INTRODUCTION TO

ADAPTIVE OPTICS

TUTORIAL TEXTS SERIES

- *Infrared Optics and Zoom Lenses,* Allen Mann, Vol. TT42
- *Introduction to Adaptive Optics,* Robert K. Tyson, Vol. TT41
- *Fractal and Wavelet Image Compression Techniques,* Stephen Welstead, Vol. TT40
- *Analysis of Sampled Imaging Systems,* R. H. Vollmerhausen and R. G. Driggers, Vol. TT39
- *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis,* Valery Tuchin, Vol. TT38
- *Fundamentos de Electro-Óptica para Ingenieros,* Glenn D. Boreman, translated by Javier Alda, Vol. TT37
- *Infrared Design Examples,* William L. Wolfe, Vol. TT36
- *Sensor and Data Fusion Concepts and Applications, Second Edition,* L. A. Klein, Vol. TT35
- *Practical Applications of Infrared Thermal Sensing and Imaging Equipment, Second Edition,* Herbert Kaplan, Vol. TT34
- *Fundamentals of Machine Vision,* Harley R. Myler, Vol. TT33
- *Design and Mounting of Prisms and Small Mirrors in Optical Instruments,* Paul R. Yoder, Jr., Vol. TT32
- *Basic Electro-Optics for Electrical Engineers,* Glenn D. Boreman, Vol. TT31
- *Optical Engineering Fundamentals,* Bruce H. Walker, Vol. TT30
- *Introduction to Radiometry,* William L. Wolfe, Vol. TT29
- *Lithography Process Control,* Harry J. Levinson, Vol. TT28
- *An Introduction to Interpretation of Graphic Images,* Sergey Ablameyko, Vol. TT27
- *Thermal Infrared Characterization of Ground Targets and Backgrounds,* P. Jacobs, Vol. TT26
- *Introduction to Imaging Spectrometers,* William L. Wolfe, Vol. TT25
- *Introduction to Infrared System Design,* William L. Wolfe, Vol. TT24
- *Introduction to Computer-based Imaging Systems,* D. Sinha, E. R. Dougherty, Vol. TT23
- *Optical Communication Receiver Design,* Stephen B. Alexander, Vol. TT22
- *Mounting Lenses in Optical Instruments,* Paul R. Yoder, Jr., Vol. TT21
- *Optical Design Fundamentals for Infrared Systems,* Max J. Riedl, Vol. TT20
- *An Introduction to Real-Time Imaging,* Edward R. Dougherty, Phillip A. Laplante, Vol. TT19
- *Introduction to Wavefront Sensors,* Joseph M. Geary, Vol. TT18
- *Integration of Lasers and Fiber Optics into Robotic Systems,* Janusz A. Marszalec, Elzbieta A. Marszalec, Vol. TT17
- *An Introduction to Nonlinear Image Processing,* E. R. Dougherty, J. Astola, Vol. TT16
- *Introduction to Optical Testing,* Joseph M. Geary, Vol. TT15
- *Image Formation in Low-Voltage Scanning Electron Microscopy,* L. Reimer, Vol. TT12
- *Diazonaphthoquinone-based Resists,* Ralph Dammel, Vol. TT11
- *Infrared Window and Dome Materials,* Daniel C. Harris, Vol. TT10
- *An Introduction to Morphological Image Processing,* Edward R. Dougherty, Vol. TT9
- *An Introduction to Optics in Computers,* Henri H. Arsenault, Yunlong Sheng, Vol. TT8
  Digital Image Compression Techniques, Majid Rabbani, Paul W. Jones, Vol. TT7
- *Aberration Theory Made Simple,* Virendra N. Mahajan, Vol. TT6
- *Single-Frequency Semiconductor Lasers,* Jens Buus, Vol. TT5
- *An Introduction to Biological and Artificial Neural Networks for Pattern Recognition,* Steven K. Rogers, Matthew Kabrisky, Vol. TT4
- *Laser Beam Propagation in the Atmosphere,* Hugo Weichel, Vol. TT3
- *Infrared Fiber Optics,* Paul Klocek, George H. Sigel, Jr., Vol. TT2
- *Spectrally Selective Surfaces for Heating and Cooling Applications,* C. G. Granqvist, Vol. TT1

INTRODUCTION TO

ADAPTIVE OPTICS

ROBERT K. TYSON

Tutorial Texts in Optical Engineering
Volume TT41

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Tyson, Robert K., 1948–
Introduction to adaptive optics / Robert K. Tyson.
p. cm. – (Tutorial texts in optical engineering ; v. TT41)
Includes bibliographical references and index.
ISBN 0-8194-3511-2 (softcover)
1. Optics, Adaptive. 2. Optical instruments. 3. Optical detectors. 4. Optical measurements. 5. Telescopes. I. Title. II. Series.
TA1522.T95 2000
621.36—dc21 99-058692
CIP

Published by

SPIE—The International Society for Optical Engineering
P.O. Box 10
Bellingham, Washington 98227-0010
Phone: 360/676-3290
Fax: 360/647-1445
Email: spie@spie.org
WWW: www.spie.org

Printed in the United States of America.

INTRODUCTION TO THE SERIES

The Tutorial Texts series was initiated in 1989 as a way to make the material presented in SPIE short courses available to those who couldn't attend and to provide a reference book for those who could. Typically, short course notes are developed with the thought in mind that supporting material will be presented verbally to complement the notes, which are generally written in summary form, highlight key technical topics, and are not intended as stand-alone documents. Additionally, the figures, tables, and other graphically formatted information included with the notes require further explanation given in the instructor's lecture. As stand-alone documents, short course notes do not generally serve the student or reader well.

Many of the Tutorial Texts have thus started as short course notes subsequently expanded into books. The goal of the series is to provide readers with books that cover focused technical interest areas in a tutorial fashion. What separates the books in this series from other technical monographs and textbooks is the way in which the material is presented. Keeping in mind the tutorial nature of the series, many of the topics presented in these texts are followed by detailed examples that further explain the concepts presented. Many pictures and illustrations are included with each text, and where appropriate tabular reference data are also included.

To date, the texts published in this series have encompassed a wide range of topics, from geometrical optics to optical detectors to image processing. Each proposal is evaluated to determine the relevance of the proposed topic. This initial reviewing process has been very helpful to authors in identifying, early in the writing process, the need for additional material or other changes in approach that serve to strengthen the text. Once a manuscript is completed, it is peer reviewed to ensure that chapters communicate accurately the essential ingredients of the processes and technologies under discussion.

During the past nine years, my predecessor, Donald C. O'Shea, has done an excellent job in building the Tutorial Texts series, which now numbers nearly forty books. It has expanded to include not only texts developed by short course instructors but also those written by other topic experts. It is my goal to maintain the style and quality of books in the series, and to further expand the topic areas to include emerging as well as mature subjects in optics, photonics, and imaging.

Arthur R. Weeks, Jr.
Invivo Research Inc. and University of Central Florida

Contents

Preface

Adaptive optics systems and components have reached a level of sophistication and simplicity where they are beginning to be used beyond traditional applications in astronomy and the military. At the turn of the millennium, the technology and principles have gone beyond the laboratory into developments in medicine, manufacturing, and communications. The operation of an adaptive optics system no longer can afford a few engineers and an army of specially trained technicians. *Introduction to Adaptive Optics* was written to reach those interested in the technology and those needing a broad-brush explanation without wading through thousands of journal articles. It bridges the gap between many well-written popular magazine articles about adaptive optics and the few detailed texts on the subject.

Introduction to Adaptive Optics follows the structure of a full-day tutorial that I teach. I have a lot of sympathy for those who sit and listen to one topic, presented by one instructor, for a full day. Thus, the book includes some humor and a few sidebars that discuss historical elements of adaptive optics. Some principles in the book are best explained by the use of mathematics, including calculus, but the book by no means relies on an understanding of anything other than the English language and the desire to be introduced to the multidisciplinary field of adaptive optics.

Robert K. Tyson
January 2000

Chapter 1

A long time ago, in a laboratory far, far, really far away

Five billion years ago, at about 9:30 in the morning, a burst of photons left a far-away star, traveling in a reasonably straight line. During the long journey of this vast wave of light, out there somewhere in the Milky Way galaxy, the planet Earth was born. On Earth, from the primordial ooze evolved protozoa. During the last four million years of the journey of the photons, the planet Earth was evolving a variety of unusual species including one which we call optical engineers. So, despite the photons' long journey, during which they were pretty much unchanged, they could finally reach the inquiring eyes of astronomers like Tycho Brahe or the clever inventor Galileo.

However, after five billion years, after literally staying on the straight and narrow, heading towards the blue planet, in the last millisecond (the last darn thousandth of a second!) the photons encountered the atmosphere of planet Earth. Air, seemingly transparent, should provide a nice soft landing environment for the photons. After all, the atmosphere slows down each photon from 186,280 miles per second to about 186,230 mps. The problem arises when part of the vast wave entering the atmosphere slows down a little more than another part. Although the difference in time is only a millionth billionth of a second, the information that the wave carries is garbled. The star's precise position, velocity, type, planetary neighbors, and most recent sports scores are distorted by the atmosphere.

During the course of human history, some people worried about the problem. (Isaac Newton, for instance.) Some people ignored it. (Most everybody else.) And only in the latter half of the 20th century did anybody do anything about it. This is their story and the story of whatever it was that they did about it.

Chapter 1 is a broad overview of practically everything about adaptive optics (AO). The later chapters describe the details. This is by no means an exhaustive review of the entire field of adaptive optics. For that, refer to the articles, book chapters, and monographs listed as footnotes or at the end of each chapter. To get the most out of this book, treat it like a course. Imagine the instructor. Now imagine the instructor ignoring your questions. After all, you are reading at your

own pace. So, for the book report — read chapter 1. For the final exam — read the whole book.

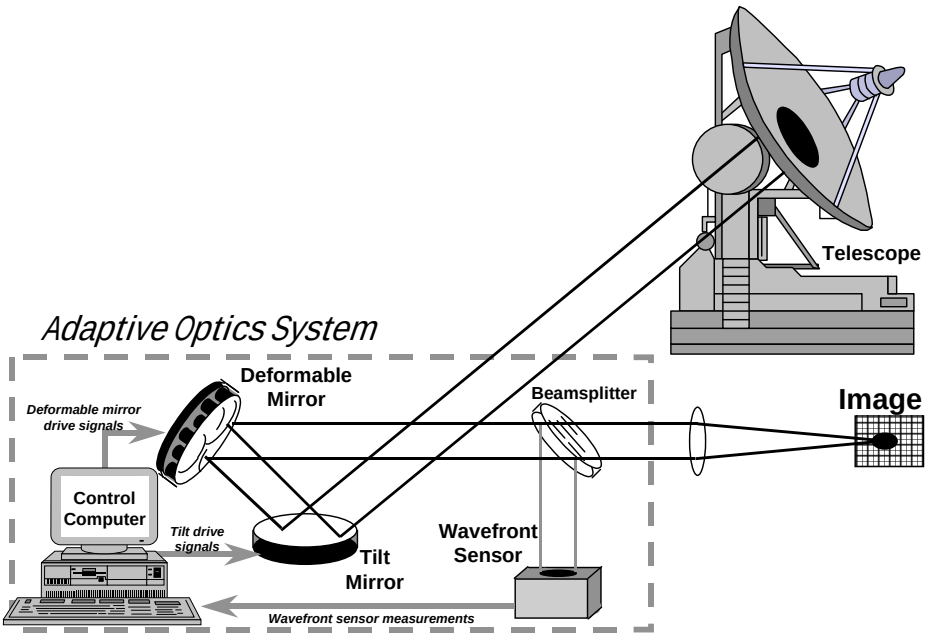

Fig. 1.1 A conventional adaptive optics system contains a deformable mirror, a wavefront sensor, and a control computer.

In the past 25 years adaptive optics systems have evolved. The most common setup with all the important parts is shown in Fig. 1.1.

This book will deal principally with "conventional, linear, inertial" adaptive optics. This is in contrast to "unconventional" systems which don't use the basic elements of Fig. 1.1, "nonlinear" optics phase conjugation systems which exploit the nonlinear characteristics of some optical materials to perform phase conjugation, and "non-inertial" systems, which is a misnomer. Inertial AO systems make use of mechanical motion (inertia) of some device to alter the optical phase. Non-inertial AO systems move something on a molecular scale, like inside crystals, to alter the phase. Although the molecular motion is really inertial (just ask the neighboring molecules) it has become commonplace to refer to these systems as non-inertial.

All systems of adaptive optics generally use the principle of phase conjugation. An optical beam is made up of both an amplitude A and a phase ϕ component and is described mathematically by the electric field $A\exp(-i\phi)$. Adaptive optics reverses the phase to provide compensation for the phase distortion. The reversal of the phase, being in the exponent of the electric field vector, means changing the sign of the term behind the imaginary number. This mathematical conjugation

corresponds to phase conjugation of the optical field, just what is needed to compensate for a distorted phase. More on that later.

The AO system in Fig. 1.1 constrains three principal elements which are the crux of all modern technology AO systems. These elements are the wavefront sensor, the deformable mirror, and the control computer. Most systems used in practice contain various supporting subsystems such as the wavefront divider, shown in Fig. 1.1 as a beamsplitter, an auxiliary tip/tilt/jitter control system, shown as a tilt mirror, and various other optical elements such as the collecting telescope, the imaging optics or science camera, pupil reimaging optics, or lasers and transmitter optics.

While many systems in use today are for astronomical imaging, the use of adaptive optics for compensating laser beam transmission through the atmosphere was the early impetus for developing the technology. The three principal elements have technologically evolved over the past 20 years to devices that are somewhat standard these days. The wavefront sensor is often a "black box" that accepts a beam of light and spits out electronic signals related to the phase of the incoming beam. The deformable mirror, generically speaking, is a mirror with a series of plungers or actuators on the back that deform the mirror (clever nomenclature) to a specified shape. The control computer is a signal processor that translates the wavefront sensor signals to deformable mirror commands to conjugate (compensate) the beam. Within the control computer there is a process often called "reconstruction" where the signals are used to reconstruct the phase of the beam from the wavefront sensor (WFS) signals.

Let me point out that contrary to the simplistic representation of Fig. 1.1, not all wavefront sensors are stand-alone devices looking at a sampled portion of an incoming beam, not all deformable mirrors are continuous faceplate mirrors, and many control computers do much more than simple translation of sensor signals. These differences and advances in technology can fill a book. This is the book.

For years, adaptive optics was under development. It still is. However, in the past few years, actual scientific results are being achieved with functioning AO systems. Examples of this abound in research publications, astronomy journals, pages on the World Wide Web, and various institutional and government reports. A few examples of recent successful imagery from operational AO systems are in Figs. 1.2 and 1.3.

The way an adaptive optics system works is fundamentally simple. (The implementation becomes the challenge and the array of sleepless nights.) For an imaging system, like that at an observatory, or a laser propagation system, like that in the movie *Star Wars*, we must measure the phase of the optical beam, determine how distorted it is, compute a compensation, apply the compensation

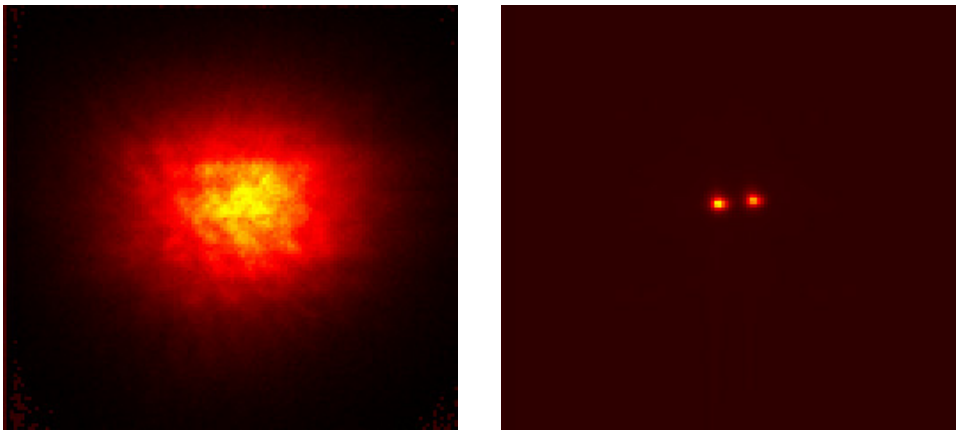

Fig. 1.2 Information improvement achievable with adaptive optics. The uncompensated image on the left is dominated by atmospheric turbulence. The image on the right, taken at the U.S. Air Force Starfire Optical Range, clearly shows the binary star.

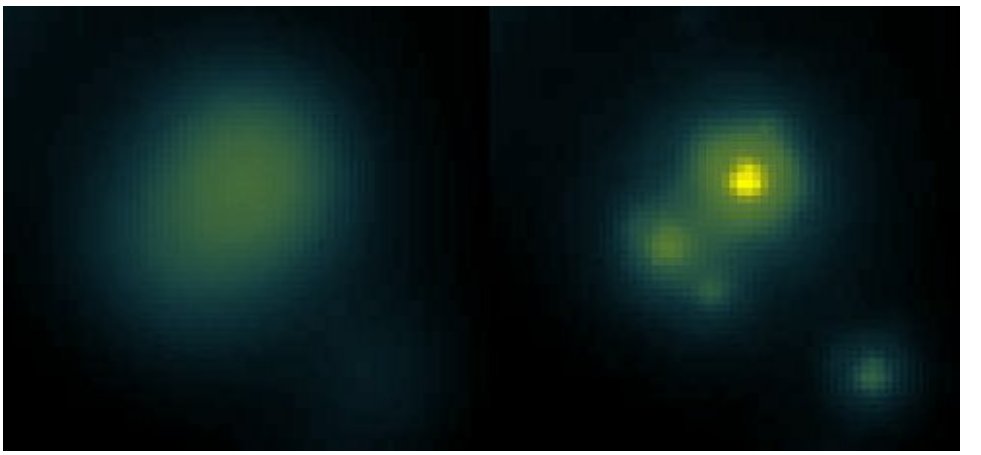

Fig. 1.3 Images taken at the Canada-France-Hawaii Telescope on Mauna Kea. The uncompensated image on the left barely shows any structure at all. With the adaptive optics system on, the multiple star image is clearly seen.

to the deformable mirror, back away and watch the fun and get ready to do it all over again when the phase distortion changes. The beam of light is represented as a wave with amplitude A and phase ϕ. The electric field has the form $U = A\mathrm{e}^{\mathrm{i}\phi}$. An adaptive optics control system will form the deformable mirror in the shape of the aberration, except with the opposite sign. The compensation that the deformable mirror imparts, in an ideal system, is the complex conjugate of the field or $A\mathrm{e}^{-\mathrm{i}\phi}$. (Thus the term *phase conjugation*.)

Because this process looks at a disturbance, applies a correction, and then looks again at the consequence of that correction, the system is termed "closed loop." This is contrasted to an open loop active system where a measurement is taken and a correction is blindly applied. Adaptive optics systems are almost always closed loop. In some circles, closed loopedness has become a defining characteristic of adaptive optics. While astronomers normally refer to their high bandwidth closed loop systems as "adaptive optics" and their low (<2 Hz) systems as "active optics," this book and a few other references make the distinction and definition between open loop and closed loop.

For atmospheric compensation systems, the incoming beam (think of it as a large sheet with wrinkles) is broken into small pieces, where the phase is measured. This is also like looking at an individual ray and figuring out its phase delay or path length from the source, with respect to the other rays. If the aperture over which we collect the rays is large, like an 8–10 m telescope, there are a lot of rays, and therefore a lot of wrinkles. We must break it into a lot of pieces to conjugate each piece. If the aperture is small, like the pupil of the naked eye, we can do a lot just by measuring the average phase of the light in the aperture and then conjugating it.

These requirements for spatial resolution have been investigated over the years and depend upon not only the size of the aperture but on the conditions and extent of the uncompensated distortion. Atmospheric turbulence becomes worse as the path through the atmosphere becomes longer. To avoid much of the atmosphere, some telescopes are placed on high mountains. The spatial resolution is represented by a term called seeing. Seeing is the atmospherically limited angular resolution of the optical system, given by the ratio of the optical wavelength and the coherence length of the atmosphere (also called r_0, "r naught," or Fried's parameter). The seeing of uncompensated telescopes on mountains (0.5 arcsec) can be more than 10 times better than the same telescope placed at sea level. Other conditions such as the zenith angle (the angle measured from zenith viewing) become important because the oblique path through the atmosphere is longer than the zenith path. An AO system at sea level would need to be sensed in about 100 pieces, or channels, to be able to compensate the system to a near-diffraction-limited 0.5 arcsec. The average amount of aberration from one side of the aperture to the other forces the deformable mirror to be able to move at least 1.5 microns or so. On Mauna Kea, Hawaii, where many of the

world's large telescopes are located because of its 14000+ ft altitude, the requirements differ. The telescopes are large and they can collect a lot of light, but a 10-m system like the W.M. Keck telescope requires thousands of channels to improve the system from its ambient 0.5 arcsec to the 0.01 arcsec visible diffraction-limited imaging system desired by the astronomers.

Spatial resolution, or number of channels, is not the only requirement. Remember, we must measure the phase or the wavefront, convert it into deformable mirror (DM) signals, drive the DM, and get ready do it all over again. To provide useful compensation, the wavefront must be measured accurately. To do this, the wavefront sensor must see a source of light called a beacon or a guide star that passes through the aberrating atmosphere. The beam should be in the direction of the object being observed, and it should be bright enough to be broken into many channels, leaving enough light for the science image camera, and still be able to provide enough photons for the sensitive detectors that will be used to measure the wavefront. Even if the wavefront sensor has enough light for all the channels, does it have enough in a short enough period of time to make all the phase changes before the atmosphere changes again?

For modern detectors, the sensitivity is good enough that signals can be computed in a few milliseconds. The atmosphere is changing at about 30 Hz [give or take a lot, depending upon high altitude winds or the telescope slew rate for imaging low-earth-orbit (LEO) satellites]. The computer must translate the sensor signals in a few milliseconds as well. Each of the actuators of the DM must then move to the proper position in a few milliseconds or so. All these delays, sometimes referred to as latency, result in a final control rate of about 300 Hz. This rate, being about 10 times the rate of the disturbance, is a rule of thumb applicable to AO. The disturbance changes only a little bit (about 1/10) before another measurement is taken. This way, the AO compensation system is sensitive enough and fast enough to keep up with the changing atmosphere. Unfortunately, books about adaptive optics do not have the luxury of high bandwidth, and therefore some of the information is a little outdated by the time the book goes to print and millions of adaptive optics groupies get to read it.

The requirements stated above are not etched in stone. They are not even etched very well in papyrus. Requirements for each system depend upon the seeing conditions, the wavefront beacon, the mission of the AO system, the technology available, and many times, the internal politics and generosity of the funding sources. Good, fast, and cheap – you get to pick only two. That cliché applies to adaptive optics.

Adaptive optics has a long and storied history. Well, maybe not as long as Egyptian civilization or as storied as the British empire, but certainly longer than Hula Hoops and Beanie Babies. Figure 1.4 shows an adaptive optics history tree with its roots being the people and institutions that developed the technology and the leaves being the operational AO systems that are the beneficiaries of the more than 30 years of technology development.

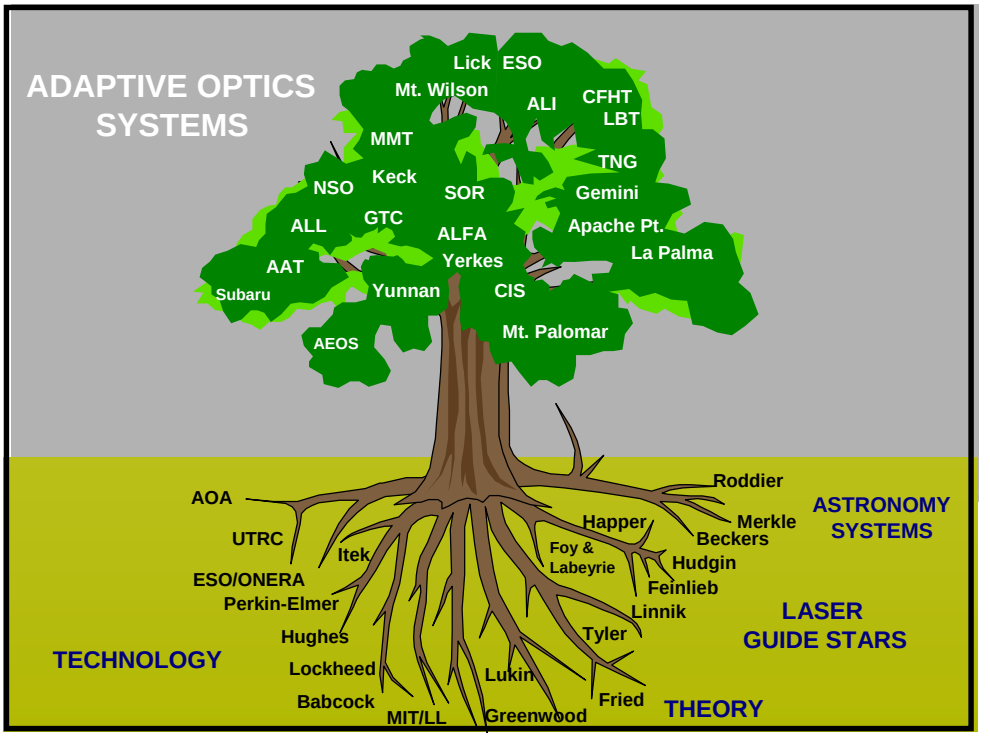

Fig. 1.4 The history of adaptive optics springs from roots in the aerospace industry and theory to many developmental and operational systems today.

The roots of AO history are within the US defense community and the international astronomy community. While US defense contractors were scurrying for 20 years to figure out a way to transmit huge quantities of photons from high energy multimegawatt lasers through the atmosphere to strategic targets, the astronomers were scurrying, somewhat more slowly, to figure out a way to keep the atmosphere from messing up the few photons that nature afforded them.

Both missions had similar problems to solve. The first published paper by Horace Babcock in 1953 paved the way by describing how the problem might be solved.[1] Although Babcock's idea was not placed directly into practice, it is considered the "first published paper on adaptive optics." Theoretical work by David Fried and Darryl Greenwood provided the seminal papers[2] with which to define the spatial and temporal requirements of the AO system (how many channels and how fast). Theoretical work by Lukin[3] and others in the Soviet Union paralleled much of the American work. It is generally known that neither Fried, Greenwood, nor the other theoreticians and experimenters of the time knew of Babcock's paper, so there honestly can be considered many "inventors" of adaptive optics.

American contractors in the optical system arena, such as Hughes, Lockheed, Itek, United Technologies (at the Research Center in Hartford CT, Adaptive Optics Associates in Cambridge MA, and Pratt and Whitney in Florida), and Perkin-Elmer (now Raytheon Optical Systems), worked to solve the technology problems of sensing the rapidly moving atmosphere, rapidly calculating the correction signals, and building deformable mirrors to apply the correction. In addition to the low power developments and demonstrations, where the driving constraints were sensor detectivity, processing speed, and actuator electromechanical limits, transfer of very high powers compounded the problem.

Sending enough energy to a boost-phase missile halfway around the world to destroy it midflight required transmitted energy much greater than the laser energy necessary to do the job. The transmitted energy would melt the very optics used for transmission. Before the development of very high power multilayer dielectric coatings, cooled mirrors were needed. These required complex fluid control systems that by themselves imparted vibration into the optical train and subsequently caused the beam to jitter around and spread out the energy.

Three concepts were employed for study and development of strategic missile defense systems. A ground-based system uses large lasers and telescopes to project the beam to orbiting relay satellites for eventual transmission and focusing on the target missile. This weapon system required adaptive optics to help get the beam to the relay through the atmosphere to maximize the deposited energy. An airborne system uses similar adaptive optics concepts to transmit energy from an onboard laser through the ensuing atmosphere between the aircraft and the target. Being airborne introduces its own set of problems such as light-weighting, crew safety, and target tracking difficulties. A space-based

[1] H. W. Babcock, *Publ. Astron. Soc. Pac.* **65**, 229, 1953.
[2] D. L. Fried, *J. Opt. Soc. Am.* **56**, 1380, 1966; and D. P. Greenwood, *J. Opt. Soc. Am.* **67**, 390, 1977.
[3] V. P. Lukin, *Atmospheric Adaptive Optics*, SPIE Opt. Engr. Press, Bellingham, WA, 1995, translation and update of the original published in 1986 by Nauka, Novosibirsk.

system seems to avoid the problem of the propagation through the atmosphere, but because of severe weight and volume issues, the self-contained laser will be vibrating, with subsequent optical aberrations imparted onto the beam. These aberrations can be measured right at the outgoing aperture and correction signals applied to the beam by an internal deformable mirror. Developers from the 1960s to 1980s addressed these problems with a series of programs, some geared to weapon system issues but most of them to technology issues.

Meanwhile, developers of the ground-based system had the problem of sensing the atmosphere out to a rapidly moving target like a relay satellite. The "point-ahead or lead-ahead problem" is stated thus: If a wavefront beacon is attached to or near the target, it must be connected far enough out in front of it so that the light from the beacon passes through the same column of air through which the upward laser beam must pass. For a LEO satellite, this distance is a few meters. For a relay satellite at geosynchronous orbit, the lead-ahead can be up to 60 m. For a ground-based satellite imaging system with adaptive optics the target's ambient reflected sunlight might not be bright enough to be used as a wavefront source. Or, for practical reasons, the beacon may not be possible. (One cannot always require another nation to place beacons on their satellites so that we might observe them. Can you say "secret?")

Investigators Julius Feinlieb and Richard Hudgin (Hutchin) and others were studying the problem. With various proposals, the Feinlieb concept caught on.[4] The idea was to shine a bright laser into the atmosphere and have enough laser light backscattered to provide a sufficient wavefront source anywhere in the sky.

Early on, people were skeptical. The laser must be sufficiently low that there is enough air, quantitatively measured by atmospheric density, to result in a large Rayleigh backscatter of the upcoming beam. But if the artificial beacon is too low, the atmosphere above it and that outside the cone between the focused beam and the receiving aperture are not sensed. A rapid series of studies in the early 1980s indicated that there was sufficient backscatter from about 20 km altitude, with available lasers, to create an artificial "laser guide star."

During the briefings on the guide star development, which remained classified by the US Department of Defense, Will Happer suggested that there might be sufficient atomic sodium in the mesosphere[5] (a layer 90–94 km high where mesos must live) that will allow resonant backscatter to create an artificial "sodium guide star." The difference between the two concepts was fundamental. A laser guide star in the lower atmosphere could be made with almost any wavelength. Lasers were available for many wavelengths. Thus, Rayleigh guide stars were rapidly possible.

[4] J. Feinleib, Proposal 82-P4, Adaptive Optics Associates, Cambridge, MA, 1982.
[5] W. Happer, G. J. MacDonald, C. E. Max, F. J. Dyson, *J. Opt. Soc. Am. A* **11**, 263, 1994.

A laser guide star from the sodium layer relied on resonant backscatter. That is, the laser light must be absorbed and then re-emitted at the resonant wavelength of the sodium atoms, about 589 nm. Lasers for this wavelength were not readily available in the early 1980s, and when they did show up at the nearest neighborhood laser lab, they didn't have enough power to get a good return. Thus, the search for efficient lasers with the right wavelength took off.

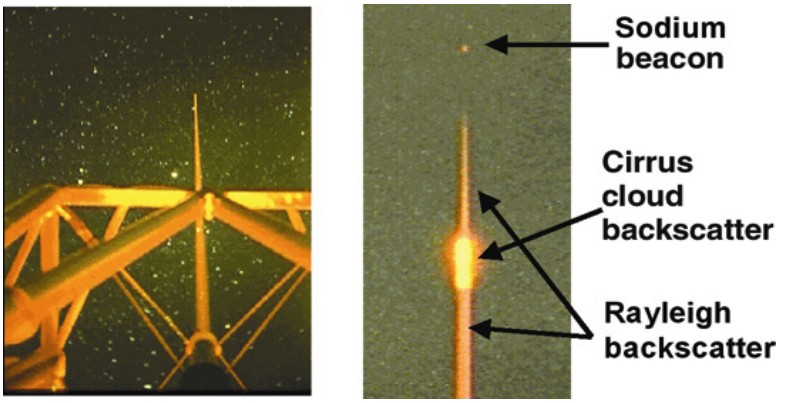

Fig. 1.5 Laser guide star produced at the Steward Observatory 6.5-m Multiple Mirror Telescope. Photos taken by D. W. McCarthy show both Rayleigh backscatter and sodium resonant scatter. Courtesy of the Center for Astronomical Adaptive Optics, Tucson, Arizona.

While the US Air Force and its contractors were carrying out experiments to prove and demonstrate the laser guide star concept for high energy laser propagation and imaging of uncooperative satellites, astronomers Foy and Labeyrie published a paper in 1985 proposing that same Rayleigh laser guide star concept,[6] which up to that time was a deeply guarded US military secret. Astronomers in the United States went to their funding sources, primarily the National Science Foundation, to gather research dollars for laser guide star research. The astronomy programs office, with staff who had been briefed by military researchers, realized that funding a program to look for answers that already were known was not a worthy way to spend precious research money. So by 1990, with the breakup of the Warsaw Pact countries, commonly called "the end of the cold war," the threats perceived at the end of World War II were fundamentally ameliorated. The Air Force, after serious investigation and proceeding along established lines for declassifying military information, released the bulk of the results of laser guide star research. The cat was out of the bag. European astronomers at the European Southern Observatory (ESO), led by Merkle and Beckers,[7] and Roddier at the Canada-France-Hawaii Telescope

[6] R. Foy, A. Labeyrie, *Astron. Astrophys.* **152**, L29, 1985.
[7] F. Merkle, J. M. Beckers, *Proc. SPIE* **1114**, 36, 1989.

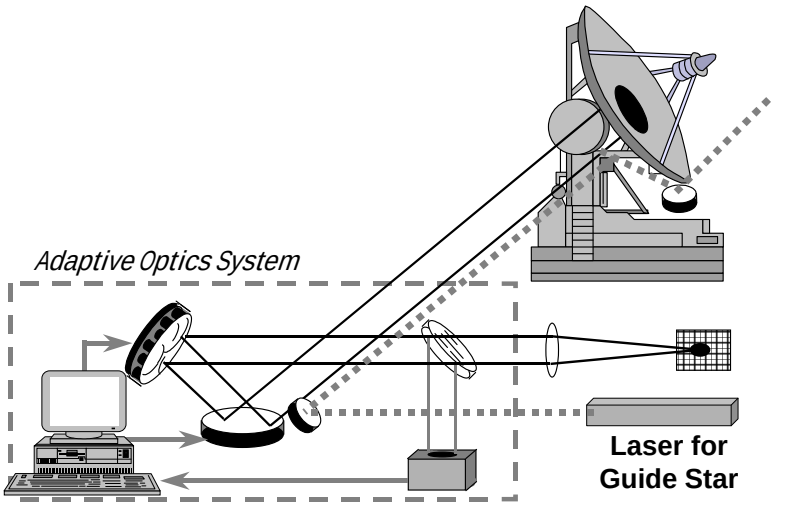

Fig. 1.6 A laser guide star adaptive optics system can incorporate a separate aperture for the projection of the laser.

(CFHT),[8] as well as US astronomers, all began plans for establishing guide stars for their adaptive optics systems in the new telescopes. By the mid-1990s, virtually all large telescopes (greater than 3 m) either were going to be retrofitted with adaptive optics or have integrated adaptive optics systems in their design stage.

Although the history is muddled, and like in many other fields there has been some controversy, the result is pretty much known. The military-industrial complex in the United States and independent contributions from France set the stage for development and implementation of adaptive optics around the world. The artificial guide star concept was not only an invention of Feinlieb or Foy and Labeyrie but most likely can be attributed first to Vladimir Linnik of the Soviet Union. His paper, published in 1957 but translated to English and disseminated only in 1992, described a high altitude beacon that could be used for wavefront sensing. Because Linnik's paper[9] precedes the invention of the laser by a few years, most credit him for describing the artificial guide star but not the laser guide star.

In the past 10 years, the roots of the tree of adaptive optics have blossomed into the leaves of operational systems used around the world. Missions for adaptive optics include everything from high energy laser propagation to deep space

[8] C. Roddier, F. Roddier, J. Demarcq, *Opt. Eng.* **28**, 66, 1989.
[9] V. P. Linnik, *Opt. Spectrosc.* **3**, 401, 1957.

astronomy, solar astronomy, surveillance, ground-to-space communications, and even "throw-away" systems in tactical weaponry.

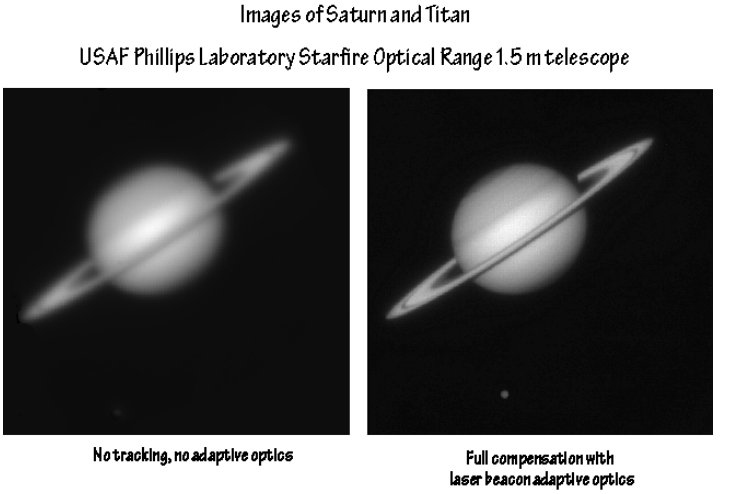

Fig 1.7 Images of Saturn from the Starfire Optical Range. The uncompensated image is on the left. On the right, the laser beacon adaptive optics system resolves the moon Titan.

The three principal components of an adaptive optics system, the wavefront sensor, the deformable mirror, and the control computer, are integrated with the telescope and optical system designed for the mission. History has seen a number of concepts and developments of each of these subsystems, driven mostly by the technology of the time. Clearly, technology has changed from 1965 to the present. Because computers and sensors and actuators are the heart of the system, the evolution of adaptive optics has followed and benefited from the startling advances of the components. No one, in 1965, could have predicted which types of sensors, mirrors, or computer control systems would survive as the preferred methods in 1999. The fallout from the early developments has resulted in a few designs that have been proven to be reliable and show high performance, with the cost coming down.

In wavefront sensors, for example, three basic configurations are now in use. Two of them (the shearing interferometer and the Shack-Hartmann sensor) measure wavefront slopes or first derivatives of the phase at various points around the optical pupil. The third type, the aptly named curvature sensor, measures wavefront curvature which is the second derivative of the phase. The three types make their measurements somewhat differently. The Hartmann sensor breaks the pupil or aperture into subapertures and measures the location of the image centroid of the subapertures in the image plane. The curvature sensor also uses the image plane, but it measures the difference in the intensity between two

slightly misfocused images. The shearing interferometer splits the beam in the pupil plane and measures the difference (the interference) between its phase and its own reference phase with a lateral or radial shift (the shear). The measurement is made in a subaperture defined by the detector area and the shift in either of two orthogonal directions. Each of the designs has various advantages and disadvantages regarding noise, measurement accuracy, sensitivity, and interface issues pertaining to the control computer and deformable mirror. It is not the purpose of this book, either through the author or the publisher, to recommend one specific type of wavefront sensor over another, but hopefully the discussion presented in chapter 7 can make it suitably less confusing so that engineering concerns can overcome preconceived notions.

A similar situation occurs in the evolution of deformable mirrors. The first such devices, built in the early 1970s, used an array of small mirror segments closely packed together, with electromechanical actuators behind each segment. This provided an active, controllable surface for wavefront control. These segmented mirrors have a number of advantages that continuous faceplate mirrors lack. For instance, the segments can be manufactured to tight tolerances, and each segment acts independently, removing concerns about coupling between adjacent regions of the deformable mirror. When a segment is optically registered with a subaperture of a shearing interferometer, for example, the control computer is simplified because of the zero coupling between subapertures and segments. Each segment can have up to three degrees of freedom for its motion, which can account for tip, tilt, and piston motion within a subaperture. A mirror with this configuration thus can control a number of channels of compensation, up to 3 times the number of segments. Because the segments are separate, some difficulties arise. The gap between segments can have an adverse effect on the optical beam because its regular pattern acts somewhat like a diffraction grating by imparting diffractive modes into the beam.

To address the optical problems with some segmented mirrors, the continuous faceplate deformable mirror was made. A mirror with good optical surface can be backed with an array of actuators. In this manner, the gaps are eliminated. The drawback, of course, is the result that the actuators are mechanically coupled. When one actuator moves, there is some finite response or influence along the entire surface of the mirror. Although these coupling effects can be passively controlled with judicious electromechanical design and manufacturing, in reality each mirror, and each actuator within each mirror, has some unique influence function which must be accounted for in the control system design. Because similar actuators can be packed as closely on a segmented mirror as they can on a continuous faceplate mirror, there is no advantage, at least in terms of spatial resolution, of one over the other.

A third active mirror type is made up of two different forms of electrostrictive material. Using the Greek word for form, or *morph*, the mirrors are cleverly

called bimorphs. When a voltage is applied to the material, the mirror, placed or polished on one side of the bimorph, is deformed. The optical surface naturally forms the second derivative, or curvature shape, which makes it very easy to integrate with a curvature sensor, as Roddier has done with the systems at the University of Hawaii.

The third major subsystem of adaptive optics is the control computer. It is fundamentally simple. It must take in the electrical signals from the wavefront sensor, which are proportional to the slope or curvature of the wavefront, and convert them to the compensation signals for each actuator of the deformable mirror. The calculation must be done faster than the wavefront is changing. The delay time, or latency, of the calculations is critical to maintaining the closed loop bandwidth. For a system like the one at the Starfire Optical Range (SOR) with over 1400 slope signals and a deformable mirror with 1000 actuators, the computer has a big job of doing the calculations in a few milliseconds to keep up with the changes of the atmosphere. Most high performance systems require dedicated digital signal processing (DSP) computers to carry out the computational tasks. Experimental systems, or those with low bandwidths and a few correction channels, can often make do with multipurpose computers and clever software.

Not all systems require a lot of channels and a high bandwidth. Considerable wavefront improvement can be achieved with just a few low-order correction modes, like tilt, focus, and astigmatism. Most of the wavefront error in the atmosphere is found in the tilt mode. The image dances around. If we just compensate for the position of the image, i.e., stabilize it, the image from a long exposure (>0.1 s, for example) can reveal much more detail than that found in the blurred image.

The spatial and temporal requirements for an adaptive optics system can be very complex. They depend not only upon the atmosphere and the telescope site conditions, but also upon the direction of the object as it moves across the sky. They depend upon the wavelength of observation and the size of the telescope. They also depend upon the intricate coupling between the number of channels of correction, the computer speed, and the brightness of the wavefront beacon.

For astronomical imaging or propagating a laser up through the atmosphere, the first-order effects are encompassed in a few rule-of-thumb equations:

$$r_0 = \left[0.423 \ \ k^2 \ \ \sec\beta \int_{\text{Path}} C_n^2(z) \ \ \mathrm{d}z \right]^{-3/5} , \qquad (1.1)$$

$$\theta_0 = \left[2.91 \;\; k^2 \;\; \sec^{8/3}\beta \int_{\text{Path}} C_n^2(z) \;\; z^{5/3}(z) \;\; \text{dz} \right]^{-3/5} , \qquad (1.2)$$

$$f_G = 2.31 \;\; \lambda^{-6/5} \left[\sec\beta \int_{\text{Path}} C_n^2(z) \;\; V_{\text{Wind}}^{5/3}(z) \;\; \text{dz} \right]^{3/5} , \qquad (1.3)$$

where β is the zenith angle. All of these relationships depend upon the atmospheric structure constant C_n^2 and the propagation path. The structure constant is a measure of the strength of turbulence. It is dependent upon altitude and it changes with temperature. Thus, the "constant" is only in a mathematical sense. It is far from a constant constant. It changes hourly, and even minutely, as well as greatly varying from day to night, month to month and throughout the year. Fried's coherence length, r_0, is a single parameter describing the spatial extent of a "turbulence cell." David Fried, in a seminal paper in 1965,[2] described the limiting distance, or size of an aperture, that would successfully pass a coherent beam. The distance is on the order of a few centimeters for visible wavelengths and goes by the 6/5 power of the wavelength for other colors. See Eq. 1.1 and recall that the wavelength is incorporated into the wave number *k*. The atmospheric "seeing" is often defined by the angle λ / r_0, often expressed in arcseconds. (4.9 μrad = 1 arcsec).

Another parameter very important to the adaptive optics requirements problem is the isoplanatic angle θ_0. Because the atmosphere is different as we look through

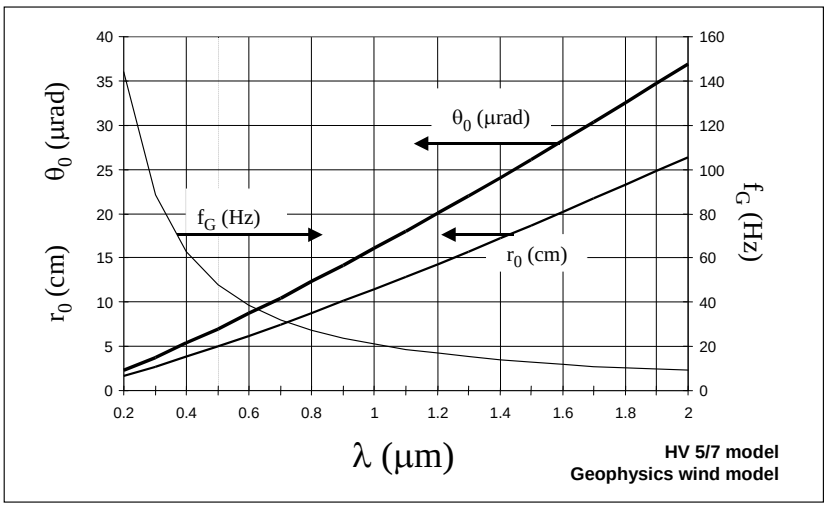

Fig. 1.8 The relationship between wavelength and coherence length, isoplanatic angle, and Greenwood frequency.

it in different directions, it is important that the atmosphere through which we are making measurements and determining the wavefront is the same as that through which we are looking. When they are different by a small angle we notice anisoplanatic effects. (The word *anisoplanatic* is an interesting construct. If we imagine the sky to be a plane that is perpendicular to the direction we are looking, the atmospheric turbulence is a collection of wiggles in the plane. If the wiggles were the same everywhere along the plane, we would call it isoplanatic, but since the atmosphere changes in only a few centimeters, it is far from isoplanatic; the atmosphere is anisoplanatic.) The maximum angle that we can look away from our object point and still measure the correct wavefront is the isoplanatic angle, θ_0. It is a function of the atmospheric structure constant *and* the altitude of the turbulence, all supported by Eq. 1.2 for θ_0.

Another parameter represents the temporal characteristic of the atmosphere. The Greenwood frequency f_G, calculated from Eq. 1.3, is the characteristic frequency of the atmosphere. This does not mean that all the atmospheric turbulence moves at this frequency. Big clumps of air move around slowly, little clumps much more quickly. For computational purposes, the Greenwood frequency helps to define how well an adaptive optics system keeps up with the ever-changing turbulence.

An adaptive optics system is not perfect. It cannot correct for all the little turbules in the air. It can't correct for turbulence it doesn't see and it can't completely keep up with the rapid changes of the turbulence. A few scaling laws have been developed for seeing how well the system works. The fitting error depends upon the number of channels in the adaptive optics system, the size of the aperture to be compensated, and the atmospheric conditions. Robert Noll in a paper in 1976 described how the residual error of an adaptive optics system is reduced by compensating spatial modes.[10] The expression for wavefront fitting error, actually the wavefront variance due to fitting, arising out of that paper is

$$\sigma_{\text{fit}}^2 = 0.29 \quad N^{-\sqrt{3}/2} \left(D \Big/ r_0 \right)^{5/3} \quad , \tag{1.4}$$

where D is the diameter of the aperture. Although Noll refers to N in the equation as the number of "Zernike modes" compensated, for our purposes, when N is greater than 7 or so, it can also mean number of channels of compensation.

Another scaling law comes from an analysis of anisoplanatism. The wavefront variance due to the misalignment of the wavefront beacon and the direction of the object is related to the ratio of the actual misalignment and the isoplanatic angle:

[10] R. J. Noll, *J. Opt. Soc. Am.* **66**, 207, 1976.

$$\sigma_{\text{iso}}^2 = \left(\theta / \theta_0 \right)^{5/3} \quad . \tag{1.5}$$

The temporal error scaling law follows a similar derivation and ends up with a similar-looking equation:

$$\sigma_{\text{temp}}^2 = \left(f_G / f_{\text{BW}} \right)^{5/3} \quad , \tag{1.6}$$

where f_{BW} is the control bandwidth. Temporal error is the residual wavefront error variance due to the system not being able to keep up with the changing atmosphere.

These scaling laws neglect second-order effects. For a detailed system design and analysis, one must consider a number of engineering and physical concerns. For example, if the wavefront beacon is not bright enough to illuminate all the subapertures in the wavefront sensor, we can back off on system performance. We can reduce the number of subapertures, or the number of correction modes, increasing the fitting error, or we can increase the integration time on the wavefront sensor detector, decreasing the bandwidth. Or we can do both. We must consider effects like noise on the detectors, since they corrupt the precision measurements of the wavefront. We have to consider the cone effect if the wavefront beacon is at a finite distance and does not deliver a plane wave into the aberrating atmosphere.

The final decisions on system performance will be driven by necessary system requirements (what you want to do), the technology (what you can do), system complexity (what you are willing to give up in terms of a meaningful social life), and cost (what you are willing to beg for).

Bibliography

Larry C. Andrews and Ronald L. Phillips, *Laser Beam Propagation through Random Media*, SPIE Optical Engineering Press, Bellingham, WA (1998).

Robert K. Tyson, *Principles of Adaptive Optics*, 2nd ed., Academic Press, Boston, (1997).

Michael C. Roggemann and Byron Welsh, *Imaging through Turbulence*, CRC Press, Boca Raton, FL (1996).

L. A. Thompson, "Adaptive optics in astronomy," *Phys. Today* **47**, (12), pp. 25-31, (Dec 1994).

H. W. Babcock, “The possibility of compensating astronomical seeing,” *Publ. Astron. Soc. Pac.* **65**, pp. 229-236 (1953).

H. W. Babcock, “Adaptive optics revisited,” *Science* **249**, pp. 253-257 (1990).

Multiple authors, *Special issue on adaptive optics*, Lincoln Laboratory J. **5**, (1) (Spring 1992).

J. W. Hardy, “Adaptive optics – a new technology for the control of light,” *Proc. IEEE* **66**, pp. 651- 697 (1978).

J. W. Hardy, “Adaptive optics,” *Sci. Amer.* **270** (6), pp. 60-65 (June 1994).

Chapter 2

Adaptive optics systems — *optics* is our middle name

Adaptive optics systems are, by definition, closed loop. This means, among other things, that any change in the optical beam performed by the AO system is sensed by the system. Consider the optical system pictured in Fig. 2.1. This system took eons to develop and probably could not have been developed so elegantly without divine intervention. Most of us possess this basic slow AO system. The eye works as a detector, the brain as a computer, and the arm as an actuator. While we all see what is being done in this system, and some of us experience it often, it is not simply reflex, an open loop operation. Our eye senses an image and transmits the information to the brain. The brain, through years of training, knows pretty well that an image is not focused correctly. We often can't define blurry, but we know it when we see it. If the image is blurry, our nervous system instructs the muscles in our hand to change the focus setting on the optical device. The focus control, acting like a deformable mirror, adjusts the focus of the beam and therefore the appearance of the image to our eyes. We sense the image getting better or worse, i.e., closing the loop. If the image gets worse, we stop our adjustment and reverse its direction. We eventually zero in on an image in its best focus.

For those who wear eyeglasses, the analogy continues. If we focus the binoculars while not wearing the eyeglasses, the binoculars compensate for the aberrations in our eye that are not corrected. Thus, the AO system compensates for all the aberrations in the system. It really doesn't know, and doesn't particularly care, where they come from. If the object moves away, like that of a rapidly moving car, we must constantly readjust the binoculars to keep the image clear. If, however, we happen to have binoculars with a large depth of focus, the adjustment doesn't have to be made often, nor very rapidly. The integrated adaptive optics system depends upon all the optics around it, the type and magnitude of the aberrations, the speed of the sensor, and the quick-wittedness of the control computer. Although large telescopes require hundreds of channels of compensation with a few hundred hertz bandwidths, our human AO system gets by with about 3 channels and 2 Hz – a good start, but still requiring a little help from technology.

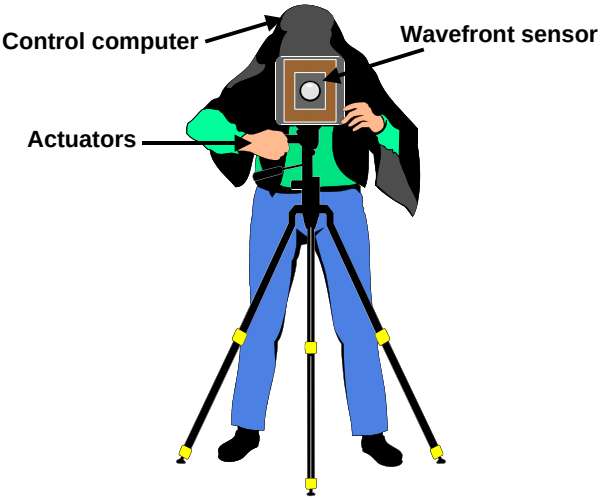

Fig. 2.1 A biophysical adaptive optics system.

Physical/Wave Optics as it applies to Adaptive Optics

Adaptive optics is a subfield of optics. The optics that we care about here are physical, or wave, optics. If we neglect the wave nature of light, we can often see how light performs by considering the path of those little particles of light, photons. Photons, as we perceive them, travel in straight lines, assuming they are going through something with a constant index of refraction. A constant index of refraction means *no* refraction. In other words, the stream of photons travels along a straight line called a ray. That is ray optics. When the ray hits a medium with a different index of refraction, the ray changes its direction according to Snell's law. When the ray hits a solid object (optically opaque), it just stops.

However, rays don't have all the properties of a wave moving through space. The "phase" of a ray is hard to define. Sometimes we think of the phase of a ray as the total length of the ray. If two rays come into contact, the difference in the ray path length between the two rays is basically a phase difference. The "amplitude" of the ray is the amount of photons that make up the ray. In most instances, ray optics approximations are not sufficient for describing adaptive optics. Because adaptive optics works by altering the phase of a beam of light, we need to know and understand wave optics.

Going back to basics, consider a beam of light propagating from a source at the left of the diagram in Fig. 2.2. Light beams *always* propagate left to right – except when they go the other way.

Assume that the source of coherent light is an infinite distance away. It is not the nature of this book to define coherence, partial coherence, or anything of the sort. A good science dictionary should be kept by your bedside. Suffice it to say, for purposes here, coherent light is one wavelength (color).

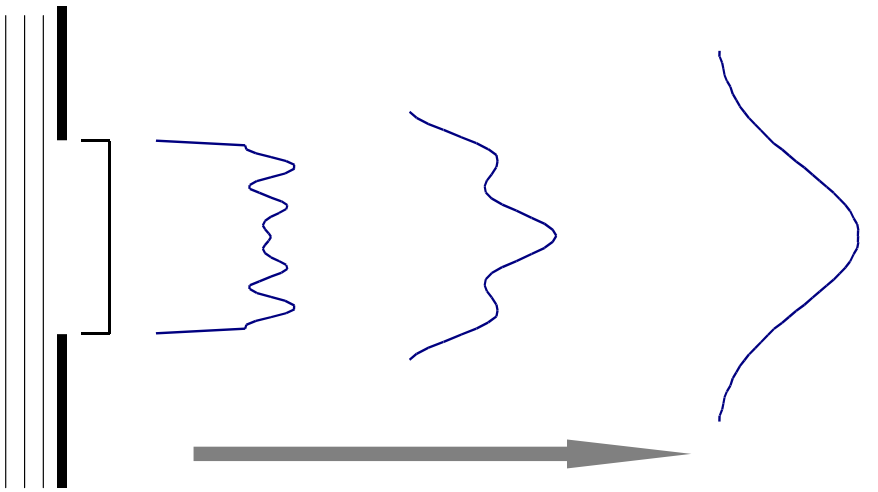

Fig 2.2 Wave optics explains diffraction, "ringing" in the intensity distribution as a wave propagates, and the far-field pattern.

It will take a long time to get to infinity, so we put in a lens to make the light think it is an infinite distance away. When the light goes through an aperture,[1] it continues on to a screen. If the light were just photons or rays and had no wave characteristics, the light on the screen would be just a shadow image of the aperture. But because the light is a wave, and because Huygen's principle says so, the light from one part of the aperture mixes with the light from all the other parts and, as it propagates, one gets an array of beamlets that seemingly interfere with each other. By the time the light gets to the screen, some of the light has spilled over the edges, and when the path distance from different parts of the aperture is an integral number of wavelengths, the beamlets can add up to produce light outside the shadow. This is diffraction.

We can calculate precisely what the wave looks like at any point on the screen at any distance from the aperture for any wavelength. The calculation involves solving the Rayleigh-Sommerfeld diffraction integral, an absolutely messy mathematical problem that has no general solution. For a few centuries, nobody seemed to care much about the solution because coherent beams produced by lasers didn't appear until the 1960s. However, there was some consideration for

[1] For this discussion, an aperture is just a hole in an opaque piece of material. We assume the light fills the aperture.

describing the physical wave nature of light, so very clever approximations were made to the diffraction integral.

The first approximation assumes that the screen is a long distance away from the aperture, meaning that the propagation distance is much greater than the size of the aperture or the size of the screen where we calculate the output wave. “Long” is related to the wavelength. For example, if we have a 1 cm size beam and we want to observe the output in an area around 1 cm on the screen, with a wavelength of 0.5 microns, the “long” distance is about 1.4 m.

The Fresnel approximation simplifies the Rayleigh-Sommerfeld so much that it gets its own name, the Fresnel integral. Although the math involved here is greatly simplified, it is not simplified greatly enough. A further approximation, that the propagation distance is really, really, a whole bunch greater than the size of the aperture, makes the integral simpler still. Since the field is calculated at a distance far away from the aperture, it is generally called the “far field.”

This latest propagation approximation, called the Fraunhofer condition, reduces the integral to the Fraunhofer integral. Fortunately, the Fraunhofer integral has an uncanny resemblance to the Fourier transform, which every second semester calculus student has loved and adored. The good news is that the Fourier calculation is very simple; it can be done manually for a whole lot of specific real conditions, and it can be done by computer for all the other, more complex conditions. In short, the Fraunhofer diffraction integral states that the “electric field,” the amplitude $A_1(\alpha,\beta)$, and the phase $\phi(\alpha,\beta)$ of the wave at the output is simply the Fourier transform of the electric field at the aperture (the pupil):

$$U_2(x,y) = C \iint_{\text{Pupil}} A_1(\alpha,\beta)\ \ \exp\left[-\mathrm{i}\phi(\alpha,\beta)\right]\exp\left[\frac{-\mathrm{i}k}{z}(\alpha x+\beta y)\right]\mathrm{d}\alpha\mathrm{d}\beta\ . \quad (2.1)$$

For many cases, the integral can be solved analytically, or in closed form. This means that we can write an equation for the answer that no longer has that pesky integral symbol. This does not rule out other pesky mathematical symbols in the solution, it just rules out integrals. For example, if the input field amplitude is a Gaussian, the output field is also a Gaussian. If the input field has a uniform amplitude over a circular aperture, the output field follows a Bessel function divided by its argument. Not something for “back of the envelope,” but not bad for hand calculators. The Fraunhofer diffraction patterns for many conditions, including phase differences, have been calculated and are shown in many books.

Strehl ratio

One of the well-known cases where Fraunhofer diffraction has an analytic solution is the circular aperture with a uniform intensity (amplitude) in the input

wave. Born and Wolf show how the peak intensity, on-axis, can be calculated from the Fraunhofer integral.[2] The peak intensity is

$$I_0 = P\frac{\pi D^2}{4\lambda^2 R^2} \tag{2.2}$$

where P is the total power (in watts) passing through the aperture, D is the aperture diameter, λ is the wavelength, and R is the propagation distance. There are no phase aberrations present. When there are phase aberrations, that is, when there is a difference between a perfect spherically focusing wavefront and the actual wavefront ("wavefront error"), the intensity on-axis in the far field is reduced. The general case is the general solution to the Fraunhofer integral, which has an analytic solution only for a few basic aberrations. However, Born and Wolf, again, show that we can calculate the reduction of peak on-axis intensity from knowledge of the wavefront error.

The *wavefront error* is the square root of the *wavefront variance* σ^2, which is the integral of the squares of the wavefront errors averaged over the whole aperture.

$$\sigma^2 = \frac{\iint_{\text{Aperture}} \left[\Phi(x,y)\text{-}\Phi_{\text{M}}(x,y)\right]^2 \,\mathrm{d}x\mathrm{d}y}{\text{Area of the aperture}} \cong \frac{\sum_{\text{All data points}} \left[\Phi(x,y)\text{-}\Phi_{\text{M}}(x,y)\right]^2 \Delta x \Delta y}{\text{Area of the aperture}} \tag{2.3}$$

where the wavefront is $\Phi(x,y)$ and the wavefront average is $\Phi_{\text{M}}(x,y)$. We assume here that the wavefront error σ is "small," meaning that σ^4 is negligible compared to σ^2, or about 1/6 wave.

When we know the wavefront error, we can calculate the on-axis intensity:

$$I = P\frac{\pi D^2}{4\lambda^2 R^2}\left[1 - \left(\frac{2\pi}{\lambda}\right)^2 \sigma^2\right]. \tag{2.4}$$

The reduction in the intensity, the expression in braces, is called the Strehl ratio S. Because the Strehl ratio definition[12] happens to have the form of the first two terms in the expansion of the exponential $\exp(-x^2)$, we can generalize the Strehl ratio definition to

$$S = \exp\left[-\left(\frac{2\pi\sigma}{\lambda}\right)^2\right]. \tag{2.5}$$

Thus, the bottom line is that we can represent how aberrated a beam is by knowing its wavefront error and expressing it in terms of the Strehl ratio. The peak on-axis intensity IS ALWAYS reduced by adding aberrations. If $\sigma > 0$, the Strehl is between 1 and 0 and the intensity is reduced. In an adaptive optics

[2] M. Born, E. Wolf, *Principles of Optics*, 5th Ed., Pergamon, Oxford, 1975.

system, where our objective is to reduce the wavefront error to better propagate a laser or transmit an image, we can say how well it performs by stating the residual Strehl ratio of the optical system.

If the wavefront error is constant but the beam is jittering,[3] we add another term to show the effect. With a single-axis angular root-mean-square jitter motion σ_{jit}, an rms wavefront error σ, a peak intensity I_0 from Eq. 2.2, and a total transmittance T, the intensity on-axis in the focal plane in the far field is

$$I_{FF} = \frac{I_o T e^{-\left(\frac{2\pi}{\lambda}\right)^2 \sigma^2}}{1 + \left(2.22 \ \ \sigma_{\text{jit}} D / \lambda\right)^2}. \tag{2.6}$$

Bibliography

Michael C. Roggemann and Byron Welsh, *Imaging through Turbulence*, CRC Press, Boca Raton, FL (1996).

Joseph W. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York (1996).

Raymond G. Wilson, *Fourier Series and Optical Transform Techniques in Contemporary Optics: An Introduction*, John Wiley & Sons, New York (1995).

Jack D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, New York (1978).

[3] Jitter is dynamic tilt.

Chapter 3

Speaking the language — a few definitions

Power in the bucket

When high energy lasers with the power to do serious damage to animals, plants, and minerals were being built, it was necessary to measure how much power was being deposited in a small region in the focal plane. The description generally grew to the point that one wanted to determine how much power could be deposited in a bucket of a certain size placed at the target. Thus rose the name *power in the bucket.*

In some cases the bucket was a physical entity, like a high power calorimeter. In the laboratory, it can be as simple as a specific size pinhole backed by an integrating detector. Without worrying too much about the details, it is just a well-known bit of trivia among optical engineers that "the power in a bucket (where the bucket is the size of the first Airy ring in the diffraction pattern of an unaberrated beam in a circular aperture) contains 84% of the total power." So it is written.

Zernike polynomials

Another less known bit of trivia is that Zernike polynomials provide a mathematically convenient way to describe the phase of an optical beam. We can represent phase in a number of ways. One way is just to say the beam has so much focus, so much astigmatism, so much coma, and so forth. Being scientists, and therefore creatures of mathematics, we decided that a series of functions with terms that represent those aberrations would be a useful thing to have around.

We can lay out the phase of an electric field (the beam of light) as a three-dimensional surface, the height of which is the phase advance (or retardation) of the beam. If the beam has just focus, we find that it has a different spherical shape than that which would propagate smoothly to a focus in the focal plane; the different, out of focus shape looks like a bowl. In the area near the axis, it has a parabolic shape, represented by the radial coordinate raised to the second power,

ρ^2. Spherical aberration, another radially symmetric term, also looks like a bowl but with steeper edges; the term that describes it is ρ^4.

When we start to incorporate the other aberrations, we have to include the effect of the azimuthal coordinate θ. Tilt, though technically not an "aberration," is represented by a plane that is not normal to the optical axis. (It is not an "aberration" because of the paradox that the line normal to the average plane of the phase is the direction in which the beam will go; thus it *is* the optical axis and the plane cannot be *not normal* to it.) Tilt is represented by the term $\rho\cos\theta$, or in the other, orthogonal direction, $\rho\sin\theta$. Other terms, such as astigmatism, have a form $\rho^2\cos^2\theta$ and are shaped like a saddle (or a potato chip). Coma, a word from the same root as *comet*, is a cubic term, $\rho^3\cos\theta$. The tear drop shape of the far-field distribution of a comatic beam is the reason for the "comet" word; the shape of the phase surface is more like a badly planned urban drainage system. We can continue this process up the food chain to higher and higher orders, many of which don't even have reasonable names.[1] To bring them all together, though, we write the phase Φ as a power series:

$$\Phi(\rho,\theta) = \sum_{n,m=0}^{\infty} K_{n,m,1}\rho^n \cos^m\theta + K_{n,m,2}\rho^n \sin^m\theta \qquad (3.1)$$

where the values of the polynomial coefficients (K) are different for different beams. This looks like a reasonable way to do things until the optical beam gets complicated. The problem with the above series is that it is not orthogonal or normalized over any specific aperture. Since optical beams are often circular, this would be nice. Because it is not orthonormal we run into problems. What if we said our phase was represented only by $K_{0,2,1} = 2.0$ and $K_{0,2,2} = 2.0$. The phase would then be

$$\Phi(\rho,\theta) = \sum_{n,m=0}^{\infty} 2\cos^2\theta + 2\sin^2\theta \quad . \qquad (3.2)$$

Because of the magic of trigonometry, this is the same as saying $\Phi(\rho,\theta) = 2$. In other words $K_{0,0,1} = 2.0$ and $K_{0,0,2} = 0.0$. We shouldn't be able to represent the same phase profile with a different set of coefficients. This whole problem is the result of the non-orthogonality of the power series representation.

[1] For a good description of Zernike polynomials with many of them visualized in surface and contour plots, see J. C. Wyant and K. Creath, "Basic wavefront aberration theory for optical metrology," Chap. 1 in *Applied Optics and Optical Engineering*, **Vol. XI**, R. R. Shannon and J. C. Wyant, eds., Academic Press, Boston, 1992.

Fortunately for us, Frits Zernike discovered, or otherwise created, a series representation so that phases could be represented by unique coefficients. The Zernike series is orthogonal over a unit circle (actually it is good over most any circle, but if the circle is radius = 1, then it is called a unit circle). No two aberrations can add up to a third like they did with the power series, and they are normalized so that the maximum of each term (except the n=0, m=0 piston term) is always +1, the minimum is always –1, and the average over the surface is always zero.

The Zernike series looks like

$$\Phi(\rho,\theta) = A_{00} + \frac{1}{\sqrt{2}}\sum_{n=2}^{\infty} A_{n0} R_n^0\left(\frac{\rho}{R'}\right) + \sum_{n=1}^{\infty}\sum_{m=1}^{n} \left[A_{nm}\cos m\theta + B_{nm}\sin m\theta\right] R_n^m\left(\frac{\rho}{R'}\right) \quad . \quad (3.3)$$

Okay, we see the ρ, and the θ, but what is the R_n^m term? To make the whole thing orthogonal over a circle (this time of radius R'), R_n^m is the radial Zernike polynomial, but it has a complicated definition and is very dependent upon the radial order n and the azimuthal order m:

$$R_n^m\left(\frac{\rho}{R'}\right) = \sum_{s=0}^{\frac{n-m}{2}} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!} \left(\frac{\rho}{R'}\right)^{n-2s} . \quad (3.4)$$

Even if you can figure out each of these polynomial terms, you still have to remember another rule: The Zernike series has terms only where n–m is even. That is, the terms that count are those like n=2,m=2 or n=3,m=1, and so forth. Terms such as n = 6, m = 3 don't exist. For example, the piston term $R_0^0 = 1$ and the tilt term $R_1^1 = \rho$ are okay, but there is no R_1^0 term.

Another important rule to remember is that each Zernike term of order n is composed of the highest power that n takes plus all the other powers of n of the same even- or odd-ness. In other words, the Zernike focus term n=2,m=0, which is

$$R_2^0 = 2\left(\frac{\rho}{R'}\right)^2 - 1 \quad , \quad (3.5)$$

contains the constant 1, or the piston term n=0,m=0. In a similar way, the Zernike spherical aberration term R_4^0 contains focus and piston. For the odd azimuthal terms the same rule holds. Zernike coma, n=3,m=1, contains the tilt term n=1,m=1. So, whenever someone says that the beam has 2 waves of coma, do they mean 2 waves of pure coma,

$$\left(\frac{\rho}{R'}\right)^3 \cos\theta$$

or two waves of Zernike coma, which would mean that there is also some overall tilt to the beam?

One of the good things about Zernike polynomials (and you were probably wondering if there were any) is that the phase variance, which is so useful to describing an adaptive optics system, is not a messy integral. When you describe the phase of the beam, or its wavefront, as Zernike coefficients, you can substitute them into a simple algebraic formula to get the wavefront variance:

$$\left(\Delta\Phi\right)^2 = \sum_{n=1}^{\infty} \sum_{m=0}^{m} \frac{A_{nm}^2 + B_{nm}^2}{2(n+1)} \quad . \tag{3.6}$$

But *be careful*: Whenever someone talks about Zernike polynomials, be aware that they may not be talking about the same series, exactly — there are at least three normalizations that are floating around out there. The $\sqrt{2}$ in Eq. 3.3 may be modified by an occasional n or $2(n-1)$, and that means that all the terms have another value. The series described in Eq. 3.6 was explained quite elegantly in the classic optics text *Principles of Optics* by Born and Wolf. The normalization that they follow is usually called the "Born and Wolf normalization." Another Zernike series, developed by Noll to describe atmospheric turbulence in terms of Zernike polynomials,[2] has a different normalization so that the simple variance equation becomes even simpler:

$$\left(\Delta\Phi\right)^2 = \sum_{n=1}^{\infty} \sum_{m=0}^{m} A_{nm}^2 + B_{nm}^2 \quad . \tag{3.7}$$

Another glitch has crept into the Zernike polynomial lexicon: Not everyone orders the polynomials the same way. What is Zernike polynomial 37, for example? Well, it depends. Some wavefront analyzers, like a Zygo interferometer, can calculate the Zernike polynomials for an input beam. Since the higher-order symmetrical polynomials n=6,m=0 or n=8,m=0 are important for optical propagation but the higher-order azimuthal polynomial n=8,m=6, a high-order astigmatism, does not have a big effect, the numbering includes all the lower order terms, sort of with increasing n, but then skips the higher-order azimuthal terms. Most confusing.

Rather than sort all this out here, just look at the references and make sure you know what Zernike series you are working on. All Zernikes are not created equal.

[2] R. J. Noll, *J. Opt. Soc. Am.* **66**, 207, 1976.

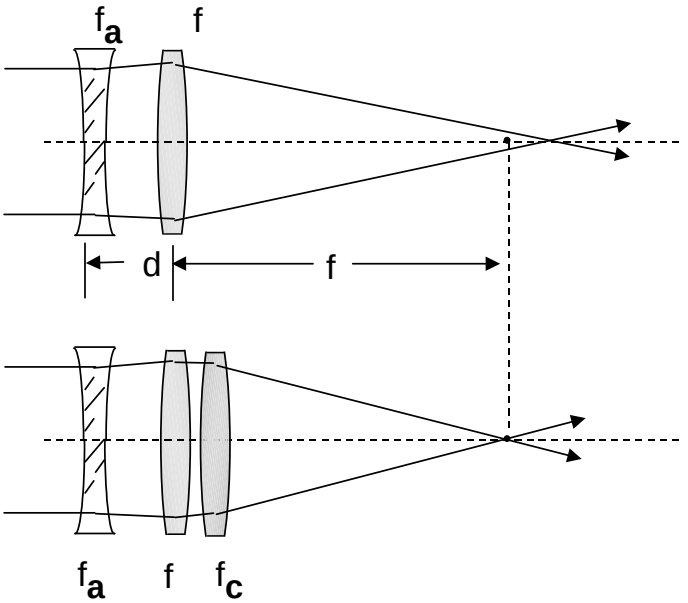

Fig. 3.1 Phase conjugation through a lens system. The aberrator changes the position of best focus. A conjugating lens, with the opposite aberrations of the aberrator, brings the focus back to the correct point.

Despite all this weeping and wailing, the Zernike polynomials are a valuable tool for describing adaptive optics systems. If we can measure the wavefront, which we will do with a wavefront sensor, we can determine the Zernike coefficients. These tell us the wavefront error directly and uniquely. Not a bad thing.

Phase conjugation

The crux of adaptive optics is phase conjugation. If we can measure the aberrations, then all we have to do is put the opposite aberration on the beam, at the right time, to cancel the aberrations.

The concept of phase conjugation is fairly simple, even as it applies in nonlinear optics. The propagating electric field (the beam of light) is an electromagnetic wave represented by an amplitude and a phase,

$$A(\rho,\theta)\ \exp\left[-i\Phi(\rho,\theta)\right].$$

If we apply a change that flattens the phase, we multiply it by $\exp\left[+i\Phi(\rho,\theta)\right]$, or the complex *conjugate* of the field. To do this physically, we add an optical aberration to the beam at precisely the right time, using a deformable mirror.

Using a simple lens analogy, consider the optical imaging system in Fig. 3.1. The lens *f* forms an image on the focal plane. If there is another lens in the beam path,

with focal length f_a, then the image is formed at the wrong place. At the original focal plane, the image appears fuzzy. If we measure the wavefront, we find that another lens, with focal length $f_c = -f_a$, placed into the beam at the right place will bring the image back into focus at the right place. This is adaptive optics. This is phase conjugation.

This simple process is performed 30 times a second in our own eyes. For those of use who have aged gracefully over a number of decades, the process often requires outside help, such as putting on eyeglasses once a day or so. If the worst complaint we have is "our bandwidth is down," we are lucky.

Imaging and MTF

The phase conjugation process described above is conceptually simple, but the physical implementation of it, in movable (adaptable) optics, took years to develop. One of the difficulties was that the entire imaging process is not easy to decode. In the early 20th century, the linear theory of image formation was developed. To sum up the basic principles of the theory, we sum up the major idea. The major idea is nothing more than the idea that an image is the sum of its parts.

Each object can be decomposed into its spectral components. That is, an image has low spatial frequency components, like large dark or light blobs, and it has high spatial frequency components, like fine detail. The information is contained in the variations in dark and light which is termed contrast, or *modulation*. An imaging system transfers the modulation from the object to the image. The mathematical function that does this transforming is cleverly called the *modulation transfer function*, or MTF. An imaging system (usually) transfers the modulation at low spatial frequencies better than the detail at high spatial frequencies. Is this not the idea of an eye chart? We can see the big E, but we can't make out the difference between the little **B** and the **R**. If the modulation is transferred perfectly, the MTF is 1.0. If the information is somehow blurred, the MTF is above 0 but less than 1. If the contrast is washed out completely, the MTF is zero. In some cases, using cleverly designed apertures for specific targets, we can amplify the modulation. This "super-resolution" is not a topic of discussion here, but I wanted to mention it.

Measuring the MTF of an optical imaging system is generally straightforward, except when there are exceptions to the rules. If we take a point source and project it into an image on the focal plane, we don't see a point, we see the spread of the point, called the *point spread function* (PSF). By mathematically taking the Fourier transform of the PSF, we get a function called the *optical transfer function* (OTF). The magnitude of the OTF, the positive part that transfers modulation, is the MTF.

Another way of measuring the MTF of a system is to look at how well it transfers modulation directly. By introducing into our imaging system objects of greater and greater detail (higher and higher spatial frequencies), we can map out the modulation, or contrast, in the image. The value of the MTF is just the ratio of image contrast to object contrast at a given spatial frequency. A good imaging system, even a diffraction-limited one, cannot transfer all the modulation exactly. How close it comes to "exact" is the measure of our imaging system quality.

We can actually calculate the image directly from the object and the PSF, without going through their spectra and all that trouble. Unfortunately, the mathematics of calculating a direct convolution, which is the equation that relates the object to the image, is usually worse than calculating their spectra and going to all the trouble of multiplying at each frequency. To keep things in perspective, we can describe the process like this. The image I is the convolution of the object O and the optical system point spread function P:

$$I(x,y) = \iint O(\xi,\eta) \quad P(\xi - x, \eta - y)\mathrm{d}\xi\mathrm{d}\eta \ . \tag{3.8}$$

The spectrum of the object is the Fourier transform of the object:

$$\mathrm{FT}[O](\alpha,\beta) = \iint O(\xi,\eta)\exp[-2\pi\mathrm{i}(\alpha\xi + \beta\eta)]\mathrm{d}\xi\mathrm{d}\eta \ . \tag{3.9}$$

The spectrum of the image is the Fourier transform of the image:

$$\mathrm{FT}[I](\alpha,\beta) = \iint I(\xi,\eta)\exp[-2\pi\mathrm{i}(\alpha\xi + \beta\eta)]\mathrm{d}\xi\mathrm{d}\eta \ . \tag{3.10}$$

The optical transfer function is the Fourier transform of the point spread function:

$$\mathrm{OTF}(\alpha,\beta) = \iint P(\xi,\eta)\exp[-2\pi\mathrm{i}(\alpha\xi + \beta\eta)]\mathrm{d}\xi\mathrm{d}\eta \ . \tag{3.11}$$

The image spectrum is related to the object spectrum by multiplying by the OTF:

$$\mathrm{FT}[I](\alpha,\beta) = \mathrm{FT}[O](\alpha,\beta) \bullet \mathrm{OTF}(\alpha,\beta) \ , \tag{3.12}$$

which is mathematically equivalent to Eq. 3.8.

The linear theory that follows this is pretty simple. If we have more than one imaging system, such as a camera lens, a zoom focusing lens system, and a detector system such as film or a CCD array, the total MTF is just the product of all the individual MTFs of the individual elements. Each object, or image, or MTF, is made up of a linear sum of its spatial frequency components. The modulation transfer, at each spatial frequency, is multiplied by the transfer for the other system at the same spatial frequency. Because MTFs range between 0 and 1, we cannot improve an imaging system by adding more optics. This seems to fly in the face of adaptive optics! All along we say how we can improve the image quality by adding deformable mirrors and sensors and so forth. This is the glory of the process. Adaptive optics, by the process of phase conjugation,

effectively removes the source of the poor MTF and leaves only a residual MTF of the adaptive optics themselves. Therefore, we CAN improve imaging by adding adaptive optics.

Bibliography

Max Born and Emil Wolf, *Principles of Optics*, Pergamon Press, Oxford (1975).

Joseph W. Goodman, *Introduction to Fourier Optics*, 2nd Ed., McGraw-Hill, New York (1996).

Jack D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, New York (1978).

Joseph S. Accetta and David L. Shumaker, exec. eds., *The Infrared and Electro-Optical Systems Handbook*, ERIM Infrared Information Analysis Center, Ann Arbor, MI, and SPIE Optical Engineering Press, Bellingham, WA (1993).

Chapter 4

Atmospheric turbulence — bad air . . . bad, bad air

We need the atmosphere to sustain life. It is made of air, a conglomerate of different gaseous elements that is pretty much transparent to the light that we normally see. However, since it is a mixture, and getting more mixed all the time, we could do really good astronomy if it weren't there at all. In fact, Isaac Newton wrote about the problem in *Opticks* in 1730:

"For the Air through which we look upon the Stars, is in perpetual Tremor ... But, these Stars do not twinkle when viewed through ... large apertures. The only Remedy is a most serene and quiet Air, such as may perhaps be found on the tops of the highest Mountains above the grosser Clouds."

In this basic and extremely perceptive observation, Newton acknowledged the problem, observed the phenomenon of aperture averaging of scintillation, and proposed a solution. Over 200 years before spaceflight, he suggested that we put telescopes on mountains to get into clearer air. We still take his advice, and we go one step further — we launch billion dollar satellites into space to really, really get above the grosser Clouds. Now, in the latter part of the 20th century, when we name computers after Newton himself, we have developed adaptive optics so that we don't really have to go to the tops of the highest mountains — we just have to be able to see around the tremors so that it looks like we are above the air.

What is wrong with the air? It has an index of refraction near 1.0, which is pretty close to a vacuum; the problem is that it is not exactly 1.0. *Pretty close* doesn't make it. The index depends upon the wavelength of the electromagnetic wave passing through it. It depends upon temperature, pressure, and density. The less air there is, the more like a vacuum it is.

Because the index of refraction is not uniform, it distorts the nice clean electromagnetic wave passing through it. While we use all sorts of lenses for many reasons, we really don't want to throw a handful of little lenses into the air in front of our telescope. The light coming through the various lenses would reach our focal plane, our eye, at different times. Being a wave, the light exhibits interference when it reaches our eye. The result is blurring or, as the intensity

wiggles around, twinkling. The analogy of a lot of little lenses wiggling around in the air is actually not too far from the actual physics.

Big masses of hot and cold air are weather phenomenon. Medium size masses of air bounce around airplanes and other less fortunate floating bodies. But small masses of air, only a few centimeters to meters, also can move around and mix. The air, as it is heated, gets less dense. The heavier, denser air falls down through the other air. Mixing is the principle behind atmospheric optical turbulence.

Structure constant

A beam of light passing through the mixing air is constantly being refracted, or bent. The mixing of the different parts of the beam results in blurring, or scintillation. The theory of turbulent air, or any turbulent fluid, is based upon the understanding of how the mixing takes place. From an optical point of view, we can compare the refractive index at one point of the atmosphere $n_1(r_1)$ with the index at another nearby point $n_1(r+r_1)$. When we average these over the whole atmosphere of interest, we get the *covariance* of the refractive index B_n ,

$$B_n(r) = \left\langle n_1\left(r + r_1\right) \;\; n_1\left(r_1\right)\right\rangle . \tag{4.1}$$

where r_1 is the position in the air and r is the separation between the points. The three-dimensional Fourier transform of the covariance is the power spectral density of the atmosphere,

$$\Phi(K) = \frac{1}{(2\pi)^3}\int_{l_0}^{L_0} B_n(r) \;\; \mathrm{e}^{-K\bullet r} \;\; \mathrm{d}^3 r \tag{4.2}$$

where the integral goes from the smallest atmospheric eddy size l_0 , the inner scale, to the largest eddy L_0 , the outer scale. K is three-dimensional spatial frequency. The inner scale, a few millimeters, is the size below which viscous effects are important and the outer scale, a few meters or so, is the size at which isotropic behavior is violated. Even though we can write it down, this integral is not easily solved in the general sense. By assuming that the inner scale was zero and the outer scale was infinity, Kolmogorov was able to come up with the *Kolmogorov spectrum*[1] $\Phi(K) = 0.033 \;\; C_n^2 \;\; K^{-11/3}$. Von Karman was able to do the integral without assuming that the outer scale was infinite.[2] The *Von Karman spectrum* has the form

$$\Phi(K) = 3.9\times 10^{-5} \;\; C_n^2 \;\; L_0^{11/3}\left(1 + \frac{L_0^2 K^2}{(2\pi)^2}\right)^{-11/6} . \tag{4.3}$$

[1] A. Kolmogorov, in *Turbulence: Classic Papers on Statistical Theory*, S. K. Friedlander, L. Topper, eds., Wiley (Interscience), New York, 1961.
[2] V. I. Tatarskii, *Wave Propagation in a Turbulent Medium*, McGraw-Hill, New York, 1961.

Both these spectra, and almost everything else about describing atmospheric turbulence, make use of the constant that appears in these last two equations. The atmospheric structure constant C_n^2 is a measure of the strength of turbulence; if it is high, the turbulence is high. The units of C_n^2 are meters to the –2/3 power, not a very easily imagined concept, but that's just the way it all came out.

The value of C_n^2 ranges from about 10^{-15} $m^{-2/3}$ to about 10^{-18} $m^{-2/3}$. It changes with the seasons. It changes on a monthly, daily, hourly, and minutely basis. But the most important variable is its change with the wind and its apparent change with altitude. The higher one goes, the colder it gets. Colder and less dense air has less turbulence because there is less air and less of it moving around. Once one gets above the high winds of the jet stream (near 10 km), the turbulence level measured by C_n^2 gets into the 10^{-18} $m^{-2/3}$ range. Although C_n^2 can vary by an order of magnitude over a few meters, there is a generally accepted average. Unfortunately, there is no *single* generally accepted average; it varies with the site and the conditions above the ground, mostly the local conditions right above the ground. For calculational purposes, there are a few models of the average C_n^2 profile that can be used. The most common one is the Hufnagel-Valley Boundary model. It has two variable parameters which can be adjusted to fit a particular site condition. The HVB model, as it is called, is an equation relating the altitude z (given in kilometers) to C_n^2 :

$$C_n^2(z) = 5.94\times10^{-23} z^{10} \mathrm{e}^{-z} \left(\frac{W}{27}\right) + 2.7\times10^{-16} \mathrm{e}^{-2z/3} + A\mathrm{e}^{-10z}. \qquad (4.4)$$

The parameters W and A in the HVB model can be adjusted. W is roughly the high altitude wind speed in meters per second and A relates to parameters near the ground boundary layer. When W = 21 and $A = 1.7\times10^{-14}$, a calculation of the coherence length results in r_0=5 cm and the isoplanatic angle θ_0 = 7 μrad. With these parameters set this way, the model is called the HV 5/7 model, but the wind speed was set at 21 m/s, it is also called the HV21 model. The two models are the same.

A number of other models are used, dependent upon the site and the amount of experimental data taken there. They have various names like Clear 1 Night, Clear 2 Night, SLC Day, and SLC Night. Details are given in other books and the literature.[3] Two model profiles are plotted against typical data in Fig. 4.1.

[3] See for example, R. R. Beland, "Propagation through atmospheric optical turbulence," Chap. 2, Vol. 2, *The Infrared and Electro-Optical Systems Handbook*, ERIM, Ann Arbor, MI, and SPIE Opt. Engr. Press, Bellingham, WA, 1993.

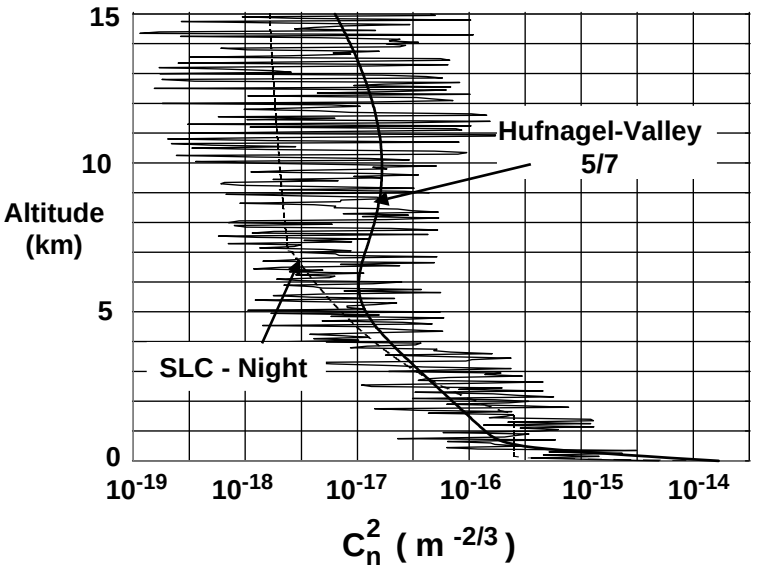

Fig. 4.1 A C_n^2 profile vs. altitude showing the Hufnagel-Valley model and the SLC Night model.

Gaussian beams

Turbulence affects a beam of light in a number of ways. It can spread out the energy of the beam, blur an image, cause speckle in the beam, or vary the intensity of the beam in the focal plane. Many lasers make beams with intensity distributions that follow a Gaussian form:

$$I(r) = \frac{w_0^2}{w_b^2} \exp\left(-\frac{2r^2}{w_b^2} \right) \quad . \tag{4.5}$$

The beam waist (radius) normal to the propagation axis is w_b; the beam waist at the transmitting aperture is w_0. Over a long distance *L*, the Gaussian beam waist grows because of the turbulence according to

$$w_b^2 = \frac{4L^2}{k^2 w_0^2} \quad + \quad 3.58 \quad C_n^2 \quad L^3 w_0^{-1/3} . \tag{4.6}$$

The factor *k* is the wave number, $k = 2\pi / \lambda$.

Fried's coherence length

The analogy that atmospheric turbulence is a lot of lenses floating around the atmosphere, moving around one another, and then occasionally merging is not too inaccurate. The eddies of the atmosphere do just that. They grow and shrink.

They wander. Some of them get together and form big wedges that just shift images around, while others break up into little pieces and just blur out the edges of an object.

In 1965, David Fried calculated the effect of all these little lenses floating around. The lenses have the effect of reducing the coherence of the beam. That is, light travels in various paths via various atmospheric lenses from the source to the detector. If the path of one part of the wave is sufficiently disturbed, that part is no longer coherent with another part of the wave. If the aperture is small so it doesn't gather in light from too many different paths through the atmosphere, you can keep the coherence of the lightwave all the way from the source to the detector. Fried determined that there is a characteristic length, called the coherence length, that is the limiting size of an aperture. Any aperture larger than this exhibits the breakdown of coherence and the distortion of the light. The coherence length r_0 is often referred to as *Fried's parameter*, or a *seeing cell size*. It is not necessary to debate the merits of the nomenclature. It is all just widely used in the literature.

For a plane wave and Kolmogorov turbulence, the coherence length is

$$r_0 = 1.68\left(C_n^2 L k^2\right)^{-3/5}, \tag{4.7}$$

where L is the propagation path length. The exponent (–3/5) appears here like it will in a lot of other places because Kolmogorov turbulence is assumed. It's not a bad assumption, but there are times when you have to consider other types of turbulence, like the airflow at a boundary layer. It also cleverly works out that r_0 is in units of distance.

We often must calculate the coherence length for a variable path, like propagating up (or down) through the atmosphere from the ground. The structure constant C_n^2 varies with altitude z. The path is also stretched out by pointing in a direction other than straight up or down. The zenith angle β, the angle between the pointing direction and normal to the ground, is a factor in a complete r_0 calculation. Repeating Eq. 1.1, we find the coherence length from

$$r_0 = \left[0.423 \;\; k^2 \;\; \sec\beta \int_{\text{Path}} C_n^2(z) \;\; \mathrm{d}z \right]^{-3/5}. \tag{4.8}$$

Fried's parameter is a widely used descriptor of the level of turbulence at a particular site. The value varies from a few meters in very good seeing conditions in the infrared to a few centimeters in difficult seeing at visible wavelengths. From the above relationships, we find that $r_0 \propto \lambda^{6/5}$. Thus, the limiting useful coherent aperture at infrared wavelengths can be twice to 10 times the aperture at visible wavelengths.

Noll relates the size of the uncompensated aperture and r_0 to the wavefront error.[4] The variance of the wavefront error caused by the atmosphere is

$$\sigma^2_{\text{Spatial}} = 1.02 \left(D \big/ r_0 \right)^{5/3} . \tag{4.9}$$

When the tilt is removed from the atmosphere by a separate control system, the variance reduces to

$$\sigma^2_{\text{Spatial}} = 0.134 \left(D \big/ r_0 \right)^{5/3} . \tag{4.10}$$

When adaptive optics is used to compensate for some of the atmosphere, the residual wavefront error is the fitting error (Eq. 1.4).

Scintillation

We most often observe the effect of atmospheric turbulence as an intensity variation, or twinkling, of the stars. This can occur because the light from various parts of the object travels through various paths before it reaches the detector (our eye). The light from these various sources and paths interferes. Constructive or destructive interference is observed as intensity variation, or scintillation. We can calculate how much variation occurs.

We actually calculate the intensity variance over time C_{I} using the parameter C_χ, which is the log-amplitude variance over time:

$$C_{\text{I}} = \exp(4C_\chi) - 1 . \tag{4.11}$$

The log-amplitude variance is a fairly simple function of the turbulence:

$$C_\chi = 0.307 \; k^{7/6} L^{11/6} C_n^2 . \tag{4.12}$$

Like the coherence length, when the path is through a region with variable C_n^2, the value must be integrated along the path.

Despite all the calculations, we can still figure out what is going on with a simple analogy. Consider Fig. 4.2 which shows four possible configurations, two with a point source and two with an extended source, each with a small or a large aperture. In the case of the point source with its rays of light passing through turbulence, some of the rays reach the small detector. This is analogous to our eye watching a star in the sky. Some of the rays, as they pass through the air, experience a ½ wavelength phase delay and some pass through cleanly. When the ½ wave rays combine with the nondelayed rays, these coherent rays interfere

[4] R. J. Noll, *J. Opt. Soc. Am.* **66**, 207, 1976.

destructively, washing out any visible light. When the rays happen to hit the detector without phase differences, the light is visible because the rays interfere constructively. Since the atmosphere is moving constantly, the rays are experiencing different phase delays all the time. The star seems to blink on and off in a random fashion. They twinkle.

If we get a pair of binoculars, or go down to our local store and buy a 4 m telescope, we see that much of the twinkling disappears. Because now we have so many rays entering our eye, it is very unlikely that most of them are exactly out of phase with the rest of them. Everything is averaged out. The light we see is more incoherent; there is less interference, and less chance of intensity variation, or twinkling. This phenomenon is termed *aperture averaging*, and it is dependent upon the wavelength, the size of the aperture, and the propagation path.

We can get the same effect by using an incoherent source to begin our experiment. An extended object, like a planet orbiting the sun or a large satellite, has light radiating from all the different parts of the surface. The light from each point is incoherent with the other points because their phases are all averaged out and random when they get started. By the time they reach the atmosphere, it doesn't really matter that the air randomizes the phase any more. The light we see is still just an average of all the rays, with very little scintillation. Scintillation is how we can tell a planet from a star even if we never took Astronomy 101.

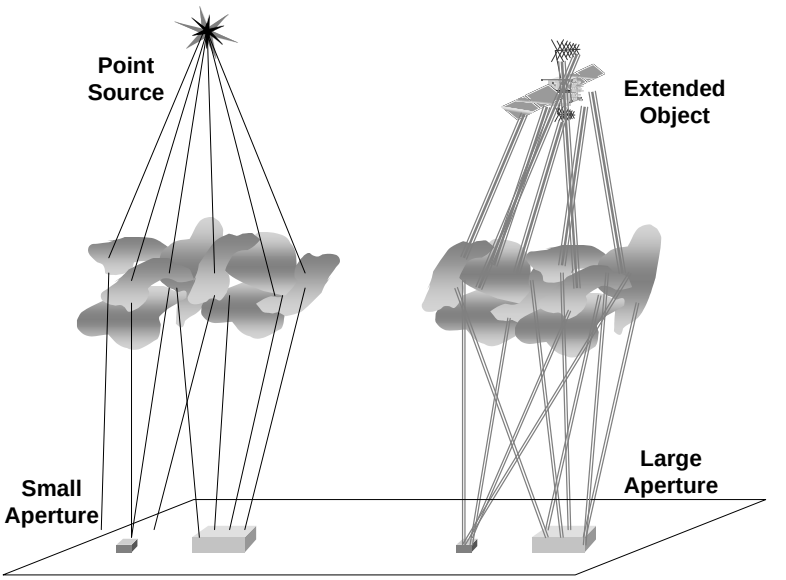

Fig. 4.2 Depiction of four possibilities for scintillation effects. A point source and a small aperture exhibit high scintillation: larger aperture averages out the interference effects. An extended object allows for multiple paths to the aperture and thus less scintillation.

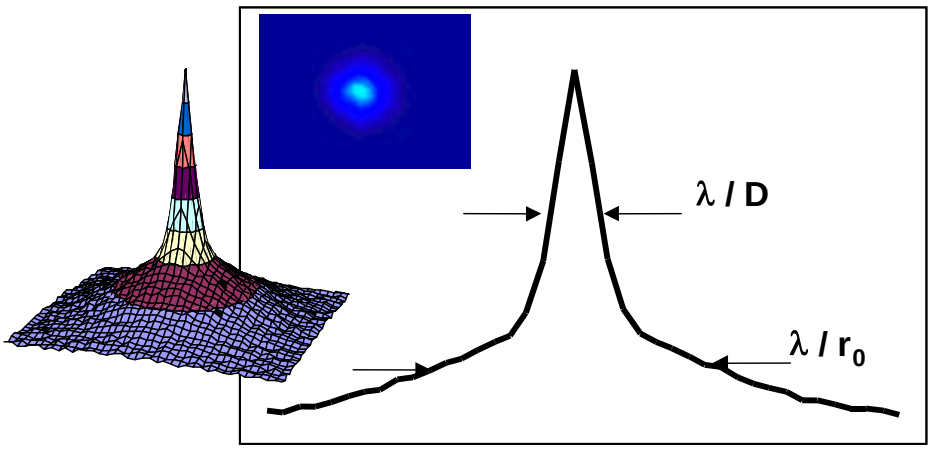

Fig. 4.3 A point spread function showing a broad halo that is due to atmospheric turbulence.

Resolution and the halo

According to J. W. Strutt, better known as Lord Rayleigh, we can generally resolve two closely spaced objects if they are separated by the angle λ/D, as measured from where we observe them. Rayleigh assumed a lot of things here, but the most important is that we are diffraction limited. That is, the diameter D of the aperture is not restricted by some sort of atmosphere or aberration. If we looked at a star (a point object) through such a system, we would see just the point spread function of the optical system. (That is the definition of the PSF, by the way.) It would have a full width at half maximum (FWHM) of about λ/D. If two objects are separated by at least this angle, then we could perceive two distinct objects. We call this the *resolution* of the system. When we look at the same star through the atmosphere with a coherence length r_0 , the image is blurred, and we see the PSF of the optics *and* the atmosphere – the sharp PSF of the optical system is lost. When we have a partially corrected optical system, there appears to be a halo around the star with a full width at half maximum of λ/r_0. Since D is generally larger than r_0, the halo masks the extent of the star image, and our resolution is reduced. Figure 4.3 shows this phenomenon.

Temporal effects: the Greenwood frequency

The air is mixing constantly. For adaptive optics to work, we have to measure the aberrations that are caused by turbulence faster than they can change. If the turbulence were frozen in time by the speed of a wavefront sensor, then everything would work. The “frozen turbulence” hypothesis is the basis for calculating a workable bandwidth for adaptive optics systems. From the viewpoint of our aperture, the tilt across the field is moving much more slowly

than the higher-order turbulent eddies. This intuitive process is based on the principle of inertia (big things move slowly; little things move quickly). The characteristic frequency of the tilt of the atmosphere is dependent upon what is meant by "big." Tilt primarily wiggles the image in the focal plane. If the wiggle is faster than the bandwidth of the tilt control loop, the image gets smeared. For an aperture of diameter *D*, the *tilt Greenwood frequency* is

$$f_{\mathrm{T}} = 0.33 D^{-1/6} \lambda^{-1} \sec^{1/2} \beta \left[\int_{\mathrm{Path}} C_n^2(z) V_{\mathrm{Wind}}^2(z) \mathrm{d}z \right]^{1/2} , \qquad (4.13)$$

where the wind velocity as well as C_n^2 is a function of altitude.[5]

Repeating Eq. 1.3, we note that for the higher-order aberrations that tend to blur the image the *Greenwood frequency* is

$$f_{\mathrm{G}} = 2.31 \lambda^{-6/5} \left[\sec \beta \int_{\mathrm{Path}} C_n^2(z) V_{\mathrm{Wind}}^{5/3}(z) \mathrm{d}z \right]^{3/5} . \qquad (4.14)$$

To make life simpler, sometimes we can assume a constant wind, *V*. Knowing the path integrated r_0 for the zenith angle, the Greenwood frequency simplifies to

$$f_{\mathrm{G}} = 0.43 \left(V \big/ r_0 \right) . \qquad (4.15)$$

For cases of interest, the Greenwood frequency of the atmosphere can range from tens to hundreds of hertz. For example, at Hawaii's Mt. Haleakala site, the Greenwood frequency is about 20 Hz. The residual wavefront variance associated with a fixed control bandwidth f_{BW} can be represented by the expression

$$\sigma_{\mathrm{temp}}^2 = \left(f_{\mathrm{G}} \big/ f_{\mathrm{BW}} \right)^{5/3} . \qquad (4.16)$$

Thermal blooming

A beam of light is an electromagnetic field. If the field is very intense, the air that it passes through can ionize. This air breakdown can cause damage and a severe reduction in the propagated intensity of the beam.

When the beam is not intense enough to cause air breakdown but is intense enough to heat the air through which it passes, a phenomenon called *thermal blooming* can occur. If a beam is too intense, or focused to a region where it

[5] D. P. Greenwood, *J. Opt. Soc. Am.* **67**, 390, 1977.

becomes intense, the medium (air, for want of a better propagation medium) heats up and becomes less dense and the index of refraction changes. The localized index change acts as a negative lens, which defocuses the beam. The thermally affected beam spreads out and appears to bloom like a flower, thus the term thermal blooming.

The intensity of the bloomed beam depends upon the absorption of the atmosphere and the strength of the blooming N_B. The latter depends upon the power in the beam, the type of beam, whether it is pulsed or continuous, and all the atmospheric parameters such as density, specific heat, and the wind that acts as a cooling engine by blowing hot parts of the air out of the path of the beam.[6]

We can quantify the intensity reduction by an empirical model which shows that the relative intensity is reduced by the factor

$$I = I_0 \frac{1}{1 - 0.0625 N_B^2} \ . \tag{4.17}$$

A number of experiments have shown that adaptive optics can compensate for the phase variations associated with thermal blooming. In systems where high power propagation is the purpose, it is necessary to follow the technology path that maximizes the efficiency of the propagation.

Anisoplanatism

If we stare up into the sky, we can imagine that the air that causes all that twinkling is really just a plane surface at a particular altitude. The assumption of *planatism* can be used in our wavefront sensor process. If the air exhibited uniform aberrations and it did not matter where we looked through the plane, it would be isoplanatic. But it isn't. The aberrations change as we look through the air in different directions; it is *anisoplanatic*.

The source of light for a wavefront sensor should be in such a position that the light propagating from it samples all the aberrations. If the source, the beacon, is offset laterally from the path we want to measure, such as the path to a science object, it exhibits *displacement* anisoplanatism. If the sensor is in one position and the source and the propagation path are separated, we experience *angular* anisoplanatism. If there is a time delay between the beacon and the propagation path due to some wind moving the air, we call it *temporal* anisoplanatism. When there is a difference between the paths of the beacon and the science object because of atmospheric dispersion which would separate the paths by wavelength, it is *chromatic* anisoplanatism. If the beacon is at a finite distance while the science object is at an essentially infinite distance, we call it *focal* anisoplanatism.

[6] D. C. Smith, *IEEE J. Quantum Electron.* **QE-5**, 1679, 1969.

Although each of these can be treated separately, most have the same consequence — what we measure is not what we want to measure.

We must look for the wavefront in the direction that we want to observe. We don't have to be exact; we can be off by a small angle called the isoplanatic angle θ_0 or isoplanatic patch and still get good images. In the direction of zenith angle β, the angle θ_0 is

$$\theta_0 = \left[2.91k^2 \sec^{8/3}\beta \int_{\text{Path}} C_n^2(z) z^{5/3} (z)\,\mathrm{d}z \right]^{-3/5}. \tag{4.18}$$

For a fixed C_n^2 along a path of length L, the isoplanatic angle is

$$\theta_0 \approx 0.6\left(\frac{r_0}{L}\right). \tag{4.19}$$

The wavefront variance associated with an angle between the beacon and the science object can be represented by Eq. 1.5, repeated here:

$$\sigma_{\text{iso}}^2 = \left(\theta / \theta_0\right)^{5/3}. \tag{4.20}$$

Bibliography

Larry C. Andrews and Ronald L. Phillips, *Laser Beam Propagation through Random Media*, SPIE Optical Engineering Press, Bellingham, WA (1998).

Michael C. Roggemann and Byron Welsh, *Imaging through Turbulence*, CRC Press, Boca Raton, FL (1996).

Vladimir P. Lukin, *Atmospheric Adaptive Optics*, SPIE Optical Engineering Press, Bellingham, WA (1995).

D. L. Fried, "Limiting resolution looking down through the atmosphere," *J. Opt. Soc. Am.* **56**, pp. 1380-1384 (1966).

D. L. Fried, "Anisoplanatism in adaptive optics," *J. Opt. Soc. Am.* **72**, pp. 52-61 (1982).

R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.* **66**, pp. 207-211 (1976).

D. P. Greenwood, "Bandwidth specification for adaptive optics systems," *J. Opt. Soc. Am.* **67**, pp. 390-393 (1977).

Chapter 5

Laser guide stars — beacons in the wilderness

Knowledge of the aberrated wavefront is the critical input for an adaptive optics system. Measuring the wavefront requires sufficient signal to keep the estimation error low. An astronomical adaptive optics system wavefront sensor requires a sufficiently bright source in the direction of the science object. If a natural star or the science object itself is bright enough then we use that. If it is not, then up until the mid-1980s we would all have just gone home. However, because of the development of artificial laser guide stars, we don't get to go home anymore; we get to shine lasers up into space. Saying the guide star is "bright enough" has some definite meaning: It must be bright enough (have enough photons that reach our sensor) so that the signal to noise ratio is high enough. “High enough” means that the residual wavefront error is low enough to make aberration compensation possible. For a Hartmann wavefront sensor, the wavefront error σ_{WFS} is inversely proportional to the signal to noise ratio:

$$\sigma_{\mathrm{WFS}} = \frac{3.7}{\mathrm{SNR}} \quad \text{(rad)} \quad . \tag{5.1}$$

If we need a wavefront sensor error less than 1/10 wave, for example, the SNR should be greater than 5.9. For a sensor that uses a focal plane array where n_{pix} is the number of pixels that are used for determining the wavefront error, we can use the expression

$$\mathrm{SNR} = \frac{N}{\left[N + n_{\mathrm{pix}} \left(\sigma_r^2 + \sigma_{bg}^2 \right) \right]^{1/2}} \quad , \tag{5.2}$$

where N is the number of signal photoelectrons, σ_{bg} is the background photoelectron count, and σ_r is the noise photoelectron count.

For 4 pixels in a subaperture and 10 electrons of noise per pixel, we need roughly 137 photons per subaperture to guarantee 1/10 wave measurement. For this example, let’s round it off to 150 photons needed. Two significant digits is pretty much how we live in scaling-law land. Continuing the example, where do we get 150 photons? Is a natural star sufficient? One nice simple, albeit slightly suspect, approximation for the number of photons we get at the Earth's surface is

$$N = 4.6 \times 10^6 10^{-m/2.5} \text{photons/cm}^2 - \text{sec} \ . \tag{5.3}$$

This is a good all-around average when we can address the star source as a visual magnitude *m*. Of course, this differs with different wavelengths since stars have various diverse and very interesting and informative spectra. But, in the enjoyment of this exercise, we can use approximations like this until we get caught. The expression accounts for a number of photons in a subaperture area during a period of time. What area should we use? What time should we use? Assume, again for simplicity, that the subaperture diameter is 1.0 r_0 and we sample it at 100 Hz. That will give us about 1/10 wave accuracy across our wavefront and let us close the loop at around 10 Hz. So, cranking all these assumptions into the bigger assumption, we find out that we need a magnitude 12 star or brighter. Continuing the assumptions, we can use the expression

$$\text{Number of stars brighter than } m = 1.45 \text{e}^{0.96m} \text{stars/rad}^2 \ . \tag{5.4}$$

For visible wavelengths, there are about 150,000 stars/rad^2 with magnitude greater than m = 12. For longer wavelengths, things get a little better, but 150,000 seems like a lot until we remember that Carl Sagan said that there are billions and billions. We not only need to have the guide star bright enough, but it must be within the isoplanatic patch (about 10 μrad or 2 arcsec at visible wavelengths) around our science object. There are about 10^9 isoplanatic patches in the sky, so 150,000 stars won't be enough to go around. Nuts! Artificial guide stars are needed.

Where do we get one? We can make one of two types. A laser can be focused in the lower atmosphere, between 16 and 20 km, and we'll get enough Rayleigh backscatter to give us 150 photons, but we will have the "cone effect" errors in our measurement. We can make a guide star up at a higher altitude, around 90 to 92 km, where there isn't enough air for Lord Rayleigh but there is enough sodium floating around to make a sodium resonant guide star. Higher is better. Lower is cheaper. Let's see.

How much laser power is needed to make a Rayleigh laser guide star (LGS) at 20 km? From the laser radar equation we can calculate the return flux F in photons per square meter:[1]

$$F = \eta T_{\text{A}}^2 \frac{\sigma_R n_R(z_0 + z_t)}{4\pi z_0^2} \frac{\Delta z \ \ \lambda_{\text{LGS}} E}{hc} \quad , \tag{5.5}$$

where

$$\Delta z = \frac{4.88 \ \ \lambda_{\text{LGS}} \ \ z_0^2}{D_{\text{proj}} \ \ r_0} \quad . \tag{5.6}$$

[1] C. S. Gardner, B. M. Welsh, L. A. Thompson, *Proc. IEEE* **78**, 1721, 1990.

Since we've made so many assumptions up to this point, let's make a few more. The detector efficiency η is 0.075, the one-way transmission up through the atmosphere T_A is 0.85. The wavelength λ_{LGS} is 351 nm since the laser is a XeF or copper vapor laser. Our laser projector optics is 1-m diameter, the site altitude z_t is 3 km, and the subaperture is 1 r_0 (we can be consistent in our assumptions). For 150 photons, we need about E = 14 mJ/pulse. This performance is available with Questek's XeF laser (150 mJ/pulse at 300 Hz), Oxford Lasers' Cu vapor laser (20 mJ/pulse at 5 kHz), and a number of others.

We have to be careful about some other things when we go shining laser beams up into the sky. Besides safety concerns, our laser will not just give us a nice point source and a small spot up in the atmosphere. We get scatter from the laser all along its path.

Fig. 5.1 A laser guide star test at Yerkes Observatory.

So that we can get light from just a specific region of the sky along the path, we range-gate our receiver. That is, we open our wavefront sensor aperture (optically or electronically) only during the time when we want to see the return. It takes a pulse of light about 66 μs to reach the height of a Rayleigh guide star and it takes another 66 μs to get back to us. So what we do is simple — we just open our receiver 132 μs after we send out the pulse. We will then see only the scatter from 20 km, not all along the path. Is this timing reasonable? Doesn't another

pulse go out and confuse our timing? No, not if the pulses are spread out. If the repetition rate of the laser is 5 kHz, the next pulse will go at 200 μs, giving us time to send one pulse out, wait 132 μs, open the range gate, close the range gate, and wait for the next pulse. If the rep rate of the laser is slower, like 300 Hz, we have plenty of time before the next pulse — a 300 Hz laser waits 3,333 μs for the next pulse, leaving time for our range-gated wavefront sensor and a nice long nap too.

If the science object is not directly overhead and our guide star also is not directly overhead, we must deal with zenith angle consequences. The zenith angle is the angle between zenith (straight up) and our pointing direction. The flux we receive is proportional to the secant of the zenith angle. Pointing near the horizon gives us almost no return since the pathlength to 20-km altitude is stretched and the path of light through the region of backscatter is stretched.

All laser guide stars are not created equal. Because the backscattered light comes from a finite volume of space, and we need a finite amount of air to scatter the light, the guide star is not a point source. In general, this is not too big a deal since the light "seems" to come from a point, because our small subaperture can't resolve its actual shape. Also, the chosen altitude of the LGS affects the returned flux. If it's higher, there is less focal anisoplanatism (cone effect), but there is also less air to scatter the light (see Fig. 5.2). If the LGS is low, we get plenty of return, but is doesn't go through enough atmosphere for a meaningful measurement. A rough estimate of the photons that we can get from a Rayleigh guide star, forgetting all that big-time equation stuff on the previous pages, is given by

$$\log(\text{photons}) = 5.28 \;\; -0.14 \;\; z_0 \;\; (\text{km}) \qquad , \tag{5.7}$$

where z_0 is the LGS altitude. If the LGS altitude is 20 km, the logarithm of the number of photons is about 2.4, or the number of photons is nearly 150. Just like we predicted, we have enough photons, but is it at a high enough altitude?

David Fried predicted that the wavefront error is proportional to $\left(D/d_0\right)^{5/6}$, where our observing aperture is D and the focal anisoplanatism parameter is d_0. From models of the atmosphere, for example the SLC Night model, and plowing through a bunch of equations, we find that for a 4-m aperture and a 1/10 wave error, we need a Rayleigh laser guide star at 143 km. This is a big problem since there is no air, and thus no Rayleigh scatter, at 143 km. For a Rayleigh guide star, we need to allow more wavefront error or have a smaller aperture system. For instance, a guide star at 90 km will give us 1/7 wave error — tolerable, but not great. Since anisoplanatism limits us, then we should consider either going to the higher altitude sodium LGS or making multiple guide stars. By making multiple laser guide stars and stitching their information together, we can reduce the effect

of anisoplanatism. One scaling law that is useful is saying that d_0, the Fried cone effect parameter for visible wavelengths (in meters), is

$$d_0 = 0.23 \quad N_{\mathrm{LGS}} + \quad 0.95 \quad , \tag{5.8}$$

where N_{LGS} is the number of laser guide stars. The more the merrier.

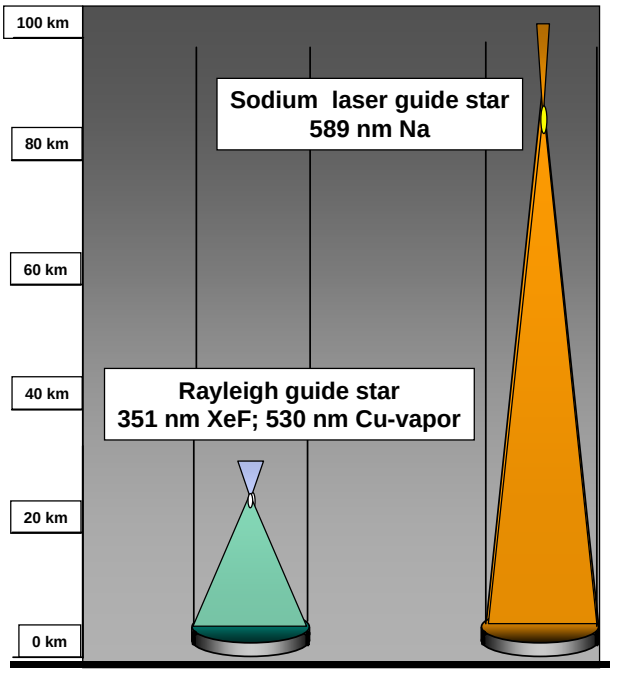

Fig. 5.2 A Rayleigh laser guide star at 20-30 km samples only a portion of the atmosphere. The light travels in a cone and exhibits focal anisoplanatism, or the "cone effect." Light traveling from a higher altitude guide star, like the sodium laser guide star, will sample more of the atmosphere.

Okay, so you don't want to make a bunch of laser guide stars and figure out where the information is coming from and decode it and mess around with it too much. I understand. The sodium laser guide star is the solution. Unfortunately, there are physical limitations and engineering problems that come along with it. First, we need a laser or a system of lasers to make sufficient power at the resonant line of the atomic sodium, way up there at 90 km. Second, we have to realize that the sodium layer sort of moves around from month to month and season to season. Sometimes it is as low as 89 km; sometimes it is up at 92 km. We can use lidar to find it, and that is the subject of a lot of ongoing research work.

Another concern arises when we worry about how much sodium is up there — there is only "so much." If we shine a laser on the sodium, some will backscatter. We can increase the power of the laser until all the sodium backscatters, but there is a physical limit to this saturation of about 1.9×10^8 photons/sec. If we add more power, we just send it off into space. A photon is a terrible thing to waste. So we are limited by how many atoms of sodium there are along our laser path. It is measured by its "column abundance," which is the number of atoms of atomic sodium per area. The column abundance can also change by an order of magnitude as the year progresses.

The thousands of smaller meteors that pummel the planet will continue to replenish the atomic sodium, which constantly is disappearing by forming molecular species. Big massive meteors that strike the Earth are good for increasing the column abundance too, but they're better for depleting the population of human species. Disaster films document this drama without even considering the wonders of laser guide stars.

Just like with the Rayleigh guide star, we can calculate the flux returning from the sodium layer for a given laser pulse energy. Using the equation

$$F = \eta T_{\mathrm{A}}^2 \frac{\sigma_{\mathrm{Na}} \rho_{\mathrm{Col}}}{4\pi z_0^2} \frac{\lambda_{\mathrm{LGS}} E}{hc} \quad , \tag{5.9}$$

we put in the sodium column abundance as 0.02, the altitude z_0 at 90 km, and the laser wavelength of 589 nm. The result, for a subaperture of 1.0 r_0, is 550 photons/J of laser energy. So now with rather uncomplicated algebra, we find that we can get our 150 photons for our subaperture with 270 mJ in a laser pulse. A number of laser sources can do the job. Sodium laser guide stars have been produced with flashlamp pumped Nd:YAG with organic dyes, excimer lasers with organic dyes, and two Nd:YAG lasers mixed in a crystal. The Monolithic Mirror Telescope (MMT) at the Steward Observatory has a 2.7 W dye laser pumped by a 25 W argon laser that produces a magnitude 9 laser guide star in the visible band.[2]

Why we can't measure global tilt from a laser guide star

We can measure all sorts of higher-order aberrations from a LGS since the path from the guide star to the wavefront sensor is different for each of the subapertures, precisely what we want to measure. However, the constant bias from a star that is exactly aligned with all the sensor subapertures is missed. We don't know where the absolute center of the sensor line-of-sight is without knowledge of a source out of our control.

[2] J. T. Murray, *ESO Workshop on Laser Technology for Laser Guidestar Adaptive Optics in Astronomy*, Garching, Germany, 1997.

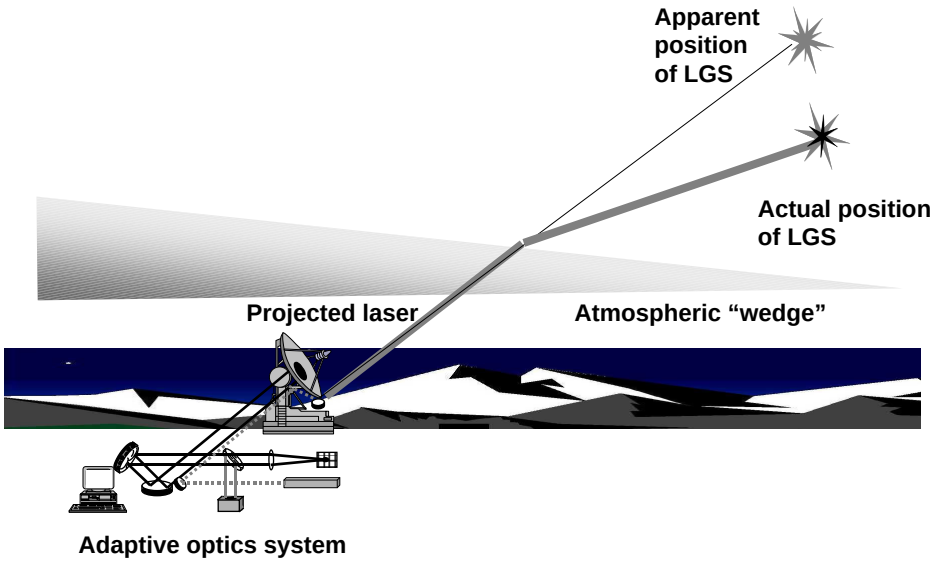

Fig. 5.3 The apparent position of the laser guide star is different from its true position. With just one laser guide star, we can’t measure global tilt.

Consider this problem: we have an adaptive optics system with a wavefront sensor on the ground looking up into space (see Fig. 5.3). Somewhere above the system, in the atmosphere, is a wedge of air that is slightly more dense than the air surrounding it; this bit of air acts like a prism. When we send a laser beam up through the air, and through the wedge, the direction of the beam changes. The laser guide star is formed at a point in the upper atmosphere. Its light scatters down toward the wavefront sensor, going along the same path as the upward laser — both paths pass through the wedge and are deflected by the same angle. Therefore, we don't know, and can't measure from the apparent laser guide star position, exactly how much wedge (or tilt) is there. If the wedge changes, the guide star still seems to be in the same position — in the direction of our laser beam as it exits our system. So, since we can't measure the position of the LGS, we can't measure the global tilt. What do we do? A few things are in our favor.

For one, to measure tilt, we can use all the photons available to us from the entire aperture, not just a subaperture. We can use a natural star if it is bright enough and if it is within the isoplanatic patch for tilt measurement. Using Eq. 5.3, for our full 4-m aperture at 100-Hz rate we find that we need a star source of magnitude 19 or brighter, and there are more than 10^8 stars/rad^2 with a magnitude brighter than m = 19. We need these sources to be close to our science object. The isoplanatic angle for our higher-order wavefront measurement was on the order of 5–10 μrad, and we needed a wavefront source within that angle. For tilt measurement, we find that we can use a star that is about 4 times that angle. In

conclusion then, we will assume that we can find within our tilt isoplanatic patch a natural guide star that is bright enough. The global tilt can then be found.

If these assumptions are not quite correct, i.e. if we have a different aperture or the atmosphere is particularly bad that night, we can try other things. Since the global tilt in the atmosphere acts like a prism, we could use two or more different color lasers to measure the relative differences in the apparent positions of the laser guide stars and deduce the atmospheric global tilt. These and other innovative techniques are being investigated theoretically and experimentally.[3]

"The galaxy is nothing else than a congeries of innumerable stars distributed in clusters." — Galileo, *Sidereus Nuncius*, 1610

Bibliography

L. A. Thompson, "Adaptive optics in astronomy," *Phys. Today* **47**, (12), pp. 25-31 (Dec 1994).

R. Q. Fugate, "Laser beacon adaptive optics," *Opt. & Phot. News* **4**, (6), pp. 14-19,(June 1993).

Multiple authors, *Special issue on adaptive optics*, Lincoln Laboratory J., **5**, (1) (Spring 1992).

R. H. Ressmeyer, "Robert Q. Fugate: Starfire's Magician Optician," *Sky & Telescope* **87**, (5), pp. 21-22 (May 1994).

B. M. Welsh, C. S. Gardner, "Effects of turbulence-induced anisoplanatism on the imaging performance of adaptive-astronomical telescopes using laser guide stars," *J. Opt. Soc. Am. A* **8**, pp. 69-80 (1991).

C. S. Gardner, B. M. Welsh, L. A. Thompson, "Design and performance analysis of adaptive optical telescopes using laser guide stars," *Proc. IEEE* **78**, pp. 1721-1743 (1990).

[3] M. S. Belen'kii, *Proc. SPIE* **3126**, 101, 1997.

Chapter 6

Systems — putting it all together

Adaptive optics are useful for compensating a wide variety of aberrations, not just those imposed by the atmosphere. Early investigation of adaptive optics included compensation for the deleterious effects of high energy lasers, mostly thermal distortions in the laser gain medium and in the optics. High energy lasers were developed so that their energy, when focused down to a small spot on an enemy missile, would melt through it and destroy it, but problems arose when the energy was being propagated out through an optical system. If the power was enough to melt enemy missiles, it was enough to melt and destroy the optics that direct it, so the field of high energy laser beam control evolved to include high reflectivity optical coatings and liquid-cooled optics. Since it was impractical for a missile to have a high reflectivity coating and really a mess to require a heavy water-cooled surface, the advantage goes to the high energy beam control system. When it's laser versus missile, it becomes laser: game, set, match.

Thermal distortions in optics come in two flavors. The first arises when a flat plate, the mirror, is constrained around its edges and is not free to expand in-plane when heated. As the mirror absorbs energy, even one with a good coating bows and its surface assumes a parabolic shape, adding defocus into the beam. The *bowing distortion* is proportional to the total power incident and is called power-induced distortion. When the mirror is hit with a nonuniform intensity pattern, like from just about every high power beam in existence, the expansion goes in an out-of-plane direction; hot intensity spots distort or bubble up more than cold spots. The *intensity mapping distortion*, or flux-induced distortion, is proportional to the incident intensity at each point and the absorption of the laser energy at each point. The amount of out-of-plane growth has a proportionality constant ξ which converts intensity to mirror surface growth. Because engineers, in contrast to physicists, don't deal with Greek letters too often, the symbol ξ, which couldn't be pronounced and which looks a little like a worm, became a measure of the thermal distortion of a high energy laser mirror — the "worm factor" was born.

One consequence of the intensity mapping distortion is that the nonuniform intensity is mapped into a nonuniform phase in the beam; the mirror distorts

more with hot spots than with cold spots, so the advance or delay of the phase as it reflects off the mirror is altered by the intensity pattern.

In addition to the mirror distortions in the high energy laser beam train, the laser itself will experience distortions that can be removed with adaptive optics. Lasers that are slightly misaligned introduce aberrations in the beam because the wavefront passing along the resonator optics keeps missing the preferred position and reflection angle of the next mirror. These add up, and over multiple passes through the resonator and through a variable gas or plasma in the laser gain medium, the wavefront leaving the laser is by no means a plane wave, nor even a nicely focused wave. The aberrations that arise from the process are often translated into a term called "beam quality," which roughly relates to how big a focused spot would be with respect to a diffraction-limited spot. Beam qualities of experimental high energy lasers have been known to be greater than 10 and even up to 20, but for operational system performance, a beam quality of 1.2 or so is sought and often achieved. The residual aberrations after the beam leaves the laser can often be further reduced by an external adaptive optics system.

Adaptive optics, although defined by "real-time correction," can be used to compensate static aberrations. Misalignments in optical systems impose correctable aberrations, as do imperfections or even damage to optical surfaces. These static aberrations can be compensated by permanently biasing the correctors in systems or by just ignoring them during construction and letting the active system measure and compensate them.

It must be remembered that large aberrations are difficult to compensate with limited deformable mirrors, but there is no physical reason why this isn't possible. Large optics, or lightweight deployed optics that have to be formed on-orbit, have special consequences. Moving from a 1-g environment to 0-g operation, the large optics can distort just because they are large. Deformations within optical tolerances appear by just turning or rotating a mirror on the ground, and polishing and aligning can remove most of the errors, but launch and deployment can put a lot back in. Adaptive optics in one form or another are needed to maintain a stable optical figure between manufacture and operation. If the Hubble Space Telescope had had a basic adaptive optics system, the spherical aberration found only after launch could have been removed, thereby avoiding a costly repair mission. (20/20 hindsight is a wonderful thing.)

Adaptive optics can be used in imaging systems and laser propagation systems. In some cases two-way light paths should be compensated. In laser communications, often a two-way path, the distortions along the path can be measured and compensated. It is possible, though not necessary at this juncture, to compensate the laser of a laser guide star. If the atmosphere is so bad or the laser guide star aperture is so small that a reasonably small point star cannot be created, the upgoing laser point can be measured and, using adaptive optics,

reduced to a manageable size that reduces wavefront error. Although this complicates the system — actually greatly complicates the system — it may be needed for some future applications.

Configurations of adaptive optics systems

An adaptive optics system made of the principal components previously stated — the wavefront sensor, the deformable mirror, and the control computer — can be configured in a number of ways. When the system is used to correct for the aberrations in a laser beam and the path to the target is not a problem (like the vacuum propagation of space), a local loop beam cleanup system is put together (see Fig. 6.1).

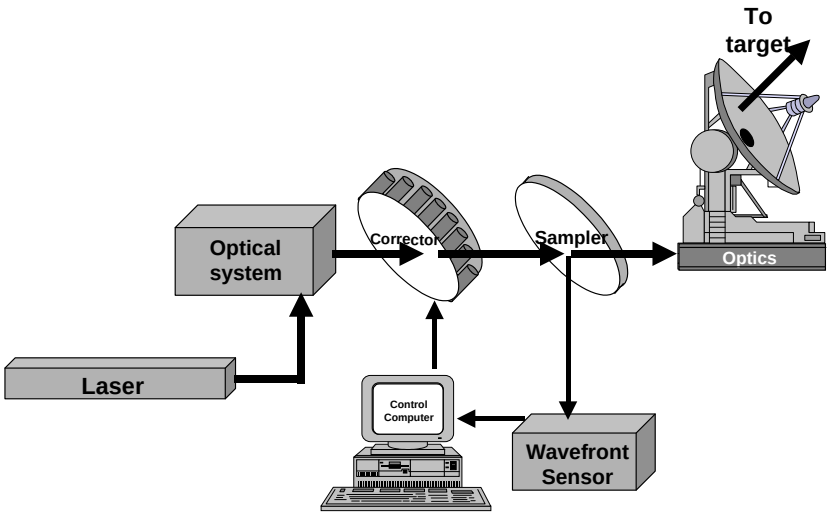

Fig. 6.1 An outgoing wave local loop beam cleanup system.

In the outgoing wave scheme, the last element in the beam train prior to propagation to the target is the wavefront sampler. This beamsplitter removes a small portion of the beam and sends it to the wavefront sensor, and then it and the computer try to reduce the aberrations that it senses by telling the deformable mirror to put the conjugate phase on the beam. When the wavefront sensor detects a plane wave, the mirror is commanded to stay there. As the laser or optics change, the wavefront sensor sees the change and *closes the loop* by applying the appropriate correction to the mirror. The outgoing wave local loop beam cleanup system is the configuration of the US Air Force's Space-Based Laser (SBL).

By turning the components around, so to speak, an adaptive optics imaging system can be configured (Fig. 6.2). The source, rather than being a laser, is

a natural star, an artificial guide star, or reflection from the target object. The light passes through the entire imaging system, but right before the focusing optics for the image sensor a small portion of the beam is split away and sent to the wavefront sensor. The efficiency of the beamsplitter can be adjusted here to balance the photons returned for the image and those needed for the wavefront sensor. The balance is often a design variable, with better phase correction possible with brighter sources but longer image sensor integration times required for dim sources.

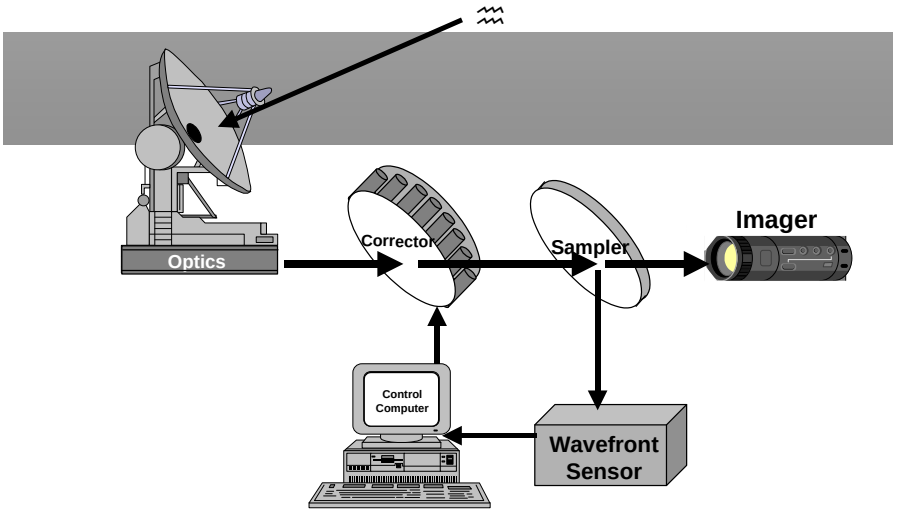

Fig. 6.2 An adaptive optics imaging system in the most common configuration.

There are some cases where the aberrations can change, but they are predictable. The thermal distortion from a laser turned on for a few minutes or hours or the air flow across an imaging window that changes with velocity, are a few possibilities. With a system like that in Fig. 6.3, a computer somewhere can determine what corrections should be applied to the mirror without having a direct wavefront sensor. These corrections can be applied open-loop to the mirror and then, just to be sure, a figure sensor can be looking at the mirror surface to make certain that the right correction is being applied.

A large class of adaptive optics systems uses the return from the target for wavefront sensing; astronomy with laser guide stars almost falls into this class, except that the target (the science object) is not the beacon for the wavefront sensor. A laser propagation system like that shown in Fig. 6.4 can sense the return from the target. If the propagation path is a vacuum, the target return can be used for sensing the distortions in the optics themselves. If the propagation path does have aberrations, all those plus the optics are sensed. The target return

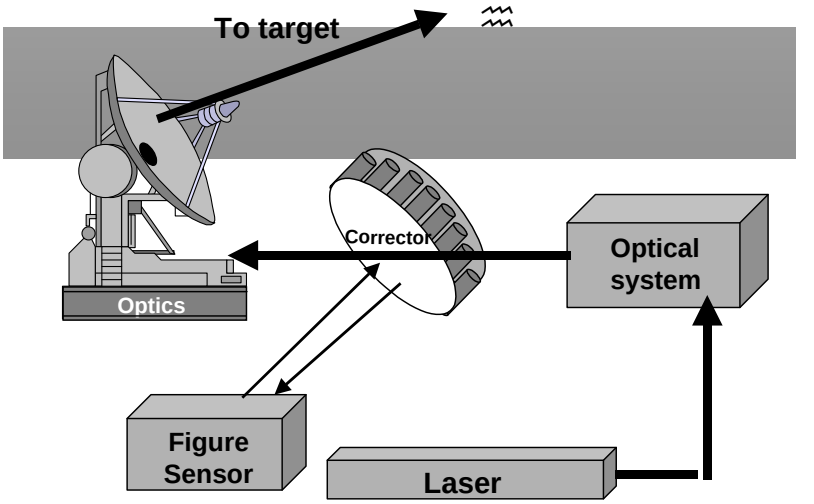

Fig. 6.3 A predictive adaptive optics laser propagation system. To monitor the performance and surface of the deformable mirror, a figure sensor is included.

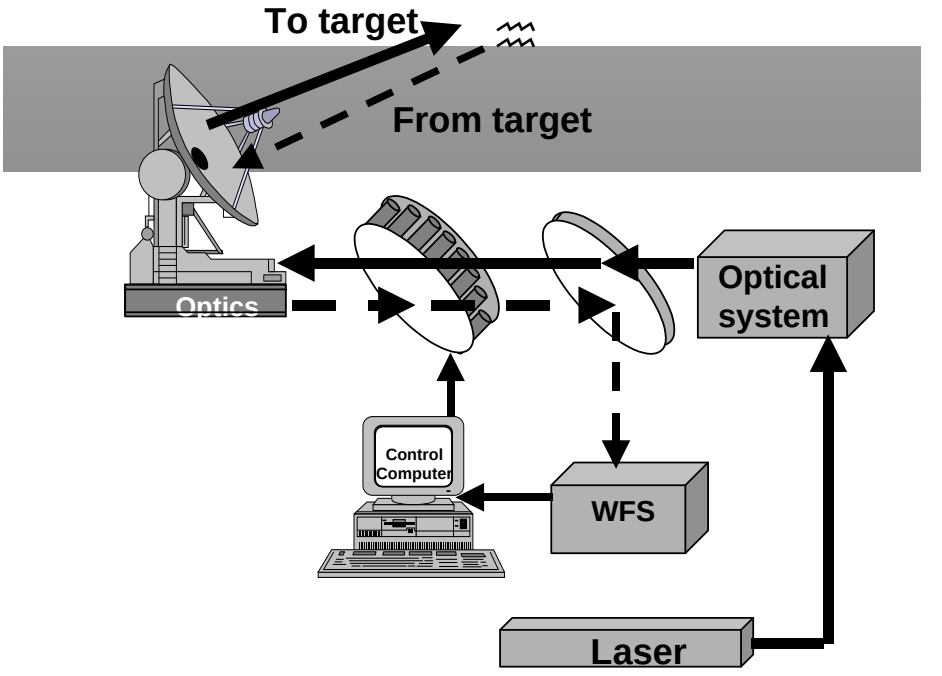

Fig. 6.4 A return wave adaptive optics laser propagation system with a shared aperture.

does not have to be the same wavelength as the laser; the corrections can have compensating gain factors in their algorithms. The target return wave can be infrared while the laser is visible; the heat from the target can be sensed in the same manner as the laser except with IR detectors.

The adaptive optics wavefront sensor path system can use the same aperture as the outgoing laser, as shown in Fig. 6.4 (shared aperture). It can share part of it (spatial sharing), use a different wavelength and split the two (spectral sharing), or flip from one path to another with movable optics (temporal sharing). It also can use different apertures for sensing and laser projection (separate aperture), a concept that was the basis of coherent optical adaptive techniques (COAT). In this method, optical tags were placed on various portions of an outgoing laser beam with a fast dither mirror and the glint (a small bright specular return) from the target was decoded to determine which part of the beam was in phase and which was out of phase. A separate control mirror was then used to correct the phase to maximize the energy deposited on the target glint.

One other variation is notable. The target, like a satellite or a distant ground object, can be illuminated with a separate laser, and the incoherent return from it can be detected through a shared aperture configuration and used for wavefront sensing (see Fig. 6.5). The quality of the correction is, of course, dependent upon the signal return, and therefore laser power is important. Most simple laser radar analyses can be applied to this situation. The conclusions are not at all surprising: the more light you send out, the more you get back.

Putting little building blocks together to make an adaptive optics system seems so easy ... on paper. But when it comes time to do it in the laboratory or the observatory, there are physical principles to worry about — real life difficulties with space, time, and funds — and because everything doesn't always work as advertised, the blessings of Murphy's law (whenever something can go wrong, it will).

Looking at Fig. 6.6, we can point out some things that need to be considered in the real world of adaptive optics systems. First, we can't ignore the conjugate pupil problem — adaptive optics systems want to apply a phase correction to a wavefront. We must measure the wavefront with one instrument at one point in the beam path and correct it at another, where we can put our deformable mirror. The phase that we correct should be the same as the one we measured.

Except by sheer accident, the only places where the phases are the same in an optical propagation path occur when the two places are optical conjugates. Sometimes, in fact a lot of times, the plane that we want to conjugate (the primary mirror aperture) is neither the wavefront sensor plane nor the deformable mirror plane. Therefore, we have to optically map the aperture, using pupil

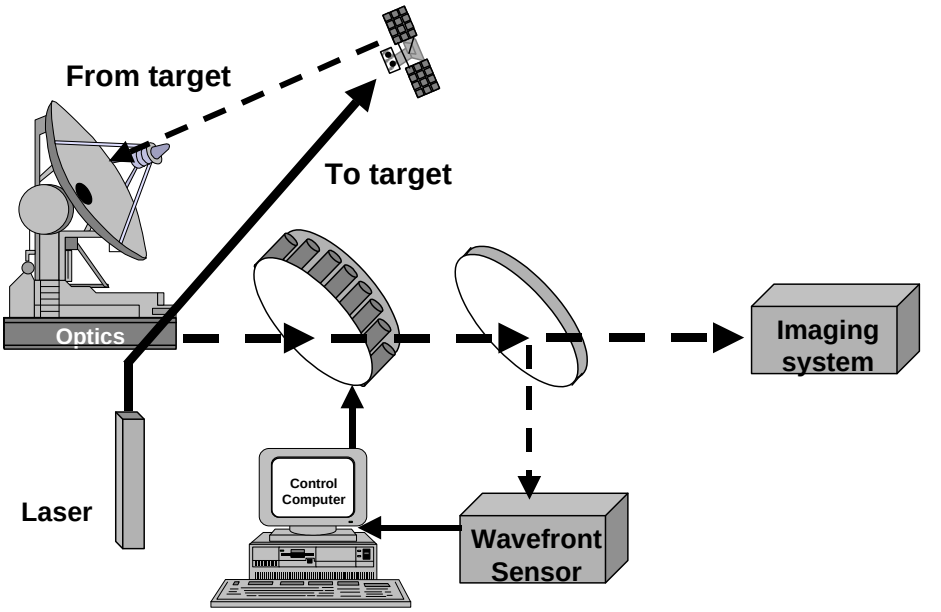

Fig. 6.5 A return wave, illuminator adaptive optics imaging system.

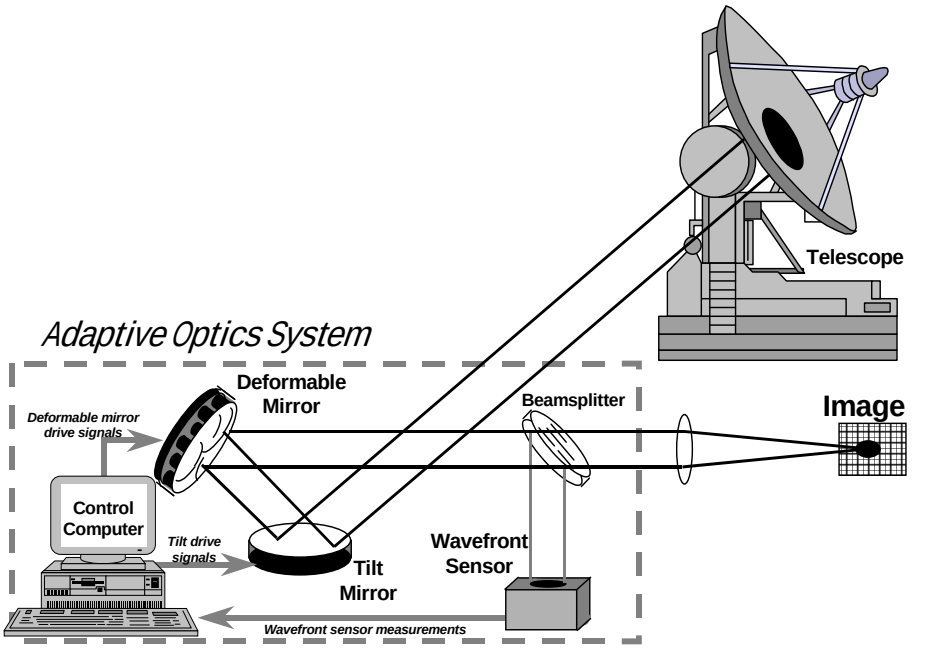

Fig. 6.6 System engineering concerns. Issues like pupil reimaging and rotation, spatial registration between the wavefront sensor and the deformable mirror, and stray light from the optics must be addressed.

reimaging optics, to a plane in the beam train where we can put the wavefront sensor and the deformable mirror. This adds complexity to the system but allows it to work. For "perfection," we should have both the sensor and the mirror at conjugate planes, but in reality this is often impractical, so we place both the sensor and the mirror as close as possible to the conjugate plane, wherever it is.

Another problem of real concern is the registration of the wavefront sensor and the deformable mirror. The subapertures of the sensor measure the phase precisely at known positions of the beam. The control computer decodes the signals and tells the deformable mirror what to do, knowing the precise positions of actuators of the deformable mirror. If the actuators are not in the known locations with respect to the wavefront sensor subapertures, errors can creep in. We can usually tolerate some misregistration error, but each additional distraction makes our system perform less perfectly. The misregistration can come from many sources. Manual initial alignment of the two subsystems is important, as is vibration or a slow drift in the alignment during operation. Sometimes, when initial alignment remains steady, the misregistration can be calibrated out in the control system.

There are some cases where the misregistration is a direct result of system operation. For example, some telescope configurations will rotate the pupil as the telescope slews. Often the conjugate pupils and the image rotate at different angles. These can be corrected optically, with a *K-mirror* for example, and it may be possible to correct the motion in software with known rotations.

Many adaptive optics systems are retrofitted to an existing telescope. In so doing, it is necessary to provide adaptive optics compensation without adversely affecting the science image path. Because many observatories have expensive imaging and spectroscopic instrumentation designed for a specific F-number in the beam path, the adaptive optics must be able to intercept the beam, provide compensation, and then pass along the beam with the same focal ratio.

In astronomy, adaptive optics are supposed to improve the imagery. If the system is difficult to use or unreliable, the observers will experience a steep learning curve or a large amount of downtime. Cloud cover is bad enough; we shouldn't add a cumbersome electro-optical system to contribute to the frustrating loss of information. To make the adaptive optics system palatable to most users, it should have a simple user interface, not require a team of specially trained engineers to operate, and be able to be bypassed if something goes wrong. Simply put, operation and maintenance should not be an issue.

Bibliography

Robert K. Tyson, *Principles of Adaptive Optics*, 2nd ed., Academic Press, Boston (1997).

Warren J. Smith, *Modern Optical Engineering*, McGraw-Hill, New York (1990).

G. Rousset, J.-C. Fontanella, P. Kern, P. Lena, P. Gigan, F. Rigaut, J.-P. Gaffard, C. Boyer, P. Jagourel, F. Merkle, "Adaptive optics prototype system for infrared astronomy, I: system description," *Proc. SPIE* **1237**, pp. 336-344 (1990).

R. Q. Fugate, W. J. Wild, "Untwinkling the Stars – Part I," *Sky & Telescope* **87**, (5), pp. 24-31 (May 1994).

R. Q. Fugate, W. J. Wild, "Untwinkling the Stars – Part II," *Sky & Telescope* **87,** (6), pp. 20-27 (June 1994).

Chapter 7

Wavefront sensors — the eyes

If life were simple, we could just go to our local hardware store and there between the mousetraps and the power tools would be an optical phasemeter. We could buy it and take it back to our lab and put it into the adaptive optics system. If only wavefront sensors could be that simple. The problem occurs when we find out that at the higher frequencies of visible and infrared radiation, the phase of an electric field does not directly interact with matter in a way that we can measure. In fact, the amplitude doesn't either (its squared magnitude does).

We can observe the phase of a beam only indirectly. It was this difficulty that limited early adaptive optics engineers. For centuries it was known that two beams can interfere with each other because they are waves. If the beams were coherent, or nearly so, the time average difference in their phases would show up as constructive or destructive interference fringes, or something in-between. These would be a good indicator of the phase difference. If one of the beam's phases were known, or arbitrarily assigned a reference zero, then the intensity from the fringes indicated the phase of the other. This has been the principle used for optical testing since the time of Galileo.

To make the jump into 20th century electro-optics, we had to have a fast way to look at the fringes, particularly from a beam that passes through the atmosphere. The development of electro-optic detectors provided just that, a way to convert the short term intensity of light into an electronic signal, and therefore its wavefront.

It would all have been easier if we could just invert the Frauhofer diffraction integral which relates the focused intensity pattern *I(x,y)* to the Fourier transform of the field [amplitude $A(\alpha,\beta)$ and phase $\phi(\alpha,\beta)$] at the input pupil:

$$I(x,y) = \left|U_2(x,y)\right|^2$$
$$= \left| C \iint\limits_{\text{Pupil}} A_1(\alpha,\beta)\exp\left[-\mathrm{i}\phi(\alpha,\beta)\right]\exp\left[\frac{-\mathrm{i}k}{z}\left(\alpha x+\beta y\right)\right]\mathrm{d}\alpha\mathrm{d}\beta \right|^2 . \qquad (7.1)$$

But it was not to be so. For any given intensity pattern (not really *any* pattern but a whole, whole lot of them), there are many different phases in the electric field that would match it. Mathematically speaking, the magnitude of the Fourier transform is not invertible, and the phase relationship is not unique. This interesting fact has been the subject of a lot of research into the field of *phase retrieval*, but the physical limits still apply. One just cannot recover the phase from a single intensity pattern, or a single modulation transfer function, or any other single relationship, in the general sense.

The breakthrough comes when we realize that we can break the beam up into smaller and smaller pieces. We don't, necessarily, physically break the beam up; we do it conceptually. The wavefront of the beam, which at each point can be described as the shape of the phase surface with a line normal to it, can be determined if we can measure all those little lines normal to it. In other words, if we can break apart the beam to where we can measure the tilt (or propagation direction) of each little beamlet, we then have the information necessary to form a wavefront. This is the principle of the Hartmann test, which we use as a basis for a Hartmann wavefront sensor.

Measuring tilt

If a beam passes through an aperture, the light can be focused down to a spot. If the beam is coming straight into the aperture, the spot is formed along the optical axis of the aperture, normal to the plane of the aperture. If there is an average tilt in the beam, or a plane best-fit to the phase of the beam is tilted with respect to the plane of the aperture, the spot position will be offset. For all intents and purposes, the spot position, actually its intensity weighted centroid, is shifted by an amount linearly proportional to the tilt.

By placing a detector in the focal plane, preferably one which responds to the position of the spot, we have a direct measurement of beam tilt. We now have a clever way of performing beam alignment or overall beam tilt measurement. When we have lots of these all looking at small portions of the same beam, we have a wavefront sensor.

The quadcell

There are a number of electro-optic detectors that can be used to measure the position of a focused spot. The silicon lateral effect photodiode is one. The electrical current output is proportional to the spot position to within 0.1%.

Another sensor configuration, using four closely spaced detectors, is called a quadcell. It can be read out to compare the signals from the four detectors to determine the centroid of the spot. When each detector of the quadcell has some energy falling on it, the algorithm is quite simple. From Fig. 7.2(a), we see that the spot energy should cover some of each detector. By averaging the energy the

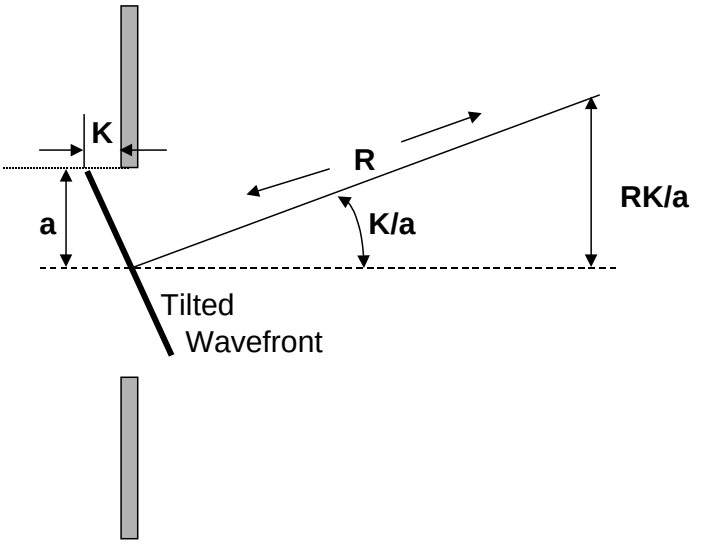

Fig. 7.1 A beam passing through an aperture with a tilted wavefront follows geometry. Even when it is focused to a spot, the centroid of the spot is positioned in proportion to the wavefront tilt.

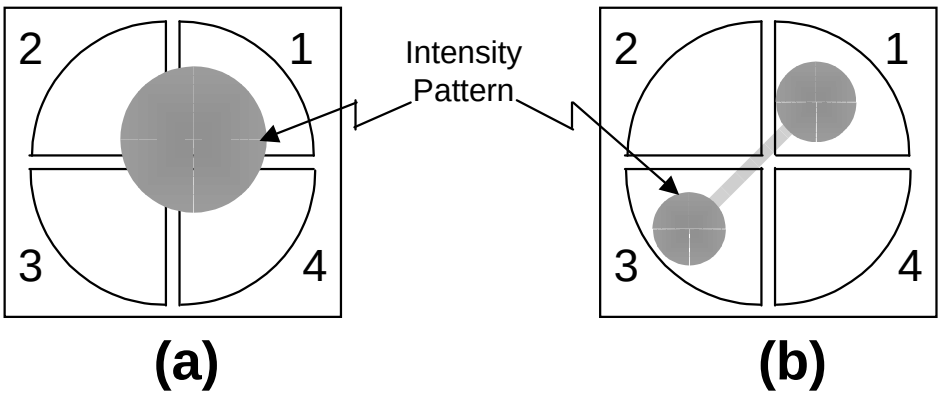

Fig 7.2 Quadcell with focused spot on it. (a) It is well-conditioned, unresolved, and covering all four detectors but not overflowing (b) The intensity distribution corresponds to a resolved image; the centroid no longer represents the wavefront tilt.

from the two leftmost detectors and subtracting it from the average of the rightmost detectors, we can determine its horizontal position, or plainly speaking, its centroid in the *x* direction. Similarly, we can average the difference of the top and bottom detector pairs and find its vertical position, its centroid in the *y* direction.

The accuracy is quite good as long as some measurable signal above the noise is detected in each of the four detector cells. When the optics that are focusing light

on the quadcell can resolve the object [Fig. 7.2(b)], the simple centroid calculation is corrupted; it no longer represents wavefront tilt. Special care must be taken for resolved objects or extended scenes.

Modern optical detectors such as CCD, CID, or CMOS arrays can serve as high resolution quadcells. By using many more than four cells, and assuming subpixel accuracy of the quadcell, we can achieve a larger dynamic range and greater accuracy with the two-dimensional focal plane arrays.

It is also possible to measure the tilt of a beam by placing a translucent mask or filter in the beam near the focused spot. The transparency of the mask should be linearly varying across its surface so that the intensity transmitted is proportional to the position of the spot on the mask. We split the beam into three arms, letting the first arm project a spot on a detector behind the horizontal mask, the second arm project a spot on a detector behind the vertical mask, and the third arm project a spot on a detector with no mask. See Fig. 7.3. The ratio of x-detector to total-detector gives the *x* position; the ratio of y-detector to total-detector gives the *y* position.

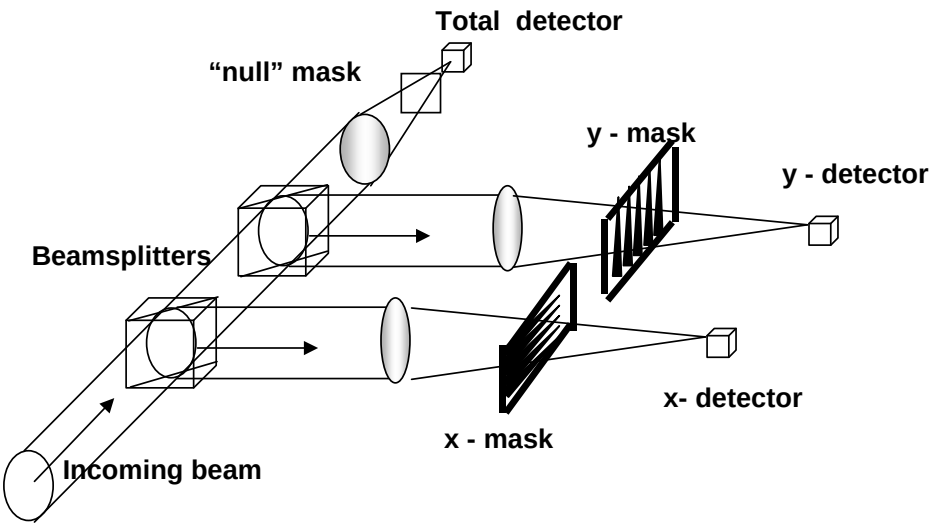

Fig. 7.3 Tilt of a beam is measured by comparing the intensity through various masks.

Measuring focus

Besides measuring tilt, the lowest-order wavefront mode, it is possible to directly measure focus as well. The Foucault knife-edge test is used for the purpose (see Fig. 7.4). In the test, we need only a bicell (two closely spaced detectors). We can take apart a finely crafted modern shaving system, formerly called a razor, and

extract the blade — or we can buy an optical knife edge designed for this purpose. In either case, with either budget, we place the knife edge so that the edge is parallel to the gap between the cells of the bicell and precisely at the desired focus of the beam. The more accurately we can do this, the more accurate our measurement will be. If the beam coming in is not focused correctly, as in (b) and (c) of Fig. 7.4, there will be a difference between the signals from the two cells that is proportional to the focus. The actual relationship between the radius of the reference spheres and the number of waves of defocus is related to the F-number of the system, the accuracy of the setup, the noise in the detectors, and the medical cost incurred when we try to shave with the used razor blade.

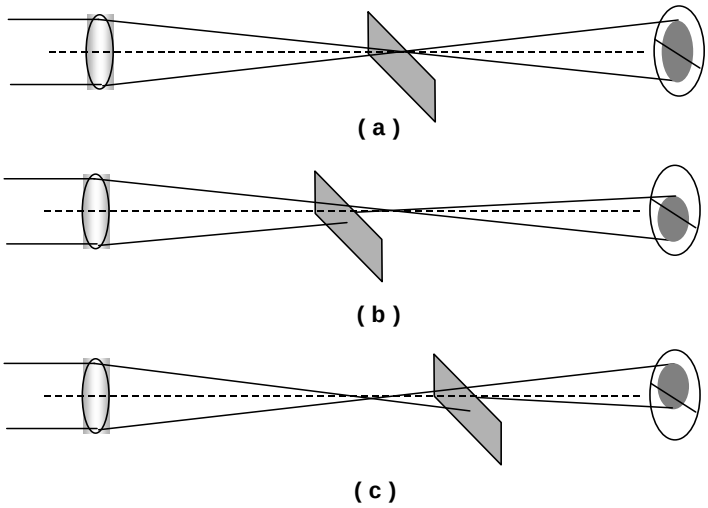

Fig. 7.4 The Foucault knife-edge test for focus measurement.

Interferometers

Instead of measuring the tilt of little pieces of the beam and then piecing them together, we can measure the phase of a beam indirectly using an optical interferometer. The intensity of each point on the interference pattern is proportional to the phase of the beam with respect to a reference.

The principle of interference was first discovered and reported by Thomas Young in about 1801. By shining a beam of light onto an opaque screen with two closely space slits, we can observe the pattern of light on a screen behind the slits (see Fig. 7.5).

Young was clever enough to figure out what was happening and by doing so made a pretty good argument for the wave nature of light. Young showed that the intensity of the light at any point in the pattern on the screen is a combination of the intensity from one slit added to the intensity of the other slit, and then added

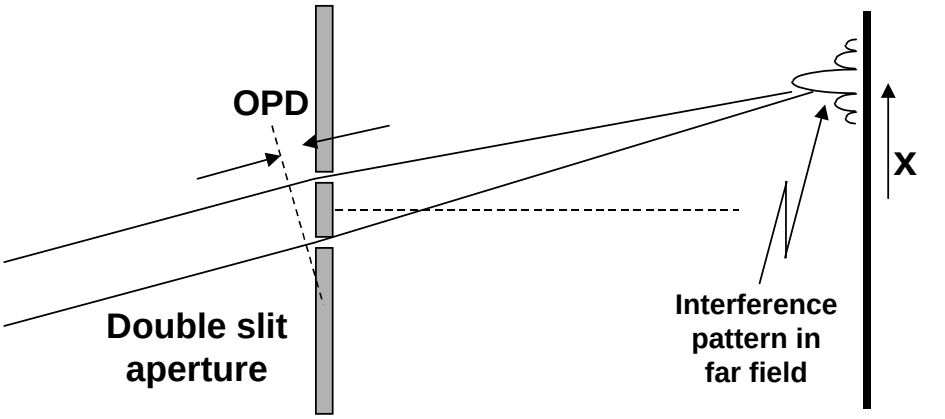

Fig. 7.5 Young's double slit experiment is used to demonstrate that light is a wave and exhibits interference effects.

to the combined intensities of the two slits multiplied by the cosine of the phase difference. Stated mathematically, the interference equation is

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \quad \cos\delta \quad . \tag{7.2}$$

When the two beams are of equal intensity and in-phase, the angle δ is zero and $I = 4I_1$. When they are out-of-phase, the angle is 180^o and $I = 0$.

Young's theory was in opposition to Newton's corpuscular, particle nature of light theory. Since we know now that both theories are correct, that light behaves as both a wave and a particle (a photon), we make use of both in adaptive optics. Light interference is used to create an intensity difference, or measure of phase, and photons are used to stimulate detectors and convert the phase information to electrical signals.

We can build an interferometer to measure phase in a number of ways. This is not a book on interferometers; there are plenty of good ones out there. This is a book on why we care about interferometers — because they can be used to measure the wavefront for our adaptive optics system.

A Twyman-Green interferometer is shaped like the T in Twyman (see Fig. 7.6). If Green had been principal author of the paper, the interferometer might be shaped like a G, but happily we don't have to explain how we curve the light.

The Twyman-Green interferometer works quite simply. A source of light is split by a beamsplitter. One part of the beam is reflected off the front surface of the splitter and goes toward mirror 1. That beam bounces back and goes through the beamsplitter toward a detector. The other half of the beam goes through the first

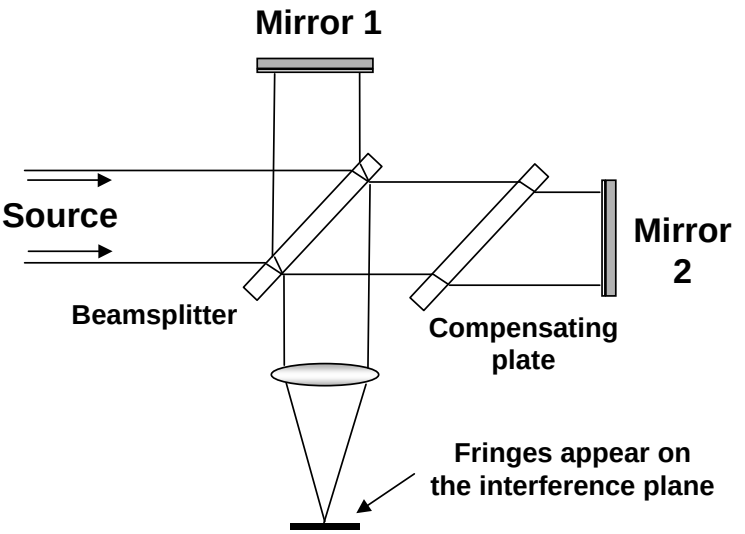

Fig. 7.6 A Twyman-Green interferometer.

beamsplitter and hits another mirror, where it reflects back to the beamsplitter and is reflected from that surface toward the detector. The intensity pattern on the detector shows the interference pattern, which measures the phase difference between the two paths, including the effects of each mirror. If one of the mirrors has some aberration and the other is essentially flat, the interferometer measures the aberration.

A lot of practical things are done to make the interference pattern a useful tool, such as precise alignment and making sure that there are no extra reflections from the beamsplitter surfaces.

Another major type of interferometer is the Mach-Zehnder interferometer. The interferometer is shaped neither like an M nor a Z but more like a squared-off circle (see Fig. 7.7). The principle of operation is similar to the Twyman-Green interferometer. A beam is split into two paths, one of which serves as a reference, and they are combined again to form an interference pattern that measures the difference in phase of the two paths.

The reference legs of the Twyman-Green and Mach-Zehnder interferometers depend upon our ability to make their mirrors very flat, at least with respect to what is being measured. Another type of interferometer, the Smartt point diffraction interferometer, makes its own reference leg by capturing a small part of the beam where the phase would be flat in the region and expanding that into a plane wave, which it then uses as the reference (see Fig. 7.8). This self-referencing interferometer makes the plane wave in the same manner as a spatial filtered beam.

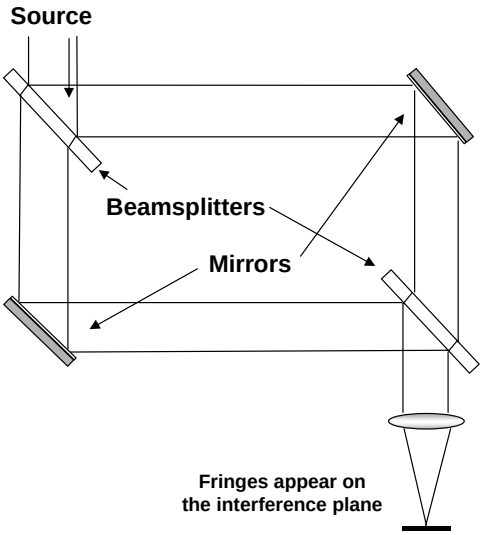

Fig. 7.7 A Mach-Zehnder interferometer.

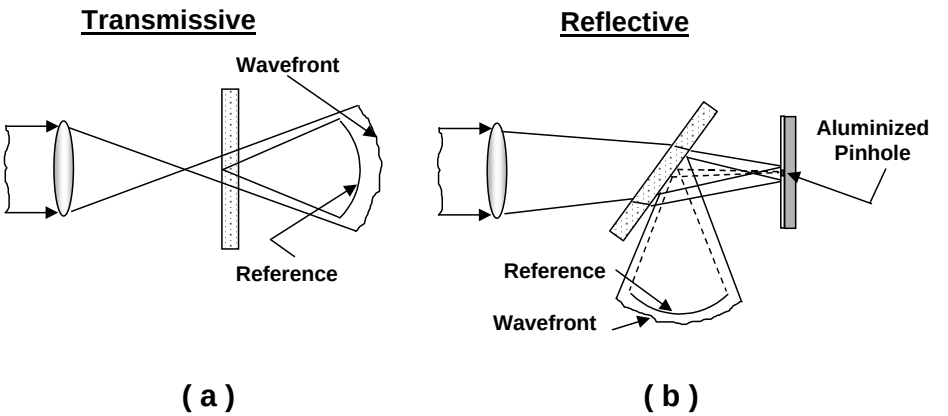

Fig. 7.8 Smartt point diffraction interferometer. A self-referencing interferometer uses part of the object wavefront as a reference source.

The principle of the shearing interferometer

The above interferometers make use of a plane wave reference and interfere the test beam with it. Although the Smartt interferometer is self-referencing, it still uses the plane wave reference so that the measured fringe intensities are proportional to the wavefront of the beam. Another useful way of self-referencing is *shearing* the beam. If we split the beam in a Mach-Zehnder configuration (see Fig. 7.9), the reference is a beam shifted in one direction. The overlapping beams form an interference pattern whose measured fringe intensities are proportional to the slope, or the derivative, of the wavefront. To implement the procedure, we actually split the beam into two beams first and shear each one in orthogonal directions. They can be *x*-*y*, or radial-azimuthal, or any other weird coordinate system that makes sense in the optical layout.

The shearing interferometer works by physically implementing the textbook definition of the derivative of a function. The interference (subtraction of the wavefronts) of two sheared beams is exactly the derivative of the wavefront in the direction of the shear, when the shear limit approaches zero. Because we can't physically produce a zero shear, or really because it wouldn't do any good, the error of the shearing interferometer is related to the shear distance. The mathematics behind this is pretty simple; Fig. 7.10 shows the situation graphically. In one dimension, the wavefront is $\Phi(x)$. A sheared wavefront, with shear distance s, is $\Phi(x–s)$. The interference pattern is normalized by the shear distance:

$$I = \left[\Phi(x) - \Phi(x-s)\right] \,/\, s \ . \tag{7.3}$$

When this function approaches the limit $s \to 0$, the result is the first derivative of the wavefront at x:

$$I = \left\{ \left[\Phi(x) - \Phi(x-s)\right] \,/\, s \right\}_{\lim\ s\to 0} \equiv \frac{\mathrm{d}\Phi(x)}{\mathrm{d}x} . \tag{7.4}$$

The practical operation of the shearing interferometer can get complex because the measurement of interference fringes and their conversion to electrical signals must correspond to the accuracies and speeds required for aberration compensation. Devices have been built with ac modulation of the signals using a rotating grating, and some devices with fixed gratings for beam separation and shearing use state-of-the-art focal plane arrays.

Hartmann sensors

A noninterferometric type of wavefront sensor was first built in the early 1970s. It uses the classical Hartmann optical testing technique, looking at the position of the spots when a beam is masked by a screen with a lot of small holes. Instead of

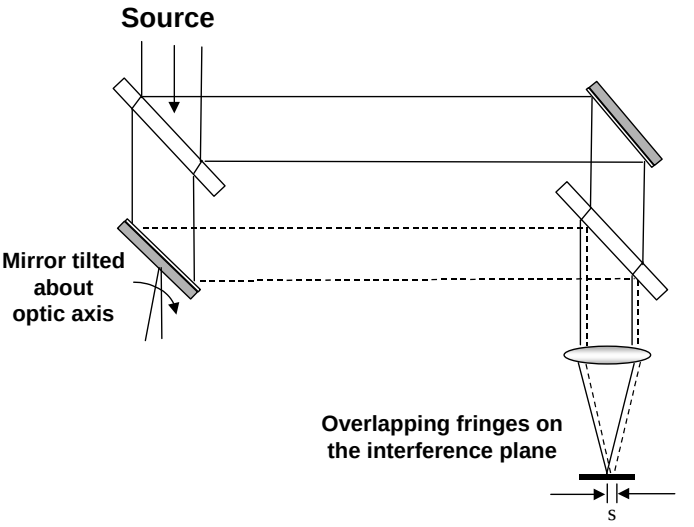

Fig. 7.9 A shearing interferometer is a modified Mach-Zehnder configuration with a mirror tilted to produce shear.

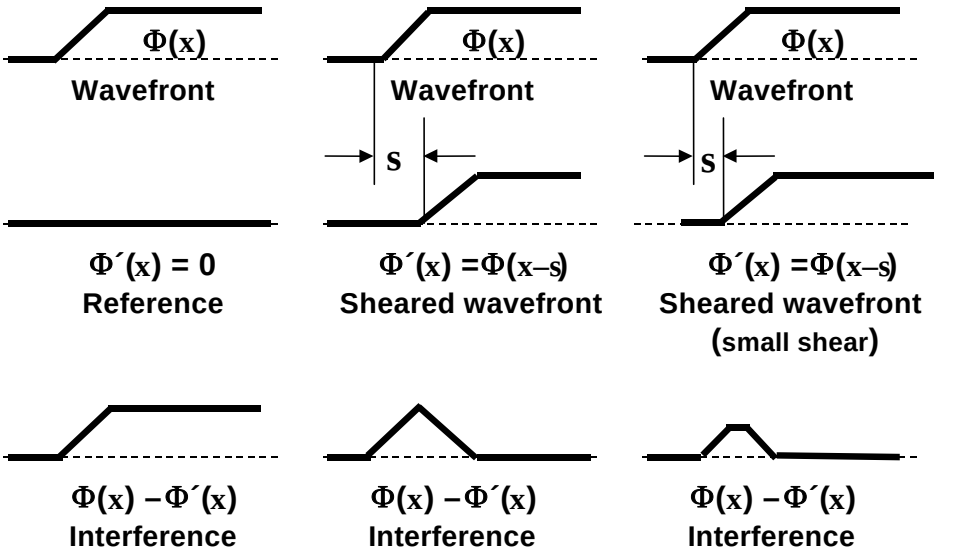

Fig. 7.10 The principle of the shearing interferometer. The interferometer optically takes the derivative of the wavefront.

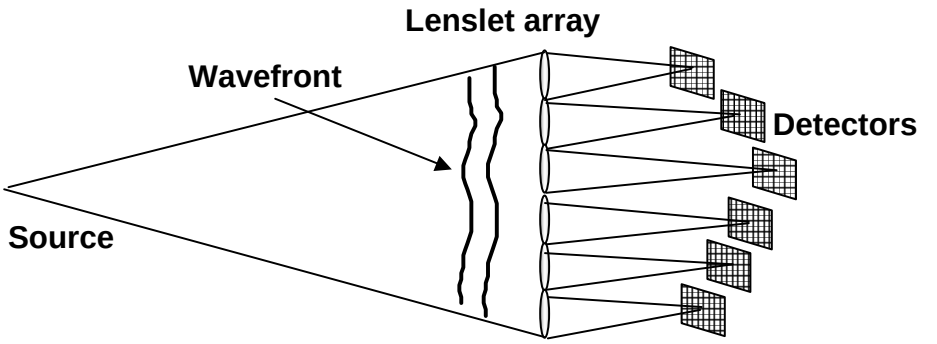

Fig. 7.11 The principle of the Shack-Hartmann sensor. The beam in each subaperture is focused onto a detector. The position of the centroid of the spot corresponds to the tilt in the subaperture.

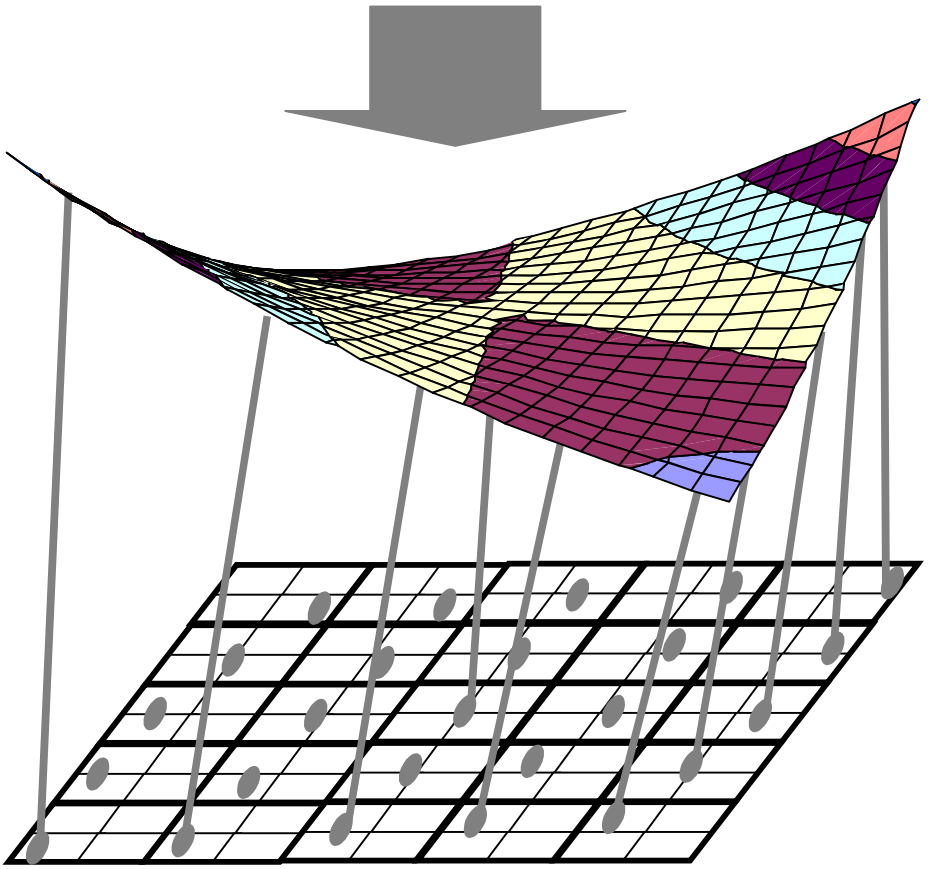

Fig. 7.12 A wavefront on a Shack-Hartmann sensor and the resulting spot diagram. Decoding the spot positions and mapping them to the wavefront corrector is the heart of the wavefront reconstructor problem.

small holes, which cause the loss of a lot of light, Shack and Platt[1] decided that a small array of lenses could serve as the divider to produce a set of spots. These subapertures each produce a spot of light on a detector, such as a quadcell or CCD array, where the local tilt of the beam is measured (see Figs. 7.11 and 7.12). If there are enough spots, covering enough of the beam, the entire wavefront can be reconstructed from the signals from the detectors. The device is variously called a Shack-Hartmann sensor, a Hartmann-Shack sensor, or just a Hartmann wavefront sensor.

Because both shearing interferometers and Hartmann sensors produce signals proportional to the local derivative (slope) of the wavefront, there have been many comparisons. Which is better? The answer depends upon the noise conditions of the sensor, the complexity of the setup and calibration, and the reliability required. For example, radial shearing interferometers have a rotating grating, while most Hartmann sensors don't have any moving mechanical parts. This may seem like a major disadvantage for the interferometer, but the alignment and calibration requirements of the Hartmann sensor, specifically the need to know where the *centered, zero tilt spot* should fall, make it more than trivial to make.

Over the years there have been various versions of the Shack-Hartmann sensor that make use of ongoing technological advancement. In some implementations, inventors have added a separate plane wave reference beam to define that ubiquitous *centered, zero tilt spot* position. Sometimes, for instance in the Integrated Imaging Irradiance ("I-cubed") sensor, the plane wave reference was modulated and nutated on a quadcell to detect the difference from the wavefront signal. In other sensors, pixels of a focal plane array are used for calibration and as a guard band, or dead band, to avoid overlap and confusion of signals. The actual implementation depends on the bottom line. How accurate does it have to be? How much dynamic range should it have? How fast does it need to be? Or the question most laboratory managers ask — how can we possibly afford *that*?

Curvature sensing

One interesting development to reduce some costs in an adaptive optics system has been the development of the wavefront curvature sensor. [It reduces cost by simplifying the control system design when used with a bimorph deformable mirror (see chapter 8).]

If an optical beam with aberrations is focused onto a detector such as a focal plane array, the intensity distribution is the magnitude of the Fourier transform of the field at the pupil. If the beam is slightly out of focus, the intensity pattern will be blurred. The intensity patterns of two beams, both out of focus by the same amount but of different signs, can be subtracted to reveal information about the

[1] R. V. Shack, B. C. Platt, *J. Opt. Soc. Am.* **61**, 656, (abstract only) 1971.

phase aberrations. In this image plane sensor, it is not the wavefront slope (first derivative), like from the Hartmann sensor, but instead the second derivative (the curvature) that can be determined. A curvature sensor, like that shown in Fig. 7.13, optically solves Poisson's equation. The difference of the two intensity patterns $I_1(r)$ and $I_2(r)$ is the difference of the wavefront curvature at r and the derivative of the wavefront $\mathrm{d}\phi/\mathrm{d}n$ at r in the radial direction:

$$I_1(r) - I_2(r) = C\left[\nabla^2\phi(r) \quad - \quad {\mathrm{d}\phi(r)}\Big/{\mathrm{d}n}\right]. \tag{7.5}$$

The slope at the edge of the subaperture $\mathrm{d}\phi/\mathrm{d}n$ is a measurable boundary condition and the constant C depends upon the response of the detector and the actual defocus of the two intensity patterns.

Curvature sensors have the advantage of being able to work with extended objects or resolved scenes — the processing is a bit more complicated, but the wavefront information is attainable.

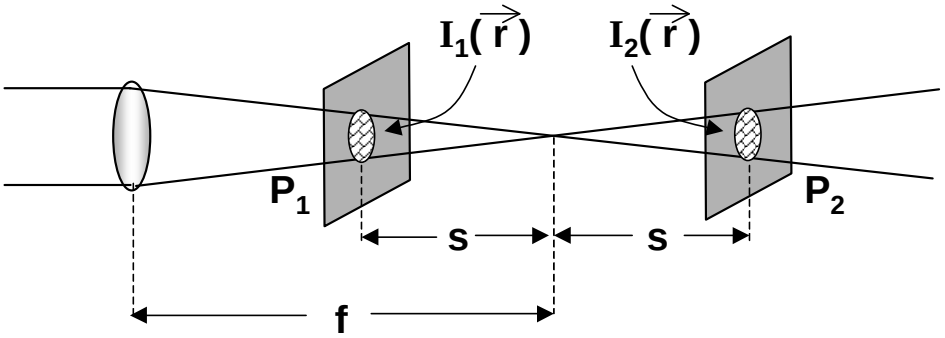

Fig. 7.13 A wavefront curvature sensor. The difference in intensities in the offset image planes corresponds to the second derivative of the wavefront.

Image sharpening

Another sensor that processes resolved scenes for wavefront information uses an image sharpening technique. This system acts much like the human eye-brain processor. When an image is blurry, fuzzy, or not sharp, it is recognized by the sensor as being incorrect. Adjustments of the deformable mirror or other active optics can bring the image into sharpness. If there is little information about the object, like a scene from a high altitude surveillance satellite, the processing can be slow. On the other hand, when the object is known, like a star being imaged from the ground, the sharpening of the image can be suitable for atmospheric turbulence compensation.

How does the sensor know the image is not sharp? It measures the sharpness from the intensity signals on the detector. The sharpness *Sh* has various definitions. Some, like the square of the integrated intensity,

$$Sh = \int I(x,y)^2 \,\mathrm{d}x\mathrm{d}y\,, \tag{7.6}$$

work quite well with a point source since it basically just maximizes the power in the bucket. The region of integration is variable so that sensor noise, in areas of low intensity, does not corrupt the measurement. There are countless other measures of sharpness. Some, like

$$Sh = -\int I(x,y)\ln\left[I(x,y)\right]\mathrm{d}x\mathrm{d}y \quad , \tag{7.7}$$

are useful for large dynamic range images.

The exact algorithm for driving the deformable mirror has many variations. Most multiparameter equation solvers, such as gradient search, Simplex, simulated annealing, or genetic algorithms will work; the problem lies in making them efficient enough to work in real time with real adaptive optical hardware. The conditions of the object and the image, the processing time, and the allowed residual wavefront error need to be considered prior to choosing a particular method.

Wavefront sensor requirements

In almost all adaptive optics systems, it is necessary to determine the wavefront sensor requirements before one begins to build or buy a sensor. This is fairly straightforward — assumptions can be made, and they are at least as valuable as wild guesses. The measurement range, the dynamic range of the wavefront sensor, is roughly 2.5 times the expected standard deviation of the wavefront error that we need to measure, $R = \pm 2.5\ \sigma_{\mathrm{m}}$. If Gaussian statistics are assumed, which is usually the case since other statistics are much more confusing, we anticipate that the wavefront error will be within 2.5 standard deviations. For atmospheric turbulence compensation, the standard deviation is

$$\sigma_{\mathrm{m}} = \sqrt{0.17\left(\frac{d_{\mathrm{sub}}}{r_0}\right)^{5/3}} \qquad [\mathrm{waves}] \quad , \tag{7.8}$$

where d_{sub} is the size of the subaperture.

Detectors

In principle, anything that detects the presence or absence of light can be used as a detector for a wavefront sensor. In reality, you might end up with a very poorly engineered system that doesn't work and costs a fortune. To do the job correctly, you must consider the noise environment, the expected intensity range of the

wavefront beacon signal and background, and the optical wavelength or band of operation.

For visible wavelengths, silicon detectors and silicon CCD arrays produce on the order of 0.6 A/W of optical power. They have fast risetimes to respond to rapidly changing wavefronts, and many have low noise characteristics. Developments change quite rapidly in the fast-growing field. Consult your favorite buyer's guide, or better yet schedule a trip to a conference that has a trade show in an exotic vacation locale. Take your family along and spend the romantic evenings discussing detectivities and noise equivalent power.

When the wavelength turns to the infrared, the detector wars heat up even further. Wavefront sensors have seen the likes of pyroelectric detectors, solid state InSb detectors, cooled HgCdTe and PbS detectors, and detector arrays. Often the weak link in an adaptive optics system is the inadequacy of the detector, which is the first element to see the light of night, so to speak. Use what your budget will bear.

Beamsplitters and samplers

Before the wavefront sensor beacon light ever hits the detector, it will have to penetrate the optical system, probably through the telescope, and find its way to the wavefront sensor. The last element in the beam train before the wavefront sensor will be some sort of sampling optical element or beamsplitter (see Fig. 7.14).

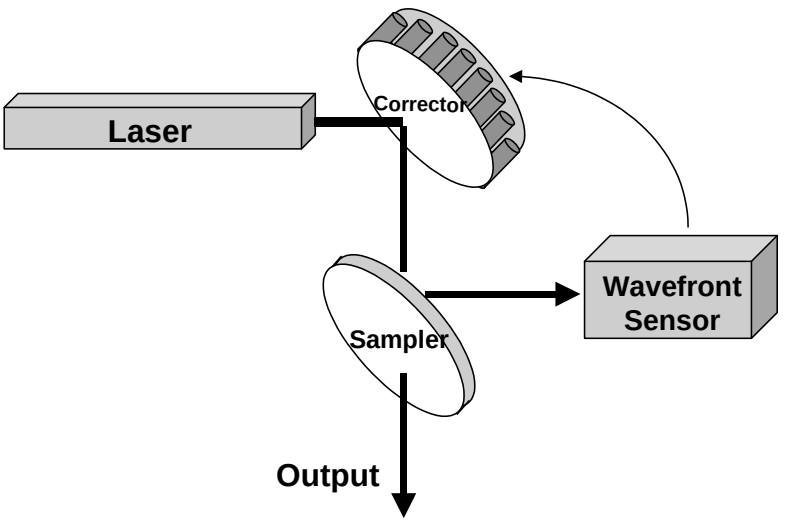

Fig. 7.14 Outgoing laser beam cleanup system showing sampler position.

Commercial beamsplitters provide a wide range of possibilities. They are usually thin glass or plastic elements, hopefully with very little distortion, that are coated such that the amount of light you want to sample is sent to the wavefront sensor while the rest is sent to the science camera, image collector, or out in the direction of the propagating laser beam.

For some systems, where a wide range of objects is being observed, electronically or mechanically variable beamsplitters that incorporate filters with the beamsplitting function can simplify the optical system.

For real sophisticated high power systems, where the laser beam will melt the beamsplitter in a millisecond, other solutions are needed. In many such systems, optical elements are used with high power coatings that reflect the laser but transmit or diffract the beacon light. These come in various forms, like wedges, buried wedges, buried gratings, and grating rhombs. All have found their place in some high power adaptive optics systems. Some also have found their place in the trash since a high power element that must also transmit other wavelengths tests the engineering capabilities of the day.

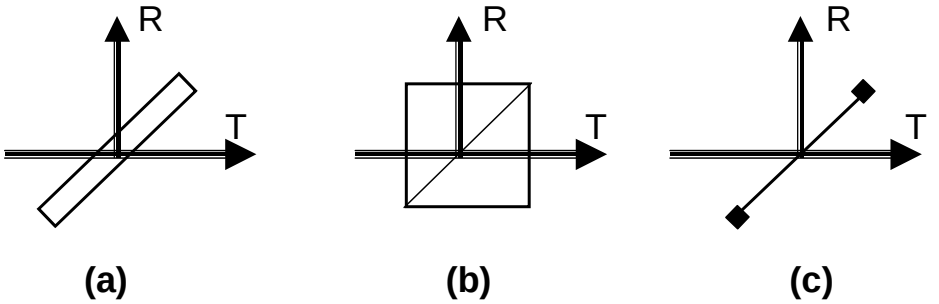

Fig. 7.15 Plate, cube, and pellicle beamsplitters.

The beamsplitting function in the system is defined by the optical path, the wavelengths, the power requirements, the stability, and the environment. However, in all cases of wavefront sensor beamsplitters (Fig. 7.15), there is one cardinal rule: Whether a beamsplitter is a flat plate or a cube, polarizing or nonpolarizing, grating or wedge, *low distortion* in both paths is a must. If the wavefront is to be corrected, we can't have distortion that isn't in the other path creeping into the wavefront sensor, nor can we have distortion in the other path that the wavefront sensor never gets to see.

"Success is never final; failure is never fatal." — Winston Churchill

Bibliography

Joseph M. Geary, *Introduction to Wavefront Sensors*, SPIE Optical Engineering Press, Bellingham, WA (1995).

Multiple authors, *Special issue on adaptive optics*, Lincoln Laboratory J. **5**, (1) (Spring 1992).

J. C. Wyant, "Use of an ac heterodyne lateral shear interferometer with real-time wavefront correction systems," *Appl. Opt.* **14**, pp. 2622-2626 (1975).

J. W. Hardy, A. J. MacGovern, "Shearing interferometry: a flexible technique for wavefront measurement," *Proc. SPIE* **816**, pp. 180-195 (1987).

F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics," *Appl. Opt.* **27**, pp. 1223-1225 (1988).

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

Chapter 8

Deformable mirrors — the hands

A lot of people use the terms *deformable mirror* and *adaptive optics* interchangeably — "Let's put adaptive optics in the beam." — with no regard for the fact that the singular form of the noun *optics*, — "the adaptive optic goes right before the beamsplitter" — is almost never used anymore. The trouble with the terminology is that there is no need for the confusion. Adaptive optics is a closed loop system; a deformable mirror is a special type of active mirror. Saying "The active mirror goes right before the beamsplitter" is okay — as long as it really doesn't go *after* the beamsplitter.

Another popular misconception is describing an adaptive optics system by referring only to the deformable mirror. "Oh, you're going to add adaptive optics. How many actuators?" Actually, if the system was designed cleverly and only one type of aberration is going to be corrected, focus for instance, you might have a system with 2000 subapertures in the wavefront sensor, a special purpose microprocessor, a million dollar beamsplitter – but one actuator! (This system is not recommended by me; it is for satirical purposes only.)

A deformable mirror is one part of an adaptive optics system. Sometimes it is the most important part. Sometimes it is the most difficult part to build. Sometimes it is the most expensive part. But in any case, without something to control the deformations on the DM, it is a rather passive part.

Types of deformable mirrors

Adaptive optics is a multidisciplinary field, encompassing optics, electronics, computers, physics, chemistry, and accounting. The physics and chemistry parts come in when we describe many types of DMs. The two main groupings are inertial and non-inertial devices. The inertia nomenclature arises from what we consider to be macroscopic to us. Some non-inertial devices, such as electro-optic modulators and acousto-optic modulators, have molecules moving around. Molecules have inertia, but we normally don't detect it on the scale of an adaptive optics system. On the other hand, reflective optical systems, like tilt mirrors and deformable mirrors, have big things that move. Mirror faceplates can

move, and we can actually see the motion, not just the effects of the motion. Clearly, the most common active mirrors in use today are inertial.

Tilt mirrors

The basic thing an adaptive optics system must do is point a beam of light. Many electro-optic systems, such as complicated trackers and missile seekers, must point a beam. Everyday devices such as supermarket scanners, CD players, and laser printers have systems to point a beam. In the adaptive optics world, when atmospheric turbulence is the most common enemy and tilt is the most common threat of that enemy, we need a stand-alone tilt mirror to compensate for the beam tilt. Many off-the-shelf devices exist that cover a wide range of tilt stroke, accuracy, and bandwidth.

Fig. 8.1 Physik Instrumente (PI) GmbH&Co. (Waldbronn, Germany) makes a number of tilt or scanning mirrors.

Many tilt mirrors have driver systems with their own mirror position sensors and feedback loops so that a command is precisely carried out. Some are even reactionless so that their motion does not show up as a source of vibration for the rest of the delicate system.

We can calculate the requirements of the tilt mirror of an adaptive optics system used for atmospheric turbulence compensation. The bandwidth, or speed and acceleration of the mirror, derive from the tilt Greenwood frequency of chapter 3. The mirror should be able to respond so that it can keep up with the commands sent to it. A tilt mirror should move to its commanded position and settle down

within its specified accuracy in less than half the time constant of the control system.

The maximum atmospheric tilt that the mirror should be able to remove is about $M_{\text{tilt}} = \pm 2.5\sigma_{\text{tilt}}$, where σ_{tilt} is the standard deviation of atmospheric tilt motion:

$$\sigma_{\text{tilt}} = \sqrt{0.184\left(\frac{D}{r_0}\right)^{5/3}\left(\frac{\lambda}{D}\right)^2} \quad . \tag{8.1}$$

D is the aperture diameter of the telescope primary mirror and r_0 is Fried's coherence length. Tilt isn't removed by moving the large primary mirror back and forth; it is removed by a smaller mirror in the beam train. Because of the telescope:beam-train magnification, and with the added factor of ½ because motion of a tilt mirror results in twice the angular tilt motion of the beam, the total stroke of the tilt mirror should be at least

$$\text{Stroke} = \frac{1}{2} M_{\text{tilt}} \left(\frac{D_{\text{telescope}}}{D_{\text{tilt mirror}}} \right). \tag{8.2}$$

Tilt mirrors can normally move very fast. Kilohertz rates are not uncommon. But they can control only one optical mode — tilt. When the aberrations are of higher order than tilt, other correction devices come into use. A telescope may have a movable secondary mirror that can presumably maintain control over some tilt and some beam translation, and by moving along the optical axis, control focus.

Large primary mirrors can behave like a highly packed deformable mirror. Primary mirrors with hundreds of actuators to control gravitational sag and other distortions can also be used to control slow, very high spatial order wavefront changes. Continuous faceplate or segmented deformable mirrors can have both high speed and high spatial resolution. They seem like the ideal solution, and they come pretty close. They must be placed at an optical pupil of the system, which often takes some clever designing. They also must fit within the environmental volume of the system and the budget of the owner.

Deformable mirrors

Deformable mirrors come in a few flavors. Segmented mirrors have individual flat segments that move either in just an up-down piston mode or, with 3 or more actuators driving each segment, in tip-tilt and piston motion.

Continuous faceplate mirrors can have behind the faceplate either force or displacement actuators that push and pull on the surface to deform it. Displacement actuators, such as piezoelectric or magnetostrictive, or force type actuators, such as electromechanical or hydraulic, are governed by their

properties of stress, strain, and stiffness. They are specified by the deformable mirror component requirements such as stroke and bandwidth.

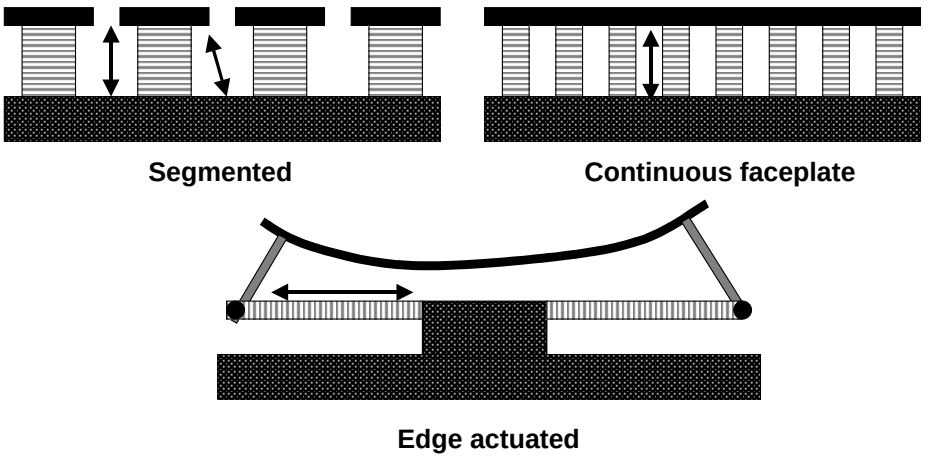

Fig. 8.2 Various types of deformable mirrors. Segments can be made to move in piston, tip, and tilt. Continuous faceplate mirrors have 100% fill-factor but are limited to electromechanical constraints on the influence function. Edge actuated mirrors structurally deform the surface in mechanical modes.

Some active mirrors use actuators that move in the same plane as the mirror surface. These actuators, when attached to lever arms (see Fig. 8.2), can induce bending into the surface and produce curved deformations that compensate a wavefront.

The most common actuators for deforming optics make use of the piezoelectric effect, which is essentially the creation of a strain-inducing stress under an applied electric field. Some materials exhibit a strong piezoelectric effect, such as lead zirconate titanate $Pb(Zr,Ti)O_3$, commonly called PZT.

Another material, lead magnesium niobate, $Pb(Mg_{1/3}, Nb_{2/3})O_3$, commonly called PMN, makes use of its large dielectric constant to achieve a large electrostrictive strain. Other materials can be used for actuators that generate a strain from an applied magnetic field. Magnetostrictive materials, such as the lanthanide-based iron alloys like Terfenol, produce a displacement proportional to the magnetic field intensity.

There is a whole list of materials that can potentially move things to optical tolerances. Shape memory alloy thermal actuators produce force and deformation with heat applied from the environment or ohmic heating. Bimetal alloy thermal actuators make use of the difference in thermal expansion between two dissimilar

metals to produce a linear displacement with an applied temperature change. Anything that can exert a force or change its shape with applied electric field, magnetic field, or heat can conceivably be used. But because adaptive optics is a field that requires reliable, useful, and inexpensive components, PZT and PMN have emerged as the leaders for conventional deformable mirrors with mirror diameters larger than 2 cm.

No matter what the actuator type or type of force train between the actuator and the mirror, the mirror surface is deformed by pressure from an actuator. To conserve energy and for all kinds of other mechanical constraints, the surface, especially a continuous faceplate mirror surface, forms a bump when an actuator is energized. The two-dimensional shape of the surface bump shows the influence of one actuator on the surrounding surface, thus it is called its *influence function.*

On many mirrors with mechanically thin faceplates, an influence function appears like that in Fig. 8.3. If a mirror facesheet is very thin, the mirror has the look of a rubber diaphragm with small bumps above each actuator; there is little surface coupling from the position of one actuator to the next. If the faceplate is thick, the motion of one actuator drags the surface along with it and there is a surface change at the position of adjacent actuators.

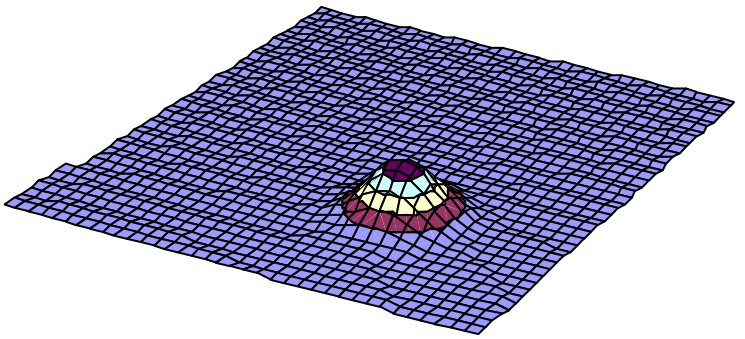

Fig. 8.3 Depiction of a deformable mirror surface with the motion of one actuator. The surface represents the actuator influence function.

The height of the surface at an adjacent actuator is called *coupling,* and it is usually expressed in a percentage. Mirrors with thin faceplates have zero to a few percent coupling. Commonly found deformable mirrors have 10%-20% coupling. A segmented mirror would have zero coupling unless the segments clunk together, which would not be a good thing anyway. Low coupling is not always good. Mirrors that primarily must control low-order wavefront aberrations need a smooth surface transition between actuators. One doesn't want a pimply surface unless there are a lot of actuators and the aberrations are very high order.

Influence functions can be modeled with any of a number of simple equations. The constants of the equations depend upon the faceplate material, the force train, and the backing structure of the mirror. For estimation purposes, the cubic form,

$$S(x,y) \propto (1-3x^2+2x^3)(1-3y^2+2y^3)\,, \tag{8.3}$$

or the Gaussian form,

$$S(r) \propto \exp\left(\frac{\ln[\text{coupling}]}{r_c^2} r^2\right), \tag{8.4}$$

closely approximate many operating deformable mirrors (see Fig. 8.4).

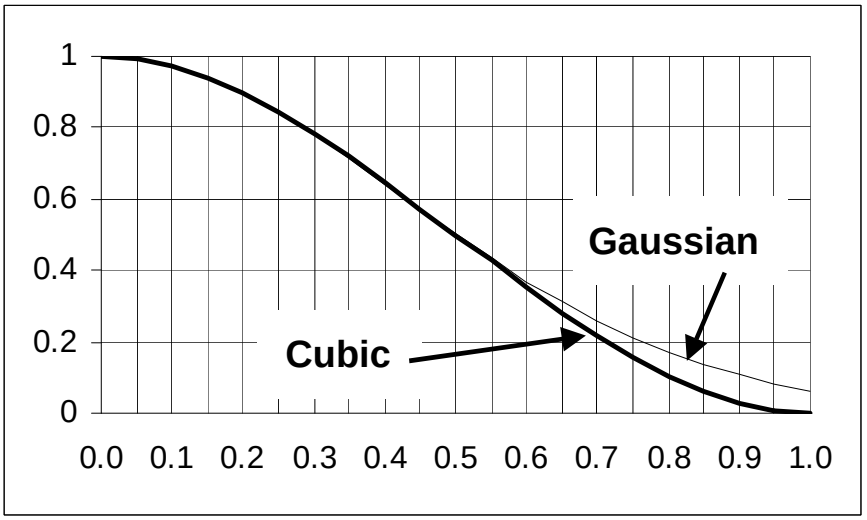

Fig. 8.4 Models of influence functions can take many forms. The horizontal axis is in units of actuator spacing.

In reality, most influence functions are not perfectly symmetric, nor are they even equal for each actuator. Those near the edge of the mirror will experience different stiffnesses than those in the middle, surrounded by other structure. It is best to make measurements of the actual deformable mirror and use that data in reconstructors or system performance predictions. If predictions must be made before the DM is finished, then constant Gaussian influence functions are usually just fine.

Another concern of actuators and their response in a deformable mirror is hysteresis, the phenomenon of the mirror surface position being dependent upon

whether the actuator was *pushed* or *pulled* into the position. Many piezoelectric materials exhibit hysteresis, and some mirror return springs or force trains contribute to the hysteresis. Hysteresis can be measured as a calibration step and compensated somewhat.

Hysteresis is acceptable if one is willing to give up control bandwidth. The mirror must cycle back and forth across the desired target point and zero in on the final position; this delays the mirror response and lowers the closed-loop bandwidth.

Many other mechanical factors are used in the design of a deformable mirror. Because the mirror is used at high speed to compensate fast disturbances, there can't be a resonance of the mirror-actuator force train in the band of operation; if this occurs, the mirror can become unstable and damage can occur.

Deformable mirrors have been built for nearly 40 years now. Because of the ease of laying out drawings on Cartesian coordinates, many mirrors have actuators laid out in a square array. To fit the array on a circular aperture, sometimes the corner actuators are removed. There are some *canonical* geometric layouts that we see in many deformable mirrors. The actuator counts that are common for square arrays are 4, 9, and 16, obviously, and then 21 (5×5 minus 4 corner actuators), 69 (9×9 minus 12 corner actuators), 241 (17×17 minus 48 corner actuators), and 941 (35×35 minus 284 corner actuators).

Hundreds of millions of years of evolutionary engineering by the honeybee has resulted in a research report. They found that we could pack more actuators into a circle, optimizing the correctability of the deformable mirror, by using a hexagonal-pack configuration. It is also very good for correcting aberrations with three axes of angular symmetry. The *canonical* layouts for hex-pack deformable mirrors contain numbers of actuators that follow the series 3, 7, 19, 37, 61, 91, 127, and so forth using $1+6\sum_{n=1}^{\infty} n$. The honeybee engineers used their 30-60-90 degree drawing tools extensively.

Deformable mirror requirements

When deformable mirrors are used in an adaptive optics system for atmospheric turbulence compensation, the requirements can be easily derived. The number of actuators can be found by considering the fitting error equation. Mathematically what we are doing is applying the linear superposition of influence basis functions. A less sophisticated way of saying that is, we want to fit the bumps of the deformable mirror surface to the aberrations of the atmosphere. The Strehl ratio *S*, the system measure of performance, is related to the actuator spacing r_c,

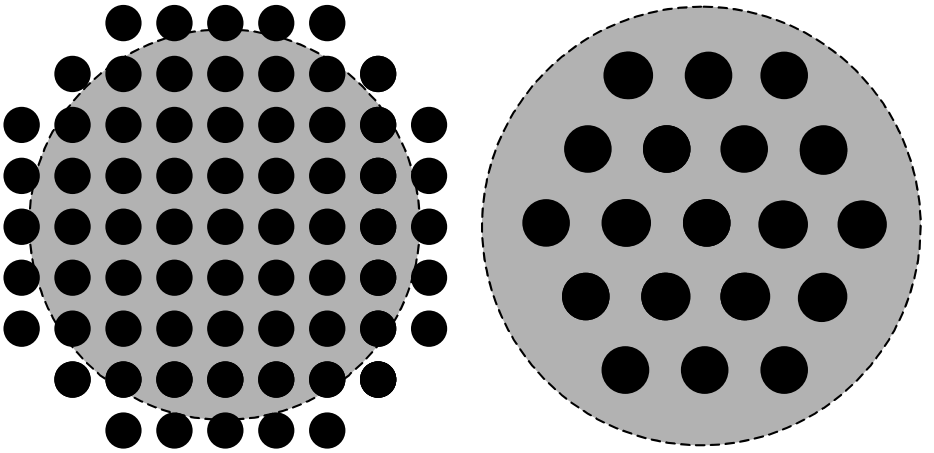

Fig. 8.5 Square and hexagonal arrays of actuators on a deformable mirror.

the atmospheric coherence length r_0, and the constant κ which depends upon the type of deformable mirror. For most continuous faceplate deformable mirrors, we can use $\kappa = 0.35$. The Strehl ratio is

$$S = \exp\left[-\kappa\left(\frac{r_c}{r_0}\right)^{5/3}\right]. \tag{8.5}$$

By putting in the desired system Strehl ratio, we can find the actuator spacing:

$$r_c = r_0\left(\frac{\ln S}{-\kappa}\right)^{3/5}. \tag{8.6}$$

The required stroke of each actuator is determined by finding the maximum amount of atmospheric wavefront error across the aperture. We assume that global tilt is corrected by a dedicated tilt mirror somewhere else in the system. The tilt corrected variance of atmospheric turbulence is

$$\sigma^2\left[\text{waves}^2\right] = 0.00357\left(\frac{D}{r_0}\right)^{5/3}. \tag{8.7}$$

The peak-to-peak aberration across the aperture is about 5 times the standard deviation σ. If a mirror surface moves 1 unit, the wavefront phase changes by 2 units. Reflect on that.

With all these factors, we end up with a simple expression for the stroke requirement of a deformable mirror:

$$\text{Stroke}[\text{waves}] = \frac{1}{2}\left(5\sqrt{0.00357\left(\frac{D}{r_0}\right)^{5/3}}\right) \quad . \tag{8.8}$$

Bimorph deformable mirrors

A bimorph mirror is made from two thin layers of material bonded together. *Bi*, meaning two, and *morph*, meaning having a specified form, could be a name for a lot of things. Here it just means a basic two-piece deformable mirror. One piece is a piezoelectric material such as PZT and the other is the optical surface, made from glass or silicon, or both pieces are PZT material, with the outer surface as the polished mirror face (see Fig. 8.6). A thin conductive film is deposited between the two layers and acts as a common electrode. Voltage applied to the electrodes leads to a variation in area of the PZT. Because the other material, which is the mirror, does not expand, the result is a local bending much like that of a bimetal strip. The PZT electrodes need not be contiguous. They can be placed in square or hexagonal arrays just like in other deformable mirrors. An applied voltage *V* at position *(x,y)* on the PZT results in a surface deformation *S(x,y)* following Poisson's equation:

$$\nabla^2 S(x,y) = \frac{\partial^2 S(x,y)}{\partial x^2} + \frac{\partial^2 S(x,y)}{\partial y^2} \propto V(x,y)\,. \tag{8.9}$$

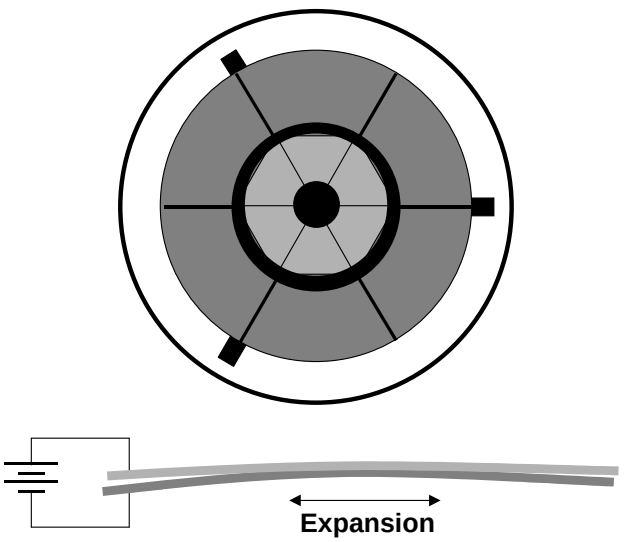

Fig 8.6 A bimorph deformable mirror configured for 13 segments.

Thus, the mirror responds with a curvature proportional to the voltage applied. This is very convenient for operation with curvature wavefront sensors because there doesn't have to be complex reconstruction circuitry to convert the wavefront sensor signals to actuator commands. The drawback, however, is the limit to which we can force a bimorph with local curving surfaces into a lot of desired aberration modes. Mirrors with 13 or 19 electrodes can cover most of the low-order modes. Bimorphs up to 61 elements have been tested and can be used in many atmospheric turbulence compensation systems.[1]

Micromachined deformable mirrors

A new class of deformable mirrors, derived from the membrane mirror concept, can be made with hundreds of actuators, high bandwidths, and low hysteresis and fit into a microchip. The devices, called micro-electro-mechanical (MEM) deformable mirrors, have the potential for production level costs of only a few hundred dollars. The actuation is in the form of electrostatic attraction and repulsion between a thin common electrode membrane that acts as the mirror surface and control electrodes (see Fig. 8.7).

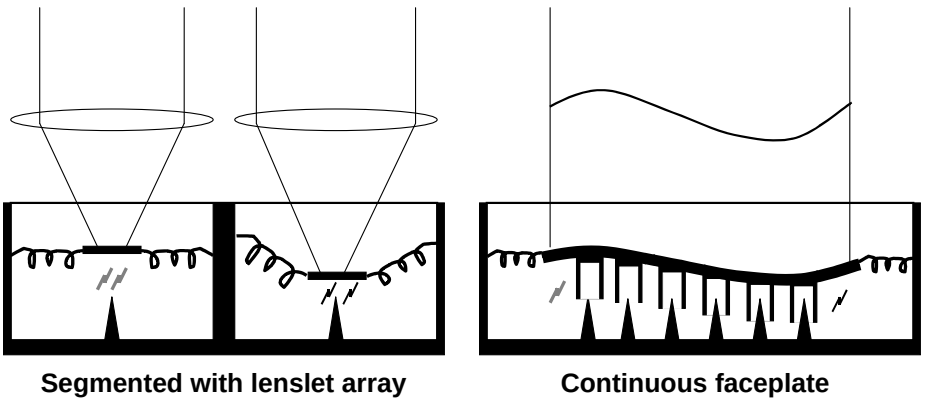

Fig. 8.7 Cross section of a micro-electro-mechanical (MEM) membrane deformable mirror. Both segmented and continuous faceplate mirrors can be configured.

A number of devices have been made that have the same basic structure as continuous faceplate deformable mirrors or segmented mirrors except that they are often only a centimeter in diameter with 300 or more active channels. Modern CMOS microchip manufacturing techniques are used to produce the devices. The actuator drivers can also be miniaturized because the voltage needed for a few microns' stroke is only about 15 V with very low current. The deformable mirrors can be integrated with driver circuitry on the same chip, and the potential exists for integration of the entire wavefront sensor, control computer, and mirror

[1] J.-P. Gaffard, P. Jagourel, P. Gigan, *Proc. SPIE* **2201**, 688, 1994.

into the same package. This possibility opens up a wide range of uses for adaptive optics for low cost medical applications and communications, as well as disposable adaptive optics for imaging in severe environments and tactical weapons.

Bibliography

Gleb Vdovin, *Adaptive Mirror Micromachined in Silicon*, Delft University Press, Delft, Netherlands (1996).

M. A. Ealey, J. F. Washeba, "Continuous facesheet low voltage deformable mirrors," *Opt. Engr.* **29**, pp. 1191-1198 (1990).

R. Hudgin, "Wave-front compensation error due to finite corrector-element size," *J. Opt. Soc. Am.* **67**, pp. 393-395 (1977).

Chapter 9

Control computers and reconstructors — the brains

"I think there is a world market for maybe five computers."
—IBM Chairman Thomas Watson, 1943

Just as the wavefront sensor is the eyes of the adaptive optics system and the deformable mirror is its fingers, the control computer is the brains of the system. While many systems can have just a few wires connecting the output of the sensor to the mirror, some systems have racks and racks of electronics to make the precise calculations necessary to drive the control mirrors with the confusing signals from the wavefront sensor detectors.

Many servo-control systems have been developed over the centuries. From simple pressure regulators to the artificial intelligence of robots, the idea of automatic control of a process was advanced.

While controls engineers spoke in tongues with block diagrams and Laplace transforms and assumed recognition of terms such as *gain*, *type II controller*, and *phase lag*, the adaptive optics scientists were going all-out to coordinate multiple feedback loops, all working in parallel, to drive a multiactuator deformable mirror.

Single-channel servo control

The first step in describing an adaptive optics control system is examining how a beam of light could be stabilized by a single-axis tilt mirror. If we can't control one beam, we can't control a thousand pieces of it individually. One of the nice things about adaptive optics controls is that most of them are handled by simple, or not so simple, single-channel control theory. Then there are just a lot of them tied together. A block diagram of a control channel is shown in Fig. 9.1.

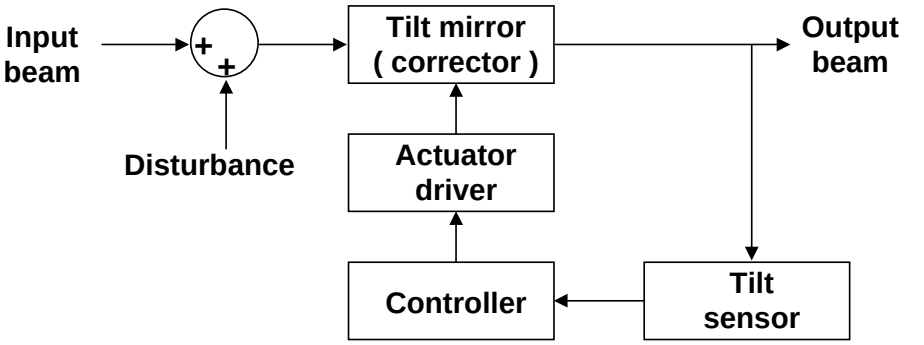

Fig 9.1 Block diagram of single channel tilt control, the basic building block for adaptive optics controls.

The optical system that corresponds to the block diagram is fairly simple. A beam of light, a laser for instance, is disturbed by some sort of vibration. For now, just consider the vibration to be a static misalignment. The beam bounces off the tilt mirror and then goes on out, with a little bit of it sent off toward a tilt sensor. The sensor can be a quadcell, or since this is a one-dimensional problem, a bicell. For the first photon run through the system, the system doesn't do anything because it isn't fast enough to catch up with the photons streaming through. But for the second photon, so to speak, the tilt sensor detects an error. The beam is not going in the right direction. It sends this signal to a controller (the control computer), which determines that the error demands a certain drive signal for the tilt mirror to respond and cancel the error. The low voltage or digital drive signal is sent to the actuator driver, where it is converted to the appropriate voltage for the mechanical actuator. The voltage applied to the actuator moves the tilt mirror and perfectly corrects the beam direction.

This simple process is repeated in a continuous fashion as the disturbance changes from being static to one that is constantly varying. The fact that the sensor sees the changes the mirror makes, and responds to them, makes this a closed-loop feedback system. In our terminology, it is *adaptive*. Since it deals with a beam of light, it is *optics*. Thus, a simple single-channel adaptive optics system is born.

The sensor can respond only so fast; the light must be integrated on the detector and then an error signal formed. The computer can only respond so fast; even analog calculations are not instantaneous, requiring electrons to move through the circuitry. The actuator can only move so fast, too. The voltage tends to make the actuator move, and the actuator gets up a head of steam (or a head of electrons) and starts moving toward the final position. Some actuators and their sensors can translate the error signals to motions quickly; these are termed high gain. Ones that are slower are termed low gain. An actuator that gets moving

quickly probably cannot stop quickly, and although we want to get to the final position and get rid of the disturbing error, we don't want to do it so fast and frantically that we overshoot the final position and have to backtrack.

Looking at Fig. 9.2, we see that the consequences of gain selection can affect how well our simple system performs. A high gain system gets to the target quickly but can't stop, so it overshoots. A low gain system gets to the target with less overshoot, but it takes longer. A unity gain system has no overshoot, but it just never really gets to the target anyway. Control system designs are principally an art of adjusting the gain (and a lot of other factors) to optimize the servo so that the goal is achieved with as little penalty as possible. One of those penalties is the slowdown of our system. If the disturbance is constantly changing, like a vibration or turbulence in the atmosphere, we can't keep overshooting and backtracking. That makes our overall control bandwidth low and the residual optical error stays around too long.

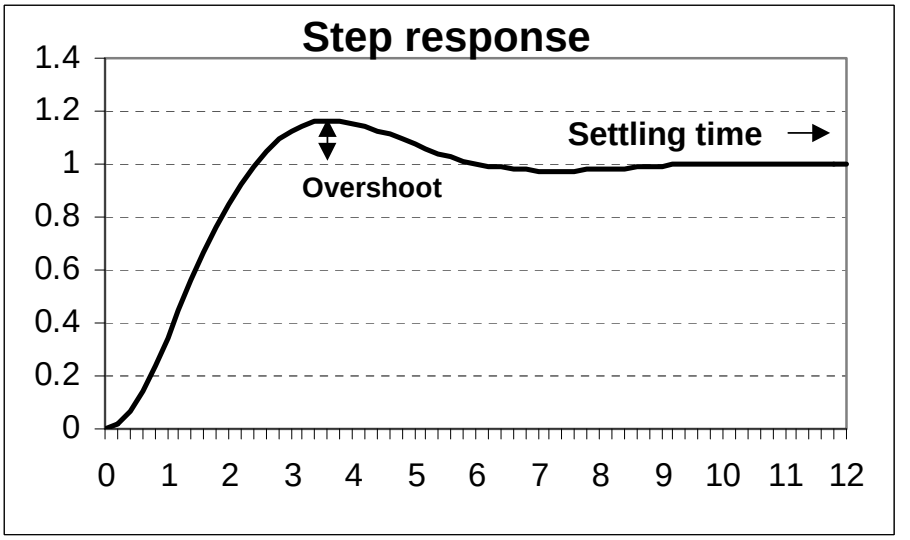

Fig. 9.2 A control system response to a step input. Depending upon the gain, the amount of overshoot and the settling time can be adjusted.

Single-channel dynamic control

The disturbances in an adaptive optics system are changing rapidly. It is necessary to have a system of sensors, computers, and actuated mirrors that keep up with the speed of the disturbance. If they don't there is no point in having the system at all; it cannot be "better late than never." In adaptive optics, late, in terms of too late to correct a disturbance, is not better. It can often make the system worse.

To figure out what is needed in a control system for a rapidly changing disturbance, we use Fourier's theory: all disturbances, no matter how messy or how rapidly changing, can be broken down into a number of individual sinusoidal disturbances, each with its own amplitude, frequency, and phase. With this knowledge, we determine what an adaptive optics servo-control loop will do with any one of the sinusoidal disturbances and then just add them all back together to find out what it will do with the whole disturbance.

If our input disturbance onto the beam of light is *r(t)*, which is a function of time *t*, following the form of a sine wave with amplitude *A*, $r(t) = A\sin(\omega t)$, we find that a servo-control analysis will result in the output *c(t)*, or the actual response of the beam of light,

$$c(t) = AM(\omega)\sin\left[\omega t + \phi(\omega)\right] . \tag{9.1}$$

$M(\omega)$ is the ratio of the output amplitude to the input amplitude and is a function of the frequency of the disturbance ω, and $\phi(\omega)$ is the phase delay of the output relative to the input, also a function of the frequency (see Fig. 9.3).

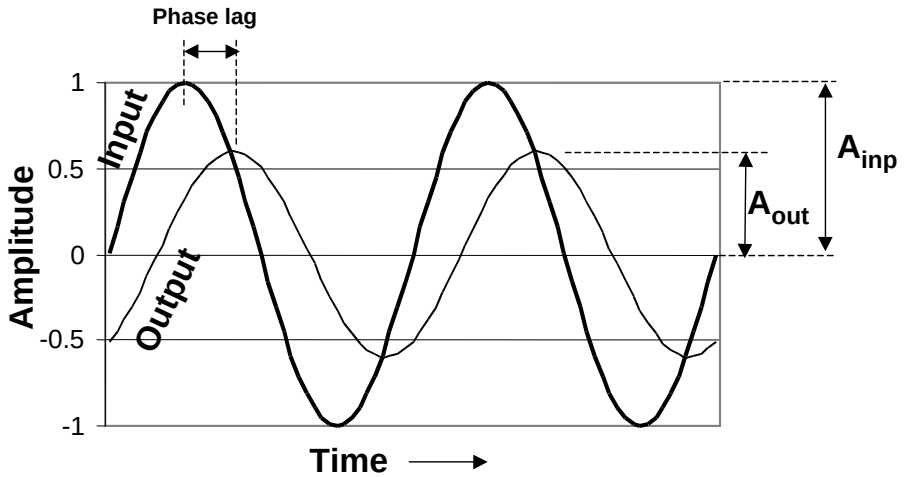

Fig. 9.3 The response of a control system to a sine wave. Note the input and output at this particular frequency. The reduction in amplitude and the delay in the phase is a consequence of the control bandwidth. A plot of the amplitude response and the phase response versus frequency is a Bode plot.

If we plot gain $M(\omega)$ and phase angle $\phi(\omega)$, we have a Bode plot. Figure 9.4 is an example of a simple lag control that has the form

$$M = 20\log 1/\sqrt{1+(\omega T)^2} \ , \ \phi = -\tan^{-1}\omega T . \tag{9.2}$$

Gain is traditionally plotted in decibels, or dB, M(dB) = 20 $\log_{10}M$. Phase is plotted in degrees. If M = 1, the response is the same as the input and M = 0 dB.

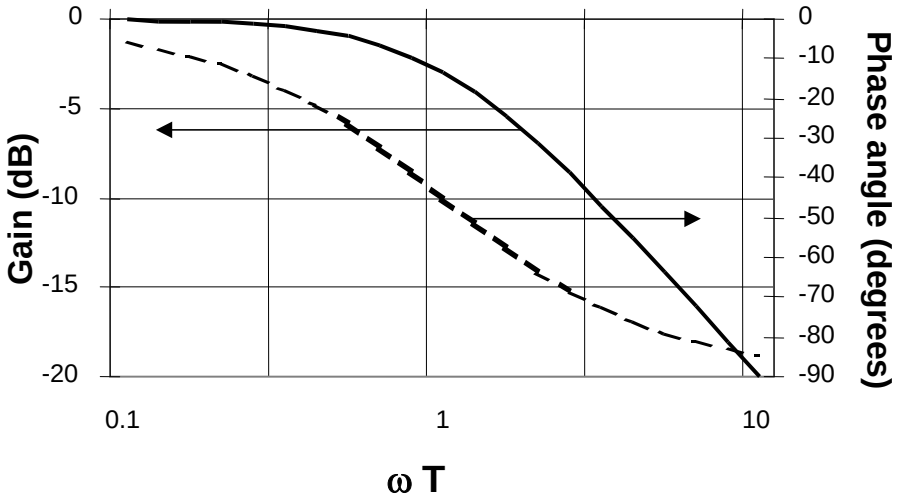

Fig. 9.4 The Bode plot for a simple lag control.

Figure 9.5 is a Bode plot of a more complicated control system, in this case a quadratic lag, which has the gain go slightly above the 0 dB response line. For those frequencies there is an amplification of the input:

$$M = 20\log\left[\left(1-\frac{\omega^2}{\omega_n^2}\right)^2+\left(\frac{2\zeta\omega}{\omega_n}\right)^2\right]^{-1/2} \quad , \tag{9.3}$$

$$\phi = -\tan^{-1}\frac{2\zeta\omega/\omega_n}{1-\left(\omega^2/\omega_n^2\right)} \quad , \tag{9.4}$$

where ζ is the damping ratio.

A few control system definitions, used (almost) universally, come from Bode plots. The *crossover frequency* is the frequency where the magnitude curve crosses the 0 dB axis. The *phase margin* is 180° plus the phase angle at the crossover frequency. The *bandwidth* is the range of frequencies where M is more than 0.707 of its dc value. Qualitatively, the bandwidth is the range of frequencies over which the response is considered to be satisfactory. Sometimes crossover frequency and bandwidth are confused. Be sure to ask the control system designer what their definition of bandwidth is. Even if you don't understand the answer, asking the question makes you look like you know what you're doing.

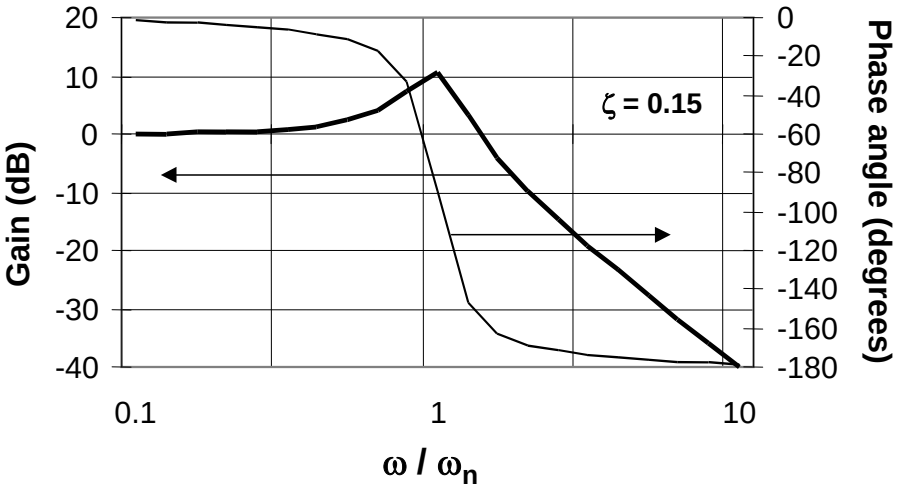

Fig. 9.5 The Bode plot for a quadratic lag control.

Bandwidth limitations

The bandwidth of a control is related to the correctability of an adaptive optics system. A theoretical treatment of the problem was presented by Greenwood. In terms of the Greenwood frequency f_G, the spectrum *F(f)* of high-order atmospheric turbulence phase disturbance is

$$F(f) = 0.32 f_G^{5/3} f^{-8/3}. \tag{9.5}$$

A control system acts as a high-pass filter, allowing the disturbances with high frequency to carry through while removing the low frequency things. The error variance (σ^2, in radians) after the filter process is

$$\sigma_{\text{temp}}^2 = \int_0^\infty \left[1 - H(f, f_{\text{BW}})\right]^2 F(f) \mathrm{d}f. \tag{9.6}$$

For a perfect step (everything above f_{BW} is removed 100% and everything below it is kept), a high-pass filter takes the form:

$$H(f, f_{\text{BW}}) = 1 \ (f < f_{\text{BW}}), \ \ \mathrm{H}(f, f_{\text{BW}}) = 0 \ (f > f_{\text{BW}}).$$

The required bandwidth for a system of residual temporal bandwidth error σ^2, is $f_{\text{BW}} \approx 0.37 \ \ \sigma^{-6/5} f_{\text{G}}$. For example, for a λ/20 wave residual temporal error, the bandwidth of the step filter is about $f_{\text{BW}} = 1.5 \ f_{\text{G}}$. Because perfect step filters are not really possible to build, other filters, such as a classical RC filter,[1] can be

[1] John Van de Vegte, *Feedback Control Systems*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ, pp. 80-85 (1990).

used. In this case, the bandwidth requirement for a λ/20 system is about $f_{BW} = 4\, f_G$.

Phase reconstruction

Now that we know how a single-channel control system works, sort of, we can move toward figuring out how to connect all the different channels from the wavefront sensor to the deformable mirror. Because some wavefront sensors measure the phase directly while others measure wavefront slopes or measure wavefront modes like tilt and focus directly, the control computer must be able to *reconstruct* the phase in a form that the deformable mirror, or set of active mirrors, can use. To complicate things, as if they aren't complicated enough already, some mirrors such as tilt mirrors respond to modal commands, and some mirrors such as continuous faceplate deformable mirrors respond in a small zone around the actuators. The result is that there are multiple reconstruction paths to follow when we go from modal or zonal sensors to modal or zonal mirrors.

Often we never need to know the phase exactly; we just need to translate the sensor signals to mirror commands. But then again, we might also want to know exactly what the phase is and use it for diagnostics. This maze of reconstruction possibilities, shown in Fig. 9.6, and the goal of translating a large number of signals to a large number of commands, lends itself to linear matrix algebra.

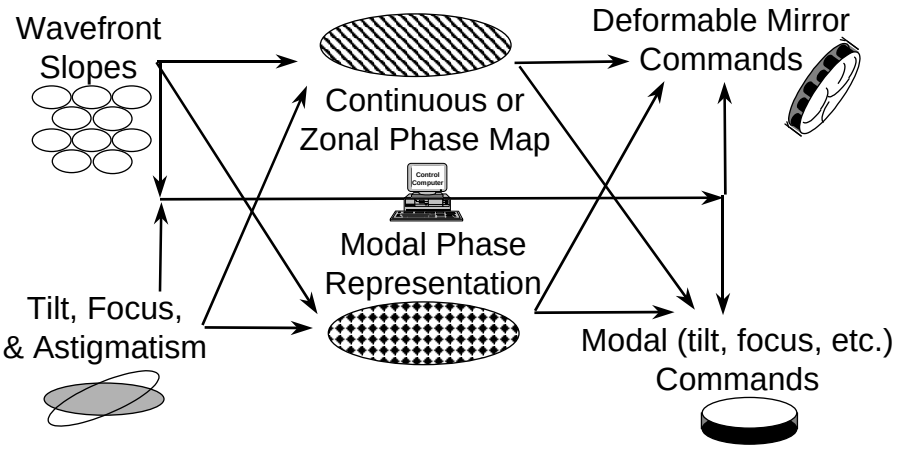

Fig. 9.6 Wavefront reconstructors can be configured in many ways. The computations are geared toward reliable translation of the wavefront sensor measurements into commands for the active optical elements.

For adaptive optics systems it is generally necessary to have more wavefront sensor measurements than actuators to move. If we don't, then the problem of determining the actuator commands cannot be solved. There would be an infinite

number of solutions, none of which would probably be the right one. A system of this type is called *underdetermined.* There are a few ways to get around this. If the measurements are in modes that can be reconstructed into a wavefront surface and the actuator influence functions can also be decomposed into a wavefront surface, the surfaces can be matched and a reconstruction solution can be found.

If there are exactly as many sensor signals as actuator commands, the system is *determined*. This can occur if, for example, a segmented mirror is precisely registered with the subapertures of the sensor. Each subaperture spills out two slope signals and the segment can slope in two directions. If the segment also can move in a piston fashion, there can be a simple circuit that zeroes the average of all the piston motions without regard to their tilt motion.

Most cases of adaptive optics reconstructors are of a third type. They are *overdetermined*, meaning that there are more wavefront sensor signals than there are actuators, or degrees of freedom, to control. Because this is so, it is important to correctly align, both optically and mathematically, the sensor subapertures and the actuators.

Historically, the alignment has followed only a few possibilities. The four most important ones, named after the developing authors or programs, are shown in Fig. 9.7. The Hudgin geometry (a) measures orthogonal slopes centered on one acutator, with no redundancy by either slope or actuator. The Southwell geometry (b) measures orthogonal slopes centered on each actuator. The Fried geometry (d) measures orthogonal slopes centered in-between actuators. The WCE geometry (c), named after the wavefront control experiment, is the same as the Fried geometry except that it is oriented at 45^{o} to the actuator grid.

From any of these configurations, and many others, one can develop the equations that relate the wavefront sensor signals to the actuator commands. We consider the string of M signals as a one-dimensional vector $\mathbf{y}$ and the N actuator commands as a vector $\mathbf{a}$. Connecting the two is an $N \times M$ rectangular matrix $\mathbf{B}$ that represents the interactions between the actuator commands, the wavefront that's modified, and the wavefront sensor signals. The equation relating the actuator commands to the wavefront sensor signals is $\mathbf{y} = \mathbf{[B]a}$, where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_M \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ a_N \end{pmatrix}, \quad \mathbf{B} = \begin{bmatrix} B_{11} & B_{12} & \cdot & B_{1N} \\ B_{21} & B_{22} & \cdot & B_{2N} \\ B_{31} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ B_{M1} & B_{M2} & \cdot & B_{MN} \end{bmatrix} \quad . \qquad (9.7)$$

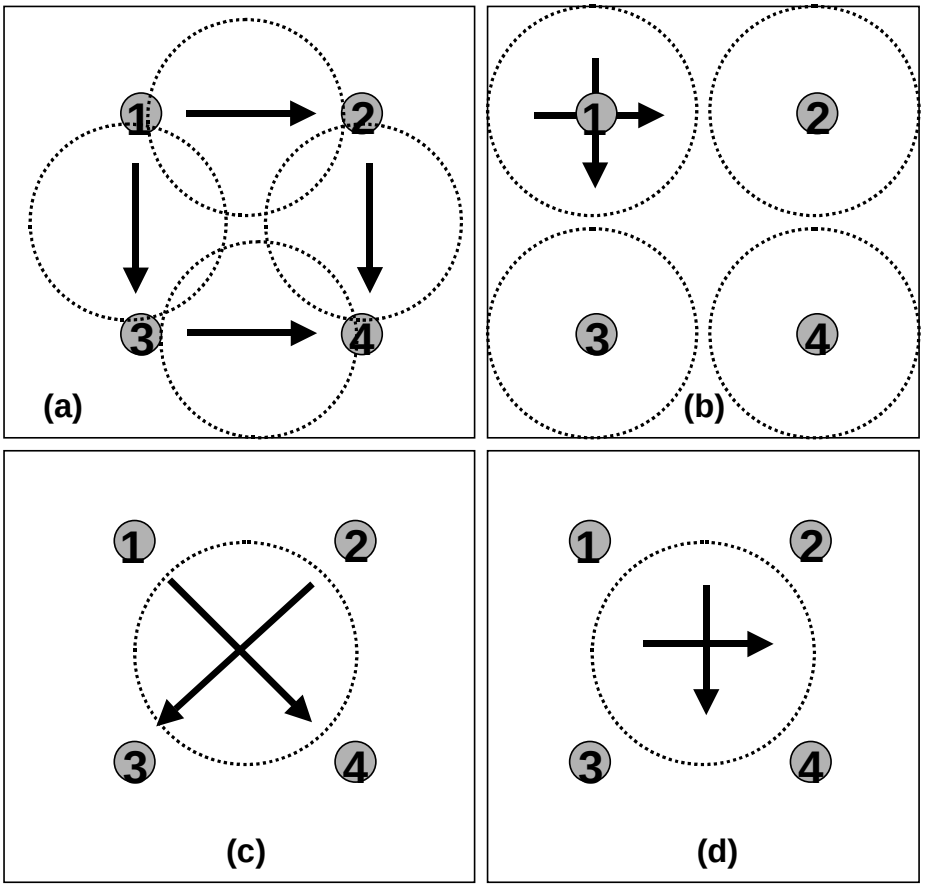

Fig. 9.7 Various wavefront sensor - actuator geometries: (a) Hudgin, (b) Southwell, (c) WCE, and (d) Fried. The circles represent the layout of the subapertures.

To find the values of **B**, it is necessary to determine how the wavefront surface (that the sensor will eventually see) changes with each actuator command. For segmented mirrors, it is pretty straightforward. The influence of each segment covers only those subapertures that overlap the segment; **B** is nearly diagonal and sparse. For continuous faceplate deformable mirrors, each actuator can affect a large part of the surface and the signals in many subapertures; **B** can be completely filled with meaningful signal-command interactions.

If there were the same number of signals as actuators, the inversion of Eq. 9.7 would be relatively easy. Inverting a square matrix is possible as long as it is not singular. Physically speaking, the matrix would be singular if some group of actuator commands would create the same wavefront sensor signals as some other combination. The matrix could not be inverted, because then there would be some group of sensor signals that would derive from the wrong combination

of actuator commands. The control system just wouldn't know which solution to choose.

But this is not normally the case. Equation 9.7 can be inverted with a little mathematical legerdemain. The method of least squares produces a solution that has a new matrix called the pseudoinverse of **B**. The result looks a little messy, but it is just matrix multiplication, performed quite simply with most calculators and computers.

The pseudoinverse matrix $\mathbf{[B]}^{-1}$ also in adaptive optics parlance the reconstruction matrix, is a product of the matrix **B** and its transpose in the form $[\mathbf{B}]^{-1} = [\mathbf{B}^{\mathrm{T}}\mathbf{B}]^{-1}[\mathbf{B}^{\mathrm{T}}]$, where $\mathbf{B}^{\mathrm{T}}$ is the transpose of matrix **B**:

$$\mathbf{B}^{\mathrm{T}} = \begin{pmatrix} B_{11} & B_{21} & \cdot & B_{M1} \\ B_{12} & B_{22} & \cdot & B_{M2} \\ B_{13} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ B_{1N} & \cdot & \cdot & B_{MN} \end{pmatrix}. \tag{9.8}$$

In virtually all adaptive optics wavefront reconstructors, this process is followed. It is not followed exactly; many adaptive optics system configurations lend themselves to different geometries, different methods of matrix inversion (not the pseudoinverse least squares method), and different reconstruction matrices. If wavefront measurements are taken modally, for example, the modes must be converted into actuator commands. This can be done by converting them to a surface map of the wavefront and then converting them to a mirror surface by adding up influence functions. The amplitudes of the influence functions become the actuator commands. In another method, the modal wavefront measurements can be converted to wavefront slopes at the positions of the wavefront sensor measurements in the Fried geometry. From this, the reconstructor can just assume that the slope measurements were made directly and go directly to actuator commands just as if a Shack-Hartmann wavefront sensor had been used.

Conversion from one set of measurements to any set of actuator commands is possible through linear algebra if the physical limits are understood. Many recent research papers describe clever techniques that can be used in reconstructors to cut down on the number of calculations, thus speeding up the process, or to make more accurate estimations of the wavefront and correction, thus reducing the residual error and improving performance.

Sometimes the matrix $\mathbf{B}^{\mathrm{T}}\mathbf{B}$ can't be inverted mathematically because there are other physical things happening. The adaptive optics system is trying to generate a conjugate of the wavefront. If only slopes or relative phase differences are measured, the solution for the actuator commands can have an infinite number of

answers, all meeting the least squares best-fit of the mirror surface to the slope values. This is because the piston component on the mirror could have any value and still match the wavefront shape. So, we have to constrain our mirror to have a certain attainable overall piston component by making sure the actuator commands result in a known average surface height.

To illustrate the process, consider a simple configuration in the Hudgin geometry (see Fig. 9.8). Let's assume that we can measure the wavefront slopes **y** between positions on the wavefront that correspond to the positions of the actuators **a**. Also, for purposes of this illustration, assume that we don't care what happens between actuators; we just want the 9 actuator commands to fit the 12 slopes that we measure.

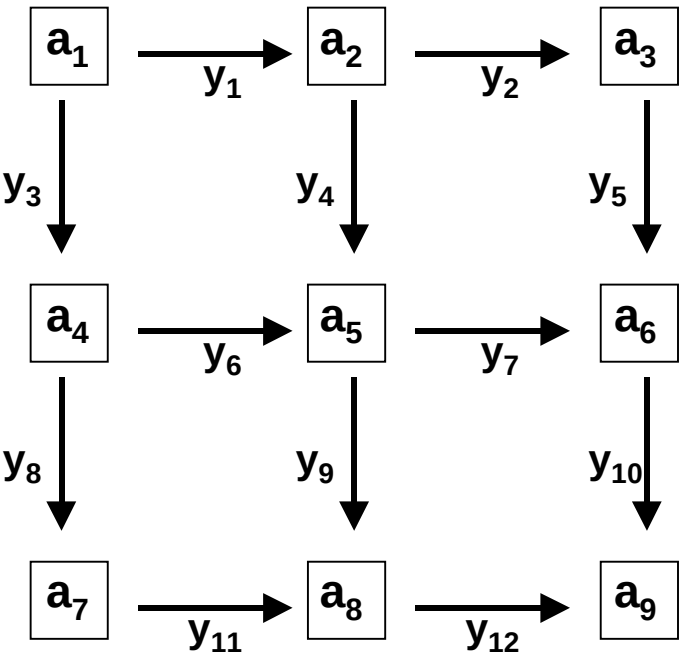

Fig. 9.8 An example configuration for analysis. The Hudgin geometry is assumed.

The relationships between an actuator pair and the slope that would be seen are

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ 0 \end{pmatrix} = \begin{pmatrix} a_1 - a_2 \\ a_2 - a_3 \\ a_1 - a_4 \\ a_2 - a_5 \\ a_3 - a_6 \\ a_4 - a_5 \\ a_5 - a_6 \\ a_4 - a_7 \\ a_5 - a_8 \\ a_6 - a_9 \\ a_7 - a_8 \\ a_8 - a_9 \end{pmatrix}, \tag{9.9}$$

and the **B** matrix then looks like

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{9.10}$$

The row of 1's on the bottom is used to force the average surface to a specific value and keep **B** from being singular.

Although the example here, with 12 slopes and 9 actuators, is fairly simple, remember that many adaptive optics systems have tens or even hundreds of actuators. Thus, the inversion of matrix **B** should be performed once, preferably before any real-time control takes place. Also, there are systems which don't use

wavefront slopes but rather modal or zonal phase measurements. Even though the matrix techniques are the same, the construction of matrix **B** will be different.

Example problem: actuator commands from wavefront slopes

Sometimes the slopes are not just differences of surface height above an actuator, but rather they are the integrated tilt over a subaperture which may experience the influence of a number of actuators. To illustrate this process, the simplest example I know is an adaptive optics reconstructor with 3 actuators and 4 slope measurements, like that shown in Fig. 9.9. Most of the math accompanying the 3×4 problem is the same as that in a 500×500 problem. It's just that the 3×4 problem can be shown in a paperback book.

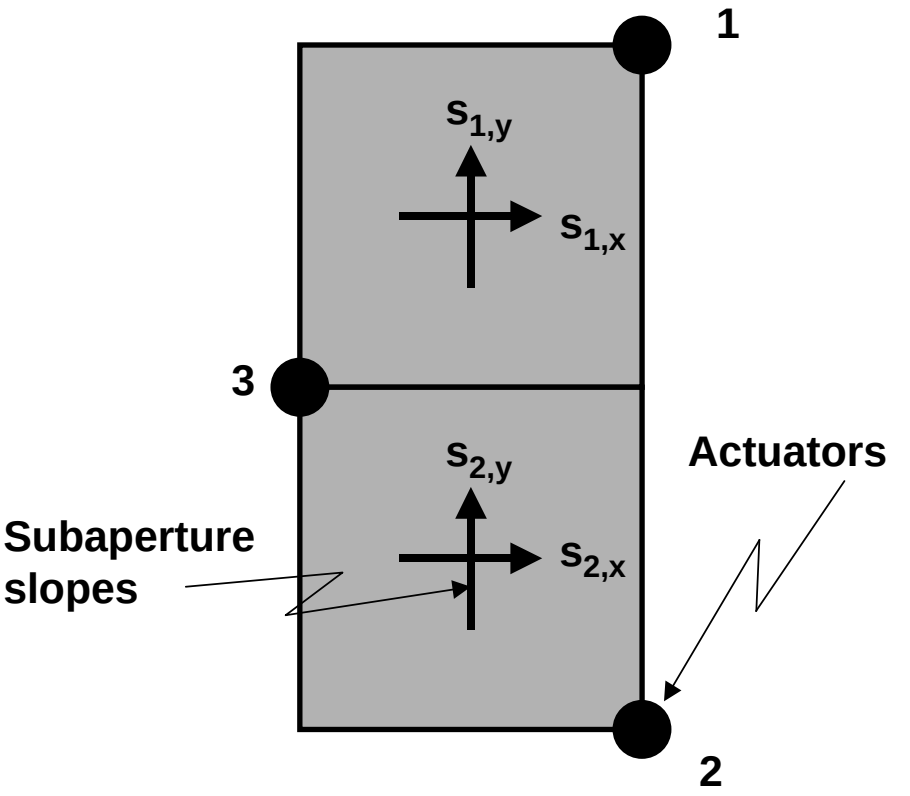

Fig. 9.9 A 3-actuator, 2-subaperture system. The subapertures are square, 1 unit of distance on a side. The actuators are in opposing corners, $\sqrt{2}$ units apart.

The most common wavefront sensor used today is a Shack-Hartmann sensor that measures wavefront slopes over a region of the wavefront within a subaperture. The most common deformable mirror is a continuous faceplate mirror with a little bit of coupling (15% or so) between actuators. So, what we want is a

reconstructor that converts the measured slopes to actuator commands. The result will be a conjugate wavefront that improves the optical system resolution.

The influence function, or the shape of the deformable mirror surface when you push one actuator, is modeled as

$$I(x,y) = \exp\left[\frac{\ln(\text{coup})}{r_c^2} r^2\right] = \exp\left\{-\frac{1.9}{\sqrt{2}^2}\left[\left(x - x_n\right)^2 + \left(y - y_n\right)^2\right]\right\} \quad , \quad (9.11)$$

where r_c is the spacing between actuators, in this case $\sqrt{2}$ units. The 3 actuators are positioned at (x_n,y_n), or (+½,+1), (+½,–1), and (–½,0). The 2 subapertures are centered at (0,+½) and (0,–½). In most reconstruction exercises, the units are arbitrary except when it comes to the final values, which include electronic gain and voltages.

The 4 slope signals are constructed by integrating the total tilt over a subaperture when the actuators are pushed. To find the interaction between slopes and actuators, we calculate the 4 slopes from each actuator pushing by itself. Thus, we have 12 equations of the form

$$s_{1x,a1} = \int_{\text{subap}} xI(x,y)_{a1}\mathrm{d}x = \int_{-1/2}^{1/2} x \; \mathrm{d}x \int_{0}^{+1} \exp\left\{-0.95\left[\left(x - \tfrac{1}{2}\right)^2 + \left(y - 1\right)^2\right]\right\}\mathrm{d}y \, . \quad (9.12)$$

This one has the influence of actuator 1 on slope 1 in the x direction. The other 11 follow along nicely. We end up with a matrix equation that relates the actuator amplitudes (a_1 , a_2, a_3) with the 4 slopes. The piston term p makes sure that the matrix doesn't become singular because the sum of all the actuators is held constant, $\mathbf{y} = [\mathbf{B}]\mathbf{a}$, or

$$\begin{pmatrix} s_{1x} \\ s_{2x} \\ s_{1y} \\ s_{2y} \\ p \end{pmatrix} = \begin{bmatrix} 0.45 & 0.088 & -0.45 \\ 0.088 & 0.45 & -0.45 \\ 0.45 & -0.26 & -0.45 \\ 0.26 & -0.45 & 0.45 \\ 1 & 1 & 1 \end{bmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad . \quad (9.13)$$

What we really want is the controlling matrix that converts slope signals to actuator commands, which is the inverse of Eq. 9.13. We use the least squares solution to invert it:

$$\mathbf{a} = [\mathbf{B}^{\mathrm{T}}\mathbf{B}]^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{y} \quad (9.14)$$

$\mathbf{B}^{\mathrm{T}}$ is pretty straightforward,

$$\mathbf{B}^{\mathrm{T}} = \begin{pmatrix} 0.45 & 0.088 & 0.45 & 0.26 & 1 \\ 0.088 & 0.45 & -0.26 & -0.45 & 1 \\ -0.45 & -0.45 & -0.45 & 0.45 & 1 \end{pmatrix} , \qquad (9.15)$$

and then cranking through the algebra we get the pseudoinverse

$$[\mathbf{B}^{\mathrm{T}}\mathbf{B}]^{-1}\mathbf{B}^{\mathrm{T}} = \begin{pmatrix} 0.52 & -0.046 & 0.70 & 0.41 & 0.34 \\ -0.046 & 0.52 & -0.41 & -0.70 & 0.34 \\ -0.43 & -0.43 & -0.35 & 0.35 & 0.30 \end{pmatrix} . \qquad (9.16)$$

So, the actuator command solution, the reconstructor, is

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0.52 & -0.046 & 0.70 & 0.41 & 0.34 \\ -0.046 & 0.52 & -0.41 & -0.70 & 0.34 \\ -0.43 & -0.43 & -0.35 & 0.35 & 0.30 \end{pmatrix} \begin{pmatrix} s_{1x} \\ s_{2x} \\ s_{1y} \\ s_{2y} \\ p \end{pmatrix} . \qquad (9.17)$$

Now, suppose our unknown wavefront comes streaming in to this adaptive optics system. The wavefront we see looks like Fig. 9.10. We measure 4 wavefront slopes (1,1,1,–1).

Putting the 4 slopes into Eq. 9.17, and making sure the average of the 3 actuators is 0 (we don't want to drive the deformable mirror beyond its limits), we find that the best positions of the 3 actuators are ($a_1 = 0.76$, $a_2 = 0.76$, $a_3 = -1.56$). The sign on each follows intuitively; actuators 1 and 2 go up to match the rising slope and actuator 3 goes down. They add, within two significant digits, to 0.

Another simple example would be a wavefront that has a minimum in the gap between subapertures which rise upward in the y direction (see Fig. 9.11).

We measure 4 wavefront slopes (0,0,1,–1). Putting the 4 slopes into Eq. 9.17, we find that the best positions of the 3 actuators are now ($a_1 = 0.29$, $a_2 = 0.29$, $a_3 = -0.70$). The sign on each is the same as before, but the actual values differ to account for the fact that the mirror doesn't want to add any tilt in the x direction since there isn't anything there to conjugate.

See, reconstructors aren't that hard. It's just that when you have things moving around, like rotating pupils, or when you have changing conditions and have to increase the size of the subapertures to collect more of the photons from a dim star, or when you have to increase the bandwidth because of a high altitude wind and start to group subapertures and actuators together within the software ... only then does it get complicated. But that doesn't mean it's impossible.

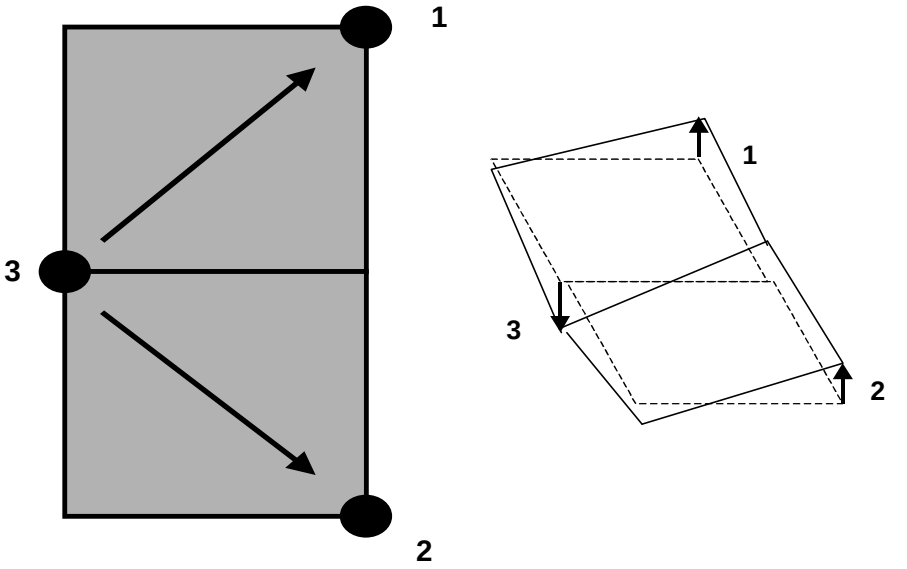

Fig. 9.10 An example of wavefront slopes and actuator commands to compensate the wavefront. The wavefront is sloping up from actuator 3.

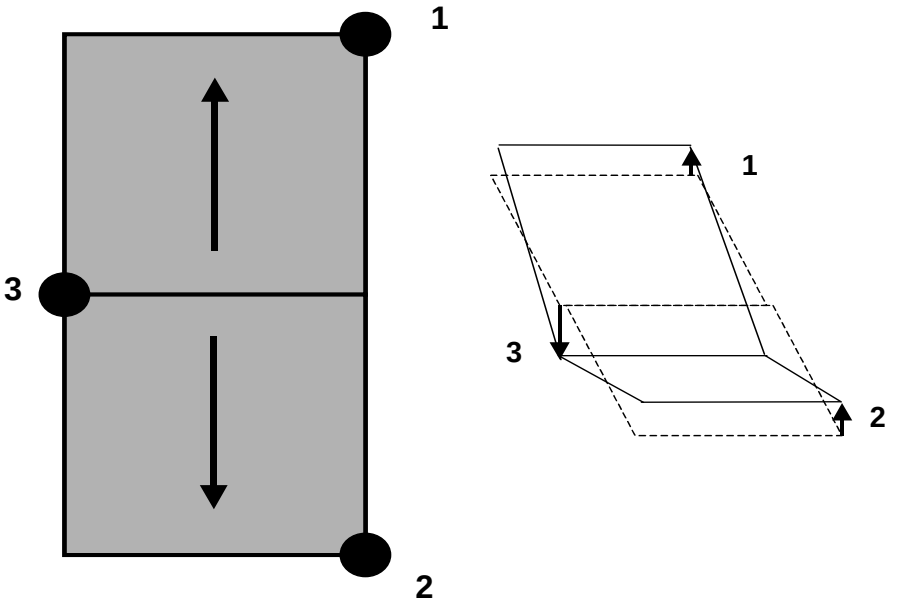

Fig. 9.11 An example of wavefront slopes and actuator commands to compensate the wavefront. The wavefront is sloping up from the line between subapertures.

Bibliography

John Van de Vegte, *Feedback Control Systems*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ (1990).

Gilbert Strang, *Linear Algebra and Its Applications*, 3rd ed., Harcourt Brace Jovanovich, San Diego (1988).

Robert K. Tyson, *Principles of Adaptive Optics*, 2nd ed., Academic Press, Boston (1997).

R. H. Hudgin, "Wave-front reconstruction for compensated imaging," *J. Opt. Soc. Am.* **67**, pp. 375-378 (1977).

R. J. Noll, "Phase estimates from slope-type wavefront sensors," *J. Opt. Soc. Am.* **68**, pp. 139-140 (1978).

Epilogue

Adaptive optics is a diverse mix of optics, atmospheric physics, and electrical, mechanical, and computer engineering, with a splattering of common sense. Whereas the concepts can be simple, the implementation can be quite complicated, quite costly, and quite frustrating. This book was meant to be a readable introduction to the varied discipline, with enough detail to be a useful stand-alone reference and enough equations to satisfy the physicist in me.

Over the years, many researchers have developed and tried many new things. Some were technology driven; some were brilliantly innovative; some were miserable failures. The poetry of this introductory survey lies in the attempt to filter out the failures, thus leaving our time to be brilliantly innovative while we ride the wave of technology.

I hope that the trend in adaptive optics continues toward new applications and low cost technology that eventually will make adaptive optics a household word. Maybe not as common as central heat and running water ... but it will be close.

"We are all faced with a series of great opportunities – brilliantly disguised as insoluble problems."

— John Gardner,
U.S. Secretary of
Health, Education, and Welfare
1965-1968

Figure Credits

Fig. 1.2: U.S. Air Force Starfire Optical Range. Available: http://www.sor.plk.af.mil/images/binary.htm. 1999.

Fig. 1.3: Canada-France-Hawaii Telescope. Available: http://www.cfht.hawaii.edu/instruments/imaging/AOB. 1998.

Fig. 1.5: Center for Astronomical Adaptive Optics. Available: http://athene.as.arizona.edu:8000/caao/caao/laser.tn.gif and http://athene.as.arizona.edu:8000/caao/caao/laser.spot.gif. 1999.

Fig. 1.7: U.S. Air Force Starfire Optical Range. Available: http://www.sor.plk.af.mil/images/exper/jpeg/satcomp.jpg. 1999.

Figs. 3.1, 4.1, 6.1, 6.2, 6.3, 6.4, 7.1– 7.11 incl., 7.15, 9.1, and 9.7: Tyson, Robert K., *Principles of Adaptive Optics*, 2nd Ed., Academic Press, Boston, 1997.

Fig. 5.1: University of Chicago. Available: http://astro.uchicago.edu/chaos/laser.html. 1999.

Fig. 8.1: Physik Instrumente (PI) GmbH & Co. Available: http://www.physikinstrumente.com/pages/pzttilt.html. 1999.

Index

Robert K. Tyson is an Associate Professor of Physics at the University of North Carolina at Charlotte. This is his second monograph on adaptive optics. He also is the author of two book chapters on adaptive optics and the editor of an engineering handbook on the subject. Dr. Tyson is a member of the Optical Society of America and SPIE-The International Society for Optical Engineering. He holds a B.S. degree from Penn State and M.S. and Ph.D. degrees in physics from West Virginia University. His interests include adaptive optics, wavefront sensors and reconstructors, diffraction theory, atmospheric propagation, motor racing, and writing humor and satire.