be strongly emphasized.

These improvements have also allowed us to resolve the highest density foveolar cones in normal, healthy eyes. Figure 3.15 gives a particular example of such, and several more will be given in the following chapter. Although the current lateral resolution of the AOSLO is sufficient for resolving the smallest foveal cones, it is still unable to do so for all eyes. This is most likely due to the amount of residual aberrations being subject dependent. One reason why this might be the case is because the statistically weighted reconstructor is currently optimized for only a 6 mm diameter pupil. This poses a problem when imaging certain high myopes when the minifying effect of a high minus power lens is placed in front of the eye to bring the initial aberration magnitude to within correctable range. In addition, the assumed input statistics are that which approximates a Kolmogorov power spectrum[122]. Therefore, AO performance will depend on how well the aberration profile for a particular eye is approximated by this model.

Acknowledgements

The C library used for the real time implementation of our AO controller was developed by David Auslander which he distributed to his students as a resource for work assigned in his course (ME 230). I acknowledge David Merino for realigning the interferometer which greatly increased the fringe contrast. I would also like to thank Curt Vogel again for the many clarifications he gave on the mathematics associated with wavefront reconstruction. Sandipan Mishra contributed to various parts of the research particularly the stability analysis.

(a)



(b)

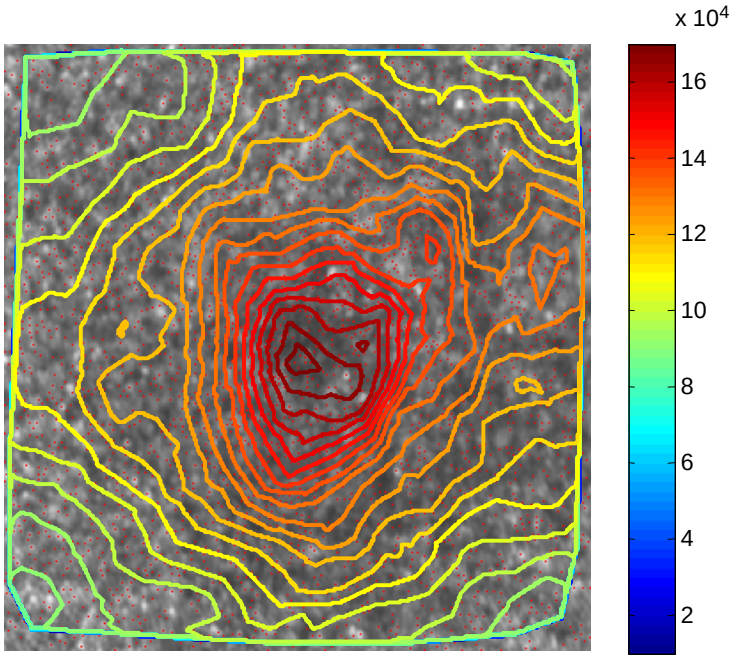


Figure 3.15: (a) Image of the cone mosaic at the very center of the fovea of a young (23 yrs), healthy emmetrope and (b) its corresponding cone density topographic map expressed in cones per square mm. The image is about 1° by 1° which for this particular subject converts to 288 µm by 288 µm.

4. Retinal imaging applications: cone density, foveal fixation and eye length

4.1 Introduction

The research described in the Chapters 2 and 3 was technical but marks some of the first work done on improving AO control systems for vision science applications. This chapter covers the scientific contributions of my dissertation: 1) the relationship between foveal fixation and cone density; and 2) inter-subject variability of foveal cone density in relation to axial length. Some basic but important refinements to estimating retinal feature size and cone density, which were applied to address these two goals, are also described in this chapter. The experiments were conducted in parallel with the developments in wavefront measurement and control described in the earlier chapters, and this partially explains why our retinal image dataset was not particularly consistent in quality. Nevertheless, I was able to image the foveal cone mosaics of more than 20 eyes with AOSLOII. As I mentioned earlier, to avoid potential bias, only 18 eyes (18 subjects) were included in the investigations detailed in this chapter. To my knowledge, this is the largest sample size in any AO related study to date. Most of the work in this chapter has recently been accepted for publication in *Investigative Ophthalmology & Visual Science*.

4.1.1 Fixation

In an AOSLO, a live video of the retina is recorded by a laser raster scan (Figure 4.1a), and the subject sees the raster scan pattern (Figure 4.1b). Therefore, an arbitrary stimulus pattern can be generated by turning the laser on and off between frames at appropriate moments of each raster scan[132]. In the Figure 4.1 example, this pattern was a blinking rectangle (≈ 0.1° wide) that served as a fixation stimulus. In measuring fixation events, an AOSLO has the advantage over other modalities, such as flood-illuminated cameras[2], for being able to isolate precise locations on the retina used for fixation. Any potential alignment or timing error is eliminated because the fixation target is generated as part of the image formation process[132]. In fact, the only additional tool required is a simple image segmentation algorithm to extract the position (average) of the black rectangle from a series of raw frames.

(a) (b)

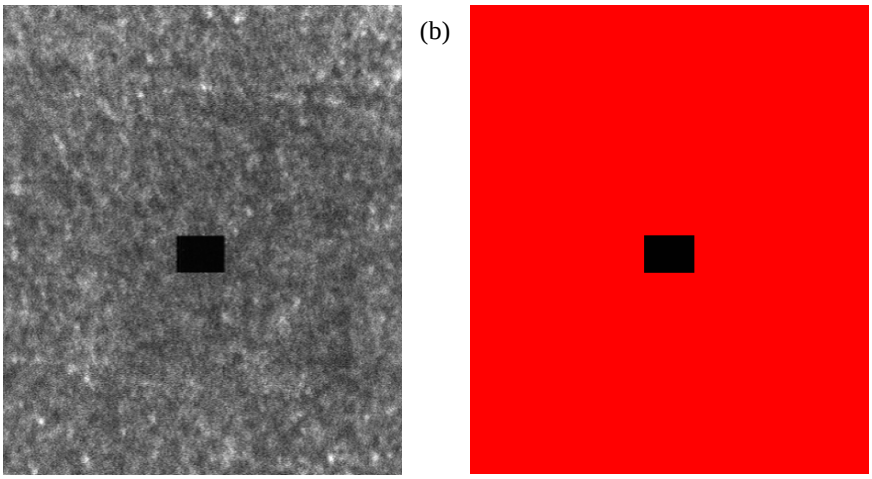


Figure 4.1: (a) Live image of the cone mosaic seen by the observer and (b) the corresponding laser scan pattern that the subject viewing through AOSLOII sees during an imaging session

4.2.2 Myopia

Myopia is most often due to elongation of the vitreous chamber rather than to changes in the cornea and lens. This is evidence by high correlations between axial length and refractive error reported in previous studies[133-141]. Reported complications that have been associated with myopia include lower best-corrected acuity[133, 134, 136, 142], reduced contrast sensitivity[143, 144], slower and/or inaccurate accommodation[137, 145, 146], and object aspect ratio misperception[147]. Retinal stretch may be a plausible explanation for some of these observations because a longer eye would require the same number of photoreceptors in the retina to tile over a larger surface area, an idea supported by cone density measurements from 1 to 2 mm eccentricity.[135] Although it may seem natural to extrapolate such findings into the foveal center, a study based on experimentally induced myopia in marmosets has shown that the opposite occurs as the longer, myopic eyes actually had significantly higher retinal cone density (cones/mm^2) than emmetropic eyes[148]. Therefore, the fact that the most dramatic changes in cone density distribution occur in the fovea begs for a more rigorous treatment of the matter.

In studies where eye length is an important variable, extra care needs to be taken when reporting cone density or spacing as a function of eccentricity because eye length directly affects the conversion between angular and retinal units. For example, a recent study reported a decrease in retinal cone density with increasing axial length at 2° eccentricity[139]. Whether or not their measurements supported the retinal stretch hypothesis, however, remains unknown because 2° eccentricity corresponded to retinal eccentricities of 0.56 and 0.72 mm away from the foveal center for the shortest and longest eyes included in that study. According to anatomic measurements provided by Curcio *et al.*[28], cone density was about 34,000 and 24,000 cones/mm^2 at 0.56 and 0.73 mm eccentricity respectively. In the extreme case where retinal cone density as a function of retinal eccentricity is preserved during eye growth, one would still expect to find a difference of about 10,000 cones/mm^2 due to how a particular angular eccentricity converts to different retinal eccentricities when eye lengths are not equal. For clarity, we

presented our results in both angular and retinal units and discussed the visual and anatomical implications associated with each approach.

4.1.3 Acuity

Since foveal cones are not easily accessible in the living human eye, investigations concerning cone spacing and myopia have generally been inferred from visual acuity measurements. A review of the main literature on this topic will reveal that myopes perform similarly or a little worse than emmetropes when refractive error is corrected near the pupil (i.e. contact lenses, refractive surgery, etc.), while they almost always perform worse with a spectacle correction due the associated minifying effects[133, 134, 136, 142, 149, 150]. Since myopes generally have higher retinal magnification due to their eye length, a likely explanation has been that myopes may have increased cone spacing due to retinal stretch. Alternatively, in studies where refractive error was corrected using spectacles and contact lenses, performance may have been compromised by optical factors such as scatter or high order aberrations[133, 136]. Optical complications are minimized when testing acuity with grating patterns generated with laser interference[134, 151]. Using this method, the level of myopia no longer seemed to affect acuity at the fovea. However, the higher retinal magnification factor (RMF) afforded by a longer eye would predict that myopes should actually perform better than emmetropes if their retinal cone densities were similar. Therefore, results from these two studies still support the idea that foveal cones become more widely spaced as myopia progresses[134, 151]. More recently in our lab, Rossi *et al.* used AOSLOI to test AO corrected visual acuity in emmetropes and low myopes (less than 4 diopters of myopia) and found that the low myope group performed significantly worse than the emmetrope group. However, there was still a distinct possibility that the AO correction was worse for the low myope group than for the emmetrope group because AOSLOI employs a 37 channel DM and ran a static correction at that time. Therefore, actual images of foveal cone mosaics for a group of individuals with different refractive errors (or axial lengths) are needed to help answer whether or not cones become physically more widely spaced as myopia progresses.

4.2 Methods

Prior to each experiment, informed consent was obtained from the subject after we explained the nature and complications of the study. Eighteen eyes from 18 healthy (not including refractive error) subjects, between the ages of 23 to 43 years, participated in the study (Table 4.1). A self-report questionnaire regarding basic health history was part of the subject recruitment process to ensure that only persons with no signs of ocular health problems were included in the study. The distinction between axial and refractive myopia was not made. Subjects who have smaller natural pupil sizes (< 6 mm diameter) were administered with 2.5 percent phenylephrine and 1 percent tropicamide prior to AOSLO imaging. Retinas were imaged from approximately the preferred retinal locus (PRL) to just beyond 1° eccentricity. Retinal imaging procedure and computations required to obtain the measurements used in this study are explained below.

4.2.1 Retinal imaging

All imaging was done using AOSLOII with 840 nm light. Most of the subjects were imaged before I upgraded to the current DM which is a BMC MEMS device with a 5.5

μm advertised stroke. The previous DM was also a BMC MEMS device but the slightly lower (3.5 μm) stroke model. I installed the high step resolution (14-bit) driver for the DM much earlier in the study (May, 2009), so almost half of our subjects were imaged using this new driver. The current default controller is the statistically weighted reconstructor under integral control while the local waffle penalty reconstructor was used in the earliest experiments. Eye alignment to the optical path and head stabilization were achieved via a chin rest and temple supports mounted on a three-axis stage. The scan field was about 0.9° by 0.9°. Since the fixation target is part of the acquired image, fixation locations are recorded to simultaneously track each subject's fixation pattern[132]. The average fixation location on the retina was defined as the PRL. The acquired retinal videos were first manually reviewed frame-by-frame for poor quality frames that may be caused by several factors (i.e. inaccurate wavefront measurement and/or correction, blink, tear film breakup, etc.). These frames were deleted prior to image post-processing.

Table 4.1: Subjects

	Eye	Gender	Age (yrs)	Axial length (mm)	Spherical equivalent refraction (D)	RMF (μm/deg)
1	OD	F	31	22.86	0	272.20
2*	OD	M	29	22.87	0.5	270.68
3	OD	F	31	23.40	-1.5	278.75
4	OS	M	42	23.50	0	280.47
5	OD	M	30	23.51	0	281.47
6‡	OD	F	24	24.08	0	288.41
7*	OS	M	43	24.18	0	288.49
8	OD	F	38	24.48	0	298.59
9†	OS	M	31	24.49	-0.75	298.61
10	OD	F	23	24.54	-3.5	298.98
11	OS	M	36	25.00	-2.5	305.06
12	OD	M	43	25.37	-2.25	310.81
13‡	OD	F	23	25.61	-5.5	316.58
14‡	OS	F	23	25.73	-5.25	320.07
15‡	OD	F	25	26.85	-6.75	335.61
16	OD	M	24	27.05	-7	341.68
17†	OD	M	34	27.46	-4.5	348.84
18	OD	M	23	28.31	-11	362.32

* No cycloplegia administered

† Inaccurate anterior chamber depth (ACD) measurement. The ACD value from Gullstrand model eye (3.585 mm) was used

‡ Entire cone mosaic resolved

4.2.2 Image processing and analysis

Due to eye movements during recording and the use of a resonant scanner for the horizontal scanning mirror of AOSLOII, the raw retinal videos contain distortions that must be corrected. The sinusoidal and eye motion artifacts were corrected using custom image registration software developed by collaborators[152, 153]. I did not take any part in the development of algorithms for registering AOSLO images. The retinal features in the

frames of registered videos are almost perfectly aligned, so the co-registered frames can be stacked and averaged to produce a high signal-to-noise image. These steps were then repeated for each acquired video, and the resultant images were stitched together in Photoshop (Adobe Systems Inc., Mountain View, CA) to create a larger montage image of the foveal cone mosaic.

In making noninvasive measurements of the human cone mosaic, the task of labeling individual cones is unavoidable We implemented custom software for identifying individual cones in the AO retinal images in C++ with calls to several MATLAB functions via the MATLAB Runtime Compiler. The interface allows the user to manually select individual cones and/or specify a region of interest for automated identification[154, 155]. MATLAB code for the current automated cone identification algorithm is given in Appendix B. The automated algorithm that this code implements is from our paper[154] but has since been updated. A combination of both manual and automated methods was employed for analyzing the cone mosaic at and near the foveal center because the current version of the automated algorithm still does not perform adequately near the foveal center where the cones are the smaller and consequently have lower contrast.

4.2.3 Retinal feature size estimation

The adjusted axial length method coined by Bennett *et al.*[156] is often used to estimate the RMF[142, 156, 157]. This method assumes that the retina and the eye's back focal plane coincide, which is not the case in myopia. Since visual angle is defined with respect to the nodal points of the eye, a more reasonable approach would be to locate the second nodal point and the resultant retinal image size subtended by the nodal ray (Figure 4.2). We specified a custom four surface model eye for each subject to carry out this calculation. The anterior radius of curvature of the cornea (r_1), anterior chamber depth (ACD) and axial length were measured using an IOLMaster (Carl Zeiss Meditec, Jena, Germany). For the cornea, we chose a fixed thickness and refractive index of 0.535 mm and 1.38 respectively[158, 159]. The corneal thickness was subtracted from the measured ACD to obtain the anatomical ACD (distance from the posterior cornea to the anterior lens). The posterior radius of curvature of the cornea (r_2) was taken to be $0.8831r_1$[160]. The Gullstrand schematic eye was used to approximate lens thickness and refractive indices of the aqueous, lens and vitreous.

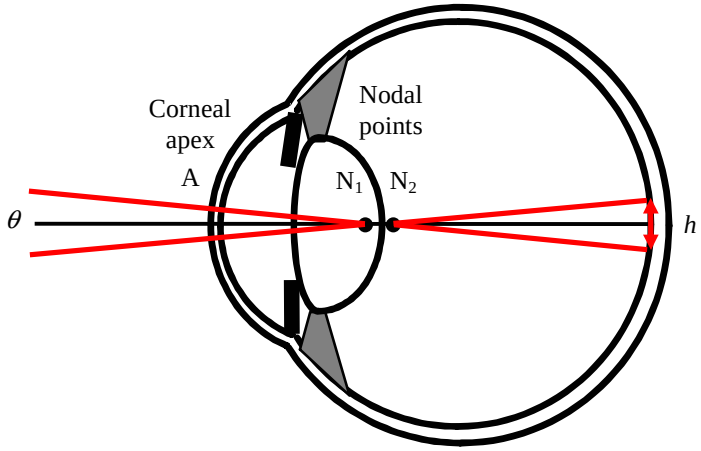


Figure 4.2: Schematic eye illustrating the relationship between visual angle, θ, and the retinal image size h

The location of the secondary nodal point was estimated for each eye via a paraxial ray trace[161]. Once determined, retinal image size is related to visual angle by the equation:

$$\begin{aligned} h &= 2\tan(0.5°)\left(x - \overline{AN_2}\right)\theta \\ &\approx 0.0175\left(x - \overline{AN_2}\right)\theta \end{aligned} \tag{4.1}$$

where h is retinal image size, x is axial length, $\overline{AN_2}$ is the distance from the apex of the cornea to the eye's second nodal point and θ is visual angle. The MATLAB code for calculating the RMF this way is given in Appendix B. Another magnification factor must be applied (multiplied) to Equation 4.1 when wavefront correction is aided with trial lenses. For example, a negative powered lens placed in front of the eye decreases the AOSLO's scan angle so the resultant retinal image size will be smaller. This magnification factor is given by the thin lens formula:

$$M = \frac{1}{1 - P(d + y)} \tag{4.2}$$

where P is the power of the trial lens, d is the spectacle vertex distance and y is the distance from the corneal apex to the entrance pupil. A fixed value of 14 mm was used for d for all subjects. The location of the entrance pupil was estimated from the ACD and the corneal radii of curvatures (r_1 and r_2).

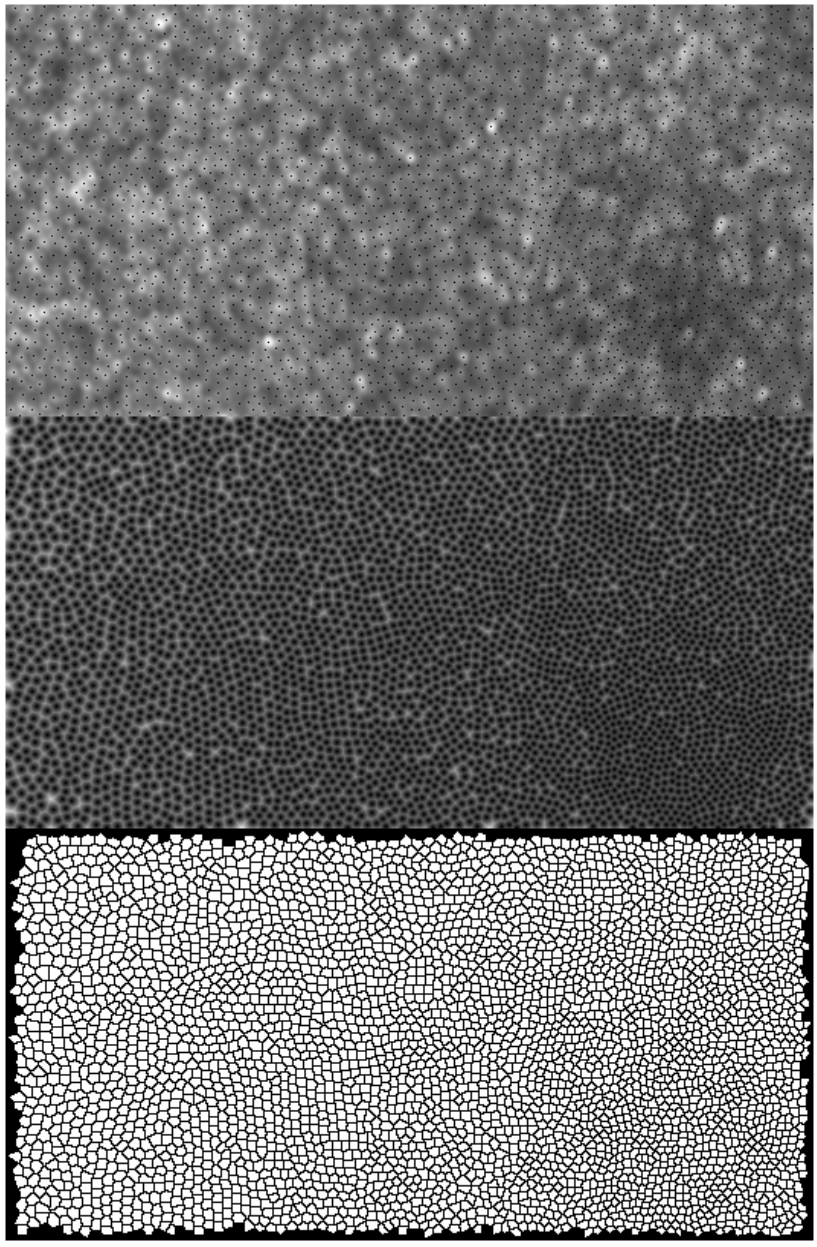


Figure 4.3: (a) A 0.5° × 1.0° (144 μm × 288 μm) section of a cone mosaic for subject 6 with identified cone locations. The foveal center is at the bottom right corner of the image. (b) Result after taking the distance transform of the (x, y) cone locations. (c) Voronoi tiles generated using the watershed transform.

4.2.4 Cone density estimation

Computations for estimating cone density were carried out using the MATLAB Image Processing Toolbox (IPT). The procedure used for estimating density from a list of *x*, *y* locations was adapted from several earlier studies where a fixed sampling window with an approximate area of 1300 μm^2 is scanned across the image to compute the mean density at each sampled location.[2, 28, 154] Using this window size, the peak cone density averaged across the reported normal eye data is about 201,000 cones/mm^2.[2, 28] Due to

relatively higher cone density gradients near the foveal center, a fixed window size will result in erroneous density estimates due to averaging over a variable number of cones at different locations. We addressed this issue by adopting an adaptive sampling window where the window size is adjusted to contain a constant number of cones instead. The constant was set to 150, which is approximately the expected number of cones in a 1300 μm^2 sampling window based on published peak density values[2, 3, 29].

Cone density is often calculated by dividing the number of cones recorded in the sampling window area. This method assumes a uniform density distribution within the sampling window and will underestimate density if the sampling window overlaps regions of missing data (i.e. image border). These limitations have motivated the adoption of what is known as Voronoi local density analysis[162-164]. Figure 4.3 illustrates several steps in this computation procedure. After identifying individual cones in an image (Fig. 4.3a), the resultant coordinates are used to construct a binary mosaic that is all zeros except at pixels corresponding to the cone centers. The nearest-neighbor distance calculations are applied to the resultant binary mosaic using the IPT function *bwdist.m*. Voronoi tiles (Fig. 4.3c) are generated from the resultant "distance" image (Fig. 4.3b) using the watershed transform (*watershed.m*), and tiles containing pixels on the image border are removed. Finally, a raw density value is calculated at each cone location by inverting the area of the corresponding Voronoi tile:

$$d(x,y) = \frac{1}{A(T(x,y)) + 0.5P(T(x,y))} \tag{4.3}$$

where $A(T(x,y))$ and $P(T(x,y))$ are the area and perimeter of the Voronoi tile $T(x,y)$. Inclusion of the perimeter adjusts for the single pixel wide boundary that separates all adjacent Voronoi tiles. Raw density values of $d(x,y)$ were first scaled by the calibrated imaging field size to obtain angular density (cones/deg^2) and then once more with the appropriate RMF to obtain the retinal density (cones/mm^2). The cone density value reported at each particular cone location is taken as the mean of the local density estimates within its associated sampling window.

We estimated the center of the anatomical fovea with the location of peak cone photoreceptor density. The two-dimensional sequence $d(x,y)$ was linearly interpolated to generate a cone density topography map for each eye. The location of peak cone density was determined from each topography map by a center of mass calculation: average of the centroid locations for regions enclosed by the first six iso-density contour lines (contour levels are separated by 5,000 cones/mm^2)[2, 3]. This method was adopted because it provides a systematic approach for estimating the peak density location when it is obviously located within the region where the cones were not resolved. This method was also applied to images without such a region because errors due to cone mislabeling and digital artifacts can produce spike-like protrusions with artificially high values anywhere across the topography map. Representative cone density measurements at particular eccentricities were computed by circular averaging of density estimates around all meridians. Center-to-center cone spacing $s(x,y)$ was calculated from density by assuming that cones are arranged in a perfect hexagonal lattice leading to an exact relationship between cone density and spacing:

$$s(x,y) = \left(\frac{2}{\sqrt{3} d(x,y)} \right)^{1/2} \tag{4.4}$$

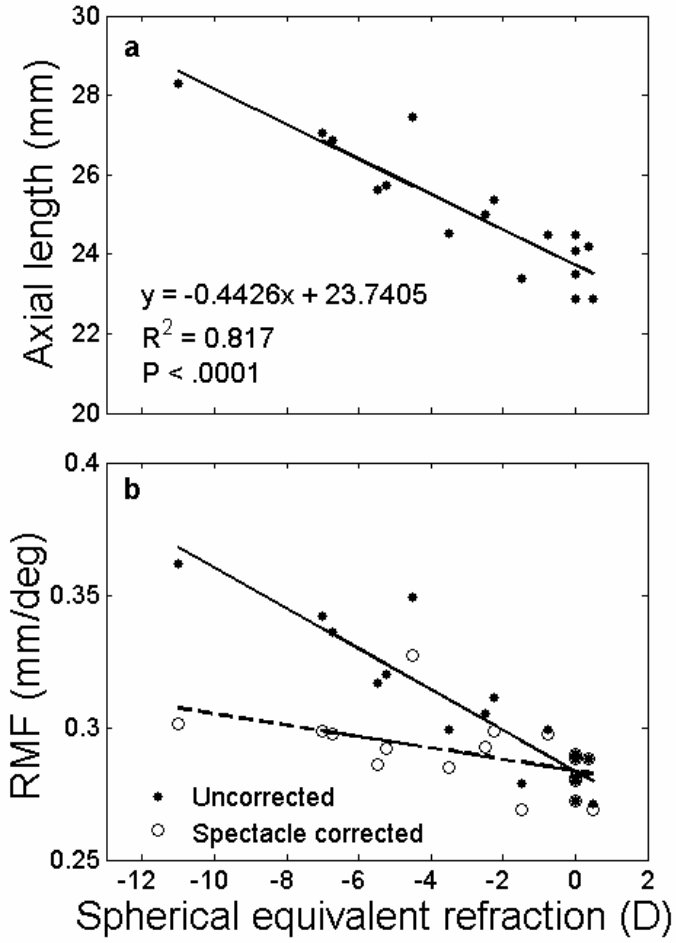


Figure 4.4: (a) Axial length plotted as a function of the spherical equivalent spectacle refraction with the solid line being a linear regression to the data. (b) Calculated retinal magnification factor (RMF) plotted as a function of the spherical equivalent spectacle refraction. Equation 4.1 was used directly to compute RMF for the uncorrected case (filled circles), while the spectacle corrected RMF values (open circles) were obtained by multiplying the uncorrected RMF values by the corresponding spectacle magnifications calculated using Equation 4.2. Lines are linear regressions to the data, and the decrease in RMF with less refractive error was significant ($P < .05$) for both the corrected and uncorrected cases.

4.3 Results

Axial length is plotted against the spherical equivalent refraction in Figure 4.4a confirming once more that refractive error is primarily due to changes in the length of the eye rather than the optics. Therefore, individuals with different eye lengths will have different size retinal images even when viewing the same scene at the same distance away. Quantitatively speaking, this is just due to differences in RMF (Equation 4.1) which is also plotted against the spherical equivalent refraction (Figure 4.4b). The trend observed for both the corrected and uncorrected RMF in relation refractive error are in

close agreement with those reported in Coletta and Watson[134]. Stable AO performance, even in the cases without cycloplegia, was achieved without build-up of unobservable deformable mirror modes (i.e. local waffling[82]). Referring to what I discussed in Chapter 3, a low computed wavefront error did not always correspond to similar retinal image quality. Even though AO performance appeared to be excellent most of the time, we only successfully resolved the entire foveal cone mosaic for four eyes. Figure 4.5 is an example such an image covering about 2° across the foveal center. For the other subjects, the hexagonal packing structure of the cone mosaic gradually faded toward the foveal center forming a region with an undesired speckle-like appearance which we do not believe to correspond to individual cone photoreceptors. This region extended to at most 0.03 mm eccentricity for the majority of the eyes imaged but went out as far as 0.10 mm eccentricity for one individual.

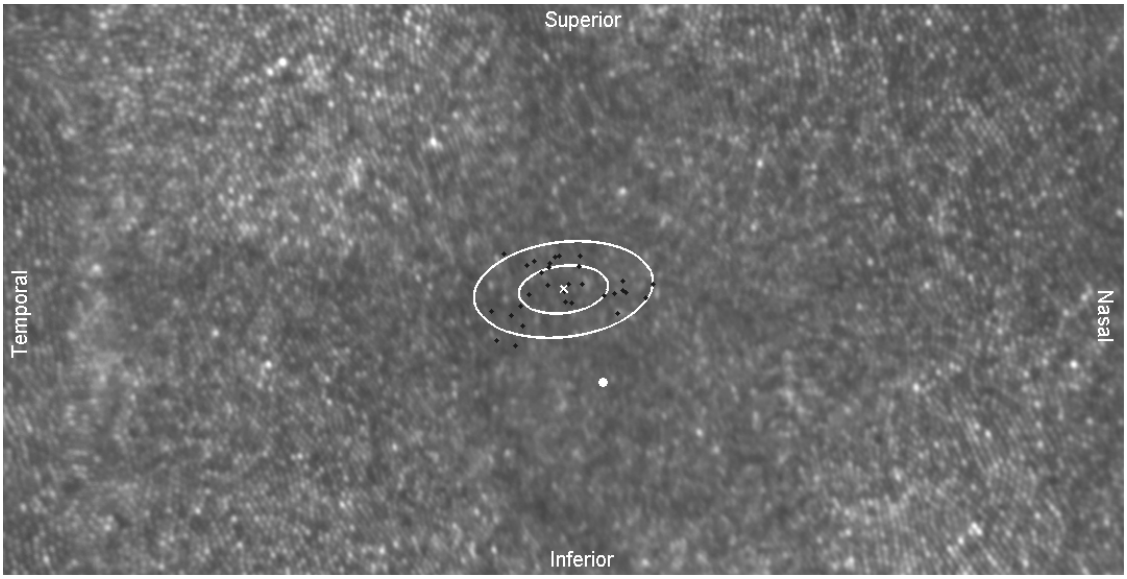


Figure 4.5: 1° by 2° (320 μm by 640 μm) foveal cone mosaic centered about the PRL (white x) for subject 13 (see Table 4.1). Fixation locations are represented by the small black dots, and the white ellipses correspond to one and two standard deviations of the fixation points. The PRL is displaced about 9.5 arcmin (50 μm) from the foveal center (white spot).

4.3.1 Foveal fixation and cone density topography

Foveal fixation events were analyzed in much of the same way as described by Putnam *et al.*[2] with the exception that the principal components of each set of fixation points were first computed to estimate the orientation of the distribution. In some cases, the principal components (semi-major and semi-minor axes) are nearly parallel to the coordinate axes (Figure 4.5), but a distribution such as the one shown in Figure 4.6 is more accurately described by a Gaussian function that is rotated by 143° from the horizontal axis. The standard deviation of fixation along the semi-major axis varied from 1.75 to 5.42 arcmin (7.89 to 29.81 μm) with the mean at 3.61 arcmin (18.74 μm). Fixation along the semi-minor axis had a significantly lower standard deviation ($P < .01$) ranging only from 1.19 to 3.88 arcmin (5.72 to 20.47 μm) indicating that angular distribution of foveal fixation is generally not uniform. Both sets of standard deviations are listed in Table 4.2. Fixation

points along both axes were verified to be normally distributed (Kolmogorov-Smirnov test, $P > .05$) for nearly all recordings; the exceptions being subjects 2 and 13 along the semi-minor axes and subject 1 and 16 along the semi-major axes. Interestingly, according to Table 4.2, the PRL was located superior to the point of peak cone density in all but two eyes (subjects 2 and 7).

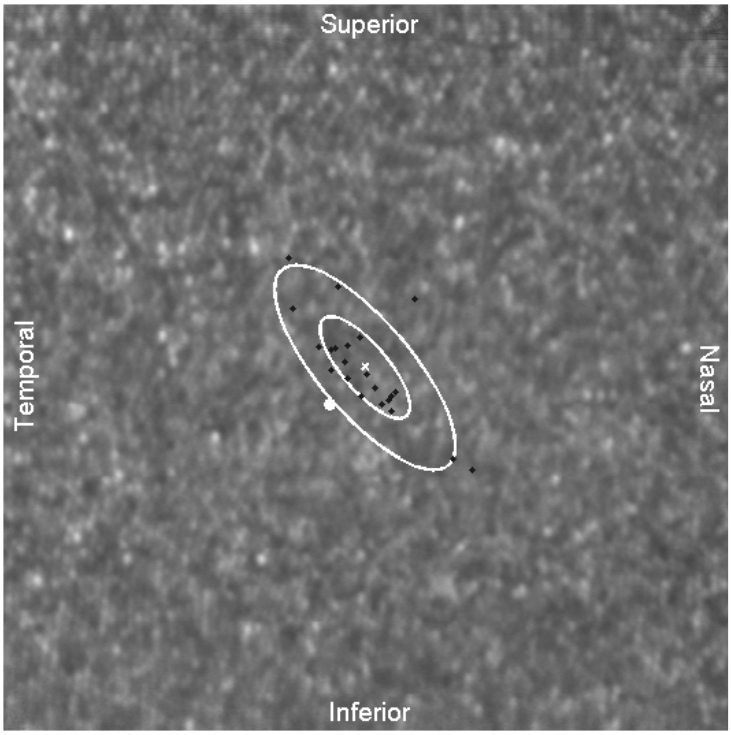


Figure 4.6: 1° by 1° (336 μm by 336 μm) cone mosaic centered about the PRL (white x). Fixation locations are represented by the small black dots, and the white dot marks the location of the anatomical foveal center. The semi-major axis angle for the distribution of fixations is about 143°. The PRL is displaced about 3.9 arcmin (22 μm) from the foveal center.

Figure 4.7 displays several topography maps with iso-density contour lines. PRLs and foveal centers are denoted by the white dots and x's respectively. Zero contour levels are regions where cones were not reliably identified or areas beyond the image support. Displacements between PRLs and foveal centers ranged from 2.98 μm (0.58 arcmin) to 92.29 μm (18.55 arcmin) with an average displacement of about 34 μm (5.62 arcmin) which is on average lower than the estimates given in Putnam *et al* [2] measured with a flood-illuminated AO ophthalmoscope. Table 4.2 lists the measured displacements for all subjects. A series of location tests on the PRLs, based on our estimates of the fixation variances described above, revealed that the PRL deviates significantly from the foveal center for all but one individual (*T*-test, two-tailed, $P < .001$). This can be qualitatively appreciated in Figure 5 by observing that the PRL generally deviates substantially from the center of mass of the corresponding contour map.

Table 4.2: Peak cone density and foveal fixation (T – temporal, S – superior, N – nasal and I – inferior). Subjects are listed in the same order as in Table 1. Standard deviations are from Gaussian fits to each set of fixation locations.

	Peak density (cones/mm^2)	Semi-major axis std. dev. [arcmin (μm)]	Semi-minor axis std dev. [arcmin (μm)]	Deviation of PRL from the point of peak cone density [arcmin (μm)]	
1	-	2.31 (10.5)	1.20 (5.4)	2.97 (13.5) N	6.94 (31.5) S
2	-	1.75 (7.9)	1.64 (7.4)	2.43 (11.0) T	1.78 (8.0) I
3	-	3.74 (17.4)	2.72 (12.6)	2.00 (9.3) N	7.51 (34.9) S
4	-	2.11 (9.9)	1.28 (6.0)	5.81 (27.2) N	6.08 (28.4 S
5	-	4.59 (21.5)	2.55 (12.0)	0.79 (3.7) T	5.62 (26.4) S
6	167,730	2.91 (14.0)	1.54 (7.4)	2.50 (12.0) T	1.62 (7.8) S
7	-	3.36 (16.2)	1.19 (5.7)	1.93 (9.3) N	7.96 (38.3) I
8	-	4.66 (23.2)	1.75 (8.7)	10.54 (52.5)T	15.26 (75.9)S
9	-	4.00 (19.9)	2.25 (11.2)	1.40 (7.0) N	6.51 (32.4) S
10	-	3.10 (15.5)	2.79 (13.9)	2.89 (14.4) T	1.98 (9.9) S
11	-	3.00 (15.3)	1.32 (6.7)	2.07 (10.5) N	0.94 (4.8) S
12*	-	3.42 (17.8)	1.63 (8.4)	0.54 (2.8) N	0.21 (1.1) S
13	116,217	5.42 (28.6)	3.88 (20.5)	2.73 (14.5) N	14.11 (74.5) S
14	167,984	4.83 (25.8)	2.51 (13.4)	3.16 (16.9) N	8.92 (47.6) S
15	149,719	5.33 (29.8)	1.95 (10.9)	2.71 (15.2) T	2.76 (15.4) S
16	-	2.85 (16.2)	2.65 (15.1)	5.07 (28.9) T	1.78 (10.1) S
17	-	4.35 (25.3)	2.38 (13.8)	0.90 (5.2) T	1.14 (6.6) S
18	-	3.19 (19.3)	2.23 (13.5)	1.62 (9.8) T	9.37 (56.6) S

* PRL was not significantly different from the estimated foveal center

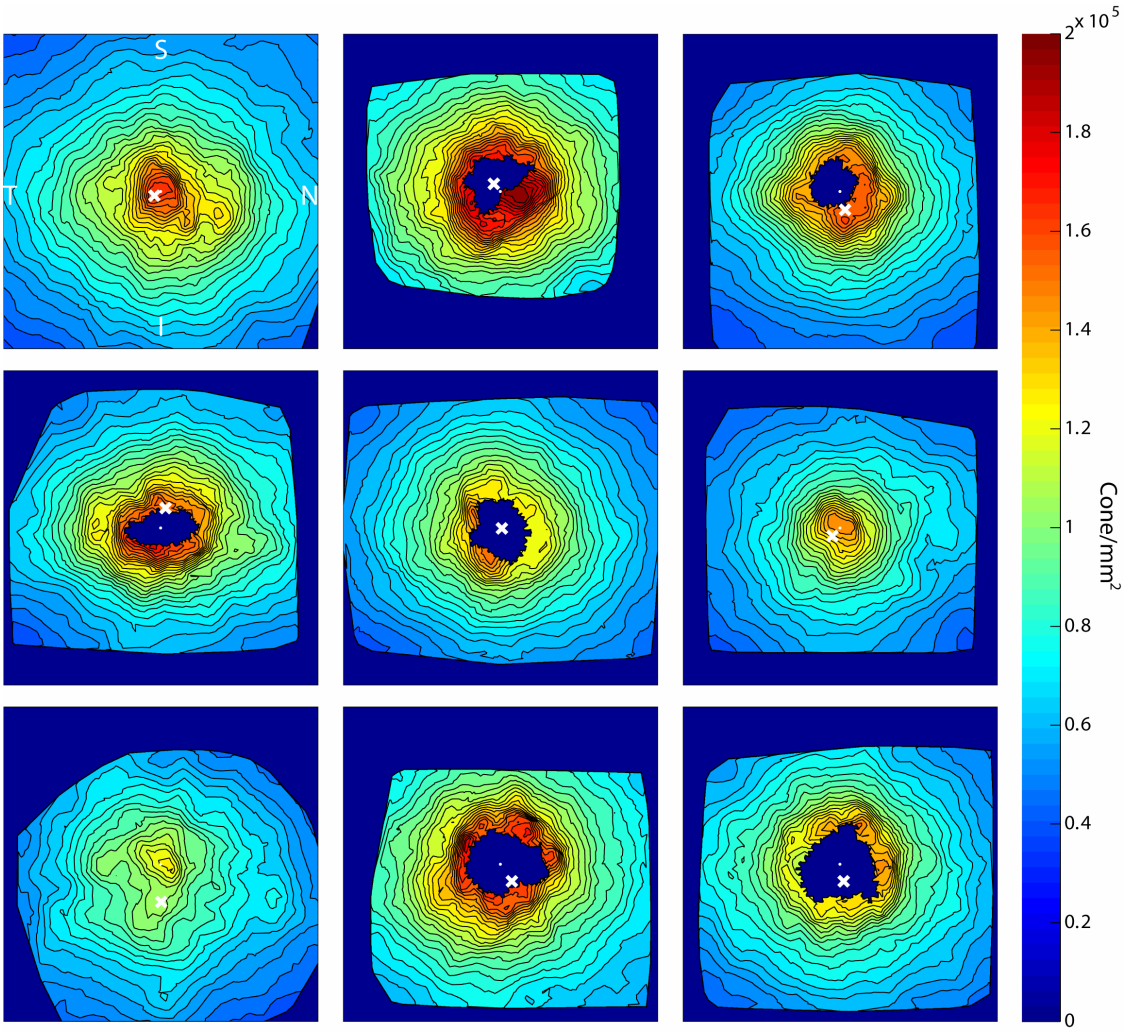


Figure 4.7: Examples of cone density topography maps. All maps are oriented as indicated on the top-left panel (T – temporal, S – superior, N – nasal and I – inferior). Locations of the foveal center and the PRL are indicated by the white dots and x's respectivity. The size of each map is 0.6 by 0.6 mm and consecutive contour lines are separated by 5,000 cones/mm^2. The dark blue areas include both the central foveal region in some eyes where cones could not be resolved and regions outside of the support of the acquired retinal images.

4.3.2 Inter-subject cone density variability

Retinal cone density is plotted against retinal eccentricity in Figure 4.8 with zero eccentricity defined at the foveal center. In the four eyes where all the foveal cones were resolved, the peak retinal cone density ranged from 123,842 to 167,730 cones/mm^2 (Table 4.2) which all fall within reported values despite subtle differences in the calculation procedure used in different studies[2, 3, 28]. The corresponding minimum center-to-center cone spacing estimates are 2.62, 3.05, 2.77 and 2.79 μm. Individual cones were resolved in most eyes beginning at about 0.03 mm eccentricity. On average, retinal cone density decreased from 151,008 to 57,312 cones/mm^2 from 0.03 to 0.30 mm eccentricity.

Variability across subjects was highest at 0.03 mm eccentricity and converged to a similar range of values beyond 0.2 mm eccentricity.

In Figure 4.9, cone density is plotted against axial length at three different retinal eccentricities. At 0.10 mm eccentricity, retinal cone density appears to decrease with increasing axial length but the effect was not statistically significant ($P > .05$) . A root-mean-squared error (RMSE) of 9,114 cones/mm^2 is a clear indication that axial length does not accurately describe retinal cone density differences near the foveal center. By 0.30 mm eccentricity however, retinal cone density decreased significantly with increasing axial length ($P < .05$), and the RMSE was reduced nearly threefold to only 4,406 cones/mm^2. Since RMF is higher in longer eyes, angular cone density actually increased significantly with axial length at all three retinal eccentricities despite the RMSE of the fit being rather high toward the foveal center. As a result, the visual angle subtended by an object along the line of sight will generally be sampled by more cones in a longer eye despite evidence of myopia induced retinal stretch as close as 0.3 mm from the foveal center.

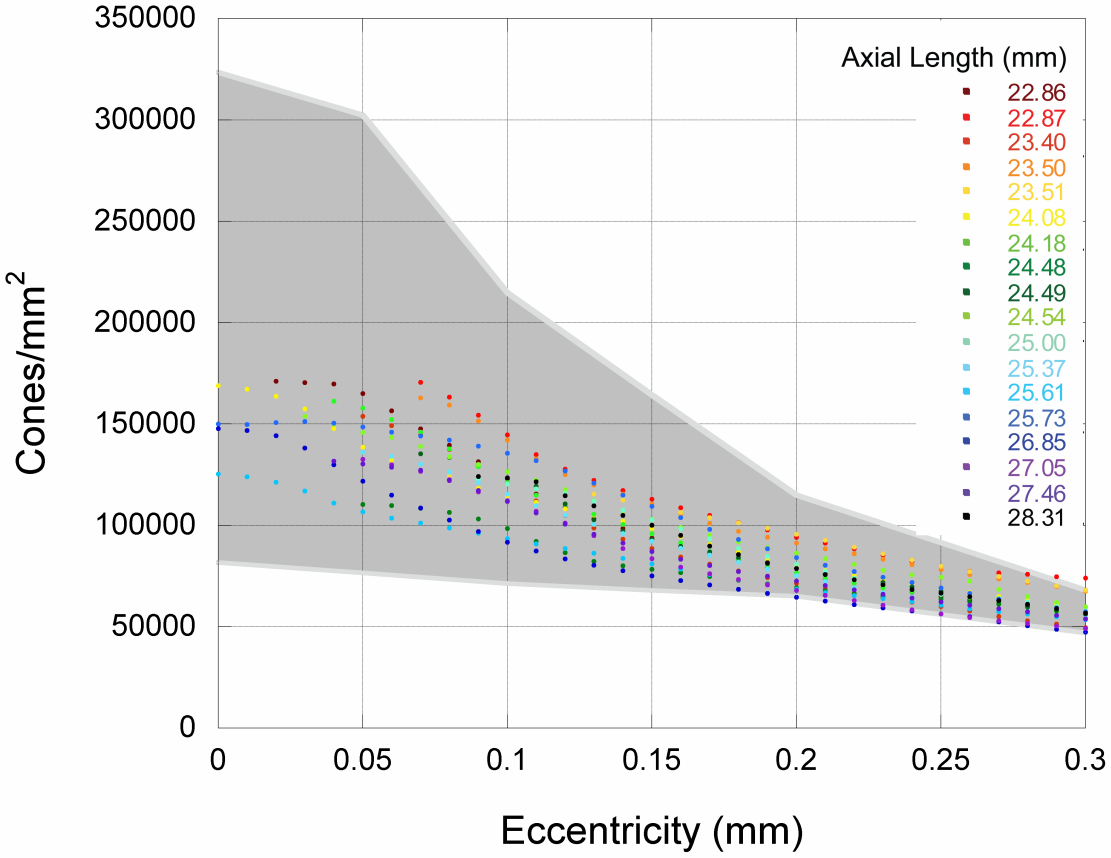


Figure 4.8: Retinal cone density as a function of retinal eccentricity. Representative cone density measurements at particular eccentricities were computed by circular averaging of density estimates around all meridians. The shaded region corresponds to the range of foveal cone density values report by Curcio *et al*.[28]

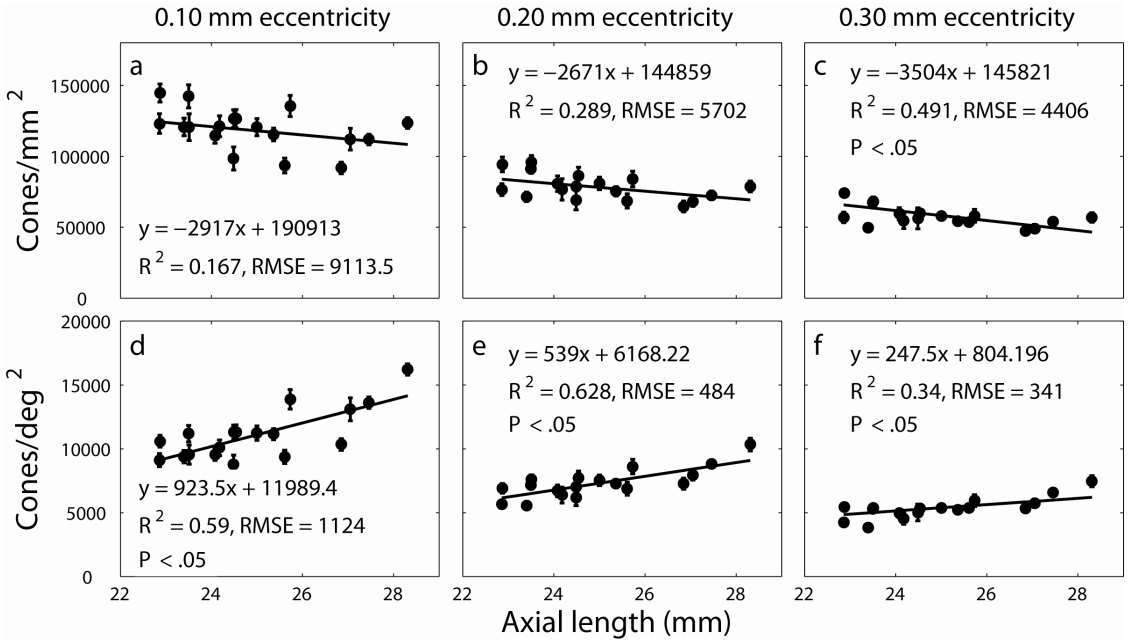


Figure 4.9: Retinal cone density (a, b & c) and angular cone density (d, e & f) as a function of axial length at three different retinal eccentricities. Error bars represent one standard deviation in the spread of cone density values at the specified eccentricities. Lines are weighted least squares linear regressions to the data.

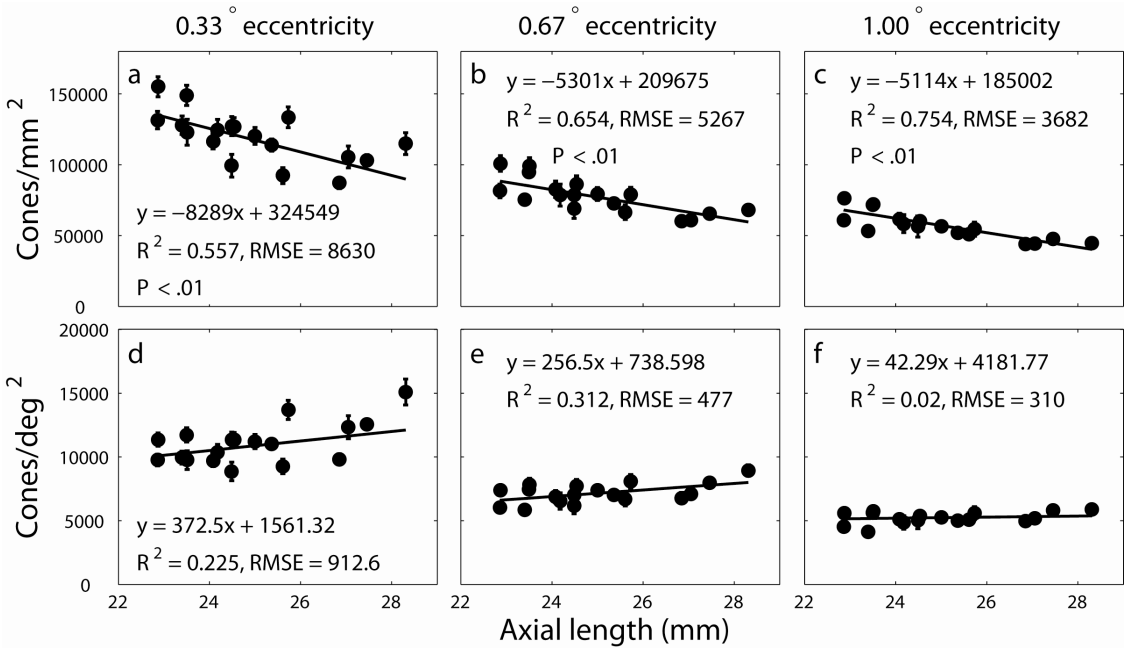


Figure 4.10: Retinal cone density (a, b & c) and angular cone density (d, e & f) as a function of axial length at three different angular eccentricities. Error bars represent one standard deviation in the spread of cone density values at the specified eccentricities. Lines are weighted least squares linear regressions to the data.

4.3.3 Foveal cone density and visual acuity

For the purpose of facilitating a comparison between our anatomical measurements and visual acuity, we also analyzed the cone density data at specific angular eccentricities. Figure 4.10 shows that retinal cone density decreased with increasing axial length at a higher rate than when eccentricity was specified in mm. However, retinal and angular eccentricities away from the foveal center have separate meanings when different eye sizes are involved. For example, 1° eccentricity converts to a distance of 0.27 mm away from the foveal center for our shortest eye and more than 0.36 mm for our longest eye. According to our measurements, we would expect a cone density drop of 14,785 cones/mm^2 due to retinal eccentricity differences alone which explains about 52 percent of the estimated decrease in Figure 4.10c. The remaining 48 percent is presumably due to retinal stretch. An increase in axial length did not seem to have any effect on angular cone density ($P > .05$). In fact, the regression line was nearly flat by 1° eccentricity assuring that the visual angle subtended by an object arriving slightly off axis will be sampled by similar numbers of cones independent of axial length. Since the PRL can deviate substantially from the foveal center, we were actually able to determine cone density at the PRL for 10 eyes despite some of the images containing a small region of cones that were not resolved. Figure 4.11 plots the retinal cone density against axial length at the PRL. Retinal cone density appears to decrease with increasing axial length primarily due to a fairly long eye in our study that had particularly low cone density values at and near the foveal center, but the standard error was very high (RMSE = 22,154 cones/mm^2), so this effect was statistically insignificant ($P > .05$).

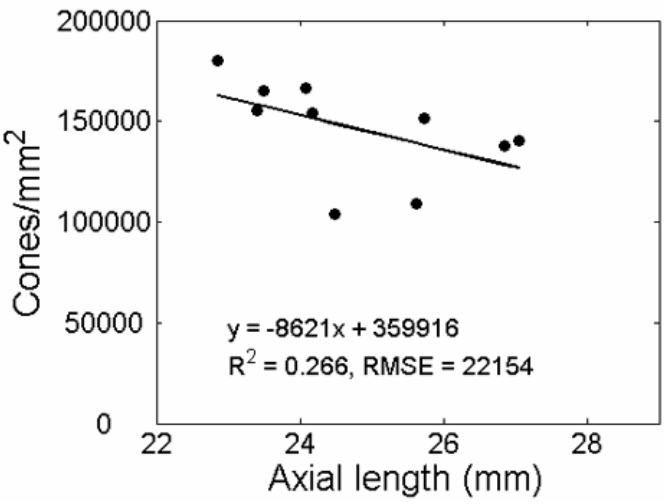


Figure 4.11: Retinal cone density as a function of axial length at the PRL. Solid line is linear regression to the data. Regression slope is insignificant ($P > .05$).

4.4 Discussion

One of the main purposes of this work was to provide baseline cone density measurements from 0 to 0.3 mm eccentricity, a region of the retina that is most important for spatial vision, but that has been rarely explored using *in vivo* imaging methods. With improved AO performance, most if not all cone photoreceptors in this central part of the

fovea can be resolved. The cone density curves plotted in Figure 4.8 are in close agreement with 6 of the 8 retinas presented in Curcio *et al.*[28] It was mentioned as a possibility by the investigators that the two retinas with much higher foveal cone density than the rest may have been due to tissue shrinkage. It is nonetheless encouraging to find that *in vivo* density measurements are in close agreement with histological data.

Recent work by Chui *et al.*[135, 165] stated that diffraction may be the limiting factor for the AOSLO for resolving cones near the foveal center. Their resolution assessment was based on the Rayleigh criterion for a 6 mm diameter pupil (2.8 μm for an emmetropic eye). For the subjects imaged in this study, the Rayleigh criterion would predict the resolution limit to be from 2.65 to 3.55 μm depending on eye length. With the exception of the two high density foveas reported by Curcio *et al.*[28], the smallest foveal cones are at least 2 μm in diameter which is approximately equal to cone spacing in the rod-free fovea[27-30]. But cone size increases rapidly with eccentricity, so that by 150 μm from the foveal center, the average cone spacing is greater than 3.5 μm according to measurements plotted in Figure 4.8. Since Chui *et al.*[135, 165] were only able to resolve individual cones at retinal eccentricities greater than 200 μm in emmetropes, it is unlikely that diffraction was the limiting factor in their measurements. Fundamentally, the more accurate description of resolution limit is known as the Sparrow criterion which predicts a resolving power that is about 22 percent higher than the Rayleigh criterion for a circular aperture[166-168]. In our study, the predicted lateral resolutions for the four eyes were 2.41 μm, 2.43 μm and 2.55 μm indicating that we were imaging close to the diffraction limit. Assuming that size is the only factor that makes foveal cones difficult to image, a diffraction limited AOSLO should be able to resolve the entire foveal cone mosaic for most eyes. Since this was not the case in the current study, a more robust AO system is required to consistently achieve near diffraction-limited image quality.

4.4.1 Peak cone density and fixation

As mentioned already, the AOSLO has the advantage over flood illuminated AO retinal imager for being able to isolate precise locations on the retina used for fixation. Nevertheless, our data are consistent with, albeit slightly lower than Putnam *et al.*[2], who found that the PRL is displaced from the point of peak cone density. Our measurements serve to confirm that the PRL deviates significantly from the foveal center and reinforces the importance of clearly defining the location of zero eccentricity whenever one is performing eccentricity-dependent measurements. Furthermore, when the angular distribution of foveal fixation is not approximately uniform, the horizontal and vertical standard deviation or the mean would not accurately describe fixation variability. Principal component analysis determines the orientation that accounts for the most variability in the data and thus provides a better overall metric for describing fixation variability.

4.4.2 Cone density and axial length

Inside the approximate foveola (0 to 0.2 mm eccentricity[26, 28]), axial length induced retinal stretch could not be verified by cone density measurements alone due to high levels of inter-subject variability. Although we were able to measure cone density only as close as 0.1 mm eccentricity if all subjects were to be included, we would expect inter-subject variability to be even greater at the foveal center based on histological data[29].

However with increasing retinal eccentricity, the tendency for all cone mosaics to converge to a state that can be characterized by axial length becomes more apparent as observed at 0.3 mm eccentricity (Figure 4.9c). Interestingly, despite the amount of inter-subject variability near the foveal center, angular cone density actually increased significantly with axial length at all retinal eccentricities (Figures 4.9d, 4.9e and 4.9f). In the interferometric acuity study by Coletta and Watson[134], the investigators generated a 1° diameter circular grating patch to measure foveal acuity in a group of subjects with various axial lengths. According to their results, all subjects performed similarly when acuity limits were specified in angular units of spatial frequency (cycles/deg). But based on their RMF estimates, the spatial frequency of the grating in retinal units (cycles/mm) for their longest eye was only half the rate as that for their shortest eye. Based on our measurements, if interferometric acuity at the fovea is indeed limited by cone spacing, then one would expect individuals with longer eyes to perform better than those with shorter eyes in terms of acuity in angular units (cycles/deg) and perform similarly in terms of acuity in retinal units (cycles/mm). This was not the case according to two separate studies[134, 151], so we can rule out retinal stretch as a possible explanation for why foveal interferometric acuity does not improve with increasing level of axial myopia.

One rather extreme interpretation of our results is that the density of the foveolar cone mosaic is completely unaffected by myopia related eye growth. This seems unlikely because the retinal surface expands globally in myopia[169], and we have little reason to believe that retinal tissue at the foveola is somehow more durable than the rest of the retina. A more reasonable interpretation would be that retinal stretch affects the foveolar cone mosaic, but a number of other developmental factors primarily govern cone density distribution there. A thorough analysis of foveal cone density and packing structure in emmetropic retinas, in tandem with other structural measures (i.e. retinal thickness, size of the foveal avascular zone and the shape of the foveal depression[170]), will be necessary to identify these potential factors. Nevertheless, because we were only able to estimate the peak cone density for four eyes, we still cannot rule out the possibility that peak cone density increases with eye growth as seen in experimentally enlarged marmoset eyes[148].

4.5 Conclusions

The lateral resolution achieved with AOSLO is sufficient for resolving the smallest cones in the foveola in some eyes and most of the foveal cones in all normal eyes. As a result, we are able to perform some of the first analyses on images of foveal cone mosaics acquired from the living human retina. AOSLO measurements of foveal fixation verified that the PRL deviates significantly from the point of peak cone density in normal eyes. Based on cone density distribution alone, myopia induced retinal stretch exists in the fovea, but near the foveal center (< 0.3 mm) these dependencies are swamped by other sources of inter-subject variability. As a result, relationships between cone density and axial length found outside this region cannot be extrapolated to infer trends at the foveal center or along the line of sight.

Acknowledgements

The authors thank Professor Christine Wildsoet for insightful conversations regarding myopia and Professor Scott Stevenson for providing us with excellent image stabilization software. I acknowledge Professor Yuhua Zhang for constructing the AOSLOII and

Michael Liu for assistance with image processing. I would like to thank Professor Martin S. Banks for seriously proofreading this chapter and making many criticisms that slipped past me, Austin and the reviewers for the *IOVS* article in press.

Bibliography

1. Liang JZ, Williams DR, Miller DT. High resolution imaging of the living human retina with adaptive optics. *Investigative Ophthalmology & Visual Science* 1997;38:55-55.
2. Putnam NM, Hofer HJ, Doble N, Chen L, Carroll J, Williams DR. The locus of fixation and the foveal cone mosaic. *Journal of Vision* 2005;5:632-639.
3. Carroll J, Neitz M, Hofer H, Neitz J, Williams DR. Functional photoreceptor loss revealed with adaptive optics: An alternate cause of color blindness. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101:8461-8466.
4. Chen DC, Jones SM, Silva DA, Olivier SS. High-resolution adaptive optics scanning laser ophthalmoscope with dual deformable mirrors. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2007;24:1305-1312.
5. Burns SA, Tumbar R, Elsner AE, Ferguson D, Hammer DX. Large-field-of-view, modular, stabilized, adaptive-optics-based scanning laser ophthalmoscope. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2007;24:1313-1326.
6. Zhang YH, Poonja S, Roorda A. MEMS-based adaptive optics scanning laser ophthalmoscopy. *Optics Letters* 2006;31:1268-1270.
7. Roorda A, Romero-Borja F, Donnelly WJ, Queener H, Hebert TJ, Campbell MCW. Adaptive optics scanning laser ophthalmoscopy. *Optics Express* 2002;10:405-412.
8. Zhang Y, Rha JT, Jonnal RS, Miller DT. Adaptive optics parallel spectral domain optical coherence tomography for imaging the living retina. *Optics Express* 2005;13:4792-4811.
9. Hermann B, Fernandez EJ, Unterhuber A, et al. Adaptive-optics ultrahigh-resolution optical coherence tomography. *Optics Letters* 2004;29:2142-2144.
10. Merino D, Dainty C, Bradu A, Podoleanu AG. Adaptive optics enhanced simultaneous en-face optical coherence tomography and scanning laser ophthalmoscopy. *Optics Express* 2006;14:3345-3353.
11. Torti C, Povazay B, Hofer B, et al. Adaptive optics optical coherence tomography at 120,000 depth scans/s for non-invasive cellular phenotyping of the living human retina. *Optics Express* 2009;17:19382-19400.
12. Zawadzki RJ, Jones SM, Olivier SS, et al. Adaptive-optics optical coherence tomography for high-resolution and high-speed 3D retinal in vivo imaging. *Optics Express* 2005;13:8532-8546.
13. Rha J, Jonnal RS, Thorn KE, Qu JL, Zhang Y, Miller DT. Adaptive optics flood-illumination camera for high speed retinal imaging. *Optics Express* 2006;14:4552-4569.
14. Thaung J, Knutsson P, Popovic Z, Owner-Petersen M. Dual-conjugate adaptive optics for wide-field high-resolution retinal imaging. *Optics Express* 2009;17:4454-4467.
15. Doble N, Miller DT, Yoon G, Williams DR. Requirements for discrete actuator and segmented wavefront correctors for aberration compensation in two large populations of human eyes. *Applied Optics* 2007;46:4501-4514.
16. Diaz-Santana L, Torti C, Munro I, Gasson P, Dainty C. Benefit of higher closed-loop bandwidths in ocular adaptive optics. *Optics Express* 2003;11:2597-2605.

17. Hofer H, Chen L, Yoon GY, Singer B, Yamauchi Y, Williams DR. Improvement in retinal image quality with dynamic correction of the eye's aberrations. *Optics Express* 2001;8:631-643.
18. Devaney N, Dalimier E, Farrell T, et al. Correction of ocular and atmospheric wavefronts: a comparison of the performance of various deformable mirrors. *Applied Optics* 2008;47:6550-6562.
19. Liang JZ, Grimm B, Goelz S, Bille JF. Objective Measurement of Wave Aberrations of the Human Eye with the Use of a Hartmann-Shack Wave-Front Sensor. *Journal of the Optical Society of America a-Optics Image Science and Vision* 1994;11:1949-1957.
20. Liang JZ, Williams DR. Aberrations and retinal image quality of the normal human eye. *Journal of the Optical Society of America a-Optics Image Science and Vision* 1997;14:2873-2883.
21. Hammer DX, Ferguson RD, Bigelow CE, Iftimia NV, Ustun TE, Burns SA. Adaptive optics scanning laser ophthalmoscope for stabilized retinal imaging. *Optics Express* 2006;14:3354-3367.
22. Zhang YH, Roorda A. Evaluating the lateral resolution of the adaptive optics scanning laser ophthalmoscope. *Journal of Biomedical Optics* 2006;11:-.
23. Hofer H, Artal P, Singer B, Aragon JL, Williams DR. Dynamics of the eye's wave aberration. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2001;18:497-506.
24. Li KY, Yoon G. Changes in aberrations and retinal image quality due to tear film dynamics. *Optics Express* 2006;14:12552-12559.
25. Gruppetta S, Lacombe F, Puget P. Study of the dynamic aberrations of the human tear film. *Optics Express* 2005;13:7631-7636.
26. Polyak S. Retinal structure and colour vision. *Documenta Ophthalmologica* 1949;3:24-56.
27. Yuodelis C, Hendrickson A. A Qualitative and Quantitative-Analysis of the Human Fovea during Development. *Vision Research* 1986;26:847-855.
28. Curcio CA, Sloan KR, Kalina RE, Hendrickson AE. Human Photoreceptor Topography. *Journal of Comparative Neurology* 1990;292:497-523.
29. Curcio CA, Sloan KR. Packing Geometry of Human Cone Photoreceptors - Variation with Eccentricity and Evidence for Local Anisotropy. *Visual Neuroscience* 1992;9:169-180.
30. Obrien B. Vision and Resolution in the Central Retina. *Journal of the Optical Society of America* 1951;41:882-894.
31. Wojtas DH, Wu B, Ahnelt PK, Bones PJ, Millane RP. Automated analysis of differential interference contrast microscopy images of the foveal cone mosaic. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2008;25:1181-1189.
32. Bigelow CE, Iftimia NV, Ferguson RD, Ustun TE, Bloom B, Hammer DX. Compact multimodal adaptive-optics spectral-domain optical coherence tomography instrument for retinal imaging. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2007;24:1327-1336.
33. Tam J, Martin JA, Roorda A. Noninvasive Visualization and Analysis of Parafoveal Capillaries in Humans. *Investigative Ophthalmology & Visual Science* 2010;51:1691-1698.

34. Martin JA, Roorda A. Direct and noninvasive assessment of parafoveal capillary leukocyte velocity. *Ophthalmology* 2005;112:2219-2224.
35. Zhong ZY, Petrig BL, Qi XF, Burns SA. In vivo measurement of erythrocyte velocity and retinal blood flow using adaptive optics scanning laser ophthalmoscopy. *Optics Express* 2008;16:12746-12756.
36. Gray DC, Merigan W, Wolfing JI, et al. In vivo fluorescence imaging of primate retinal ganglion cells and retinal pigment epithelial cells. *Optics Express* 2006;14:7144-7158.
37. Gray DC, Wolfe R, Gee BP, et al. In vivo imaging of the fine structure of rhodamine-labeled macaque retinal ganglion cells. *Investigative Ophthalmology & Visual Science* 2008;49:467-473.
38. Morgan JIW, Dubra A, Wolfe R, Merigan WH, Williams DR. In Vivo Autofluorescence Imaging of the Human and Macaque Retinal Pigment Epithelial Cell Mosaic. *Investigative Ophthalmology & Visual Science* 2009;50:1350-1359.
39. Roorda A, Zhang YH, Duncan JL. High-resolution in vivo imaging of the RPE mosaic in eyes with retinal disease. *Investigative Ophthalmology & Visual Science* 2007;48:2297-2303.
40. Zhang YH. AOSLO: from benchtop to clinic. In: Giles MK (ed), *SPIE*; 2006.
41. Williams DR, Yoon GY. Wavefront sensor with off-axis illumination. USA: University of Rochester; 2001.
42. Venkateswaran K, Roorda A, Romero-Borja F. Theoretical modeling and evaluation of the axial resolution of the adaptive optics scanning laser ophthalmoscope. *Journal of Biomedical Optics* 2004;9:132-138.
43. Fernandez EJ, Iglesias I, Artal P. Closed-loop adaptive optics in the human eye. *Optics Letters* 2001;26:746-748.
44. Porter J, Guirao A, Cox IG, Williams DR. Monochromatic aberrations of the human eye in a large population. *JOSA A* 2001;18:1793-1803.
45. Pantanelli S, MacRae S, Jeong TM, Yoon G. Characterizing the wave aberration in eyes with keratoconus or penetrating keratoplasty using a high-dynamic range wavefront sensor. *Ophthalmology* 2007;114:2013-2021.
46. Iglesias I, Artal P. High-resolution retinal images obtained by deconvolution from wave-front sensing. *Optics Letters* 2000;25:1804-1806.
47. Prieto PM, Vargas-Martin F, Goelz S, Artal P. Analysis of the performance of the Hartmann-Shack sensor in the human eye. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2000;17:1388-1398.
48. Liu HX, Thibos L, Begley CG, Bradley A. Measurement of the Time Course of Optical Quality and Visual Deterioration during Tear Break-Up. *Investigative Ophthalmology & Visual Science* 2010;51:3318-3326.
49. Koh S, Tung C, Aquavella J, Yadav R, Zavislan J, Yoon G. Simultaneous Measurement of Tear Film Dynamics Using Wavefront Sensor and Optical Coherence Tomography. *Investigative Ophthalmology & Visual Science* 2010;51:3441-3448.
50. Gao WH, Jonnal RS, Cense B, Kocaoglu OP, Wang Q, Miller DT. Measuring directionality of the retinal reflection with a Shack-Hartmann wavefront sensor. *Optics Express* 2009;17:23085-23097.
51. Goodman GW. *Introduction to Fourier optics*. 2nd ed: McGraw-Hill Science/Engineering; 1996:441.

52. van Dam MA, Lane RG. Wave-front slope estimation. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2000;17:1319-1324.
53. Dai GM. *Wavefront optics for vision correction*. Bellingham, WA: Society of Photo-Optical Instrumentation Engineers; 2008.
54. Forrester J, Dick A, McMenamin P, Lee W. *The Eye: basic sciences in practice*. London, England: W. B. Saunders Company Ltd.; 1996:409.
55. Ellerbroek BL. Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2002;19:1803-1816.
56. Christou JC, Roorda A, Williams DR. Deconvolution of adaptive optics retinal images. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2004;21:1393-1401.
57. Thibos LN, Applegate RA, Schwiegerling JT, Webb R, Members VST. Standards for reporting the optical aberrations of eyes. *Journal of Refractive Surgery* 2002;18:S652-S660.
58. Noll RJ. Zernike Polynomials and Atmospheric-Turbulence. *Journal of the Optical Society of America* 1976;66:207-211.
59. Herrmann J. Least-Squares Wave-Front Errors of Minimum Norm. *Journal of the Optical Society of America* 1980;70:28-35.
60. Southwell WH. Wave-Front Estimation from Wave-Front Slope Measurements. *Journal of the Optical Society of America* 1980;70:998-1009.
61. Cubalchini R. Modal Wavefront Estimation from Phase Derivative Measurements. *Journal of the Optical Society of America* 1979;69:973-977.
62. Poyneer LA. Scene-based Shack-Hartmann wave-front sensing: analysis and simulation. *Applied Optics* 2003;42:5807-5815.
63. Singer B. Adaptive optics software for vision research. In: Porter J (ed), *Adaptive optics for vision science: principles, practices, design and applications.* Hoboken: Wiley-Interscience; 2006:139-154.
64. Mujat M, Ferguson RD, Patel AH, Iftimia N, Lue N, Hammer DX. High resolution multimodal clinical ophthalmic imaging system. *Optics Express* 2010;18:11607-11621.
65. Baker KL, Moallem MM. Iteratively weighted centroiding for Shack-Hartmann wave-front sensors. *Optics Express* 2007;15:5147-5159.
66. Herrmann J. Cross coupling and aliasing in modal wave-front estimation. *Journal of the Optical Society of America* 1981;71:989-992.
67. Yoon G. Wavefront sensing and diagnostic uses. In: Porter J (ed), *Adaptive optics for vision science: principles, practices, design and applications.* Hoboken: Wiley-Interscience; 2006:63-81.
68. Yoon G, Pantanelli S, MacRae S. Comparison of Zernike and Fourier wavefront reconstruction algorithms in representing corneal aberration of normal and abnormal eyes. *Journal of Refractive Surgery* 2008;24:582-590.
69. Shahidi M, Blair NP, Mori M, Zelkha R. Optical section retinal Imaging and wavefront sensing in diabetes. *Optometry and Vision Science* 2004;81:778-784.
70. Lundstrom L, Gustafsson J, Svensson I, Unsbo P. Assessment of objective and subjective eccentric refraction. *Optometry and Vision Science* 2005;82:298-306.

71. Llorente L, Marcos S, Dorronsoro C, Burns SA. Effect of sampling on real ocular aberration measurements. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2007;24:2783-2796.
72. Arines J, Bara S. Hybrid technique for high resolution imaging of the eye fundus. *Optics Express* 2003;11:761-766.
73. Shahidi M, Yang YR. Measurements of ocular aberrations and light scatter in healthy subjects. *Optometry and Vision Science* 2004;81:853-857.
74. Roorda A, Glasser A. Wave aberrations of the isolated crystalline lens. *Journal of Vision* 2004;4:250-261.
75. Vilupuru AS, Roorda A, Glasser A. Spatially variant changes in lens power during ocular accommodation in a rhesus monkey eye. *Journal of Vision* 2004;4:299-309.
76. Hansen PC, Nagy JG, O'Leary DP. *Deblurring images: matrices, spectra, and filtering*. 1 ed. Philadelphia: SIAM; 2006.
77. Dai GM. Matlab code for zonal reconstruction. *Wavefront optics for vision correction.* Bellingham, WA: Society of Photo-Optical Instrumentation Engineers; 2008:122-123.
78. Hudgin RH. Wavefront Reconstruction for Compensated Imaging. *Journal of the Optical Society of America* 1977;67:375-378.
79. Dai GM. Comparison of wavefront reconstructions with Zernike polynomials and Fourier transforms. *Journal of Refractive Surgery* 2006;22:943-948.
80. Wang L, Chernyak D, Yeh D, Koch DD. Fitting behaviors of Fourier transform and Zernike polynomials. *Journal of Cataract and Refractive Surgery* 2007;33:999-1004.
81. Ou JI, Manche EE. Zernike versus Fourier treatment tables for myopic patients having CustomVue wavefront laser in situ keratomileusis with the S4 excimer laser. *Journal of Cataract and Refractive Surgery* 2007;33:654-657.
82. Poyneer LA, Gavel DT, Brase JM. Fast wave-front reconstruction in large adaptive optics systems with use of the Fourier transform. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2002;19:2100-2111.
83. Freischlad KR, Koliopoulos CL. Modal Estimation of a Wave-Front from Difference Measurements Using the Discrete Fourier-Transform. *Journal of the Optical Society of America a-Optics Image Science and Vision* 1986;3:1852-1861.
84. Roddier F, Roddier C. Wave-Front Reconstruction Using Iterative Fourier-Transforms. *Applied Optics* 1991;30:1325-1327.
85. Bahk SW. Band-limited wavefront reconstruction with unity frequency response from Shack-Hartmann slopes measurements. *Optics Letters* 2008;33:1321-1323.
86. Dai GM. Orthonormal polynomials and their merits. *Wavefront optics for vision correction.* Bellingham, WA: Society of Photo-Optical Instrumentation Engineers; 2008:34-35.
87. Dai GM, Mahajan VN. Orthonormal polynomials in wavefront analysis: error analysis. *Applied Optics* 2008;47:3433-3445.
88. Dai GM. Wavefront reconstruction methods. *Wavefront optics for vision correction.* Bellingham, WA: Society of Photo-Optical Instrumentation Engineers; 2008:105-119.
89. Llorente L. Optical aberrations in ametropic eyes and their change with corneal refractive surgery. *Optometry & Visual Science.* London: City University; 2009:286.

90. Li KY, Mishra S, Tiruveedhula P, Roorda A. Comparison of control algorithms for a MEMS-based adaptive optics scanning laser ophthalmoscope. *Proceedings of the 2009 American Control Conference* 2009;3848-3853.
91. Bifano T, Bierden P, Perreault J. Micromachined deformable mirrors for dynamic wavefront control. *Proceedings of SPIE* 2004;5553:1-16.
92. Zhang Y, Cense B, Rha J, et al. High-speed volumetric imaging of cone photoreceptors with adaptive optics spectral-domain optical coherence tomography. *Optics Express* 2006;14:4380-4394.
93. Doble N, Yoon G, Chen L, et al. Use of a microelectromechanical mirror for adaptive optics in the human eye. *Optics Letters* 2002;27:1537-1539.
94. Stewart JB, Diouf A, Zhou YP, Bifano TG. Open-loop control of a MEMS deformable mirror for large-amplitude wavefront control. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2007;24:3827-3833.
95. Morzinski KM, Gavel DT, Norton AP, Dillon DR, Reinig MR. Characterizing MEMS deformable mirrors for open-loop operation: High-resolution measurements of thin-plate behavior. *SPIE*; 2008.
96. Wallace BP, Hampton PJ, Bradley CH, Conan R. Evaluation of a MEMS deformable mirror for an adaptive optics test bench. *Optics Express* 2006;14:10132-10138.
97. Vogel CR, Yang Q. Modeling, simulation, and open-loop control of a continuous facesheet MEMS deformable mirror. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2006;23:1074-1081.
98. Blain C, Conan R, Bradley C, Guyon O. Open-loop control demonstration of Micro-Electro-Mechanical-System MEMS Deformable Mirror. *Optics Express* 2010;18:5433-5448.
99. Zou WY, Burns SA. High-accuracy wavefront control for retinal imaging with Adaptive-Influence-Matrix Adaptive Optics. *Optics Express* 2009;17:20167-20177.
100. Zawadzki RJ, Choi SS, Jones SM, Oliver SS, Werner JS. Adaptive optics-optical coherence tomography: optimizing visualization of microscopic retinal structures in three dimensions. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2007;24:1373-1383.
101. Pallikaris A, Williams DK, Hofer H. The reflectance of single cones in the living human eye. *Investigative Ophthalmology & Visual Science* 2003;44:4580-4592.
102. Chen DC, Jones SM, Silva DA, Olivier SS. High-resolution adaptive optics scanning laser ophthalmoscope with dual deformable mirrors. *JOSA A* 2007;24:1305-1312.
103. Zou W, Qi X, Burns SA. Wavefront-aberration sorting and correction for a dual-deformable-mirror adaptive-optics system. *Optics Letters* 2008;33:2602-2604.
104. Chen L. Control algorithms. In: Porter J (ed), *Adaptive optics for vision science*: Wiley-Interscience; 2006:119-137.
105. Evans JW, Zawadzki RJ, Jones SM, Olivier SS, Werner JS. Error Budget Analysis for an Adaptive Optics Optical Coherence Tomography System. *Optics Express* 2009;17:13768-13784.
106. Evans JW, Zawadzki RJ, Jones SM, Olivier SS, Werner JS. Performance of a MEMS-based AO-OCT System using Fourier Reconstruction. *SPIE*; 2009.

107. Liang JZ, Williams DR, Miller DT. Supernormal vision and high-resolution retinal imaging through adaptive optics. *JOSA A* 1997;14:2884-2892.
108. Zhu LJ, Sun PC, Bartsch DU, Freeman WR, Fainman Y. Adaptive control of a micromachined continuous-membrane deformable mirror for aberration compensation. *Applied Optics* 1999;38:168-176.
109. Boyer C, Michau V, Rousset G. Adaptive optics: interaction matrix measurement and real-time control algorithms for the Come-On project. *Proceedings of SPIE* 1990;1542:46-61.
110. Jiang W, Li H. Hartmann-Shack wave-front sensing and wave-front control algorithm. *Proceedings of SPIE* 1990;1271:82-93.
111. Hofer H, Porter J, Yoon G, Chen L, Singer B, Williams DR. Rochester adaptive optics ophthalmoscope. In: Porter J (ed), *Adaptive optics for vision science: principles, practices, design and applications.* Hoboken: Wiley-Interscience; 2006:395-414.
112. Hecht E. Interference. In: Black A (ed), *Optics.* Reading, MA: Addison-Wesley; 2002:385-442.
113. Malacara D, Servin M, Malacara Z. *Interferogram analysis for optical testing*. 2nd ed. Boca Raton, FL: Taylor & Francis Group, LLC; 2005.
114. Malacara D, Servin M, Malacara Z. Phase-detection algorithms. *Interferogram analysis for optical testing.* Boca Raton, FL: Taylor & Francis Group, LLC; 2005:259-358.
115. Malacara D, Servin M, Malacara Z. Phase unwrapping. *Interferogram analysis for optical testing.* Boca Raton, FL: Taylor & Francis Group, LLC; 2005:493-524.
116. Zhou Y, Bifano TG. Characterization of contour shapes achievable with a MEMS deformable mirror. *Proceedings of SPIE* 2006;6113:123-130.
117. Zhou Y, Bifano TG. Adaptive optics using a MEMS deformable mirror. In: Jiang W (ed), *Proceedings of SPIE*; 2005.
118. Perreault JA, Bifano TG, Levine BM, Horenstein MN. Adaptive optic correction using microelectromechanical deformable mirrors. *Optical Engineering* 2002;41:561-566.
119. Moore KL, Dahleh M, Bhattacharyya SP. Iterative Learning Control - a Survey and New Results. *Journal of Robotic Systems* 1992;9:563-594.
120. van Dam MA, Le Mignant D, Macintosh BA. Performance of the Keck Observatory adaptive-optics system. *Applied Optics* 2004;43:5458-5467.
121. Wallner EP. Optimal Wave-Front Correction Using Slope Measurements. *Journal of the Optical Society of America*
1983;73:1771-1776.
122. Cagigal MP, Canales VF, Castejon-Mochon JF, Prieto PM, Lopez-Gil N, Artal P. Statistical description of wave-front aberration in the human eye. *Optics Letters* 2002;27:37-39.
123. Law NF, Lang RG. Wavefront estimation at low light levels. *Optics Communications* 1996;126:19-24.
124. Bechet C, Tallon M, Thiebaut E. Comparison of minimum-norm maximum likelihood and maximum a posteriori wavefront reconstructions for large adaptive optics systems. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2009;26:497-508.

125. Lessard L, West M, MacMynowski D, Lall S. Warm-started wavefront reconstruction for adaptive optics. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2008;25:1147-1155.
126. Ellerbroek BL. Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques. *JOSA A* 2002;19:1803-1816.
127. Makidon RB, Sivaramakrishnan A, Perrin MD, et al. An analysis of fundamental waffle mode in early AEOS adaptive optics images. *Publications of the Astronomical Society of the Pacific* 2005;117:831-846.
128. Poyneer LA, Gavel DT, Brase JM. Fast wave-front reconstruction in large adaptive optics systems with use of the Fourier transform. *JOSA A* 2002;19:2100-2111.
129. Le Louarn M, Tallon M. Analysis of modes and behavior of a multiconjugate adaptive optics system. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2002;19:912-925.
130. Diolaiti E, Arcidiacono C, Ragazzoni R, Fedrigo E. Identification and rejection of waffle modes in layer-oriented adaptive optics. In: Wizinowich PL, Bonaccini D (eds), *SPIE*; 2003:1001-1010.
131. Vogel CR, Tyler G, Lu Y, Bifano TG, Conan R, Blain C. Modeling and parameter estimation for point-actuated continuous-facesheet deformable mirrors. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2010;27:A56-A63.
132. Poonja S, Patel S, Henry L, Roorda A. Dynamic visual stimulus presentation in an adaptive optics scanning laser ophthalmoscope. *Journal of Refractive Surgery* 2005;21:S575-S580.
133. Chui TYP, Yap MKH, Chan HHL, Thibos LN. Retinal stretching limits peripheral visual acuity in myopia. *Vision Research* 2005;45:593-605.
134. Coletta NJ, Watson T. Effect of myopia on visual acuity measured with laser interference fringes. *Vision Research* 2006;46:636-651.
135. Chui TYP, Song H, Burns SA. Individual variations in human cone photoreceptor packing density: Variations with refractive error. *Investigative Ophthalmology & Visual Science* 2008;49:4679-4687.
136. Strang NC, Winn B, Bradley A. The role of neural and optical factors in limiting visual resolution in myopia. *Vision Research* 1998;38:1713-1721.
137. Bullimore MA, Gilmartin B, Royston JM. Steady-State Accommodation and Ocular Biometry in Late-Onset Myopia. *Documenta Ophthalmologica* 1992;80:143-155.
138. Lam AKC, Wong S, Lam CSY, To CH. The effect of myopic axial elongation and posture on the pulsatile ocular blood flow in young normal subjects. *Optometry and Vision Science* 2002;79:300-305.
139. Kitaguchi Y, Bessho K, Yamaguchi T, Nakazawa N, Mihashi T, Fujikado T. In vivo measurements of cone photoreceptor spacing in myopic eyes from images obtained by an adaptive optics fundus camera. *Japanese Journal of Ophthalmology* 2007;51:456-461.
140. Adams AJ. Axial Length Elongation, Not Corneal Curvature, as a Basis of Adult Onset Myopia. *American Journal of Optometry and Physiological Optics* 1987;64:150-152.
141. Mcbrien NA, Millodot M. A Biometric Investigation of Late Onset Myopic Eyes. *Acta Ophthalmologica* 1987;65:461-468.

142. Rossi EA, Weiser P, Tarrant J, Roorda A. Visual performance in emmetropia and low myopia after correction of high-order aberrations. *Journal of Vision* 2007;7:-.
143. Liou SW, Chiu CJ. Myopia and contrast sensitivity function. *Current Eye Research* 2001;22:81-84.
144. Jaworski A, Gentle A, Zele AJ, Vingrys AJ, McBrien NA. Altered visual sensitivity in axial high myopia: A local postreceptoral phenomenon? *Investigative Ophthalmology & Visual Science* 2006;47:3695-3702.
145. Abbott ML, Schmid KL, Strang NC. Differences in the accommodation stimulus response curves of adult myopes and emmetropes. *Ophthalmic and Physiological Optics* 1998;18:13-20.
146. O'Leary DJ, Allen PM. Facility of accommodation in myopia. *Ophthalmic and Physiological Optics* 2001;21:352-355.
147. Vera-Diaz FA, McGraw PV, Strang NC, Whitaker D. A psychophysical investigation of ocular expansion in human eyes. *Investigative Ophthalmology & Visual Science* 2005;46:758-763.
148. Troilo D. Changes in Retinal Morphology Following Experimentally Induced Myopia. *Optical Society of America, Technical Digest Series, Vol 1, Vision Science and Its Application*; 1998:206-209.
149. Atchison DA, Jones CE, Schmid KL, et al. Eye shape in emmetropia and myopia. *Investigative Ophthalmology & Visual Science* 2004;45:3380-3386.
150. Bradley A, Hook J, Haeseker J. A Comparison of Clinical Acuity and Contrast Sensitivity Charts - Effect of Uncorrected Myopia. *Ophthalmic and Physiological Optics* 1991;11:218-226.
151. Atchison DA, Schmid KL, Pritchard N. Neural and optical limits to visual performance in myopia. *Vision Research* 2006;46:3707-3722.
152. Arathorn DW, Yang Q, Vogel CR, Zhang Y, Tiruveedhula P, Roorda A. Retinally stabilized cone-targeted stimulus delivery. *Optics Express* 2007;15:13731-13744.
153. Stevenson SB, Roorda A. Correcting for miniature eye movements in high resolution scanning laser ophthalmoscopy. In: Manns F, Soderberg P, Ho A (eds), *Ophthalmic Technologies XI.* Bellingham, WA; 2005:145-151.
154. Li KY, Roorda A. Automated identification of cone photoreceptors in adaptive optics retinal images. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2007;24:1358-1363.
155. Mujat M, Ferguson RD, Iftimia N, Hammer DX. Compact adaptive optics line scanning ophthalmoscope. *Optics Express* 2009;17:10242-10258.
156. Bennett AG, Rudnicka AR, Edgar DF. Improvements on Littmann Method of Determining the Size of Retinal Features by Fundus Photography. *Graefes Archive for Clinical and Experimental Ophthalmology* 1994;232:361-367.
157. Martin JA, Roorda A. Pulsatility of parafoveal capillary leukocytes. *Experimental Eye Research* 2009;88:356-360.
158. Doughty MJ, Zaman ML. Human corneal thickness and its impact on intraocular pressure measures: A review and meta-analysis approach. *Survey of Ophthalmology* 2000;44:367-408.
159. Patel S, Marshall J, Fitzke FW. Refractive-Index of the Human Corneal Epithelium and Stroma. *Journal of Refractive Surgery* 1995;11:100-105.

160. Williams TD. Determination of the True Size of an Object on the Fundus of the Living Eye. *Optometry and Vision Science* 1992;69:717-720.
161. Hecht E. Analytical ray tracing. In: Black A (ed), *Optics*: Addison Wesley Longman, Inc.; 2002:246-253.
162. Breu H, Gil J, Kirkpatrick D, Werman M. Linear-Time Euclidean Distance Transform Algorithms. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 1995;17:529-533.
163. Costa LD, Rocha F, de Lima SMA. Characterizing polygonality in biological structures. *Physical Review E* 2006;73:-.
164. Costa LD, Bonci DMO, Saito CA, Rocha FAD, Silveira LCD, Ventura DF. Voronoi analysis uncovers relationship between mosaics of normally placed and displaced amacrine cells in the thraira retina. *Neuroinformatics* 2007;5:59-77.
165. Chui TYP, Song HX, Burns SA. Adaptive-optics imaging of human cone photoreceptor distribution. *Journal of the Optical Society of America a-Optics Image Science and Vision* 2008;25:3021-3029.
166. Sparrow CM. On spectroscopic resolving power. *Astrophysical Journal* 1916;44:76-86.
167. Grimes DN, Thompson BJ. 2-Point Resolution with Partially Coherent Light. *Journal of the Optical Society of America* 1967;57:1330-&.
168. Hecht E. Fraunhofer diffraction. In: Black A (ed), *Optics*: Addison Wesley Longman, Inc.; 2002:452-485.
169. Atchison DA, Pritchard N, Schmid KL, Scott DH, Jones CE, Pope JM. Shape of the retinal surface in emmetropia and myopia. *Investigative Ophthalmology & Visual Science* 2005;46:2698-2707.
170. Dubis AM, McAllister JT, Carroll J. Reconstructing foveal pit morphology from optical coherence tomography imaging. *British Journal of Ophthalmology* 2009;93:1223-1227.

Appendix A: Proof of Equation 3.3.4

Equation 3.3.4 is given by:

$$\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\left[\mathbf{T}\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}+\Lambda_{\mathbf{vv}}\right]^{-1}=\left[\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}\mathbf{T}+\Lambda_{\mathbf{uu}}^{-1}\right]^{-1}\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}$$

In order to verify that these two expressions are indeed equivalent, we will make use of analytical formulae for blockwise matrix inversion which makes heavy use of the Schur complement. For bookkeeping purposes, I will assume the matrix dimensions that correspond to a 6 mm diameter pupil measured and compensated with the Shack-Hartmann sensor and BMC DM in AOSLOII[6]:

$$\begin{cases}\mathbf{T}\in\mathbb{R}^{426\times144}\\ \Lambda_{\mathbf{uu}}\in\mathbb{R}^{144\times144}\\ \Lambda_{\mathbf{vv}}\in\mathbb{R}^{426\times426}\end{cases}$$

I will also assume that the covariance matrices $\Lambda_{\mathbf{uu}}$ and $\Lambda_{\mathbf{vv}}$ are nonsingular (or Equation 3.3.4 would not make any sense) which is technically not guaranteed in Wavefront reconstruction but nevertheless reasonable for the current purpose. First, consider the block matrix:

$$\mathbf{M}\triangleq\begin{bmatrix}\Lambda_{\mathbf{uu}}^{-1} & \mathbf{T}^{\mathbf{T}}\\ \mathbf{T} & -\Lambda_{\mathbf{vv}}\end{bmatrix}\in\mathbb{R}^{566\times566} \tag{A.1}$$

The corresponding Schur complements of the upper-right and lower-left partitions are defined by:

$$\begin{cases}\Delta\left(\Lambda_{\mathbf{uu}}^{-1}\right)\triangleq-\Lambda_{\mathbf{vv}}-\mathbf{T}\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\\ \Delta\left(-\Lambda_{\mathbf{vv}}\right)\triangleq\Lambda_{\mathbf{uu}}^{-1}+\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}\mathbf{T}\end{cases} \tag{A.2}$$

We can obtain a different analytical expression for the inverse of matrix $\mathbf{M}$ depending on which Schur complement is used:

$$\mathbf{M}^{-1}=\begin{bmatrix}\Lambda_{\mathbf{uu}}+\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\Delta\left(\Lambda_{\mathbf{uu}}^{-1}\right)^{-1}\mathbf{T}\Lambda_{\mathbf{uu}} & -\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\Delta\left(\Lambda_{\mathbf{uu}}^{-1}\right)^{-1}\\ -\Delta\left(\Lambda_{\mathbf{uu}}^{-1}\right)^{-1}\mathbf{T}\Lambda_{\mathbf{uu}} & -\Delta\left(\Lambda_{\mathbf{uu}}^{-1}\right)^{-1}\end{bmatrix} \tag{A.3}$$

$$\mathbf{M}^{-1}=\begin{bmatrix}\Delta\left(-\Lambda_{\mathbf{vv}}\right)^{-1} & \Delta\left(-\Lambda_{\mathbf{vv}}\right)^{-1}\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}\\ \Lambda_{\mathbf{vv}}^{-1}\mathbf{T}\Delta\left(-\Lambda_{\mathbf{vv}}\right)^{-1} & -\Lambda_{\mathbf{vv}}^{-1}+\Lambda_{\mathbf{vv}}^{-1}\mathbf{T}\Delta\left(-\Lambda_{\mathbf{vv}}\right)^{-1}\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}\end{bmatrix} \tag{A.4}$$

Since Equations A.3 and A.4 are equal, their respective partitions must also be equal. Consider the upper-right partitions:

$$-\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\Delta\left(\Lambda_{\mathbf{uu}}^{-1}\right)^{-1}=\Delta\left(-\Lambda_{\mathbf{vv}}\right)^{-1}\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}$$

Substituting in the appropriate Schur complements (Equation A.2), we get:

$$-\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\left[-\Lambda_{\mathbf{vv}}-\mathbf{T}\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\right]^{-1}=\left[\Lambda_{\mathbf{uu}}^{-1}+\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}\mathbf{T}\right]^{-1}\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}$$

and with some trivial rearranging, we get:

$$\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}\left[\mathbf{T}\Lambda_{\mathbf{uu}}\mathbf{T}^{\mathbf{T}}+\Lambda_{\mathbf{vv}}\right]^{-1}=\left[\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}\mathbf{T}+\Lambda_{\mathbf{uu}}^{-1}\right]^{-1}\mathbf{T}^{\mathbf{T}}\Lambda_{\mathbf{vv}}^{-1}$$

which is exactly Equation 3.3.4 ∎

Appendix B: Matlab code

B.1 Zonal reconstruction matrices for circular pupils

The following Matlab script generates matrices **D** and Γ, saving them in a mat file which I called "DandG.mat". Only when the wavefront slope measurements are computed (Equation 2.36), vectorized and stacked with the x-slopes on top of the y-slopes can the wavefront actually be computed using the Equation 2.71:

$$\hat{\phi} = \left(\Gamma^{\mathbf{T}}\Gamma + \mathbf{V}\mathbf{V}^{\mathbf{T}}\right)^{-1}\Gamma^{\mathbf{T}}\mathbf{D}\mathbf{y}$$

Generating matrix **V** is trivial (see Figure 2.12) since it is just a column of ones and another column of alternating ones and negative ones, so I will not waste space and time to include code for that. The resultant estimated wavefront is a vector, so each entry must be moved to the right location in the pupil in order to plot properly.

B.1.1 Main script

```
% Adjust these parameters as needed
PupilDiameter_mm = 6;              % Eye's pupil diameter (mm)
LensletPitch_um = 328;             % Diameter of subaperture (microns)
magnification = 0.889;             % Pupil to lenslet array magnification

%% Do not modify any code below this line
% Generate LensletMatrix
w = LensletPitch_um/1000;
MAXCOLS = 2*floor(ri_ratio) + 1;    %Number of lenslets across the pupil
limit = floor(ri_ratio);
LensletMatrix = zeros(MAXCOLS);
CounterInsidePupil = 1;     %lenslet ID number inside pupil
col=0;
for x = -limit:1:limit
   col = col + 1;
   row = 0;
   for y = limit:-1:-limit
      row = row + 1;
      rad = sqrt(x*x + y*y);
      if rad <= ri_ratio
         LensletMatrix(row, col) = CounterInsidePupil;
         CounterInsidePupil = CounterInsidePupil + 1;
      end
   end
end
cols = MAXCOLS;
clear CounterInsidePupil MAXCOLS x y rad row col

% GetDandGMatrices
N = max(max(LensletMatrix));
In = cell(1, cols); %cell of empty identity matrices
for col = 1 : cols
   n = GetNumLensletsInColumn(LensletMatrix, col);
   if(n > 0)
      In{ n } = eye(n, N);
      if(n > 1)
         In{ n-1} = eye(n-1, N);
```

```
      end
   end
end

% x derivative operator (between columns: c1:1 vs 2, c2:2 vs 3, c3:3 vs 4...)
NonEmptyCols = [];
for col = 1 : cols
   n = GetNumLensletsInColumn(LensletMatrix, col);
   if(n > 0)
      NonEmptyCols = [NonEmptyCols; [col n]];
   end
end
Cn    = cell(1, cols-1);
A = [];
[firstID neighbourID] = GetFirstLensletIDAndNextNeighbourID(LensletMatrix, 1);
for i = 1 : length(Cn)
   col        = NonEmptyCols(i,1);    %column number
   n          = min(NonEmptyCols(i,2),NonEmptyCols(i+1,2));    %number of lenslets in the column
   [firstID neighbourID] = GetFirstLensletIDAndNextNeighbourID(LensletMatrix, col);
   C0 = circshift(- In{ n },[0 , firstID-1]);
   C1 = circshift(In{ n },[0 , neighbourID-1]);
   Cn{ i } =  C0 + C1;
   A = [A; Cn{ i }];
end

% Y derivative operator
Dn    = cell(1, cols);
B     = [];
for i = 1 : length(Dn)
   col        = NonEmptyCols(i,1);    %column number
   n          = NonEmptyCols(i,2);    %number of lenslets in the column
   firstID    = GetFirstLensletID(LensletMatrix, col);
   Dn{ i }       = circshift(In{ n-1 }, [0 ,firstID-1]) - circshift(In{ n-1 }, [0, firstID]);
   B          = [B; - Dn{ i }];
end

% Properly scale the matrices
G = [A;B]/(LensletPitch_um);        %Gradient operator, Hudgin geometry
D = 0.5*abs(blkdiag(A,B));          %Southwell to Hudgin interpolator
save('DandG.mat','D','G')
```

B.1.2 Supplementary functions

```
% GetFirstLensletID
function lensID = GetFirstLensletID(LensletMatrix, col)
[rows cols] = size(LensletMatrix);
lensID = 0;
if(col <= cols)
   for row = 1 : rows
     if( LensletMatrix(row,col) > 0)
        lensID = LensletMatrix(row,col);
        break;
     end
   end
end

% GetFirstLensletIDAndNextNeighbourID
```

```
function [firstlensID NeighbourID] = GetFirstLensletIDAndNextNeighbourID(LensletMatrix, col)
[rows cols] = size(LensletMatrix);
NeighbourID            = 0;
firstlensID            = GetFirstLensletID(LensletMatrix, col);
firstlensIndex         = find(LensletMatrix == firstlensID);
estimatedNeighbourIndex = firstlensIndex + rows;
if(estimatedNeighbourIndex <= rows*cols)
   estimatedNeighbourID    = LensletMatrix(estimatedNeighbourIndex);
   nextColFirstlensID      = GetFirstLensletID(LensletMatrix, col+1);
   NeighbourID             = max(estimatedNeighbourID, nextColFirstlensID);
   if(NeighbourID ~= estimatedNeighbourID)
      firstlensID = LensletMatrix(find(LensletMatrix == NeighbourID) - rows);
   end
end

%GetListNumLensletsInColumnsInPupil
function NonEmptyCols = GetListNumLensletsInColumnsInPupil(LensletMatrix)
NonEmptyCols = [];
[rows cols]     = size(LensletMatrix);
for col = 1 : cols
   n = GetNumLensletsInColumn(LensletMatrix, col);
   if(n > 0)
      NonEmptyCols = [NonEmptyCols; [col n]];
   end
end

% GetNumLensletsInColumn
function numLens = GetNumLensletsInColumn(LensletMatrix, col)
[rows cols] = size(LensletMatrix);
numLens = 0;
if(col <= cols)
   for row = 1 : rows
     if( LensletMatrix(row,col) > 0)
        numLens = numLens + 1;
     end
   end
end
```

B.2 Automated cone photoreceptor identification

This Matlab script originated from what was describe in my paper published in the *Journal of the Optical Society of America A*[154] which was mainly on an automated cone identification algorithm for analyzing AO retinal images. My original routine did not perform adequately near the foveal center where the cone mosaic has lower contrast (assuming the cones are resolved). Mujat *et al.*[155] added background estimation and subtraction steps to supposedly my method, and I have confirmed that these commands do indeed result in more accurate cone location estimates. However, while they opted for a 2 by 2 median filter, I found that a Gaussian filter does a better job in terms of the number of wrongly identified pixels as cone centers. The following function for automated cone identification uses a 5 by 5 Gaussian kernel, generated by Matlab function *fspecial.m*, with a standard deviation of 3 pixels which is kept constant for all cases. Some Matlab IPT functions are called, so it is needed to run the function. The input arguments are a hard threshold for the minimum intensity count that will be

considered a cone center and the original grayscale image as a two-dimensional array. The output is a binary image that is *true* at each identified cone center and false everywhere else. The total number of cones identified it outputted as well. The performance of this simple routine is good overall as one can see in Figure B.1, but there are always a few misidentified cones that I manually adjust or remove as sort of a post-inspection process.

(a) (b)

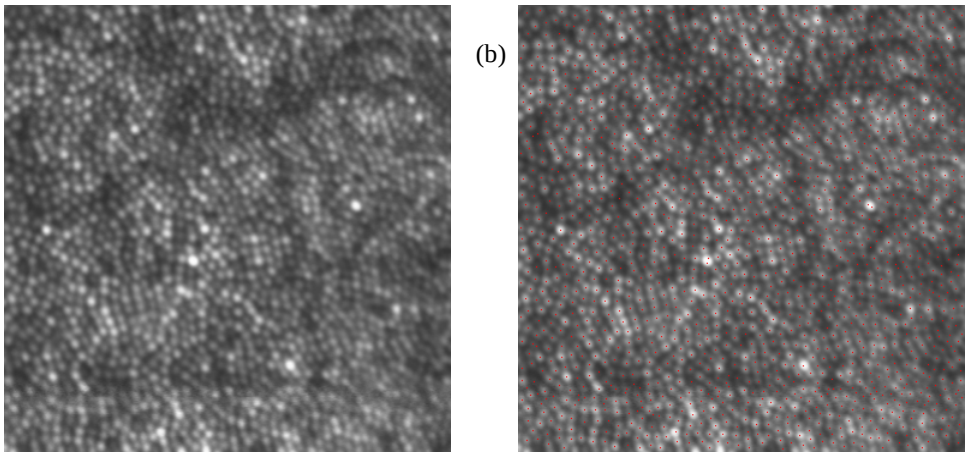


Figure B.1: (a) Image of a healthy cone mosaic at about 1° eccentricity and (b) the same image with the cone locations identified completely using the automated routine.

```
function [I no_cones] = auto2(Threshold, im_original)

f = double(im_original);
[M N] = size(f);
gauss_filt = fspecial('gaussian',5,3);          %Low pass Gaussian filter
bg = imfilter(f,gauss_filt,'replicate','conv');
f = f - bg;                                     %Subtract background from the original image
h = ones(3)/9;                                  %3 by 3 average filter
f = imfilter(f, h, 'replicate', 'same');

%Morphological markers generation
LocalMins = imregionalmax(f, 4);    %binary: 1s at all local minima
se = strel('disk', 1, 0);                       %Group multiple finds 1 pixel apart
ConeMark = imdilate(LocalMins, se);
ConeMark = imclearborder(ConeMark); %Get rid of finds touching ROI border
[L no_cones] = bwlabel(ConeMark);

Y = zeros(no_cones,1);
X = Y;
for ind = 1:no_cones;
   [y x] = find(L == ind);
   Y(ind) = sum(y)/length(y);
   X(ind) = sum(x)/length(x);
end
Y = round(Y);
X = round(X);
index = sub2ind([M N], Y, X);
I = false(M,N);
```

```
I(index) = true;
I(im_original < Threshold) = 0;
no_cones = sum(sum(double(I)));
% End of function auto2
```

B.3 Retinal magnification factor

The retinal magnification factor is the scaling factor that relates visual angle subtended by some object in the world to the actual size of the image formed by the object on the retina. The calculation I requires three arguments which are measurements given by an IOLMaster (Carl Zeiss Meditec, Jena, Germany). They are the axial length, front radius of curvature of the cornea, and the anterior chamber depth. These three parameters should be updated for different eyes assessed. The spectacle power is generally kept at zero because it affects the final positions of the nodal points, and the visual angle redefined by the presence of spectacle lenses was less important to us. All the other parameters from the Gullstrand four-surface eye model except for the corneal thickness and posterior radius of curvature (see Chapter 4, section 4.2.3 for details).

```
% Measurements from the IOLMaster in meters
x = 24.18e-3;               % Axial length
r1 = 7.8e-3;                % Front radius of curvature of cornea
acd = 3.585e-3;             % Anterior chamber depth

P_spec = 0;                 % Spectacle power

% Radii of Curvature of the lens specific to Gullstrand mode
lens_roc_f = 10.2e-3;
lens_roc_b = -6e-3;

% Refractive indices
n(1) = 1;                   % air
n(2) = 1.38;                % cornea (Patel et al. 1995)
n(3) = 1.3374;              % aqueous (Gullstrand)
n(4) = 1.42;                % lens (Gullstrand)
n(5) = 1.336;               % vitreous (Gullstrand)

% Thicknesses
t(1) = 14e-3;               % Distance from spectacle to apex
t(2) = .535e-3;             % Cornea (Doughty and Zaman, 2000)
t(3) = acd - t(2);          % Aqueous (Anatomical ACD)
t(4) = 4e-3;                % Lens (Gullstrand)
t(5) = inf;                 % Vitreous and beyond
T = acd + t(4);

% Powers
phi(1) = P_spec;                    % Spectacle (placeholder, keep at 0)
phi(2) = (n(2)-n(1))/r1;            % Cornea front surface power
phi(3) = (n(3)-n(2))/(.8831*r1);    % cornea back surface power (Littmann)
phi(4) = (n(4)-n(3))/lens_roc_f;
phi(5) = (n(5)-n(4))/lens_roc_b;

% Forward ray trace
N = 5;                      %number of surfaces
for ind = 1:N-1
```

```
    nu = nu - y*phi(ind);
    y = y + nu*t(ind)/n(ind);
end
nu = nu - y*phi(N);

bfl = -y*n(N)/nu;                   % back focal length (wrt last surface) includes refractive index (n/V)
H2F2 = -n(N)/nu;                    % second focal length
F1H1 = H2F2/n(N);                   % = N2F2
N2 = T + bfl - F1H1;                % uses relationship that F1H1 = N2F2
N2_retina = 1000*(x - N2);          % in mm
q = 1000*N2_retina*tand(1);
fprintf('Retinal magnification is %5.4f microns per degree\n',q);
```