ADAPTIVE OPTICS for ASTRONOMICAL TELESCOPES

OXFORD SERIES IN OPTICAL AND IMAGING SCIENCES

Editors

1. D. M. Lubman (ed.): *Lasers and mass spectrometry*
2. D. Sarid: *Scanning force microscopy with applications to electric, magnetic, and atomic forces*
3. A. B. Shvartsburg: *Non-linear pulses in integrated and waveguide optics*
4. C. J. Chen: *Introduction to scanning tunneling microscopy*
5. D. Sarid: *Scanning force microscopy with applications to electric, magnetic and atomic forces,* revised edition
6. S. Mukamel: *Principles of nonlinear optical spectroscopy*
7. A. Hasegawa and Y. Kodama: *Solitons in optical communications*
8. T. Tani: *Photographic sensitivity: theory and mechanisms*
9. D. Joy: *Monte Carlo modeling for electron microscopy and microanalysis*
10. A. B. Shvartsburg: *Time domain optics of ultrashort waveforms*
11. Solymar, D. J. Webb, and A. Grunnet-Jepsen: *The physics and applications of photorefractive materials*
12. R. M. Haralick: *Mathematical morphology: theory and hardware*
13. J. R. Schott: *Remote sensing: the image-chain approach*
14. P. Mouroulis and J. Macdonald: *Geometrical optics and optical design*
15. S. M. Barnett and P. M. Radmore: *Methods in theoretical quantum optics*
16. J. W. Hardy: *Adaptive optics for astronomical telescopes*

ADAPTIVE OPTICS *for* ASTRONOMICAL TELESCOPES

JOHN W. HARDY

New York Oxford
Oxford University Press
1998

Oxford University Press

Oxford New York
Athens Auckland Bangkok Bogota Bombay
Buenos Aires Calcutta Cape Town Dar es Salaam
Delhi Florence Hong Kong Istanbul Karachi
Kuala Lumpur Madras Madrid Melbourne
Mexico City Nairobi Paris Singapore
Taipei Tokyo Toronto Warsaw

and associated companies in
Berlin Ibadan

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

Oxford is a registered trademark of Oxford University Press

Library of Congress Cataloging-in-Publication Data
Hardy, John W., 1926-
Adaptive optics for astronomical telescopes / by John W. Hardy.
p. c. — (Oxford series in optical and imaging sciences, 16)
Includes bibliographical references and index.
ISBN 0-19-509019-5
1. Telescopes. 2. Optics, Adaptive. I. Title. II. Series.
QB88.H33 1998
681'.4123—DC21 97-5552

9 8 7 6 5 4 3

Printed in the United States of America
on acid-free paper

Preface

Adaptive optics is the technology for correcting random optical wavefront distortions in real time. One of its major applications is to compensate for turbulence in Earth's atmosphere, freeing ground-based telescopes from the limitations imposed by the blurred atmospheric window and thereby enabling their full optical performance to be achieved. Although originally proposed for astronomical telescopes in 1953, adaptive optics did not become a reality until the 1970s, when it was developed for national defense applications, specifically laser beam compensation and satellite imaging. This work was sparsely reported in the literature, and the progress made on one of the key technologies, the generation of laser beacons high in the atmosphere, was not revealed until 1991. The astronomical community now shows considerable interest in applying adaptive optics to the next generation of Earth-based astronomical telescopes, as well as in upgrading existing instruments.

The purpose of this book is to provide a unified and accessible account of the principles and practice of adaptive optics as they apply to astronomical telescopes. It is intended to provide astronomers, scientists, and engineers with a convenient source of information on the basic technology, recent achievements, and future potential of this important new field. This book is based on 25 years of experience in adaptive optics, which include the development of the first system capable of compensating a large astronomical telescope at visible wavelengths, as well as many other pioneering techniques and devices. During this period, adaptive optics has evolved from a highly specialized technology into one of the basic tools of ground-based astronomy. Several collections of technical papers that cover this period of intensive development in adaptive optics have recently been issued. While this material is a valuable resource for specialists in the field, it lacks the systematic approach and consistent treatment that is possible in a book. I hope that this book will fulfill the need for a comprehensive and up-to-date reference source on adaptive optics in astronomy.

The development of adaptive optics is presented as a logical step in the long history of astronomical observations; for example, the techniques presently used for wavefront sensing have origins in optical sighting and testing devices that have been in use for centuries. A major goal of this book is to convey an understanding of the physical principles on which adaptive optics is based, in the hope that this will stimulate new ideas, new devices, and new applications. Adaptive optics is an interdisciplinary activity that draws on many related sciences and technologies, such as optical design, wave propagation, photon detectors, servo systems, lasers and computer technology, to name a few. Relevant results from these disciplines have been summarized in this book; the reader is referred to the specialized texts listed in the bibliography for more detailed development and information.

Several people have assisted me in the preparation of this book. I am grateful to Horace Babcock for sharing the experiences that led to his original proposal for compensating astronomical telescopes. My

colleague Edward P. Wallner has generously contributed many ideas and analyses on various aspects of adaptive optics, and also reviewed sections of the text. Mary Latham, librarian of Litton/Itek Optical Systems, provided valuable access to the technical literature contained in a large number of specialized publications. Peter Nisenson reviewed the manuscript and suggested several improvements. Finally, this book could not have been written without the support and encouragement of my wife, Ethel, who also prepared the bibliography and the subject index.

Lexington, Massachusetts
September, 1997 J.W.H

Contents

List of Symbols

The symbols used in this book are based on those conventionally employed in the adaptive optics literature. Because of the large number of parameters involved, many symbols have more than one meaning. Symbols are generally defined in the sections in which they appear, and their meaning should always be clear from the context. Symbols printed in **bold type** represent multidimensional quantities, vectors, or matrices.

A	Area; amplitude; baseline
$\mathscr{A}$	Complex amplitude
a	Radius or side of an aperture; actuator spacing; amplitude; decay factor
a_F	Fitting error coefficient of a deformable mirror
$B(\ )$	Correlation or coherence function; transfer function
C	Contrast; coupling coefficient; capacitance
C_P	Contrast reduction ratio
$C_N^2(\ )$	Structure parameter for refractive index variations
$C_T^2(\ )$	Structure parameter for temperature variations
C_W	Wavefront curvature
$C(\ )$	System output function
c	Velocity of light in vacuum, $2.998 \times 10^8\ \mathrm{m\,s^{-1}}$
D	Diameter of optical aperture
$D(\ )$	Structure function
$\mathscr{D}$	Star density, stars per rad^2
d	Distance; subaperture diameter
d_0	Diameter over which the phase error due to focal anisoplanatism is 1 rad^2
d_1	Diameter over which the conic tilt error is 1 rad^2
$d_{m,n}$	Piezoelectric constant of order m,n
E	Energy; error propagation factor; elastic (Young's) modulus
$E(\ )$	Electromagnetic field
e	Numerical constant, 2.718; read noise of a CCD array, electrons per pixel
F	Multiplication or scaling factor
$F(\ )$	Filter function; power spectrum
$\mathscr{F}$	Fourier transform
f	Frequency; bandwidth
f_G	Greenwood frequency
G	Gain factor
$G(\ )$	Gain or transfer function; frequency spectrum; filter function
g	Period of optical grating; gain factor; acceleration due to Earth's gravity
$g(\ )$	Gradient function
$\boldsymbol{g}$	Gradient vector
H	Height above sea level; magnetic field intensity
H_S	Irradiance
H_λ	Spectral irradiance
h	Height above telescope aperture; Planck's constant (energy per Hz of a photon), $6.6256 \times 10^{-34}\ \mathrm{J\,s}$
$\bar{h}$	Mean turbulence height
h_c	Beam clearance height
I	Intensity; moment of inertia
$I(\ )$	Intensity function
i	Index number; complex operator, $\sqrt{-1}$
J_n	Bessel function of first kind and order n
j	Index number
K	Contrast reduction factor; mechanical stiffness
k	Wave number, $2\pi/\lambda$; index number
L	Optical pathlength; baseline of an interferometer
$\mathscr{L}[\ ]$	Laplace transform
$\mathscr{L}_0$	Outer scale of turbulence
ℓ	Length; scale size of turbulence eddies

ℓ_0	Inner scale of turbulence
M	Bending moment
$M(\ )$	Modulation (transfer) function
m	Electrostriction constant
m	Number of elements or terms; azimuthal frequency of Zernike polynomial
m_V	Visual magnitude (stellar)
N	Mode number; number of elements; photon flux
N_λ	Spectral radiance
$N(\lambda)$	Refractivity, (n — 1) x 10^6, at wavelength X
n	Refractive index
n	Number of elements or terms; number of photons; radial degree of Zernike polynomial
OTF	Optical transfer function
P	Pressure; power
$P(\ )$	Point spread function
p	Probability
q	Quantum efficiency
R	Radius of curvature; resistance, ohms
Re	Reynolds number
$\mathscr{R}$	Resolution or resolving power
r	Radius; separation of points
r_0	Turbulence coherence length (Fried's parameter)
r	Position vector
S	Strehl ratio Fractional sky coverage (of an adaptive optics system)
S( )	Optical transfer function
SNR	Signal-to-noise ratio
s	Signal value; shear distance; wavefront slope; complex variable a + $j\omega$
T	Temperature; tension
T_A	Optical transmission of the atmosphere
T_O	Transmission of optical components
$T(\ )$	Modulation transfer function
t	Time interval; pulse length; thickness
t_0	Turbulence change time
$U(\ )$	Complex amplitude of an optical wave; system input function
u, v	Coordinates of the spatial frequency (Fourier) plane
V	Velocity in a turbulent flow; electrical potential (voltage)
V_0	Characteristic velocity
v	Velocity; wind speed
W	Irradiance
$W(\ )$	Wavefront function; intensity weighting function
w	Weighting factor; wavefront value; effective radius of a Gaussian beam
x, y	Position coordinates
x,y	Position vectors
Z	Axial distance; lens focal distance
Z_j	Zernike expansion term or mode
$Z[\ ]$	Z-transform
z	Position coordinate; distance along optical path; optical range
Δz	Range gate interval; change in position
α	Angle of displacement (tilt); angle of arrival; coefficient of thermal expansion
β	Field angle
γ	Attenuation factor; contrast (or modulation) of interference fringes
$\Gamma(\ )$	Gamma function
Δ	Displacement; incremental value
δ	Delta function or small angle
$\delta(t)$	Dirac delta function
ε	Rate of energy input
$\varepsilon(\ )$	Error function
ζ	Zenith angle
η	Efficiency
θ	Field angle; orientation angle; angular size
θ_0	Isoplanatic angle
θ_E	Effective isoplanatic angle
κ	Spatial wavenumber; coefficient
λ	Wavelength
$\Delta\lambda$	Spectral bandwidth
μ_m	Full turbulence moment of order m
$\mu_m^{\pm}(H)$	Partial turbulence moment, above (+) or below (~) height H
ν_0	Kinematic viscosity of fluid; Poisson's ratio
ρ	Spatial dimension or vector; mass density
$\rho(h)$	Density of the earth's atmosphere at height h
σ	Standard deviation; stress; backscatter cross-section
σ_p, σ_w	Standard deviation of wavefront phase error
σ_α	Standard deviation of angular tilt error
τ	Time constant; time delay
τ_n	Natural lifetime (of a sodium atom)
τ_s	Saturation time (of stimulated emission)
Φ_N	Power spectrum of refractive index variations
Φ_T	Power spectrum of temperature variations
$\Phi(\ )$	Spectral density
ϕ	Optical phase angle
$\phi(\ )$	Phase function
χ	Log-amplitude
ψ	Electromagnetic field
$\psi(\ )$	Wavefront phase function
$\Omega(\ )$	Wavefront surface or slope function
ω	Angular frequency
$\prod$	Product
∇^2	Two dimensional Laplacian = $\partial^2/\partial x^2 + \partial^2/\partial y^2$
T	(superscript) Transpose of a matrix
Tr	Trace of a matrix
$\oplus$	Generalized inverse of a matrix

ADAPTIVE OPTICS *for* ASTRONOMICAL TELESCOPES

1 The Short, Eventful History of Adaptive Optics

A man that looks on glass,
On it may stay his eye;
Or if he pleaseth, through it pass,
And then the heaven espy.

George Herbert (1593-1633)

1.1 Introduction

Earth's atmosphere is our window on the universe. This fragile, translucent shell of gases nurtures life on Earth, providing essential air and moisture, as well as giving protection from harmful radiation and the destructive impact of meteoric debris. Its transparency allows us tantalizing glimpses of the cosmos from the security of our planet, but the atmosphere is in constant motion and is far from being a perfect window. Light rays that have traveled unchanged over vast distances, bringing information from the distant reaches of the universe, are distorted in the final moments of their long journey through space.

Human curiosity about the mysterious objects seen through the atmospheric window gave birth to the science of astronomy. The opening quotation of this chapter was written shortly after Galileo made his discoveries with the newly invented telescope. For most of human history, however, astronomers relied on the unaided eye to view the sky and to keep track of its changing aspect. The Sun and Moon provided a daily and monthly calendar and marked religious celebrations; the rising and setting of special stars guided seasonal activities, such as the planting of crops. The atmosphere was revealed by the colors of sunset and the twinkling of stars; these were natural effects to be admired, adding to the enjoyment of a starry night.

The present era of astronomy is dominated by huge Earth-based telescopes, probing the furthest reaches of the universe. The atmosphere, through its properties of absorption, dispersion, and turbulence, limits our ability to make scientific measurements with these instruments. Smaller instruments orbiting above Earth's atmosphere will play an important role in twenty-first century astronomy, but, because of the greater size and flexibility of Earth-bound instruments, it is likely that they will continue to be the main source of astronomical information for the foreseeable future.

Great efforts have been made in the last half-century to devise means of compensating atmospheric effects. Spectral windows have been found at infrared (IR) and radio frequencies that enable observations of the cosmos at wavelengths greatly different from the visible band. Radio waves will penetrate clouds and are little affected by turbulence, but their longer wavelength limits the angular resolution obtainable. At optical wavelengths, the systematic properties of the atmosphere, such as refraction and dispersion, are easily compensated, but the random variations produced by atmospheric turbulence have proved to be a difficult obstacle. The purpose of this book is to review the effects of turbulence on astronomical observations and to describe the technology that has been developed to overcome it. There are two aspects to this endeavor, foreshadowed in the opening quotation. We may look upon a telescope as an object in its own right, to be crafted and admired for its technology, but its true purpose is to reveal the truth about the universe.

To place adaptive optics in historical perspective, the changing character of astronomical observations through the ages will be outlined briefly, starting with the ancient goal of determining the positions and

motions of the heavenly bodies. The regularity of the motions was recognized in ancient times, and, for most of recorded history, astronomy has been regarded as a utilitarian activity with timekeeping as its main purpose. Early models of the universe were proposed to explain the motions of the heavenly bodies and not as a way of describing their true nature. The invention of the telescope in the early seventeenth century revealed the structure of the solar system, but it was not until the founding of astrophysics in the mid-nineteenth century that astronomy was broadened into an investigation of the character, composition, and history of the universe.

The essence of adaptive optics is its ability to measure, and ultimately to correct, the direction of light rays disturbed by atmospheric turbulence. In a simpler context, accurate measurement of the positions of heavenly bodies has been a vital part of astronomy since its earliest days. It is appropriate, therefore, to trace the development of astronomical position measurement through the ages as an introduction to the technology of adaptive optics.

1.2 Astronomy with the Unaided Eye

The most noticeable features of the night sky are the patterns produced by the fixed stars, among which wander the planets. The stars can be differentiated by color and brightness, but they all appear to move as if they were attached to a shell that rotates consistently about Earth. Before 600 B.C., Babylonian astronomers, using simple sighting instruments to establish direction, kept records enabling them to determine the motions of the Sun, Moon, and planets with surprising accuracy. Great importance was attached to the prediction of eclipses. This information was passed on to the Greeks, among them Plato (427-347 B.C), who described a simple Earth-centered model of the universe based on concentric spheres. This model did not explain all the known phenomena, and indeed Plato's followers tended to disregard observational experience, proposing abstract and idealistic models of the universe. To better explain the observed erratic motions of the planets, Eudoxos (408-355 B.C.) evolved a complex system of 27 geocentric spheres, four of which were assigned to each planet. It was recognized that this was a purely mathematical device. Pedersen [1974] notes that heavenly bodies were not regarded as permanent material objects until about 400 B.C.

Aristotle (384–322 B.C.), a pupil of Plato, had a more objective view of nature. He valued mathematics, but considered it auxiliary to direct observation, which he believed to be the only way to acquire knowledge of the physical world. He taught that "Mathematical relations are to be discovered in nature, but not imposed upon it." According to King [1955], Aristotle extended the Eudoxoan system into a consistent physical model that pictured Earth as a tiny ball standing motionless at the center of 55 transparent, rotating, crystalline spheres to which were attached the planets and stars.

Around 300 B.C., a university and library were founded at Alexandria, which became the center of Greek astronomy, both theoretical and observational. Aristarchos of Samos (310–230 B.C.) created a consistent heliocentric theory of the universe, but regarded this only as a mathemetical hypothesis and not as representing physical reality.

The apex of Greek astronomy was reached with Hipparchos (190–120 B.C), who combined observational and theoretical skills, compiling the first catalog of the positions and magnitudes of the fixed stars. He used graduated measuring instruments, including armillary spheres. His discovery of the precession of the equinoxes, which causes stellar coordinates to shift less than 1 arc minute each year, indicates the accuracy with which measurements could be made at that time. Hipparchos' observations of planetary positions revealed the shortcomings of Eudoxoan theory, resulting in his adoption of a system of epicycles, first suggested by Apollonius. The careful and practical approach used by Hipparchos marks him as the first observational astronomer to display the modern spirit.

The legacy of Hipparchos was summarized and expanded by Ptolemy (AD. 90-160) in his Almagest or "Great Composition," which remained the prime reference for astronomy for more than 1000 years. His Earth-centered universe was a geometric scheme that explained the motions of the heavenly bodies in terms of uniform, circular motions, enabling their positions to be predicted far into the future. Ptolemy also described observational instruments, including the quadrant and its simpler relative the triquetrum, in which zenith angles were measured in terms of chordal distances. This must have been the first instrument in which calibration "software" was used to support the optical hardware.

During ancient times, there was probably little awareness of the optical properties of Earth's atmosphere. The phenomena of optical reflection and refraction were known in the fifth century B.C. through the use of burning mirrors and glasses that focused the Sun's rays, but the first quantitative description did not appear until the second century AD., in Ptolemy's book on optics. All the main features of planetary motion, including retrograde motion, were known in Plato's time, around 400 B.C.. Such observations do not require instruments, so atmospheric effects were insignificant.

The first astronomical instruments were used in Egypt and Mesopotamia around 1500 B.C, to measure the daily and yearly motion of the Sun. They included the gnomon, a vertical stick, and the polos, a bowl that surrounds the gnomon and on which its shadow falls. Even with an instrument as simple as the gnomon, figure 1.1(a), many fundamental measurements were made, including the

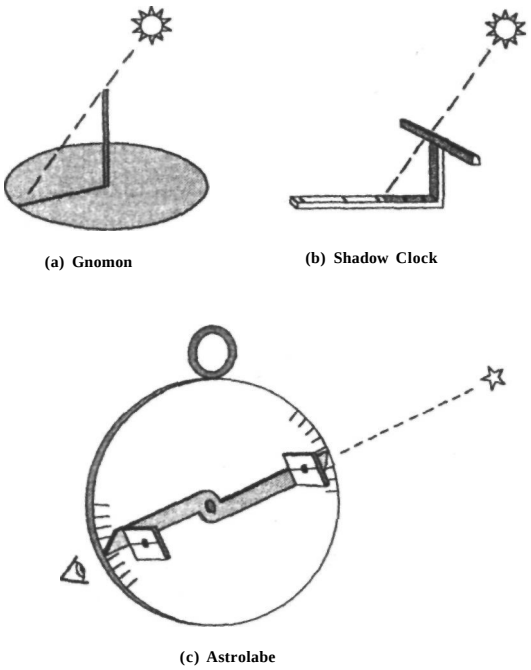


Figure 1.1 Early astronomical instruments. (a) The gnomon was used around 1500 B.C. to determine the position of the Sun, which is deduced from the length and angle of the shadow thrown by a vertical stick. (b) The Egyptian shadow clock, used since 1000 B.C., has a horizontal bar divided into six hours. It is reversed at noon to record the afternoon hours. From these instruments, the familiar sundial was developed. (c) The astrolabe dates from the third or fourth century AD. and combines two instruments. The alidade, shown above, measures the altitude angle of the Sun or a bright star. On the other side of the disk is a calculator that converts the measured angle into local time.

local meridian and cardinal points, the dates of the solstices and equinoxes, and the length of the year in days. It is interesting to note that such astronomical observations preceded formal theories of the solar system by about 1000 years. The daylight hours were recorded by shadow clocks, such as that shown in figure 1.1(b). They were calibrated with six hours and reversed at noon to record the afternoon hours.

By the fourth century B.C, the gnomon had been developed into a primitive sundial by dividing the polos into hours [Pedersen, 1974]. For ordinary activities, day and night were each divided into 12 hours. At the equinoxes, the hours were all equal, but at other times of the year they varied in length, the daytime hours being long in summer and short in winter. For astronomical measurements, hours of constant length were preferred. Short periods of time were measured by water clocks, in which water was released from a vessel through a small orifice. Accurate timekeeping remained a major problem in astronomy until the eighteenth century.

Cleomedes (first century A.D.) is credited with the first reference to atmospheric refraction, in relation to a paradoxical eclipse during which both the Sun and the eclipsed Moon were visible. This should not occur if light rays travel in straight lines. Cleomedes explained the phenomenon correctly as being due to the bending of light rays in the atmosphere, so that both disks appear to be raised above the horizon [Cohen and Drabkin 1948]. Ptolemy also studied atmospheric refraction, concluding that one of its effects was to increase the apparent polar distance of stars. An account of the history of the measurement of astronomical refraction, with an extensive list of references, has been given by Mahan [1962].

After the disintegration of the Roman Empire and the burning of the library at Alexandria in A.D. 641, the Arabian cities of Baghdad, Cairo, and Damascus became centers of astronomy. Ptolemy's work was studied and his idea of using quadrants for measuring the altitudes of stars was put into practice on a large scale. In the eleventh and twelfth centuries, huge observational instruments that used visual sightlines were built in Europe, Arabia, and India. It was at this time that the great Arabian physicist Ibn al-Haitham (Alhazen, 965-1039) first described atmospheric effects on astronomical observations — the changing shape of the Sun and Moon as they approach the horizon and the need to account for atmospheric refraction in determining the positions of stars.

Around this time, Arabian craftsmen perfected the astrolabe, a small instrument that enables time to be determined from the position of the Sun or stars, giving it great value for surveying and navigation. The basic theory of the astrolabe had been worked out by Ptolemy. The flat Arab astrolabe, which probably originated in Egypt in the third or fourth century A.D., consists of a circular disk that is held in a vertical position by suspending it from a ring. On one side of the disk is the alidade, a pivoted arm equipped with two sights that enable it to be aimed at a bright star or the Sun, the altitude of which can be found by reading a scale calibrated in degrees, as shown in figure 1.1(c). The other side of the instrument consists of a calculator that converts the measured angle into the local time. A fixed circular disk, the tympanon, is engraved with a stereographic projection of the heavenly spheres for the latitude at which the instrument is to be used. Superimposed on the tympanon is a perforated disk called the rete, which contains the projected positions of the ecliptic and a number of bright stars; the rete is cut away so that the tympanon is visible beneath it. To use the calculator, the rete is turned until the point marking the observed star is aligned with the altitude circle on the tympanon that corresponds to the measured altitude angle, whereupon the time cor-

responding to the date is read off the outer scale. The astrolabe calculator (an early analog computer) solves problems of spherical geometry, making it generally useful in astronomy and geodesy. The astrolabe probably had the longest useful life of any scientific instrument, remaining in use for more than 1000 years. Obsolescence was not a significant factor in those days.

In the present context, the method of angular measurement used in the alidade is of primary interest. The accuracy of a simple angular measuring instrument of this type depends on two factors:

1. the precision with which the instrument can be pointed at the star or other object;
2. the accuracy with which the scales are engraved on the instrument.

The sighting precision depends both on the baseline (the separation between backsight and foresight) and on the design of the sighting devices themselves. Many sighting devices were devised for angular measurement with the naked eye; some of these are shown in figure 1.2. Astrolabes usually employed pinhole sights at each end of the alidade. The problem

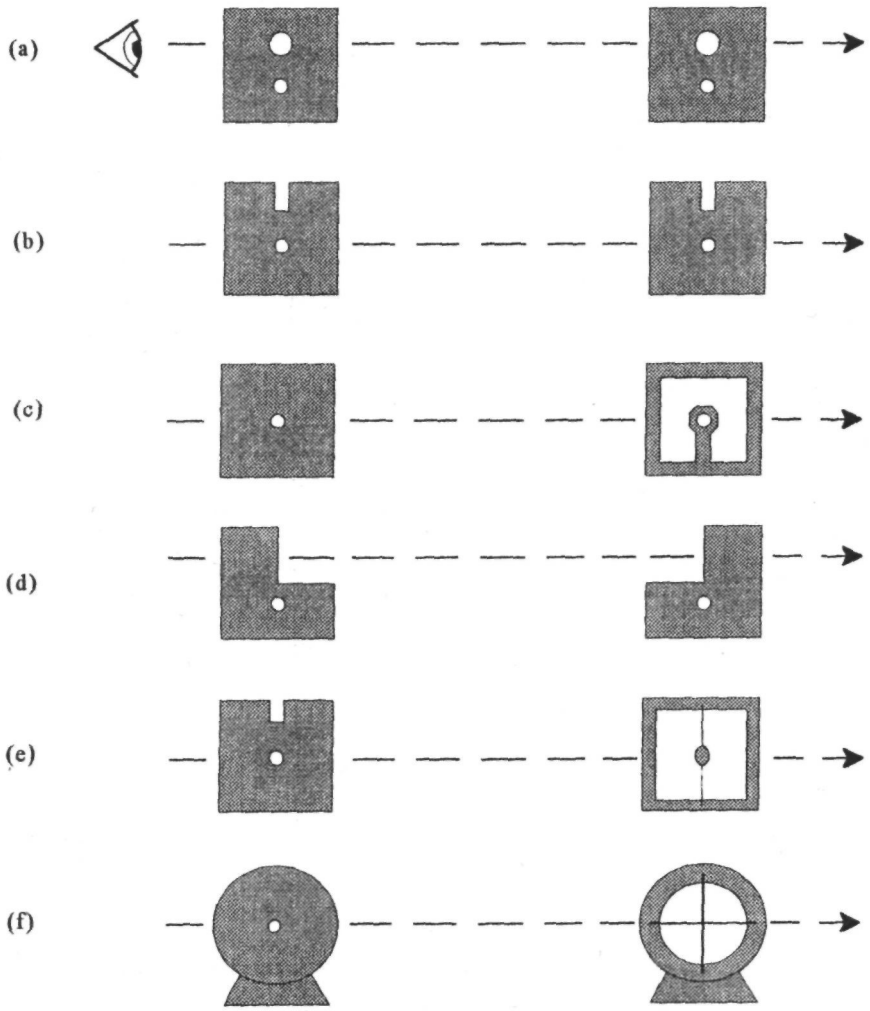


Figure 1.2 Sighting systems used in early instruments. (a) Arab astrolabes (ca. 1000–1500 A.D). (b) Graphometer, Philippe Danfrie, 1597. (c) Alidade, Hans Christoph Schissler Jr, 1591. (d) Graphometer, Matteo and Giovanbattista Botti, undated. (e) Quadrant, Tobias Volckmer, Braunsweig, 1608. (0 Quadrant, Carlo Rinaldini, 1667.

with pinholes is that to obtain high angular accuracy they need to be very small, which makes stars difficult to see. Light is lost when the pinhole in the backsight is smaller than the pupil of the eye, which may be as large as 7 mm at night. Using a large pinhole increases visibility but reduces angular accuracy. In many astrolabes, two sets of holes were provided, as shown in figure 1.2(a), the small ones for use with the Sun and the large ones for stars.

In other instruments, such as quadrants or graphometers, which make angular measurements in a horizontal plane, two slots or a combination of slots and pinholes were often used, as depicted in figure 1.2(b). In some cases, an aperture was cut away around the foresight, as shown in figure 1.2(c)–(f), to provide a larger field for finding the star. The use of slots or pinholes of finite size inevitably introduces an error due to parallax, which can be reduced by using sighting devices consisting of a single edge. The sight shown in figure 1.2(d) is an early example of the use of single knife-edges for determining the angle of optical rays.

The culmination of naked-eye astronomy came late in the sixteenth century, with Tycho Brahe's magnificent Uraniborg observatory in Denmark (established in 1576), which used large quadrants and armillary spheres to measure the positions of heavenly bodies. The accuracy of direct sighting instruments became an important issue at this time. It must be remembered that this was before the invention of the telescope and no optical aids to viewing were in use for the observation of celestial bodies. The acuity of the human eye is about 1 minute of arc (1/60 degree) in daylight, when the pupil diameter is about 2.5 mm. At night, although the pupil may expand to 7 mm, the acuity is usually far worse. With sighting instruments, two additional factors must be taken into account: the precision achievable in lining up the pinholes or other sighting devices, and the accuracy with which the instrument itself is calibrated. Before Tycho Brahe, the overall accuracy of most astronomical measurements was no better than ±10 minutes of arc, and with the passage of time the tables of the positions of planets were in error by several days.

By careful construction and calibration of his large instruments, Tycho was able to reduce the sighting errors to about 2 minutes of arc. He then devised an improved sighting system that eliminated the parallax errors encountered with conventional pinholes. The eyepiece consisted of four slits defining a square, as shown in figure 1.3. At the object end of the instrument was placed an opaque square of exactly the same size. The instrument was adjusted so that the star just touched each side of the square when viewed through the corresponding slit. A variant of this method for single-axis measurements used an opaque cylinder as the foresight. By this means, Tycho further reduced the error in his observations to less than 1/2 minute of arc. Even this angle is considerably greater than the angle-of-arrival variations due to turbulence, so atmospheric effects are not apparent, although the angular measurements still have to be corrected for atmospheric refraction. To avoid wind buffeting, some of Tycho's instruments were placed in an underground observatory. The accuracy to which star positions could be measured was not improved until the telescopic sight was invented about 50 years later.

Tycho's sighting scheme for establishing the direction of a star may be regarded as a forerunner of the optical testing technique developed by Foucault in the 1850s. This test employs a knife-edge at the focus of an objective lens or mirror to determine the directions of the light rays emanating from each part of the objective.

Copernicus (1473-1543) had meanwhile advanced the idea of a heliocentric system as a better model for determining the motions of the planets, but, even with the information available from his superior instruments, Tycho rejected this in favor of an Earth-centered universe. It is astonishing to realize that after 20 centuries of astronomy using the unaided eye, during which the positions and motions of heavenly bodies had been recorded in great detail, astronomers could not even agree on a physical model of the solar system, let alone the nature of the stars beyond it. Early in the seventeenth century, using Tycho's observations, Kepler formulated his laws of planetary motion and produced a working model of the solar system, but, even at this recent date in human history, the grand design of the universe was not even suspected.

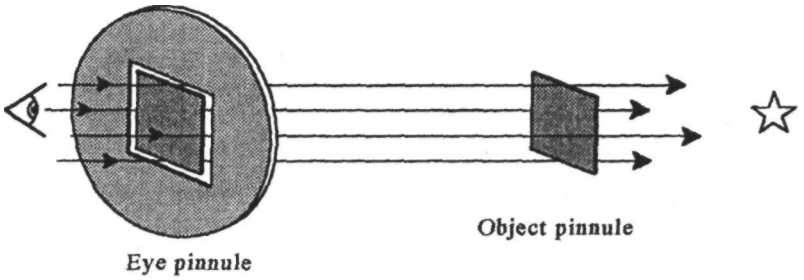


Figure 1.3 Tycho Brahe's sighting system to eliminate parallax, used in the late sixteenth-century just before the invention of the telescope. The star was lined up simultaneously with four slits on all four sides of a square pinnule.

1.3 Telescopes and Atmospheric Turbulence

It was the invention of the telescope that gave astronomers the key to understanding the universe. Galileo heard of the existence of the new instrument in 1609, but did not know the prescriptions for the lenses. He quickly found a combination that worked, using a concave eyepiece with a convex objective, and confirmed that it truly made distant objects appear closer. Over the next few months, he made better instruments. On January 7, 1610, he pointed his fourth and best telescope at Jupiter and saw three stars near the planet. He thought, at first, that they were fixed stars, but the following night was amazed to find that they had changed their positions relative to the planet. Further observations revealed a fourth object. He soon realized that they were satellites orbiting the planet; his telescope had revealed the basic structure of the solar system. In the same year, he observed the changing size and phases of Venus, confirming that it must revolve around the Sun and not the Earth. The geocentric model of the universe, accepted for more than 2000 years, was demolished. The theories that Copernicus and Kepler had laboriously constructed were confirmed beyond doubt in a few nights' observation.

Galileo's discoveries demonstrated the value of using new instruments to extend human knowledge of the natural world through scientific observation, rather than attempting to deduce the truth from arbitrary assumptions, or by logical arguments, as the early Greek philosophers had attempted. A basic principle of science had been established.

The telescope provides both a source of new information about the universe and a means of confirming theories that explain previous discoveries. This relationship between instrumentation and theory is fundamental to science. Observations produce information; analysis of the information suggests a new theory; new instruments are developed to confirm the new theory; new facts are discovered; and so the process is repeated. Each step in the process increases the total store of knowledge and provides the foundation for the next.

In the first centuries of astronomical telescopes, the instruments were simple. The problems to be overcome were basic technical limitations, such as the fabrication of lenses and mirrors, and mechanical tracking of the heavenly objects. Although determining the positions of stars and planets was the prime objective of astronomy at this time, the telescope was not immediately applied to sighting instruments. The telescopic sight was invented around 1640 by William Gascoigne, who relates that an obliging spider spun a thread, which, when placed at the focus of a convex eyepiece, became superimposed on the telescopic image as a sharply defined line [King 1955]. The principle is shown in figure 1.4. In practice, crossed hairs were found to be more durable. Gascoigne went on to invent the eyepiece micrometer, in which the separation of two metal knife-edges located at the focal plane could be adjusted precisely by calibrated screws. This device enabled the diameters or separations of objects to be measured with an accuracy limited only by the telescope optics and the atmosphere. Gascoigne's pioneering work on graduated instruments was cut short when he was killed at the age of 24 in the Civil War in England.

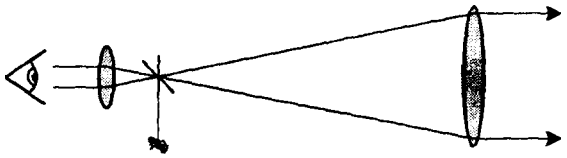


Figure 1.4 William Gascoigne's telescopic sight (1640). Fine cross-hairs are located at the focus of the eyepiece, enabling the telescope to be accurately pointed at a star.

Galileo's telescopes had a resolving power of only 10-15 seconds of arc, on which the effect of atmospheric turbulence was scarcely visible. As the optical quality of telescopes improved, astronomers became more aware of the limitations imposed by the atmosphere, not just in bending the light rays to change the position of stars, but also in blurring the images of stars and planets. The first astronomer to appreciate the importance of atmospheric conditions on telescopic images was Christian Huygens, who, around 1656, was using an aerial telescope of 123-feet focal length, the small objective of which was mounted on a high pole. He noted that stars twinkled and that the edges of the Moon and planets trembled in the telescope, even when the atmosphere appeared calm and serene. So frequent were nights of poor seeing that Huygens warned observers against too hastily blaming their telescopes.

Isaac Newton, from his studies of optical prisms, was well aware of the chromatic problems of refractive telescopes. His measurements suggested that all optical media had the same dispersive power, leading him to the erroneous conclusion that it was impossible to correct the chromatic aberration of refractive telescopes. In his search for better instruments, Newton, in 1668, built a small telescope using a concave metal primary with a flat diagonal secondary to reflect the beam to an eyepiece at the side of the tube. Newton's main contributions to science are his laws of motion, his understanding of gravity, and his work on the nature of light. It is somewhat ironic that his telescope, conceived to avoid a problem that does not exist, has become one of the classic configurations for astronomical telescopes.

Newton's instrument was not, in fact, the first reflecting telescope to be invented. In 1663, James Gregory had described an instrument consisting of a large primary paraboloid to collect the light, with a small concave secondary ellipsoid to produce the

final image. The center of the primary was perforated to allow the insertion of an eyepiece. Gregory was unable to obtain mirrors with the necessary aspheric surfaces and the design was not implemented until 1674, by Robert Hooke. Meanwhile, in 1672, Cassegrain had invented a reflecting telescope using a paraboloidal primary and convex hyperboloidal secondary, the light being reflected back on its axis through a hole in the primary to the eyepiece. Although this design was derided by Newton, it has become the basis of most large modern astronomical instruments.

Newton was familiar with the problems caused by atmospheric turbulence, and in the following passage from *Opticks*, published in 1704, he gave a vivid description of the effects of turbulence as well as some good advice on the siting of telescopes:

> If the Theory of making Telescopes could at length be fully brought into Practice, yet there would be certain Bounds beyond which Telescopes could not perform. For the Air through which we look upon the Stars, is in a perpetual Tremor; as may be seen by the tremulous Motion of Shadows cast from high Towers, and by the twinkling of the fix'd Stars. But these Stars do not twinkle when viewed through Telescopes which have large apertures. For the Rays of Light which pass through divers parts of the aperture, tremble each of them apart, and by means of their various and sometimes contrary Tremors, fall at one and the same time upon different points in the bottom of the Eye, and their trembling Motions are too quick and confused to be perceived severally. And all these illuminated Points constitute one broad lucid Point, composed of those many trembling Points confusedly and insensibly mixed with one another by very short and swift Tremors, and thereby cause the Star to appear broader than it is, and without any trembling of the whole. Long Telescopes may cause Objects to appear brighter and larger than short ones can do, but they cannot be so formed as to take away that confusion of the Rays which arises from the Tremors of the Atmosphere. 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.

The possibility of compensating atmospheric turbulence was not even dreamed of in Newton's time.

At the end of the seventeenth century, the search for a reliable method of determining longitude obsessed the seafaring nations of Spain, France, and Britain, all of which offered substantial rewards for solving the problem. Galileo had suggested observations of the Jovian satellites as a method of establishing a standard time over the entire globe. This could be used on land, but observations were too difficult at sea, where it was most needed for navigation. Observing the position of the Moon among the stars had been suggested by Johannes Werner in 1514 as a method of keeping time, but it required an immense catalog of star positions. In London, on John Flamsteed's advice, King Charles II established the Royal Observatory at Greenwich expressly to compile a star catalog for the determination of longitude by the method of lunar distances. The 1714 Longitude Act offered £20,000 for a method of finding longitude to a $\frac{1}{2}$degree, corresponding to 30 nautical miles at Earth's equator. Flamsteed's star catalog was completed in 1725, but, as it turned out, another technology eventually claimed the prize: timekeeping with a marine chronometer, the brainchild of John Harrison [Sobel 1995]. The lunar distance method continued to be used and tables were published until 1907.

During the eighteenth century, the quality of astronomical telescopes gradually improved. The reflector reigned supreme, as there appeared to be no way to achromatize a refracting objective. Mirrors with diameters up to about 18 inches (0.46m) were cast and polished from speculum metal, composed of copper and tin. Although their reflectivity was poor, the shape of these mirrors was good enough to resolve stars as close as 1 arc second when atmospheric conditions permitted. In 1778, William Herschel polished "a most capital speculum" of only 6.2-inch (15.7-cm) aperture with which he compiled a catalog of double stars and, about 2 years later, discovered the planet Uranus. Herschel went on to make reflectors with apertures of 18.8 inches (0.48 m) and, finally, 48 inches (1.2 m), the latter being by far the largest-aperture telescope that had been built up to that time. Herschel was well aware of the limiting effects of Earth's atmosphere, especially as the 48-inch instrument was set up next to his house in the damp and misty valley of the River Thames, near London. Although the weather was frequently severe, he found that high humidity did not preclude good seeing. He wrote:

> By enlarging the aperture of the telescope, we increase the evil that attends magnifying the object without magnifying the medium . . . However, in beautiful nights, when the outside of our telescope is dripping with moisture discharged from the atmosphere, there are now and then favourable hours in which it is hardly possible to put a limit to magnifying power. But such valuable opportunities are extremely scarce; and, with large instruments, it will always be lost labor to observe at other times.

This lesson went unheeded and had to be relearned several times before finally taking root at the end of the nineteenth century.

It could be said that the huge reflectors built by Lord Rosse in Ireland in the 1840s were intended more as technology demonstrations than astronom-

ical instruments. Rosse produced speculum metal mirrors of 36- and 72-inch apertures, successfully grinding and polishing them to a high degree of perfection. The 72-inch mirror was capable of resolving 0.5 seconds of arc, but could be used on few occasions because of the weather. In common with Herschel's large telescope, it employed a very crude mount in which pointing was achieved by ropes and pulleys. No major discoveries were made with this instrument, which served as a warning that large apertures are wasted under indifferent seeing conditions.

William Lassell pioneered the use of equatorial mounts in large reflecting telescopes, and was also one of the first astronomers to move his telescope to a site having good seeing, rather than waiting for good seeing to come to him. Lassell built reflecting telescopes of 24-inch and 48-inch apertures, which he took to Malta in 1852 and 1862, respectively. He drew up a catalog of 600 new nebulae and commented "I was never more struck with the conviction how necessary a pure tranquil sky is to the just performance of a very large telescope".

Although John Dolland had discovered in 1757 how to make achromatic lenses by combining elements made of flint and crown glass (a feat that Newton had deemed impossible), these were only available in small sizes up to about 5-inch diameter because of the difficulty of casting flint glass disks of the required optical quality. It was not until the end of the century that large flint blanks were available, resulting in the production of many large refracting telescopes during the second half of the nineteenth century. Alvan Clark made many objectives for these telescopes, including the 26-inch refracting telescope for the U.S. Naval Observatory in Washington, which was completed in 1873. The imaging capabilities of these large telescopes were masked by the generally poor seeing at the observatories, most of which were at low altitudes, in or near cities. Most observations were still made visually, and the large apertures enabled considerable detail to be seen on the Moon and planets, as well as in the brighter star clusters. There are usually short periods of good seeing that enable a visual observer to catch details that are normally obscured, hence the use of drawings to record the observations. In the last decades of the century, refractors of 30 inches and more were produced, culminating in the 36-inch refracting telescope at Lick Observatory in 1888 and the Yerkes Observatory 40-inch refracting telescope, dedicated in 1897.

The technique of silvering glass substrates was developed by von Steinheil and by Foucault in 1856. Glass mirrors were lighter and easier to work than metal mirrors. In England, a 36-inch glass mirror made by Calver was silvered and mounted in a Newtonian telescope by Common in 1879. With this instrument, he obtained excellent photographs of planets and nebulae, using exposures as long as 1 1/2 hours. A similar 36-inch reflecting telescope was made by Common for Edward Crossley's private observatory in the north of England. This telescope has a long and varied history. Realizing that such a telescope deserved a site with better seeing, Crossley donated the telescope to Lick Observatory in 1895. After the mirror was refigured and the mount was strengthened, Keeler obtained a brilliant series of photographs showing vast numbers of extragalactic nebulae, although their identity was not realized at the time. The value of reflecting telescopes for photography had already been convincingly demonstrated in 1887 when Roberts obtained the first photograph showing the spiral structure of the Andromeda nebula, using a 20-inch mirror at prime focus. The success of the Crossley telescope reinforced the value of operating a large telescope at a site with excellent seeing, and provided the stimulus for the construction of even larger reflecting instruments, which soon came to dominate the field.

1.4 The Emergence of Astrophysics

In the middle of the nineteenth century, the direction of astronomy was changed forever by two inventions. The first of these was photography, which, with its ability to integrate the light from long exposures, could reveal faint stars and nebulae that were invisible to the human eye. Photography was first used to record images of the Sun and Moon in the 1840s and came into general use when the dry-plate process was perfected in the 1880s. The second invention was the spectroscope, which enables chemical elements to be identified by examination of the wavelengths of light emitted at high temperatures. Between them, these new methods produced a revolution in astronomical instrumentation comparable to that of the telescope itself. Until that time, the main goal of astronomers had been to determine the positions, distances, and motions of the stars. Most of the telescopes in use were visual refractors, which were limited to apertures of less than 1 m. To obtain spectrograms of stars and galaxies, achromatic telescopes with large light-gathering power are required.

The foundations of astrophysics were laid in the 1860s by William Huggins, in his pioneering visual comparison of the spectra of the Sun and stars with those of earthly elements. In 1864, he discovered the composition of the "unresolved" nebulae that had puzzled astronomers for so long: they showed not a spectrum, but a single bright line. "The riddle of the nebulae was solved." he wrote, "The answer, which had come to us in the light itself, read: Not an aggregation of stars, but a luminous gas."

The rapid development of astrophysics around the turn of the century was largely due to the efforts of George Ellery Hale, who shared Huggins' enthusiasm about spectroscopy. Hale's lifetime goal was to establish the link between the Sun and the stars, which he

hoped could lead to the discovery of the underlying structure of the universe. He not only developed much of the instrumentation required for this task, including the spectroheliograph — which produces a photograph of the Sun's disk at a single wavelength of light, but he also succeeded in raising the funds for the largest telescopes built in the first half of the twentieth century. To extend the spectral study to stars and to reach further into the universe, it was necessary to collect more light and to concentrate it into the smallest possible spot on a photographic plate or spectrographic slit. The two essential requirements were a large aperture and good seeing.

The chosen site was Mount Wilson, at an altitude of 5886 feet, near Pasadena, California. In 1903, Hale obtained a grant from the Carnegie Institution to build the 60-inch reflecting telescope. The first observations were made in 1908. An early result was the finding that the spectrum of Betelgeuse closely resembled those of sunspots, confirming its (low) temperature. Photographic tests produced perfectly round star images of about 1-arc-second diameter after an exposure of 11 h. Stars of twentieth magnitude were recorded with 4-h exposures. Soon after, work started on the 100-inch reflector, which was put into regular use in 1918. This instrument brought enormous numbers of faint stars, star clusters, and nebulae within photometric and spectrographic range, providing a mass of information on the temperatures, composition, motions, intrinsic brightness, distribution, and distances of stars. The development of the Mount Wilson Observatory marks the beginning of the modern age of observational astronomy. Since that time, astronomical seeing conditions have assumed prime importance in the choice of sites for astronomical observations.

In 1914, V. M. Slipher, using the 24-inch refractor at Lowell Observatory, discovered that the M31 galaxy in Andromeda is receding from our own. By 1925, he had determined the radial velocities of 41 galaxies, establishing that the galaxies are moving apart. In the following years, using the 100-inch Mount Wilson telescope, Hubble determined the relation between the distance and velocity of nebulae, establishing the scale and the expansion of the universe. For over 40 years, the 100-inch telescope was in the forefront of astronomical research, its productivity aided by the excellent seeing conditions at Mount Wilson. It fell victim to light pollution from Los Angeles in the 1960s, limiting its use for deep-sky research, which is now carried out at more remote sites in the Chilean Andes and at Mauna Kea, Hawaii. The 100-inch telescope has recently been reconditioned and fitted with adaptive optics, giving this veteran instrument a new lease on life.

This brief review of the prehistory of adaptive optics has tried to show how greatly our ability to fathom the nature of the physical world depends on having suitable instruments. Centuries of painstaking measurements made with the unaided eye could not equal a few minutes' observation with a small telescope in resolving the nature of the solar system. Photography revealed faint nebulae and other distant objects that were invisible to the human eye, while spectrography enabled the measurement of their composition and radial velocities, leading to the discovery of the expanding universe. New technology is the key to new discoveries. But, by 1980, astronomical telescope design had reached a plateau.

Conventional methods of building astronomical telescopes (using passive components) reached the limit of their capability with the 5-m Mount Palomar instrument, dedicated in 1948. In recent years, the sensitivity of astronomical instruments has been increased 100-fold by the replacement of photographic film with photoelectric detectors, but even these are now near their limit, with efficiencies of over 90%. What is the next step in the design of astronomical telescopes?

Observations made by space telescopes at ultraviolet (UV) and x-ray wavelengths outside the atmospheric windows have produced new discoveries enlarging our knowledge of the universe. But, most astronomy is carried out with ground-based telescopes, and two improvements are necessary to maintain the pace of discovery: larger primary mirrors to collect more light, and removal of the effects of atmospheric turbulence that restrict the angular resolution. Both of these advances are now possible. Lightweight primary mirrors in which the surface figure is maintained by active control can be made with diameters of over 8 m. Atmospheric turbulence can be compensated by adaptive optics. These developments should set the stage for another golden age of astronomy.

1.5 The Importance of Optical Wavefront Measurements

An essential step in upgrading the performance of astronomical instruments is the ability to measure the optical quality of the components, especially the telescope objective, which is invariably the largest and most difficult component to manufacture. For the first 250 years of their existence, the quality of astronomical telescopes was assessed by their visual performance: for example, the ability to split double stars. In reality, this is as much a test of the atmospheric seeing and the observer's eyesight as the quality of the telescope itself.

Herschel studied the resolving power of the eye and of telescopes and was aware that there was a lower limit to the angular size of an object beyond which it could not be resolved. Objects smaller than this limit appeared as a spurious disk, the size of which varied inversely with the diameter of the telescope aperture. In 1831, the exact relationship was established mathematically by George B. Airy, who found that the diffraction of light spreads the image

of a star formed by a circular aperture into a bright disk surrounded by a series of faint rings of rapidly diminishing brightness. For a perfect optical system, the angular radius of the Airy disk, as it is known, is $1.22\,X\backslash D$ where λ is the wavelength of the light and D is the diameter of the objective. The effect of small random distortions (up to about 1/4 wave rms) is to reduce the peak intensity of the image, without significantly enlarging the disk. At that time, it was not known how to measure such small optical errors.

William Dawes made a practical study of the visual resolving power of telescopes and, in 1867, he published a table showing the apertures required to resolve double stars of known separation. Tests using double stars do not indicate the location of the errors on the surface of a mirror or lens. Without this information, the final polishing of objectives is a hit-or-miss affair.

A crucial step in optical measurement technology was Foucault's invention of the knife-edge test in 1859. This technique not only reveals the location and magnitude of imperfections in the figure of a primary mirror, but it also allows random wavefront variations, such as those due to atmospheric turbulence, to be visualized. As will be shown, it played a role in the evolution of adaptive optics a century later. From this perspective, it can be seen that the concept of compensating atmospheric turbulence did not suddenly appear, but was the inevitable result of experiments and technical improvements made over a long period of time.

The key to producing large reflecting mirrors is the ability to test the optical accuracy of their surfaces during the final polishing process. In the laboratory, the knife-edge test is made at the center of curvature of concave mirrors, using a small spot of light as the source. A knife-edge located at the reflected image converts small imperfections in the figure of the mirror into visible variations in brightness. In an operating telescope, the same test can be performed by pointing the telescope at a bright star and placing the knife-edge at the focus. In this case, both the figure errors and the continually changing wavefront distortion due to the atmosphere can be seen, giving a good appreciation of the effects of turbulence in relation to the optical quality of the telescope.

The two sources of wavefront error can be separated by photography. Using short exposures of 1/20 second or less, photographs of the primary mirror made with a knife-edge show the combination of the figure errors and atmospheric turbulence patterns. With long exposures of tens of seconds, the rapidly changing turbulence errors average out, revealing only the imperfections in the figure of the mirror. Photographs of the primary mirror of the 200-inch telescope of the Palomar Observatory, made in this way by Ira S. Bowen in 1949, are shown in figure 1.5.

Another method of testing large mirrors is the Hartmann test, first described in 1900, in which a mask perforated by precisely located holes is placed near the focus of the mirror. The light spots are recorded photographically and their positions are measured. The mirror aberrations can then be computed from the displacements of the spots from their expected positions. The Foucault and Hartmann techniques, originally used for testing large mirrors, are the direct forerunners of the wavefront sensing methods now used in adaptive optics. The shearing interferometer uses optical gratings rather than a knife-edge, but can be regarded as a modified Foucault test. The Shack-Hartmann sensor is based on the Hartmann test, modified to obtain high optical efficiency.

1.6 Early Ideas on Wavefront Compensation

The preceding sketch of the development of astronomical telescopes up to the middle of the twentieth century has attempted to provide the background for the evolution of adaptive optics. In 1953, Horace W. Babcock suggested a possible method of compensating atmospheric turbulence [Babcock 1953]. For various reasons, this idea was not put into practice at the time. Indeed, it was almost another 20 years before real-time turbulence compensation was actually achieved.

The forerunners of adaptive optics technology were the automatic guiders used to maintain the image of a star on a spectrographic slit or photographic plate during long exposures. These devices were first used in 1936, when sensitive photodetector tubes using electron multiplication became available. Autoguiders compensate random changes in the position of an image, whether such changes are due to the atmosphere or to mechanical strain in the telescope. At that time, there does not appear to have been any thought of extending this idea to the correction of defocus or higher order wavefront disturbances. Even in those early days, the most important component in a wavefront sensing system was the photodetector. This is just as true in today's high-performance wavefront sensors as it was with autoguiders 60 years ago.

Babcock [1990, 1992] relates that his interest in the problem of atmospheric seeing was stimulated by regular use of the Foucault (knife-edge) technique for precise focusing of reflecting telescopes. When using a bright star, this test conveys the reality of the turbulent elements in the atmosphere: their size, rate of change, and motion due to wind (see figure 1.5). In this way, an observer becomes familiar with the source and character of imperfect seeing. In the late 1940s, while working on high-dispersion spectroscopy at the Mount Wilson and Palomar observatories of the Carnegie Institution of Washington, Babcock built an automatic guider using a rotating knife-edge to guide the telescope during long expo-

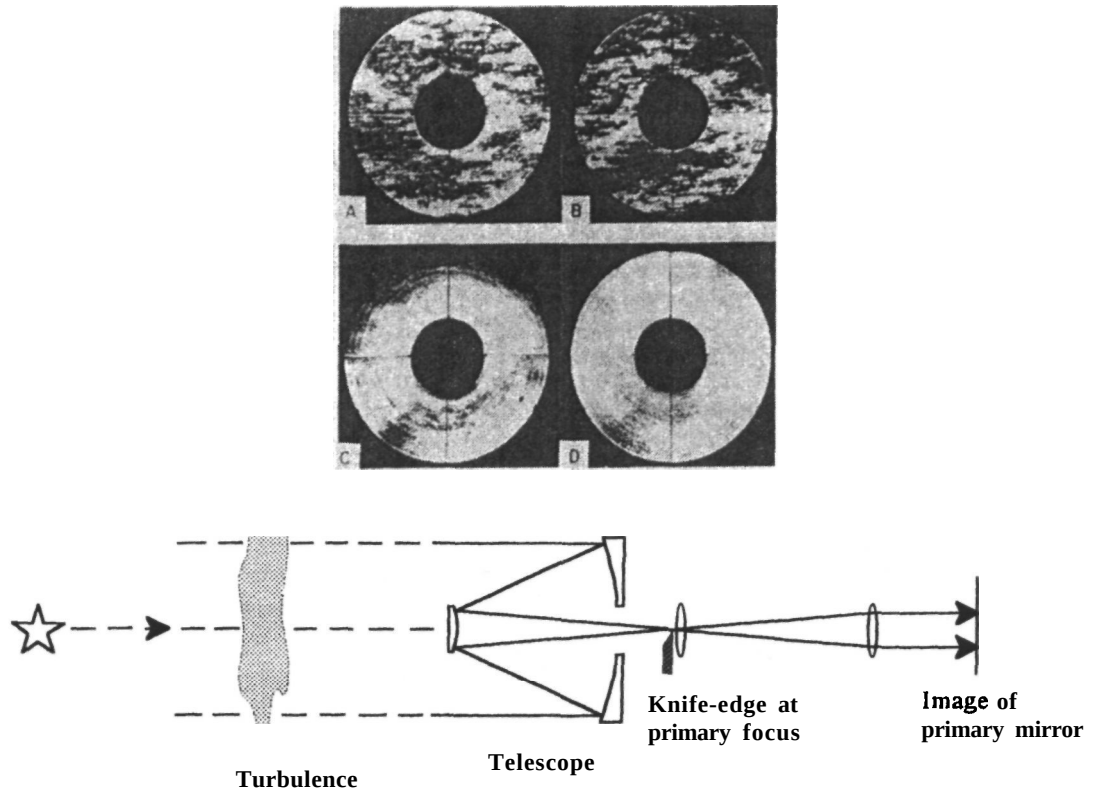


Figure 1.5 Knife-edge photographs of the primary mirror of the 200-inch telescope at Palomar Observatory [Bowen 1950]. In (A) and (B), short exposures show random wavefront distortion due to atmospheric turbulence. In (C) and (D), the turbulence is averaged out using 80-s exposures to show two stages in the figuring of the mirror surface. The principle of the knife-edge test is shown in (E).

sures. Although this operated at a low speed, making a correction every 3 s, it guided **the** 100-inch telescope well and eliminated a tedious job [Babcock 1948]. A few years **later, the** benefits of eliminating atmospheric turbulence were demonstrated by **a** night of exceptional seeing at the 100-inch telescope at Mount Wilson, when, for several hours, the stellar seeing disk was no larger than 1/4 arc second and essentially motionless. At such times, which are extremely rare, the coherence length of the turbulence may attain the extraordinary value of 1 m, the changes occurring slowly, on a time scale of several seconds. This experience was a natural demonstration of the gains to be realised if ordinary seeing could be compensated.

In 1952, Babcock saw an article on the Eidophor, an electronic device developed by Fischer in Europe for projection of television pictures in theaters, and realized that it could be used for correcting optical wavefronts distorted by atmospheric turbulence. He analyzed a possible design using a rotating knife-edge similar to that employed in the automatic guider, determining the size of the subapertures and the frame frequency for scanning. This led to an estimate of the magnitude of the reference star required. The size of the isoplanatic area was also determined.

In October 1953, Babcock published a paper, "The possibility of compensating astronomical seeing," which described a system combining a seeing sensor and wavefront corrector — the first account of what is now known as adaptive optics. The system that he proposed is shown in figure 1.6. Two active elements are employed: a fast guider using a tiltable parallel plate for correcting image motion and an Eidophor device for correction of wavefront distortion. The Eidophor employs a mirror covered with a thin film of oil in a vacuum enclosure, in which it is scanned by an electron beam. The beam is modulated to control the electric charge deposited on the oil film, which induces local slope changes that modify the wavefront reflected from the mirror. The field lens reimages the primary mirror of the telescope on the Eidophor. The control signals for the electron beam are obtained from an image tube located at the

far side of a rotating knife-edge centered on the image of the star. The knife-edge converts wavefront slope changes perpendicular to its edge into intensity variations that are detected by an image tube and processed with reference to the rotation angle. Signals correlated with the sine and cosine functions of the rotation frequency, corresponding to overall image displacement, are fed to the tip-tilt corrector, which keeps the reference star accurately centered on the rotating knife-edge. Higher frequency signals, corresponding to subaperture displacements, are used to modulate the electron beam, which controls the thickness of the oil film on the Eidophor, and produces the required correction to the wavefront.

Babcock's description of a possible means for compensating atmospheric turbulence was the first recognition that astronomers no longer had to tolerate atmospheric seeing; here was a possible way to correct it. He was also well aware of the two main limitations of seeing compensation: the need for a bright reference star and the small isoplanatic angle limiting the field of view — problems that are still not fully solved. The idea was not pursued because, at the time, the most important program for the Mount Wilson and Palomar observatories was the extension of Hubble's velocity-distance relation for remote galaxies, for which distance estimation depended on precise measurement of magnitudes. The most pressing need in instrumentation at that time was to improve photometric accuracy by developing photoelectric image tubes, a program that took about 10 years to complete.

During the 1950s, the efforts of the astronomical community to improve seeing were restricted to correcting the most basic problem, image motion. The automatic guider developed in 1947 corrected the average pointing of the telescope every 3 s. In 1953, a much faster automatic guider was developed, also employing a rotating knife-edge, but with the tip-tilt correction implemented by a gimbal-mounted quartz plate driven by two low-inertia motors and inserted in the optical beam ahead of the image plane [Babcock et al. 1956]. The knife-edge consisted of a 3-mm steel ball rotating within a nonmagnetic race whose radius was equal to the diameter of the ball. The ball was driven by a magnetic field to rotate at 60 revolutions per second. A working bandwidth of 10 Hz was achieved. This was the first "fast guiding" system in which the correction was made by a small tip-tilt corrector and not by moving the whole telescope. It was used successfully for spectroscopy at the coude focus of the 200-inch Hale telescope.

Other methods of stabilization were also tested. An electromagnetically operated tip-tilt mirror was used by Leighton [1956] for planetary photography on the 60-inch Mount Wilson telescope. A different approach to image stabilization, using an electronic image tube in which control signals were applied to the magnetic deflection coils, was described by DeWitt et al. [1957]. This " image tranquilizer, " as it was called, successfully stabilized television pictures of Jupiter obtained with a 24-inch reflecting telescope.

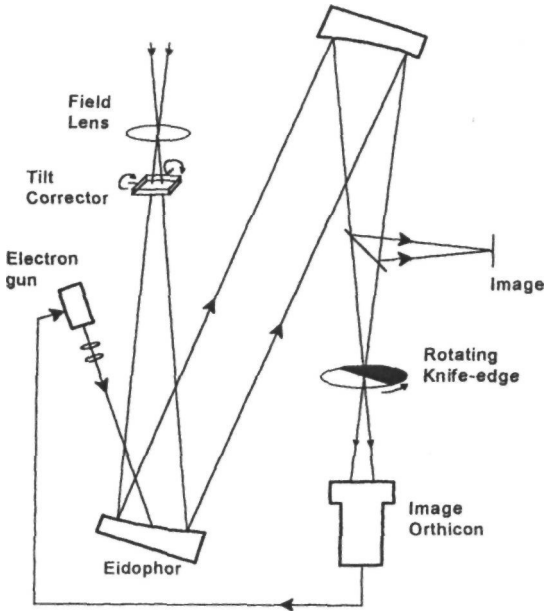


Figure 1.6 Seeing compensator proposed by Babcock [1953]. The wavefront received from a star is analyzed by a rotating knife-edge and an image orthicon detector tube. Correction signals are fed back in real time to change the optical thickness of an electronically deformed oil film on the Eidophor mirror, thereby compensating the wavefront disturbances.

The idea of using a segmented correction mirror to compensate atmospheric turbulence was first proposed by V.P. Linnik of the Astronomical Observatory of the [then] U.S.S.R. Academy of Sciences in 1957 [Linnik 1993]. Linnik's idea is shown in figure 1.7. Piston motion of the segments would be electrically controlled using a white-light interferometer to measure the phase errors within 20-cm^2 segments of the aperture. No details were given of how this system would be implemented. Linnik concludes his paper by proposing the use of an artificial light beacon mounted on an airplane at 8- to 10-km altitude, its position being controlled by the observer to coincide with the observed object. This is almost certainly the first reference to the use of artificial guide stars, a technique that did not become a reality until the 1980s.

The possible use of an electrostatically deformed membrane mirror in a feedback system for wavefront compensation was suggested by Babcock [1958]. The back of the mirror was covered with a mosaic of target elements on which charges could be deposited by an electron gun. This device is a forerunner of

modern membrane mirrors that use an array of electrodes to deform the mirror surface with electrostatic forces.

The invention of the laser in 1960 provided a new arena for optical science, and for the next decade the technical development of adaptive optics was applied to laser systems. This effort is outlined in the next section.

1.7 The Development of Laser Beam Control Systems

The first adaptive optics systems capable of compensating atmospheric turbulence in real time were developed in the late 1960s for laser beam control. Laser systems normally operate at relatively high power levels with coherent radiation, thus removing two of the major limitations that apply to astronomical adaptive optics: namely, the lack of photons from natural star reference sources and the need for operation over a wide spectral band. The use of coherent radiation allows wavefront compensation to be made on a modulo 2π basis, so the maximum phase correction is $\pm 1/2$ wave, independent of the actual path length error (provided it is within the coherence length of the source). Not only does this reduce the excursion required in deformable mirrors, but it also allows the use of electro-optical devices that generally have a limited phase-shifting capability.

It is interesting to compare the development of laser adaptive optics with that of adaptive antennas in microwave radar. The adaptive antenna concept for microwave radiation, developed in the 1950s, is a forerunner of phase-conjugate adaptive optical systems, using the same system concept, but with a difference factor of about 10,000 in the electromagnetic wavelengths. The principle of beam tagging, which originated in adaptive antenna systems, has been applied to adaptive optics systems where it is known as "multidither" or "image-sharpening." Because of the need to separate the tagging signals, either in time or in frequency, this technique is limited to relatively small optical apertures.

The basic problem in laser propagation is being able to maximize the intensity of a laser beam propagating through the turbulent atmosphere. Such systems are generally referred to as coherent optical adaptive techniques (COAT). The propagation of laser beams through the atmosphere originally had little in common with astronomy, but in recent years the development of laser-generated beacons has made this a subject of interest to astronomers. There are two main approaches: phase conjugation and multidither. The elements of a phase-conjugation system [Cathey et al. 1970, Hayes et al. 1977] are shown in figure 1.8. Phase conjugation employs the principle of optical reciprocity, in which the phase shift in the optical return path between a specular reflection (glint) in the target and the transmitter is measured in real time at multiple locations in the transmitting aperture. A compensating phase shift is then inserted at the corresponding locations in the aperture to make the transmitted laser wavefront match the wavefront received from the target glint. This process maximizes the power density in the vicinity of the target glint. Because the phase measurements can be made in parallel, there is no intrinsic limit on the number of elements that can be compensated.

Multidither systems employ the principle of measuring the intensity of the radiation returned from a target glint while making trial phase perturbations of the transmitting aperture, each section of which is

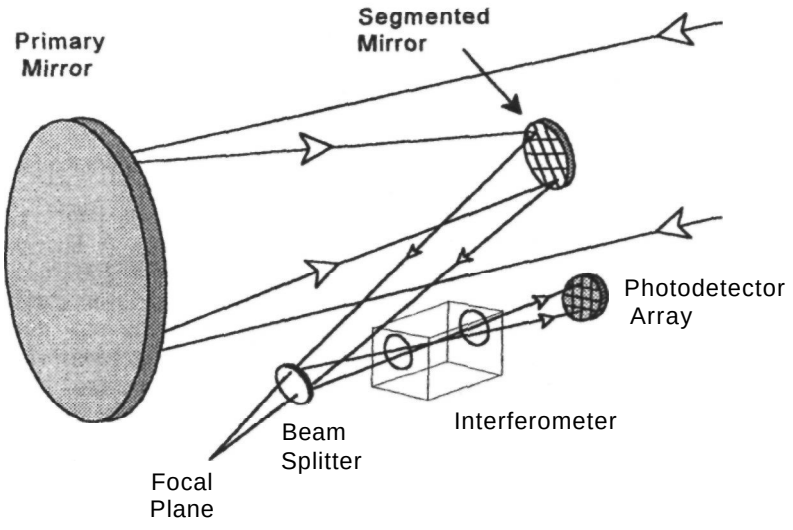


Figure 1.7 Seeing compensator proposed by Linnik in 1957 [Linnik 1993], using a segmented mirror for wavefront correction. A white-light interferometer measures the wavefront error within each segment of the mirror. The errors are compensated by piston motion of the segments.

tagged with an identifiable signature. In this case, the atmospheric path is effectively measured by the outgoing beam; the glint intensity may be monitored by a separate detector [O'Meara, 1977]. A multidither COAT system of this type, shown in Figure 1.9, employed an 18-element phase corrector [Pearson et al. 1976]. Each controlled element was identified by modulating or "dithering" the phase by about ±30° at a different temporal frequency within the range 8-32 kHz. To ensure correct operation of the servo system, the lowest modulation frequency must be at least six times the control loop bandwidth, and the modulation frequencies must be spaced by at least twice the control loop bandwidth. These constraints typically limit the number of correction elements to less than 100, so that, in practice, this type of system is limited to relatively small apertures. Both of these systems were successful for compensating laser beams through turbulent paths, but they could not be directly applied to the imaging of incoherently radiating, extended objects.

1.8 First Successes with Image Compensation

In 1972, the Advanced Research Projects Agency (ARPA), whose mission was to develop new technology for the U.S. Department of Defense, was wrestling with the problem of identifying newly launched Soviet satellites, preferably by obtaining high-resolution photographs of them on their first or second orbit. At that time, the only method of space object

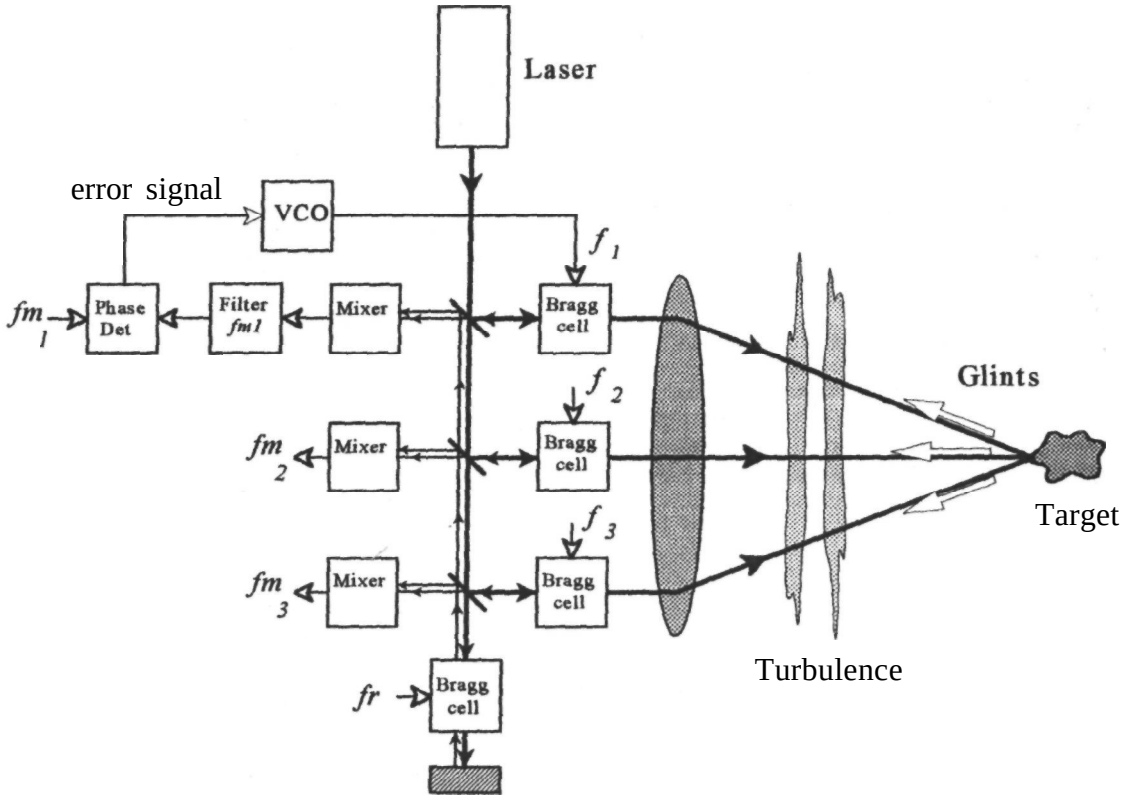


Figure 1.8 Phase-conjugate Coherent Optical Adaptive Technique (COAT) for compensation of an outgoing IR laser beam (1970). The optical aperture is divided into zones, each containing a Bragg cell that shifts the phase and frequency of the radiation passing through it. In each channel, the radiation returning from a glint on the target (modulated at frequency *2fn)* is mixed in a nonlinear photodetector with a reference beam from the laser, modulated at frequency 2/r. The electrical difference signal at $2(fn - fr)$ is selected by a filter and compared in phase with a fixed modulation signal *fm*. If turbulence in the atmospheric path disturbs the phase of the radiation returning from the target glint, an error signal appears at the output of the phase detector. The error signal shifts the frequency generated by the voltage-controlled oscillator VCR, thereby closing the loop by producing a phase shift in the Bragg cell that compensates the atmospheric phase. A system of this type operated successfully over a 7.9 km horizontal path at a wavelength of 10.6 μm.

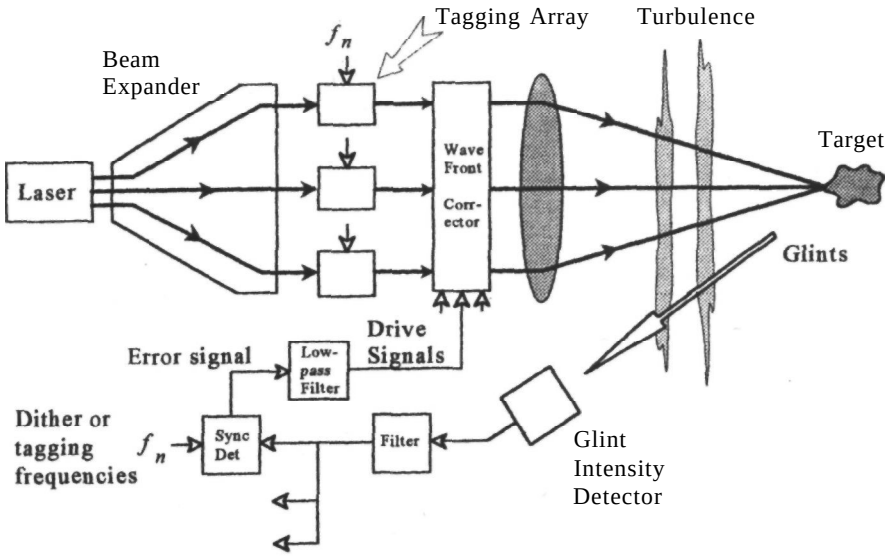


Figure 1.9 Multidither Coherent Optical Adaptive Technique (COAT) for compensation of an outgoing IR laser beam (1976). The aperture is divided into zones, each of which is 'dithered' or tagged in phase at a different frequency. Optical phase errors are not measured directly, but are inferred from the intensity of the target glints collected by a single photodetector. The intensity signals from each channel are separated by synchronous detection. If the glint intensity varies in phase with the dither signal, this indicates that a positive phase correction is required, whereas an antiphase variation indicates a negative phase correction. Wavefront corrections are made by applying control signals of the appropriate polarity to the actuators in each channel. In some systems, the tagging and correction functions are performed by a single deformable mirror.

surveillance from the ground was to obtain short-exposure (1/100 s or less) images of the space objects using satellite-tracking telescopes of 1.2- to 1.6-m aperture, and then to process the somewhat fuzzy images digitally to bring up the desired detail. This "postdetection processing" technique did not give useful results, mainly because the desired information was obscured by noise on the films and videotapes. The signal-to-noise ratio was so low that no amount of computer processing could restore the image. The main reason for the low signal-to-noise ratio was the image degradation produced by atmospheric turbulence.

At Itek Optical Systems in Lexington, Massachusetts, Program Development Manager J. Richard Vyce became aware of this problem and, in June 1972, I met with him to discuss possible solutions. For several years, Itek and other companies had been working on the problem of detecting satellites against a background of fixed stars, using methods for discriminating their motion relative to stars, over large areas of the sky. For that task, satellites can be considered as single points of light and atmospheric turbulence is relatively unimportant. However, to obtain sharp images of these tiny objects only a few arc seconds in extent, and to reveal some details of their configuration, it was essential to overcome the blurring due to the atmosphere. The telescopes themselves were capable of resolving angular detail of 1/10 arc second, but their resolving power was degraded by turbulence to between 1 and 2 arc seconds. A new approach was called for.

It was decided that the best solution was to compensate the atmospheric distortion *before* the image was recorded. This process would enable high-resolution information to be obtained even with relatively long exposures. Implementation of a system of this kind presented formidable problems. Babcock's 1953 and 1958 proposals for compensating atmospheric seeing had never been put into practice and it was evident that the compensation of satellite images would be far more difficult than compensating stars, because of their rapid motion across the sky.

As space objects traveling in low Earth orbits are tracked by a telescope, their high angular velocity causes the atmospheric turbulence structure to move across the beam at a high speed, resulting in the need for a compensation bandwidth of about 1 kHz. To achieve this speed, the wavefront would have to be measured at a rate about 10 times faster.

In addition, for space surveillance, the adaptive optics would have to work under all seeing conditions; it is not possible to wait for the best conditions, as with scientific observations. At the AMOS (formerly ARPA Maui Optical Site, now Air Force Maui Optical Site) Observatory in Maui, Hawaii, the seeing is usually between 1 and 2 arc seconds, requiring a subaperture size of about 10 cm. For real-time compensation of satellite images at visible wavelengths, a telescope of 1.6 m aperture would therefore require about 200 subapertures, sampled at a rate of 10 kHz. These requirements were far beyond the capability of any wavefront sensor or corrector known in 1972. Another critical problem was how to convert the measured wavefront slopes into suitable drive signals for a deformable mirror. The reconstruction process had been implemented for individual wavefronts [Rimmer 1974], but the data processing rate required for real-time compensation in a telescope of reasonable size far exceeded the capability of any serial processor then available.

Three key components were required: a wavefront corrector and a wavefront sensor, each with about 200 subapertures, coupled together by a fast reconstructor. At that time, the state of the art in deformable mirrors was represented by piston-type segmented mirrors designed for IR laser beam correction at 10.6 mu m wavelength. For real-time compensation at visible wavelengths of about 0.5 mu m, the precision of deformable mirrors would have to be improved by a factor of 20, the number of actuators would have to be multiplied by 10, and they would have to respond 100 times faster to input commands.

Real-time wavefront sensors of the required precision did not exist in 1972. The standard method of wavefront measurement was to photograph a laser interference pattern and scan it with a manually operated coordinate densitometer. The data were then punched onto cards that were fed into a mainframe computer to reduce the data and produce a wavefront map. With luck, the wavefront information was available in about 24 h. For real-time adaptive optics, it was needed in less than 1/1000 of a second. Furthermore, it was necessary to make the atmospheric wavefront measurements using the broadband sunlight reflected from the satellites, rather than the monochromatic light employed in conventional interferometry.

Fortunately, some experiments were in progress at Itek on a new method of wavefront measurement using a shearing interferometer. In 1972, James Wyant invented a device that used a moving grating to convert wavefront gradients to electrical phase shifts that could be rapidly measured. Using a reciprocating drive, I devised a modification of this shearing interferometer that would measure the complete wavefront over an optical aperture 20 times per second. This wavefront sensor was improved by Chris Koliopoulos using a rotating radial grating, which greatly increased the sensor's measurement rate and allowed operation with broadband light. Later versions were capable of measuring the wavefront over a full aperture at a rate of 10,000 frames per second, each measurement frame consisting of several hundred wavefront gradients, thereby fulfilling all the requirements for real-time compensation.

Itek had also developed the Pockels Readout Optical Memory (PROM), which was an electro-optic crystal configured to store optical images as an electrical charge pattern [Aldrich et al. 1971]. The bismuth silicon oxide crystal stored the data as a pattern of optical phase shifts, and therefore was capable of correcting phase errors in an optical wavefront, although its correction range was very limited. In spite of its many shortcomings, the PROM crystal was pressed into service as a wavefront compensator.

The final link needed for the real-time atmospheric compensator was a rapid means of reconstructing the individual wavefront slope measurements made by the shearing interferometer into a continuous wavefront map covering the optical aperture. The first idea was to use a digital computer, but the high data rate would have overwhelmed most of the machines then available. The solution that I developed was an analog computer in which electric currents representing the wavefront slopes were added in a two-dimensional resistor network having the same configuration as the subapertures in the wavefront corrector. The voltages appearing at the nodes of the network represented the reconstructed wavefront values, which were then applied to the wavefront corrector. This wavefront reconstructor was very fast, with a settling time on the order of 1 μs. Because of its parallel structure, it could be expanded to cover large apertures without sacrificing speed.

I described these ideas in a technical proposal submitted to the Strategic Technology Office of ARPA in November 1972; this resulted in a contract to develop a 21-actuator feasibility model of the adaptive optics system for laboratory tests. It soon became evident that a better wavefront corrector than the PROM crystal was needed, preferably a deformable mirror with several hundred controllable zones. Within a short time, the research team of Julius Feinlieb, Steven Lipson, and Peter Cone developed the Monolithic Piezoelectric Mirror (MPM), a very successful device that was used extensively in adaptive optics systems for the next decade.

A schematic diagram of the real-time atmospheric compensator (RTAC) is shown in figure 1.10. There are fundamental differences between this instrument and Babcock's seeing compensator. The RTAC was a fully parallel system with multiple feedback loops, one for each wavefront compensation subaperture. It was the first adaptive optics system to use a wavefront reconstructor, which restores the absolute phase values that are lost when local wavefront gradients are measured. The reconstruction process converts the two-dimensional array of measured wavefront gradients into a corresponding array of

drive signals for the deformable mirror, taking into account both the spatial and temporal characteristics of the turbulence-degraded wavefront; it is the key to maximizing the performance of all adaptive optics systems. This architecture provides all the basic capabilities needed to implement adaptive optics for the largest telescopes, in the most efficient way.

The RTAC using this new approach was ready for laboratory tests in December 1973. It employed a 21-element MPM with a wavefront sensor making 32 simultaneous slope measurements. Assisting me in the design and testing of this system were electronics engineer Joseph Lefebvre and technician Steven Moody. Questions had been raised over the potential stability of so many interacting feedback loops. An error in the design or construction might have produced large oscillations that could have destroyed our monolithic deformable mirror, the only one of its type in existence. It was with some trepidation that the 21 feedback loops were closed for the first time. We were elated to find that the compensation system was perfectly stable. When optical distortions were introduced into the beam, the RTAC produced well-corrected images, some of which are shown in figure 1.11.

This was the first test of an adaptive optics system using a wavefront sensor, reconstructor, and deformable mirror, similar to those in standard use today. The tests were the culmination of more than a year of intense development at Itek Optical Systems, yielding new approaches to wavefront sensing, parallel data processing, and deformable mirrors. The ultimate application of this new technology was to obtain compensated images of extended sunlit objects (Earth satellites). One question remained to be answered: Could the extended object itself be used as a reference for the wavefront sensor, or did a separate point source have to be provided near the object? This question was answered by Wyant [1974], who found that when an optical diffraction grating is used to produce the sheared wavefronts, interference fringes are generated by an extended source such as a solar illuminated satellite. This important fact is still not widely appreciated; it is not necessary to use a monochromatic point source to produce interference fringes, as will be explained in chapter 5. Laboratory tests on the RTAC in the spring of 1974 proved that "self-referencing" will work on extended objects such as low-earth-orbit satellites. This technique enables fine detail to be observed when using the whole object as the reference source for the wavefront sensor.

The RTAC underwent field tests at Rome Air Development Center's (RADC) optical test range at Verona, New York in the summer of 1974, under the direction of Donald Hanson, the RADC program manager. The tests were made over a 300 m horizontal path near the ground, using a helium-neon laser as the light source. The collecting aperture was a 1 m telescope mounted in a vertical tower, with a coelostat for beam steering sitting at the top of the tower. During the day, the RTAC worked perfectly, producing a good image of the laser source through the turbulent path. At dusk, it worked even better, when the daytime turbulence had subsided. Then, suddenly, the image disintegrated as strong turbulence overwhelmed the correction system. After some initial consternation, it was realized what had happened: at dusk, with a clear sky, the air near the ground cooled, while the telescope tower remained warm, causing a strong updraft that created considerable turbulence in the optical path. The tower was subsequently modified with doors and optical windows to prevent its functioning as a chimney.

These field tests convinced ARPA and the Air Force RADC, who were administrating the Itek contract, that the principles employed in the RTAC were practical and could be scaled up for use on a much larger telescope.

The first public description of the RTAC was given in July 1974 at a meeting of the Optical Society of America in Boulder, Colorado [Hardy et al. 1974]. United States patents covering the analog data processor used for wavefront reconstruction and the basic configuration of the real-time wavefront correction system were issued in November and December 1975, respectively [Hardy 1975a, 1975b].

In spite of the success of the RTAC, the investigation of other methods for compensating turbulence-degraded images continued. Variations of the "multidither" system used for laser beam compensation were analyzed by a team from Hughes Research Laboratories [Miller et al. 1974]. The system known as image sharpening was investigated by Muller and Buffington [1974]. In this approach, wavefront errors are detected by their effects on the full-aperture image of the reference star, rather than by directly measuring wavefront slope in a number of r_0-sized zones as in the RTAC. Using an image sharpness criterion, which can be as simple as measuring the amount of light passing through a pinhole somewhat smaller than the diffraction-limited core of the image (the Airy disk), it can be shown that the criterion is maximized only when the input wavefront is error free. The image sharpness function does not, however, identify uniquely the type of wavefront error, so it is necessary to make trial perturbations in the aperture to determine the correction required. A simple system of this type using a linear array of six actuators was built and tested in 1976. Images of bright stars were successfully sharpened in one dimension. An image-sharpening system using a 19-actuator deformable mirror was tested on a 36-cm telescope by McCall et al. [1977], and it succeeded in reducing the seeing disk of Sirius from its uncorrected size of several arc seconds to less than 1 arc second.

As a technique for real-time turbulence compensation, image sharpening becomes increasingly diffi-

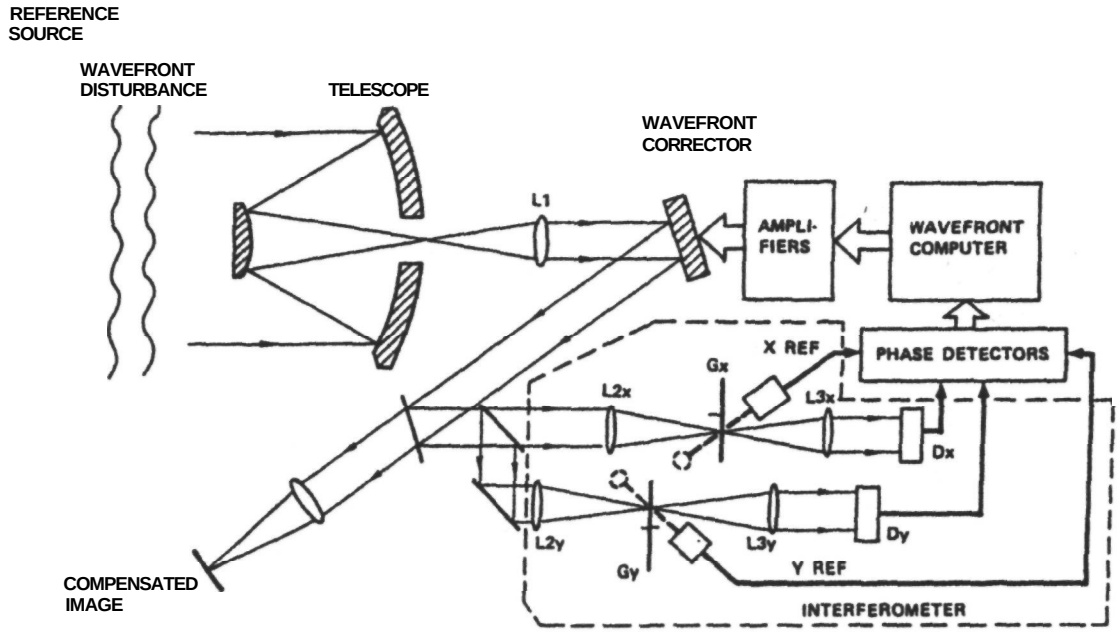


Figure 1.10 Real-Time Atmospheric Compensator (RTAC). This pioneering adaptive optics system used a 21-actuator deformable mirror, a white-light shearing interferometer, and an analog wavefront reconstructor. It first operated in the laboratory in December 1973.

cult to implement as the aperture size gets larger, because the number of trial perturbations that must be made in each operating cycle increases in proportion to the area of the aperture. Dyson [1975] compared the limiting magnitudes of RTAC-type adaptive optics systems (using parallel feedback paths) with those of image-sharpening systems (using the serial approach) and showed that parallel systems have a significant advantage for large apertures. The parallel approach, in which the wavefront is corrected simultaneously over the whole aperture, is now universally employed in adaptive optics systems.

1.9 The Evolution of Large Adaptive Optics Systems

The laboratory and field tests conducted on the Real Time Atmospheric Compensation System confirmed that the principles employed were sound

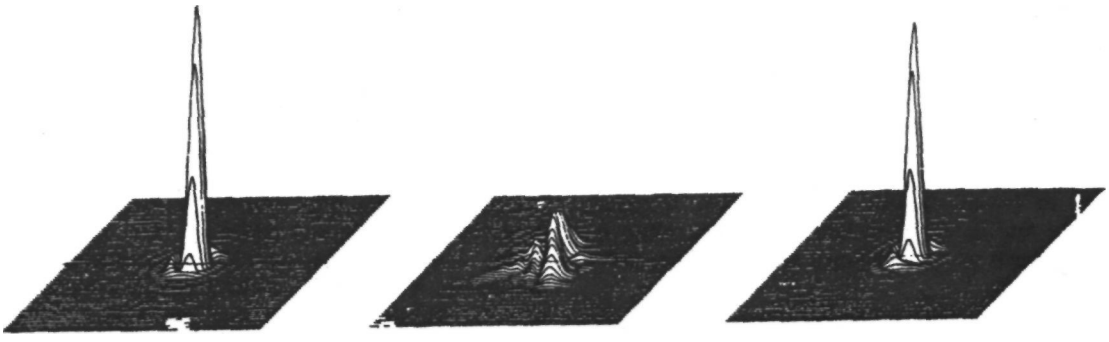


Figure 1.11 First compensated images obtained with RTAC (1973). These plots show the image intensity distribution from a point source for three cases:
(a) no added distortion, deformable mirror flat, Strehl ratio 0.94;
(b) 1.28 waves peak-to-peak distortion, no compensation, Strehl ratio reduced to 0.18;
(c) same distortion, RTAC operating, Strehl ratio restored to 0.85.

and could be scaled up to much larger telescopes. The prime candidate for adaptive optics was the 1.6-m satellite-tracking telescope at the AMOS Observatory on Mt Haleakala, then operated by ARPA. The function of this installation was to develop and evaluate new methods of obtaining data on orbiting space objects. Atmospheric turbulence imposed a limit on the angular resolution that could be obtained. Postdetection processing had not been successful due to the very low signal-to-noise ratios of the recorded data. Because low-Earth-orbit satellites move at a high angular rate, their aspect angle is continually changing, making it difficult to apply multiple-frame addition techniques to improve the image quality. Real-time compensation of atmospheric turbulence appeared to be the ideal solution.

Low-Earth-orbit satellites are, in many ways, ideal objects for real-time compensation: they are compact in angular size, usually a few arc seconds, so they are generally contained within a single isoplanatic region. They are also bright enough to act as the reference source for wavefront sensing, with many having a visual magnitude between 4 and 8. There were, of course, many difficult problems still to be solved, such as increasing the number of actuators by a factor of almost 10, to match the larger aperture, and improving the efficiency of the wavefront sensor so that individual photons could be counted. Mechanical packaging was also a challenge, as it was necessary to mount the entire optical system of the compensator on the telescope without restricting its tracking ability.

The Compensated Imaging System (CIS) successfully solved all of these problems (figure 1.12). First operating in 1982, it was a major technological breakthrough. It used a deformable mirror with 168 actuators, with a separate two-axis mirror for tilt correction. The wavefront sensor was a shearing interferometer using arrays of end-on photomultiplier tubes for photon detection. The control loop bandwidth was adjustable up to 1000 Hz. A photograph of the equipment mounted on the AMOS 1.6-m telescope is shown in figure 1.13. The adaptive optical components of the CIS were all contained in a rigid structure mounted at the Cassegrain focus.

The CIS was equipped with an operator interface having full diagnostic and control capability. During operation, the following parameters were continuously monitored and displayed at the CIS control console: turbulence strength r_0, wavefront sensor photon count, residual wavefront error, reconstructed wavefront values, and deformable mirror drive voltages. The residual error at the output of the wavefront sensor is the prime indicator of system performance. Controllable parameters, such as integration time and shear values, can be modified quickly by the operator to optimize performance as conditions change. Operating parameters may be recorded on magnetic tape for postmission analysis and diagnostics.

The Compensated Imaging System involved a considerable effort in mechanical, optical, electrical, and computer engineering. As technical director of the project at Itek Optical Systems, I was fortunate to have Richard A. Hutchin (formerly Hudgin) and Edward P. Wallner on the design team. Their analytical skills provided much-needed guidance in previously uncharted areas of technology, such as wavefront sensing using the light from faint, rapidly moving satellites, the reconstruction and prediction of random wavefronts, and methods of optimizing the performance of adaptive optics systems.

The Itek design team was supported by additional scientists with expertise in wave propagation problems, including DARPA consultants David L. Fried and Marvin King, Ben McGlamery of Scripps Oceanographic Institute, and Darryl Greenwood of Lincoln Laboratory, Massachusetts Institute of Technology (MIT). Much of the analysis generated during the design of the CIS and successive systems is now a part of the standard technical literature on adaptive optics. In the early 1980s, after development of the CIS, much of the effort in adaptive optics was directed to the use of laser beacons, a subject which is covered in chapter 7.

Some of the data obtained during tests on the CIS are shown in figures 1.14 and 1.15. The compensated image of a double star was obtained by Itek during acceptance tests in the summer of 1982. The image of the Hubble Space Telescope was made in 1992 by U.S. Air Force Phillips Laboratory, which now operates AMOS. For full compensation at visible wavelengths, the limiting magnitude was found to be about $m_v = 7$. Tilt correction (image stabilization) was obtained for dimmer objects. With double stars, it was found that some high-order wavefront compensation occurred, even when the separation was much larger than the expected isoplanatic angle, sometimes as much as 1 arc minute. This is believed to be due to a persistent layer of low-altitude turbulence at AMOS, compensation of which is effective over a large field of view. The isoplanatic angle due to high-altitude turbulence still limits the ultimate compensation possible when the guide star is separated from the object viewed by more than a few arc seconds. The images shown are taken from a videotape made with a television format ISIT (Intensified Silicon Intensified Target) camera operating at 30 frames per second. No postprocessing was employed. Average values of r_0 during these tests were between 5 and 8 cm, whereas the CIS was designed for an r_0 value of 10 cm. Analysis has shown that the performance of the CIS is limited mainly by the small stroke of the deformable mirror rather than by the number of actuators.

For many years, the CIS was the largest adaptive optics system in existence. It was still producing useful data in 1994, making it the grandfather of

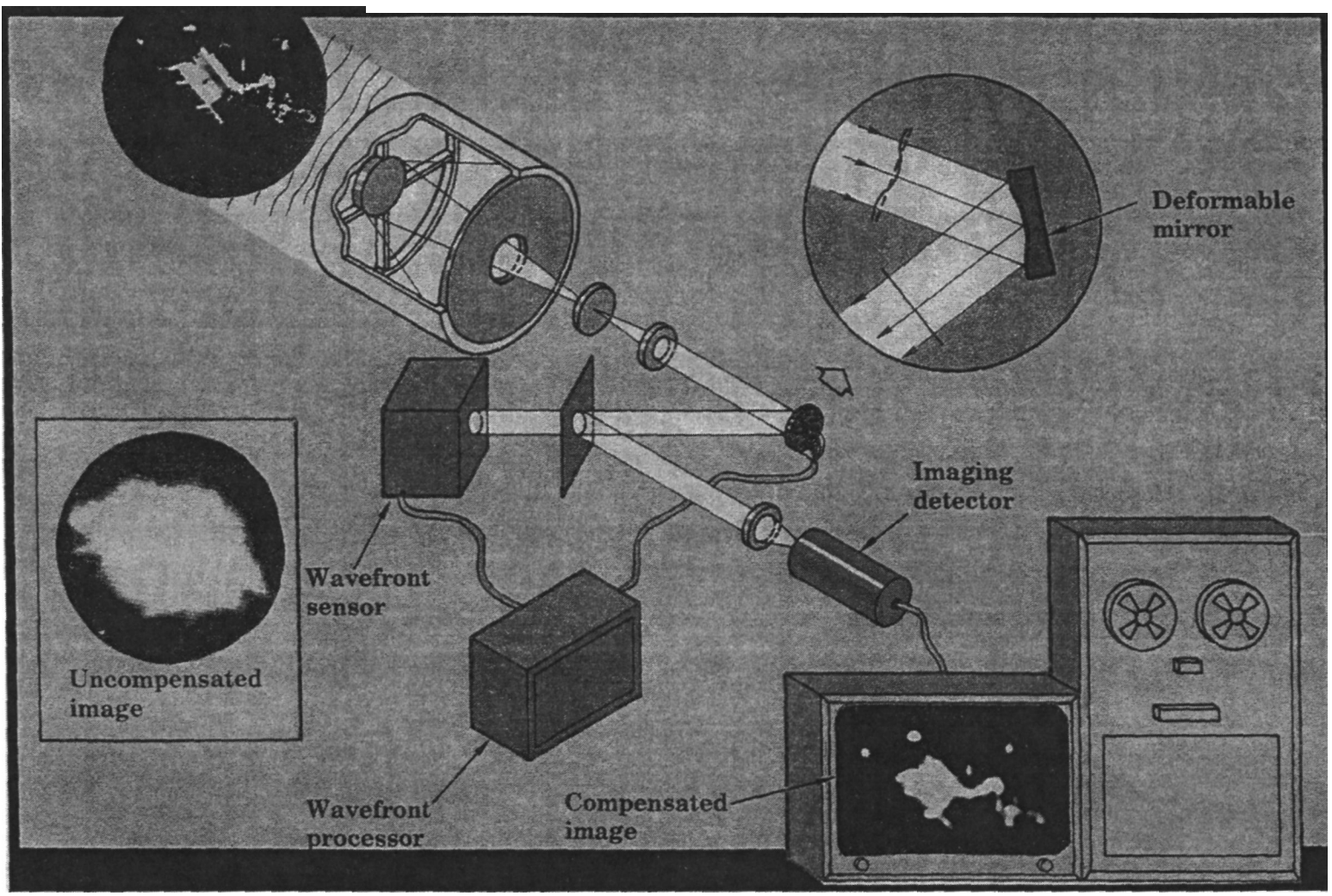


Figure 1.12 Schematic diagram of the Compensated Imaging System (CIS). The CIS was the first adaptive optics system capable of fully compensating a large astronomical telescope. It operated at visible wavelengths and was installed on the AMOS 1.6-m telescope in 1982.

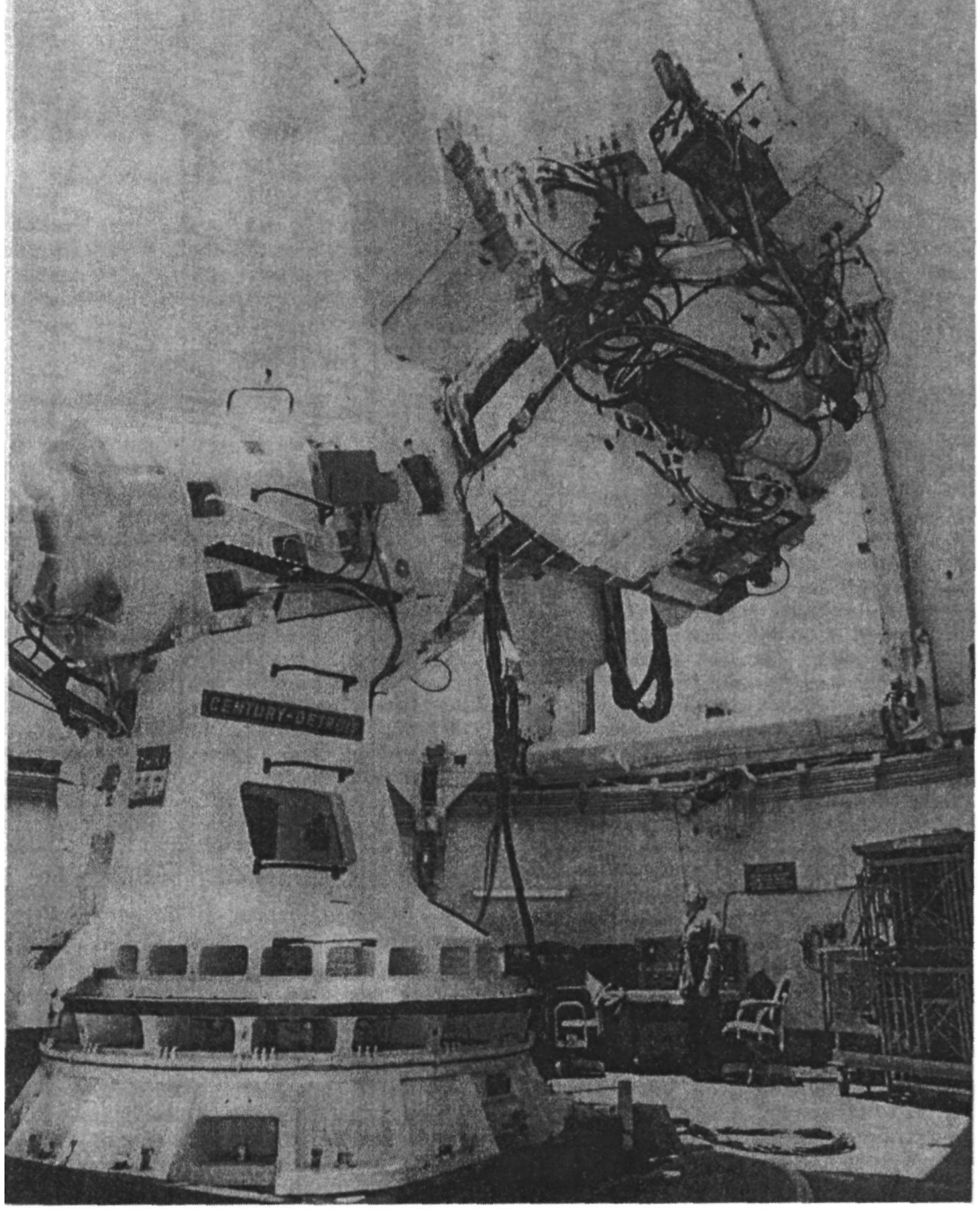
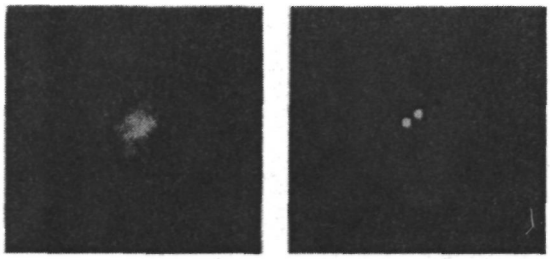
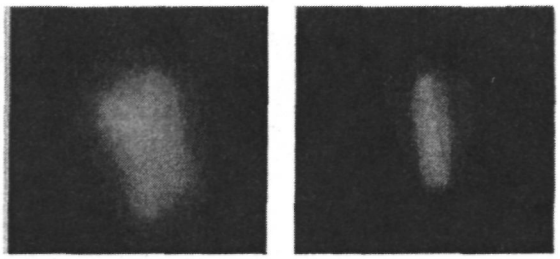


Figure 1.13 Compensated Imaging System mounted on the 1.6-m telescope at Air Force Maui Optical Site (AMOS), at Mt Haleakala, Maui, Hawaii.



Uncompensated Compensated

Figure 1.14 CIS Test data, 18 May 1982: Real-time compensation of a double star in the visible band. The separation of the components is 0.9 arc sec, and their visual magnitude is 5.2.



Uncompensated Compensated

Figure 1.15 CIS Test data, 24 July 1991: Real-time compensation of an Earth-orbiting satellite (Hubble Space Telescope). The spacecraft was at a slant range of 911 km, elevation angle 38 degrees. Its angular size is about 3 arc seconds.

adaptive optics systems. The adaptive optics technology developed for the CIS formed the basis for the Atmospheric Compensation Experiment (ACE), a program performed by the MIT Lincoln Laboratory between 1981 and 1985 [Greenwood and Primmerman, 1992]. The purpose of this work was to develop methods for compensating laser beams transmitted from the ground into space, through the turbulent atmosphere. The principle used is known as optical reciprocity: if the turbulence is measured in the downward direction using a beacon at the desired aim point of the laser and corrected by a deformable mirror near the ground, then the same corrections are valid for a laser beam propagating upward through the atmosphere to the aim point.

The ACE employed a 69-actuator MPM and a shearing interferometer wavefront sensor similar to those used on CIS. The wavefront reconstructor was implemented digitally, using an array of multiplier-accumulator processing elements. To visualize the corrections produced by the deformable mirror, Lincoln Laboratory built a "flying carpet" display, which provided a real-time image of the shape of the mirror faceplate with the excursions greatly magnified to make them visible. The ACE was installed on the 60-cm laser beam director at AMOS in 1982, and, during the next few years, laser beam compensation experiments were made to aircraft flying over the site, to the space shuttle *Discovery,* and finally to rockets at altitudes of about 600 km. These tests were the first to demonstrate atmospheric compensation of a beam propagating from the ground to space.

After 10 years as an adaptive optics workhorse for defense programs, the ACE was loaned to Mount Wilson Observatory to conduct astronomical observations under various meteorological conditions [Shelton and Baliunas 1993]. It was mounted at the coude focus of the 60-inch telescope and achieved extraordinary results [Shelton et al. 1993], considering the ripe old age of both the telescope (the mirror was cast in 1894) and the adaptive optics, which first operated in 1980. After making some minor improvements to the adaptive optics, the compensated 60-inch telescope was capable of producing long-exposure star images having a full-width at half maximum (FWHM) of about 0.15 arc seconds, for uncorrected seeing of 1.2 arc seconds or better. The improvement in Strehl ratio was about 10-fold. Some of the results obtained are shown in figure 1.16. The ACE employed photomultiplier detectors having relatively low quantum efficiency, giving a limiting magnitude for full compensation of about $m = 6$. An updated adaptive optics system has now been built for the Mount Wilson 100-inch telescope [Shelton et al. 1995] using very-low-noise charge-coupled device (CCD) detectors which improve the limiting magnitude to about $m = 14$.

1.10 The Next Generation of Adaptive Optics

The events leading up to the development of adaptive optics have been reviewed, from Cleomedes' discovery almost 2000 years ago that Earth's atmosphere could bend light rays, to Newton's appreciation of how atmospheric turbulence limits the resolving power of astronomical telescopes, to Foucault's method for visualizing optical phase and the effects of turbulence, and finally to the development of the first generation of adaptive optics systems for large telescopes. Although the limitations due to Earth's atmosphere became apparent soon after the invention of the telescope, 350 years passed before an astronomer realized for the first time, in 1953, that turbulence could be compensated directly by means of controlled deformation of optical components.

Although adaptive optics has a short history, recent progress has been rapid. In the 1980s, laser beacons were developed by the defense community,

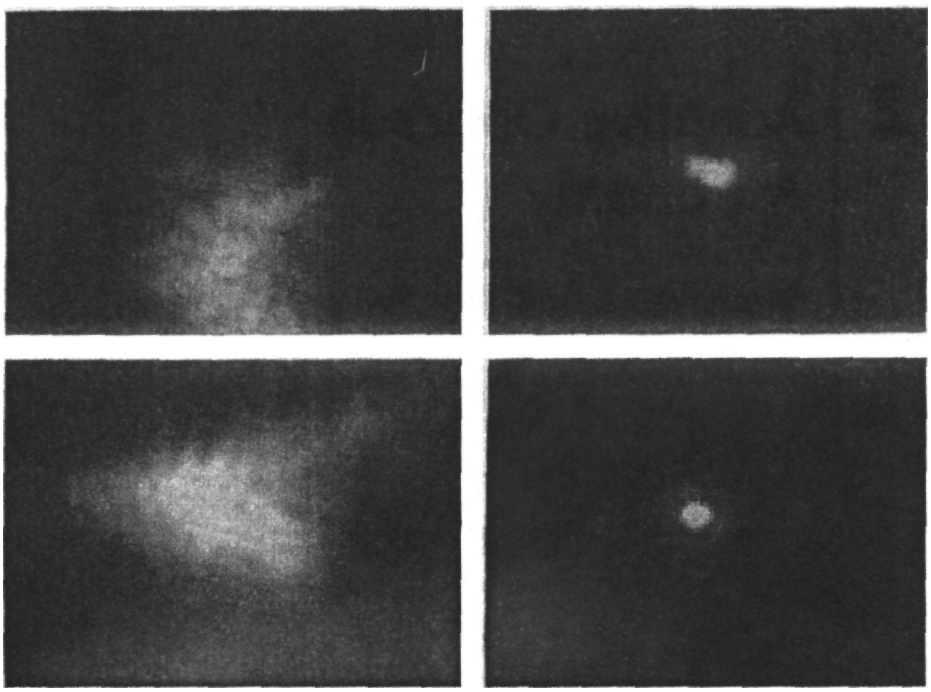


Figure 1.16 Compensation of natural stars using the ACE adaptive optics system with the 60-inch telescope at Mount Wilson Observatory (Shelton et al. 1993). Uncompensated and compensated images are shown of the stars Regulus (Alpha Leo) and Altair (Alpha Aquilae).

initially for the purpose of directing laser beams. This effort, carried out by the MIT Lincoln Laboratory and the U.S. Air Force Phillips Laboratory is summarized in chapter 7. This technology is now used to generate reference sources for observational astronomy.

Several scientific organizations started the development of astronomical adaptive optics in the mid-1980s, including the National Optical Astronomy Observatory (NOAO), the European Southern Observatory (ESO), and Office National d'Etudes et de Recherches Aerospatiales (ONERA) in France. In 1992, much of the adaptive optics development funded by the U.S. Government was made available to the scientific community and many observatories are now planning to incorporate adaptive optics into the latest generation of astronomical telescopes. These programs are summarized in chapter 10.

Adaptive optics is now recognized as an essential component of ground-based astronomical telescopes. In the following chapters, the theory and practice of adaptive optics are explored in greater depth.

2 Adaptive Optics in Astronomy

"A good idea is worth more than a large telescope."

R. O. Redman

2.1 Introduction

The purpose of this chapter is to provide an overall perspective on the application of adaptive optics to ground-based astronomy. It addresses questions such as: What is the function of adaptive optics? How does it improve astronomical observations? What effect does it have on the design of telescopes and instrumentation? How is adaptive optics implemented and what are its limitations? By providing concise answers to these questions, I hope to establish a framework for the detailed discussion of techniques and system design that is contained in the following chapters.

The function of adaptive optics in astronomy is to remove aberrations from the optical path between a celestial object and the imaging device. When this is achieved, the quality of the image is limited only by the size of the telescope aperture. For astronomical observations, the aberrations occur near the end of the optical path, in Earth's atmosphere, and in the telescope itself. If uncorrected, the image of a star is spread into a "seeing disk", which may be 10-100 times larger than the diffraction limit of the telescope itself, degrading the angular resolution and greatly reducing the peak intensity. Most of the image degradation is caused by random phase variations due to turbulence within the beam path. An adaptive optics system measures these phase disturbances and compensates them in real time, restoring the image quality to a useful fraction of the diffraction-limited capability of the telescope. For astronomical applications, the required peak intensity of the image is typically between 0.1 and 0.8 times its diffraction-limited value, depending on the scientific task. In certain cases, such as detecting dim companions of stars, a much higher degree of compensation is required to reduce scattered light from the parent star.

Adaptive optics is capable of improving the performance of most optical instruments used in astronomy, including spectrographs and interferometers as well as imaging devices. It is not limited to compensating atmospheric turbulence, although this is usually the most serious problem in ground-based observations. Adaptive control can be used to correct figure errors in the primary mirror as well as optical errors arising from any random disturbance, ranging from slow variations in temperature and the gravity vector, which may have a time scale of hours, up to the rapidly moving turbulence eddies transported by high-velocity winds, producing a bandwidth that approaches 1000 Hz.

The atmospheric disturbances that affect astronomical telescopes have characteristic frequencies above 1 Hz. Although, in principle, adaptive optics will work down to the lowest temporal frequencies, it has proved convenient, in practice, to compensate the lower frequency disturbances due to temperature and gravity by a separate control system known as "active optics." The primary mirror is potentially the largest source of low-frequency wavefront errors in astronomical telescopes; the main task of active optics is to control its optical figure, eliminating large, low-frequency phase errors. This allows the adaptive optics system to employ a small deformable

mirror to achieve the high temporal bandwidth necessary to compensate atmospheric turbulence. Another difference is that active optics systems usually employ a local reference (which may be optical or mechanical) to control the primary mirror, while adaptive optics systems must, of necessity, employ distant reference sources to measure the wavefront. Some active optics installations use a natural star to calibrate the control system.

The two important properties of an astronomical telescope are its collecting area and its angular resolution. A larger aperture D not only increases the amount of light collected, but also reduces the diameter of an image of an unresolved star so that ideally its peak intensity increases as D^4, while the angular resolution improves as λ/D (smaller is better). During the 400 years of the existence of astronomical telescopes, improving optical technology has enabled their apertures to be steadily increased. Although great light-gathering power has been achieved (the two Keck telescopes have 10-m apertures), the performance of large telescopes long ago reached the limit imposed by atmospheric turbulence. For apertures larger than the turbulence coherence length r_0, which is generally between 0.1 and 1 m depending on wavelength and atmospheric conditions, the angular resolution is limited to λ/r_0 and the peak image intensity increases only as the square of the telescope aperture. These restrictions have been removed in the last decade by the use of active primary mirrors and adaptive optics, opening the way for a huge increase in the performance of ground-based telescopes.

In principle, adaptive optics removes the atmospheric seeing limit, enabling ground-based telescopes to achieve their full performance. As will be shown, adaptive optics enables relatively small telescopes to equal or exceed the performance of much larger uncompensated instruments. This is not an argument for building smaller telescopes, but it suggests that adaptive optics has great potential in upgrading the performance of existing smaller instruments. Observatories throughout the world possess many telescopes with apertures between 1 and 3 m that could be upgraded by the addition of adaptive optics. Several "old" telescopes have already been rejuvenated in this way by installation of adaptive optics systems, including the 60- and 100-inch telescopes at Mt Wilson [Shelton et al. 1993, 1995] and the 3-m Shane Telescope at Lick Observatory [Olivier et al. 1995].

Spectroscopy is probably the most important observation technique in astronomy. Conventional spectrographs require the available light to be concentrated into a narrow slit. By reducing optical aberrations due to turbulence, adaptive optics increases the peak intensity of unresolved images, or increases the spatial resolution of extended objects. This improvement in efficiency can be used either to reduce the integration time of the spectrograph for objects of a given brightness, or to improve its spatial and spectral resolution.

Direct imaging benefits in two ways from adaptive wavefront compensation. Not only is the angular resolution of the telescope improved, but the peak intensity of the image is increased with respect to the sky background. This leads to a considerable improvement in the limiting magnitude for detection of faint objects. With multiple or extended objects, wavefront compensation increases the contrast of detail within the object. A particular case of interest is the direct detection of planetary systems around nearby stars, in which adaptive optics plays a critical role in reducing the scattered light from the parent star.

In long-baseline interferometry, adaptive optics is used to point the telescopes accurately, to correct wavefront errors within the optical apertures, and to equalize the optical pathlengths of the connecting arms. The fringe contrast of interferometers is degraded seriously by wavefront errors in the interfering beams; adaptive optics has an important role in these instruments. Multiple-aperture interferometry with large telescopes, which is becoming an important method of achieving high angular resolution, would not be feasible without adaptive optics technology.

Adaptive optics also has a role in space telescopes, to compensate for deformation of optical surfaces and structures due to thermal variations and attitude changes. It also provides correction for image motion (jitter) caused by vibration from attitude control or other equipment on board the spacecraft.

Originally developed as an exotic solution for special-purpose imaging and laser weapons, adaptive optics has blossomed into a technology that is useful in astronomical telescopes of all sizes and purposes.

2.2 Observing Through the Atmosphere

2.2.1 Atmospheric Limitations

The information obtained about celestial objects is carried by light waves that travel over vast distances in space for tens to millions of years until, in the last millisecond of their journey, the waves are distorted as they plunge through Earth's atmosphere. The atmosphere has three main characteristics that restrict astronomical observations:

- opacity, due to the absorption of light at certain wavelengths;
- luminosity, due to scattered light, broadband radiation, and airglow emission;
- turbulence, due to solar heating and winds.

Atmospheric transparency is good at wavelengths between 0.3 and 1.0 μm, covering the ultraviolet (UV), visible, and near-infrared (IR) bands. It is

through this clear window that much of human knowledge of the universe has been obtained. Between about 1 and 10 μm, atmospheric transmission varies considerably due to absorption by water (H_2O) and carbon dioxide (CO2) molecules, but there are good windows at 1.25 μm (J-band), 1.6 μm (Hi-band), 2.2 μm (K-band), and 3.4 μm (L-band). These near-IR windows are very important in the world of adaptive optics. The atmosphere is opaque to radiation at wavelengths below 0.3 μm and above 30 μm, but becomes transmissive to radio waves longer than about 1 mm. That, however, is another story.

The energy absorbed by water vapor and carbon dioxide molecules in the atmosphere is emitted as blackbody radiation, mainly at wavelengths longer than 3 μm. Measurements in the K-band and shorter wavelengths are little affected by thermal radiation from objects at ambient temperatures around 300 K, but in the L-band and above, the situation is different. Considerable care must be taken in the design of IR telescopes and instruments to prevent local thermal radiation from entering the detector.

Scattered sunlight is the largest source of sky radiation during the day and on moonlit nights, but on moonless nights most of the atmospheric radiation between 0.5 and 2.5 μm is due to airglow emissions from hydroxyl (OH) molecules. Background radiation from the sky sets a practical limit to the detection of faint objects from Earth's surface, because the exposure time is generally limited by Earth's rotation. Within this constraint, the limiting magnitude for a single exposure is determined by the contrast between the object and the background radiation, and also by the characteristics of the detector. Adaptive optics concentrates the available light into the smallest spot, improving the limiting magnitude of most instruments.

Turbulence in Earth's atmosphere is caused by energy input from solar heating, which causes air masses of various sizes to move with different velocities and in different directions. The maximum turbulence usually occurs in the surface layer, especially during the day, producing thermal plumes that extend up to about 3 km. Above this altitude, turbulence tends to occur in layers, usually associated with wind shear. High wind velocities are encountered throughout the troposphere, especially in the region of the tropopause at altitudes between 10 and 15 km. The turbulence peak often present at these altitudes is due to wind shear.

Atmospheric turbulence affects all astronomical observations made at wavelengths below about 10 μm. The wavefront distortion caused by turbulence spreads the energy received from a point source into a diffuse disk, considerably reducing its peak intensity and making its characteristics more difficult to measure. The optical pathlength changes produced by turbulence are almost independent of wavelength, because the spectral dispersion of air is very small over the visible and IR spectral bands used in ground-based astronomy. Most astronomical objects emit a broad spectrum from which a large amount of information can be deduced. Astronomical adaptive optics systems are therefore designed to operate over wide spectral bands, for which it is essential to compensate the optical pathlength, independent of wavelength. Deformable mirrors do this in the most direct way by adjusting the pathlength mechanically. Phase-shifting devices, such as liquid crystals, may be used over spectral bands in which they have low dispersion.

One result of the small value of atmospheric dispersion is that the optical effects of turbulence decrease at longer wavelengths. A given change in optical pathlength represents a smaller phase error as the wavelength increases, and it is optical phase relationships that govern the structure of the image. While observations in the visible band at 0.5 μm are greatly impaired by turbulence, the effects at 10 μm are negligible for all except the very largest apertures. Wavefront distortion is more easily compensated at near-IR wavelengths than in the visible, which is the main reason that astronomical adaptive optics has been most successful at wavelengths between 1 and 3.4 μm. The potential improvement in angular resolution is much greater at visible wavelengths.

2.2.2 Imaging Through Turbulence

The optical effects of turbulence are due to local temperature variations that produce changes in the refractive index of air. The magnitude of the refractive index fluctuations depends on the air density as well as on the range of the temperature variations. Air density is greatest at sea level and decays exponentially with height. Optical effects of turbulence therefore generally decrease with altitude, which is the reason for locating astronomical observatories on mountain peaks, well above the surface layer. At good sites, most of the turbulence tends to be concentrated in several thin layers associated with wind gradients. It is fortunate for Earth-bound astronomers that the characteristics of light waves and of the atmosphere allow near-vertical beams to be compensated in a straightforward way. In most cases, geometric optics can be assumed: wavefront errors are summed along ray paths and diffraction effects can be ignored. If the strength of the turbulence was much higher (as it is in horizontal paths near the ground), then multiple scattering would make real-time compensation using adaptive optics far more difficult, or even impossible.

The effects of atmospheric turbulence on the image of a point source are depicted in figure 2.1. The radiation emitted by a stellar source may be represented as an initially spherical wavefront that is treated as a plane wave when it reaches Earth. The telescope collects part of the wavefront and changes its curvature to produce an image at the focal plane. In the absence of wavefront distortion,

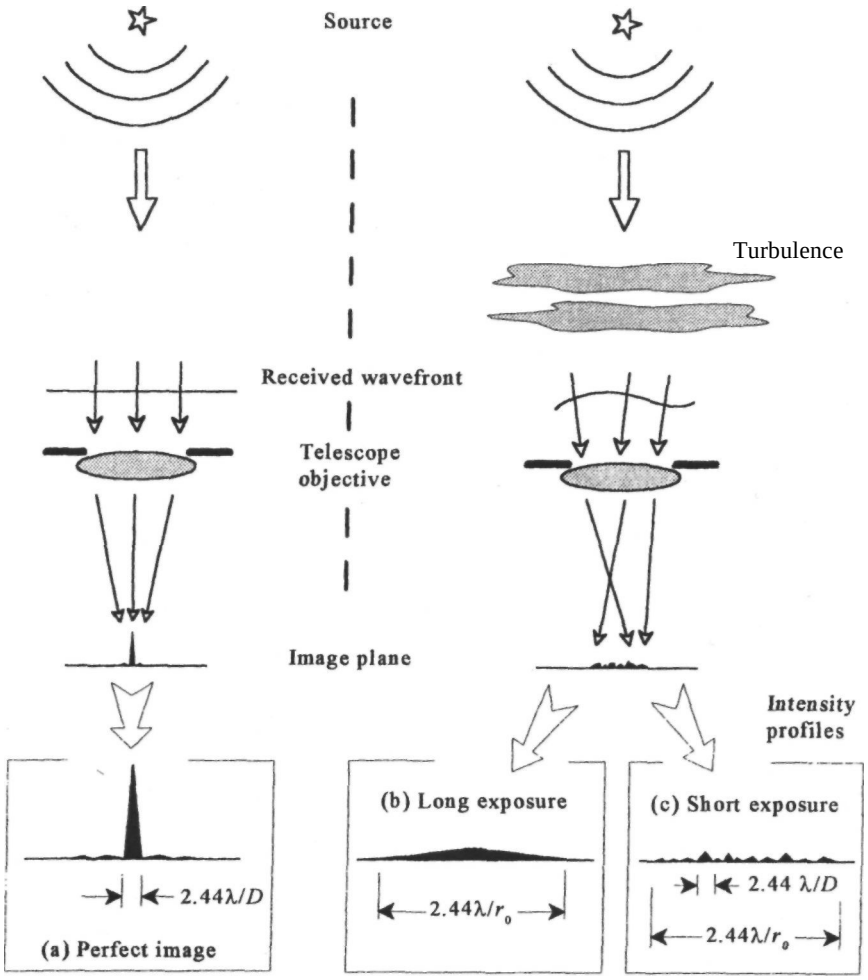


Figure 2.1 Effects of turbulence on the image of a star.

the angular diameter of this image is determined by the diffraction of light and is proportional to lambda/D, where l is the mean wavelength of observation and *D* is the telescope aperture, as shown in figure 2.1(a). The fact that the angular resolving power of a telescope is limited, even under ideal conditions, may be explained by the fact that only a very small part of the total energy emitted by a celestial object is collected by the telescope aperture, so that our knowledge of the true nature of the source is very restricted. As the telescope aperture is increased, more information is gathered and the angular resolution improves.

In the case of ground-based imaging, the incoming radiation is distorted by atmospheric turbulence, so that the converging wavefronts forming the image are no longer spherical. A useful way of quantizing the effects of wavefront distortion on optical images is to define a turbulence scale size in terms of the lateral distance over which the wavefront phase is highly correlated. A standard measure of the optical strength of turbulence is Fried's parameter r_0, the diameter over which the optical phase distortion has a mean-square value of 1 rad^2 at a wavelength of 0.5 mm. A phase error of 1 rad (l/2p waves) is the threshold above which the image quality deteriorates rapidly. For an undistorted plane wave, r_0 is infinity, while for atmospheric distortion it typically varies between 5 cm (poor seeing) and 20 cm (excellent seeing). Although sometimes regarded as a constant, r_0 is a statistical parameter that varies considerably over short periods of time, sometimes changing by a factor of 2 within a few seconds. It is also a function of wavelength, as discussed later in this section.

When r_0, is less than the telescope aperture (usually the case for ground-based observations), the angular size of the image for long exposures, referred to as the seeing disk, is determined by the

ratio λ/r_0 rather than λ/D, as depicted in figure 2.1(b). The result is that the angular resolution of even the largest uncompensated ground-based telescopes at visible wavelengths is about 1 arc second, and it is often 2 arc seconds or more. The median seeing disk at Mauna Kea, one of the best astronomical sites in the world, is about 0.5 arc second at visible wavelengths [Racine and Ellerbroek 1995]. To put these results in perspective, it is useful to recall that a diffraction-limited 4-m telescope is capable of producing an image with a full-width at half-maximum (FWHM) of only 0.03 arc second in the visible band, more than 10 times better than even the best uncompensated image. The potential gain produced by adaptive optics is therefore considerable, even for modest-sized telescopes at the best sites. With or without adaptive optics, it is always an advantage to have the best possible seeing conditions, as this makes it easier to achieve a given level of performance.

The size of the largest disturbances produced by atmospheric turbulence (the "outer scale") varies considerably, with reported values ranging from 1 m to over 1 km. Such disturbances produce a time-varying overall wavefront tilt that causes fluctuations in the angle-of-arrival of the light from a star, randomly displacing the image. If the exposure time is very short (less than about 1/50 second), the blur due to image motion is negligible and the fine structure of the image becomes visible, as shown in figure 2.1(c). This structure, which varies with time, is due to high-order wavefront distortions that occur within the aperture. The short-exposure image consists of a large number of speckles of diameter λ/D, produced by interference between rays separated by the telescope aperture D. The speckle pattern is the basis for a special imaging technique known as speckle interferometry.

The effects of turbulence on the image produced by an uncompensated telescope depend both on its aperture D and on the coherence length of the turbulence r_0. This relationship was investigated by Fried [1966a], and is described in section 3.4.4. The analysis is based on the standard Kolmogorov model of atmospheric turbulence. The results are illustrated neatly by plotting the normalized angular image size in $\lambda/$ units against the normalized telescope aperture D/r_0 as shown in figure 2.2. This chart identifies the three components that determine the size of any turbulence-degraded image:

1. the diffraction limit of the aperture, determined by λ/D;
2. the short-exposure image spread, determined by λ/r_0;
3. the image motion component, determined by both D and r_0.

The diameter of a long-exposure image is the sum of these three components.

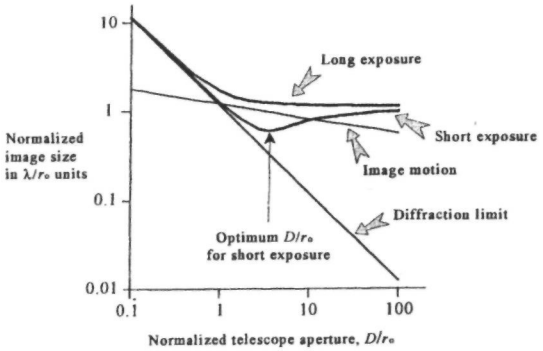


Figure 2.2 Imaging through atmospheric turbulence: normalized image characteristics as a function of D/r_0.

The chart can be divided into three main regions, according to the value of D/r_0. For very small apertures, when D/r_0 is less than 1, atmospheric turbulence has little effect on the image size, which is determined by D. At D/r_0 values between 1 and 10, image motion is significant and there is a notable difference between short-exposure and long-exposure images. This difference gradually disappears when D/r_0 exceeds 10. Astronomical telescopes usually operate at D/r_0 values between 5 and 50.

When $D < r_0$, the wavefront distortion over the aperture is mainly overall tilt. There is little distortion within the aperture because of its small size, so the instantaneous image diameter is determined primarily by the aperture D. The effect of turbulence is to produce random displacements of the centroid of this image, but, for values of $D/r_0 < 0.5$, even this effect becomes negligible. Adaptive optics has no role in this regime.

When $D > 10r_0$, conditions are entirely different. The uncompensated image size is now determined mainly by the turbulence coherence length r_0. The magnitude of the wavefront error due to overall tilt increases as $D^{5/6}$, but on an angular basis it is proportional to $D^{1/6}$; that is, image motion decreases as the aperture gets larger. As a result, the difference between (uncompensated) short-exposure and long-exposure images becomes less significant as the aperture increases, both being asymptotic to λ/r_0. The potential improvement from the use of adaptive optics is shown by the difference between this value and the diffraction-limited curve on the chart. With a typical D/r_0 value of 40 at visible wavelengths (4-m telescope, 0.1-m turbulence), adaptive optics is capable of reducing the image diameter by a factor of about 30, increasing the peak intensity of a point source 900 times.

The transition region, where D/r_0 is between 1 and 10, is the region in which a useful improvement in telescope performance can be obtained by compensating image motion. As shown in figure 2.2, the size

of an uncompensated short-exposure image is minimized when D/r_0 is equal to 3.8, at which its diameter is only half that of a long-exposure image. Image motion compensation is a simple form of adaptive optics that is particularly useful at IR wavelengths, where r_0 values of 1 m or more are encountered, as explained in section 2.4.4. The improvement obtained by compensating the image motion due to turbulence is modest, a factor of two being the best that can be obtained. With large telescopes having D/r_0 ratios of 10 or more, the theoretical improvement obtained by image stabilization alone is of marginal value. In practice, several other sources of image motion, such as wind buffeting and telescope tracking errors, may be reduced significantly by tilt compensation.

It is apparent from the preceding discussion that the potential gain from full wavefront compensation far outweighs the limited improvement obtained by removing only the image motion. This is especially true with large apertures; for this reason, adaptive optics is now included in the design of most new astronomical telescopes.

2.2.3 Wavefront Compensation

Adaptive wavefront compensation is based on the use of a reference source or beacon, the light from which traverses the same optical path as that from the object being observed. The simplest case to consider is the compensation of a single point source, as shown in figure 2.3. For a point source at infinite distance, the wavefront reaching Earth's atmosphere is a plane wave that becomes distorted in its passage through the turbulent air. The optical pathlength disturbances in each part of the aperture are measured by comparing the received wavefront with a local reference (equivalent to a plane wave) in the wavefront sensor, to produce the pathlength error $W(\mathbf{x})$, where x is the position in the optical aperture. The original wavefront is then restored by inserting equal and opposite pathlength corrections C(x) in the optical path, producing a compensated image of the source.

Although a point source is, by definition, spatially coherent, there is no requirement for it to be temporally coherent (monochromatic), providing that the wavefront sensor measures optical pathlength (rather than optical phase). As mentioned earlier, the dispersion of air is very small over the spectral range employed in optical astronomy, and wavefront distortion may be measured in terms of optical pathlength variations, independent of wavelength.

The point source model is valid for multiple layers or distributed turbulence, providing that the turbulence level is consistent with the use of geometrical optics. The optical path differences in each layer are summed along the ray paths from the object to the telescope aperture, and are compensated at a single location, making the appropriate adjustment in each zone of the aperture, as before. The practical result of compensating a point source (such as a star) is that the peak intensity of the image is improved by the factor of $(D/r_0)^2$, which may be several hundred in the case of an astronomical telescope. Point sources contain spatial information (their positions), but of at least equal importance is the spectral information carried in their light. Wavefront compensation dramatically improves the spatial resolution and signal-to-noise ratio in spectrographic work, probably the most important tool in astronomy.

The combination of turbulent layers, telescope optics, and adaptive optics may be considered as an extended imaging system, which converts a plane wavefront reaching the top of the atmosphere into a perfectly spherical beam that forms the image. The optical pathlength variations produced by atmospheric turbulence are a form of encoding applied to the received wavefront. The function of the wavefront sensor is to find the key to this code, which is the distribution of wavefront distortion over the aperture. The wavefront corrector is then able to perform the decoding operation, retrieving the original information. Adaptive optics system design includes

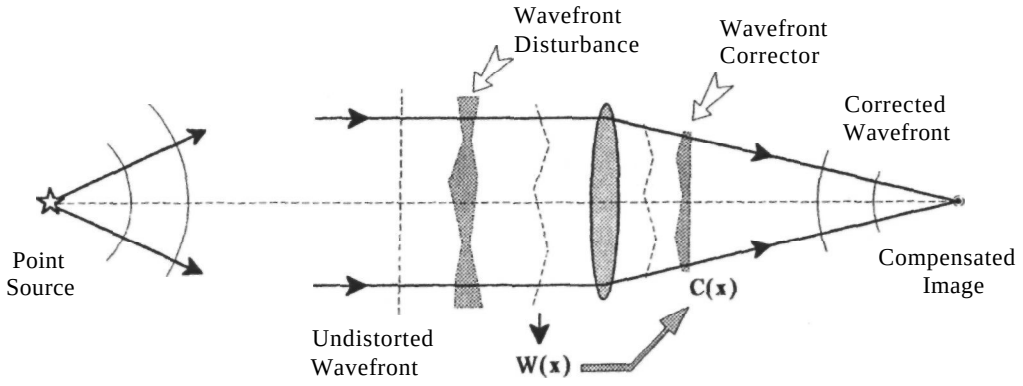


Figure 2.3 Principle of adaptive wavefront compensation.

consideration of the whole of this extended imaging system.

2.2.4 Anisoplanatism

The wavefront measured with a point reference source is strictly valid only for an object in exactly the same direction as the reference. If the turbulence is concentrated in a thin layer at the entrance pupil of the telescope, then the wavefront error can be compensated completely at all field angles by placing a corrector at a conjugate image of the pupil. In reality, turbulence is distributed along the propagation path through the atmosphere, with the result that the wavefront error becomes decorrelated as the field angle increases. This effect is known as angular anisoplanatism and is one of the major limitations to astronomical adaptive optics.

The basis of angular anisoplanatism is shown in figure 2.4. The simplest case is a single layer of turbulence at distance h from the entrance pupil of the telescope. The wavefront error varies as a function of the angle between two rays. If the coherence length of the turbulence is r_0, the angle over which the wavefront disturbance is well correlated must be a function of r_0/h. The isoplanatic angle θ_0 is defined as the angle at which the mean-square wavefront error, averaged over a large aperture, is 1 rad^2. For a single layer of turbulence, this angle is found to be $\theta_0 = 0.31\ r_0/h$. The isoplanatic angle is extremely small, typically about 2 arc seconds at visible wavelengths, increasing to about 10 arc seconds in the near IR.

To make adaptive optics work, a bright reference source of about $m_V = 10$ is required within a few arc seconds of the object being observed. In some cases, the science object may include a bright star that can be used as a reference, but, in general, the small size of the isoplanatic angle severely limits the sky coverage of adaptive optics systems. The solution is to create beacons high in the atmosphere, using the backscattered light or fluorescence from lasers located near the telescope. In this way, a reference source can be placed at any position in the sky, enabling the use of adaptive optics with the faintest objects. Even when using laser beacons, a natural guide star is still required to stabilize the position of the compensated image, but much dimmer stars ($m_V \approx 16$–20) can be used for this purpose. Tilt is correlated over relatively large areas, so the sky coverage is greatly improved, although still not complete at visible wavelengths. The issues involved in the use of laser beacons are summarized in section 2.9, and the subject is treated in detail in chapter 7.

2.3 The Role of Adaptive Optics

2.3.1 Science Objectives

Adaptive optics had little impact on astronomy until the 1990s, but it is clear that it will have a major role in the coming decades. There are many scientific areas in which the increased angular resolution and image contrast made possible by adaptive optics should greatly expand our knowledge of the universe. Specific fields of interest have been described by Bahcall [1991] and colleagues, Beckers [1993], Lena [1994], and Chanan et al. [1996]. Some general areas of interest are summarized in this section. They may be broadly divided into *galactic* and *extragalactic* astronomy.

Galactic Astronomy

Imaging of Small Extended Objects Improved angular resolution will enable detailed images to be obtained of solar system objects, such as asteroids, comets, and satellites, leading to the determination of their shapes, rotation, and surface structure. It

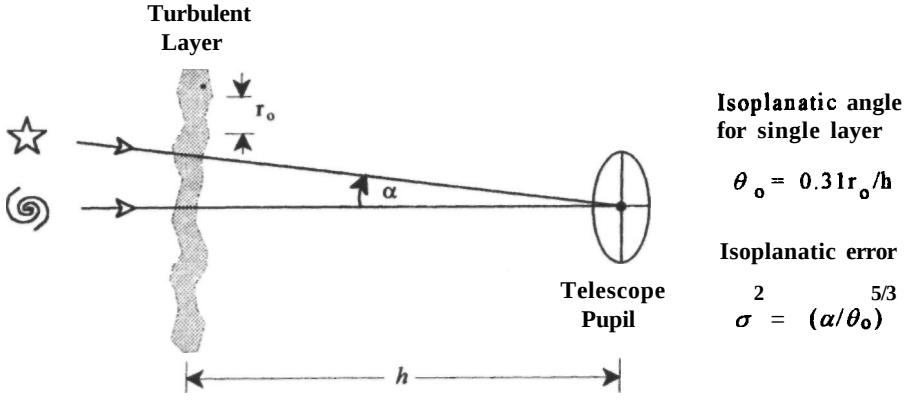


Figure 2.4 Basis of angular anisoplanatism.

should also provide better images of small areas of the Sun, the Moon, and the planets, within the isoplanatic angle. Ultimately, it is expected that the development of wide-field (multiconjugate) compensation will enable high-resolution images to be obtained over large areas. Many of these objects are bright enough to function as their own reference sources for adaptive optics.

Star and Planet Formation The processes that lead to the formation of stars and planetary systems are difficult to observe and consequently are not well understood. Young stars are embedded in dense dust and gas clouds, and the gaseous disks and jets associated with star and planet formation are best seen at IR wavelengths. Large telescopes equipped with adaptive optics will provide the spatial and spectral resolution necessary to analyze these objects and their dynamics at various stages of their evolution. When used in conjunction with a coronagraph to suppress scattered light, adaptive optics provides the high angular resolution and image contrast needed to see the structure of protoplanetary disks and to obtain direct evidence of companion objects. With these objects also, a star bright enough to function as the reference source is often present, eliminating the need for a laser beacon.

Detection of Faint Companions The ability to detect brown dwarfs and planetary systems around nearby stars is one of the biggest incentives for the use of adaptive optics. Brown dwarfs are cool objects about the same size as Jupiter (1/10 of the Sun's diameter) with a mass of less than 1/16 of the Sun. Such objects cannot sustain fusion and have a low surface temperature, requiring a search at near-IR wavelengths. They are difficult to detect, but if found in large numbers they may explain the missing mass in the universe. The direct detection of extrasolar planets is a good application for adaptive optics because the parent star provides the reference source and the objects searched for are usually within the isoplanatic angle [Angel 1994]. Adaptive optics can improve detection of faint companions in two ways: low-order compensation concentrates the energy of the objects into a smaller radius, while high-order compensation reduces the halo due to scattered light from the central star. Some objects have already been detected using a combination of adaptive optics and coronagraphic techniques [Nakajima 1994, Nakajima et al. 1994]. To make a survey of stars within 30 light years with the goal of detecting Jupiter-size planets at a separation of 0.5 arc second, Sandler et al. [1995b] found that a very high-resolution adaptive optics system is required. Using a 6.5-m telescope with an observation wavelength of 1 μm, about 10,000 correction elements are needed, with an update interval of 0.5 ms. Such a system appears to be within the state of the art.

Separation of Close-Packed Objects Adaptive optics will enable closely packed stars in the central regions of globular clusters to be resolved. A census of stellar types as a function of radius is needed to determine the dynamics of globular clusters. Another area of study is the structure of the galactic center, which, because of the high extinction due to interstellar dust, can be penetrated only at IR wavelengths. High spatial resolution is needed to separate individual objects and determine their spectra.

High-Resolution Spectroscopy Adaptive optics is capable of improving the performance of spectrographs by concentrating the beam in slits as small as 0.1 arc second, increasing the spectral resolution to over 100,000. This high resolution will be of great value for research in stellar physics, specifically to measure magnetic fields and convective motions. It has been suggested that the physical and chemical structure of dark interstellar clouds could be determined using high-resolution IR absorption spectroscopy of background stars.

Extragalactic Astronomy

It is not known how the universe evolved from its initial state to the large-scale lumpy structure containing the clusters of galaxies that are observed today. To find the answers it is necessary to trace the evolution of galaxies back to early epochs corresponding to large redshifts. These studies will require a significant increase in the light grasp of astronomical telescopes, using high-resolution imaging and spectroscopy at visible and IR wavelengths. The combination of large telescopes of the 8- to 12-m class augmented with adaptive optics is expected to bring significant discoveries in the following areas:

- the formation of galaxies, with spectral analysis of their composition and radial motion;
- galaxy counts and morphology at high redshifts;
- active galactic nuclei;
- stellar populations in distant galaxies;
- quasar environments and spectra.

There are many specialized areas of astronomical research in which the huge increase in the measuring power of instruments brought about by adaptive optics will bring new discoveries about the universe. Reasenberg [1990] has reviewed the special science opportunities that are presented by long baseline interferometers. They include measurement of stellar diameters, plotting luminosity-temperature curves for Cepheid variables, and determination of the orbits of binary stars. The structure of protoplanetary disks surrounding young stellar objects should be observable. It should also be possible to determine the diameter and shape of the luminous shells of novae as a function of time. Observations of giant

stars in the visible and the IR spectral bands would enable the tracking of starspots and the measurement of rotational and magnetic distortions of their shape.

2.3.2 Direct Imaging and Spectroscopy

Astronomical instruments measure the intensity of radiation received from celestial objects in a three-dimensional space, defined by two spatial (or angular) coordinates, together with a third spectral dimension. The two classic instruments used in astronomy are the imaging camera, which provides two spatial dimensions with a fixed spectral band determined by a whole-aperture filter, and the long-slit spectrograph in which one axis of the detector is used to record a continuous spectrum. By concentrating the light from each element of the object into a smaller core or slit, adaptive optics improves the resolution in all three dimensions. This capability vastly increases the data-gathering capacity of observational instruments. To take full advantage of adaptive optics, the observational instruments must be designed to accommodate the full diffraction-limited resolution of the telescope rather than the seeing-limited resolution.

Contemporary astronomical instruments are variations on a single theme: photons collected by the telescope are routed through various optical pathways that analyze the light, and the results are recorded on a two-dimensional detector array. The analyzing elements are often mounted on wheels for easy interchangeability: they include spectral filters, grisms, slits, field stops, gratings, polarization optics, and reimaging optics. Instruments are still designed for specific telescopes and focal stations, but the modular approach provides flexibility for optimizing the measurements for the scientific task and imaging conditions.

A single instrument may be capable of being configured as an imager/photometer, a long-slit spectrograph, a multiobject spectrograph, or a polarimeter, by suitable arrangement of internal components, using the same detector array in each case. Examples of this approach are the FORS and CONICA instruments designed for the European Southern Observatory Very Large Telescope (ESO VLT). The FORS (FOcal Reducer/low dispersion Spectrograph) is a multipurpose instrument capable of direct imaging, long-slit spectroscopy, multiobject spectroscopy, polarimetry, and medium-dispersion echelle spectroscopy [Appenzeller and Rupprecht 1992, Seifert et al. 1994]. It has a spectral range of 0.33–1.1 μm, using a 2048x2048 charge-coupled device (CCD) array. The CONICA is a high-resolution near-IR camera with capability for low-resolution spectroscopy (at high spatial resolution) and moderate-precision polarimetry. The detector is a 1024 x 1024 InSb array with a spectral range of 1-5 μm. This instrument has the capability of inserting coronagraphic masks in the focal plane, making it suitable for measurements near bright objects. Similar instruments are being designed for all of the new generation of 8- to 12-m telescopes. Their high precision and flexible design allows them to be optimized for use with adaptive optics.

Because of their superior performance, CCD cameras and their equivalents at IR wavelengths are now used almost universally in astronomical instruments. The availability of large-area low-noise CCD arrays has revolutionized the design of these instruments. Because of the commonality of the detector arrays and supporting software, the instruments have tended to become multipurpose. They are now segregated by wavelength rather than purpose, because optical materials and sensors are different for visible, near-IR, and thermal-IR bands. For example, mirrors are efficient at wavelengths below about 2.5 μm, but they contribute to the thermal background and so are avoided in instruments operating at over 3 μm.

Adaptive optics improves the spatial resolution of imaging devices by concentrating more light in the diffraction-limited core. To take full advantage of this, the pixel size of the CCD cameras now used universally for astronomical imaging must be compatible with the diffraction limit of the telescope, rather than with the seeing limit. The diffraction-limited image diameter (i.e., the FWHM) of a 4-m telescope at 0.5 μm (and of an 8-m telescope at 1 μm) is 0.03 arc second, requiring a pixel size of about 0.015 arc second. With a 1k $\times$ 1k array, the field of view is about 15 arc seconds, not all of which will be compensated fully at short wavelengths, due to anisoplanatism. It is anticipated that the development of multiconjugate compensation will enable larger fields to be corrected in future.

In the case of the long-slit spectrograph, Beckers [1993] has noted that the resolution of seeing-limited instruments is coupled linearly to the telescope aperture, because of the practical limit on the size of diffraction gratings. As a result, the resolution of conventional spectrographs is limited to about 100,000. This restriction is removed if the telescope delivers a diffraction-limited image, because the image size is no longer constant but becomes inversely proportional to the telescope aperture. The slit size can then be reduced and still capture most of the light, resulting in much higher spectral resolution. The smaller slit also allows physically smaller instruments.

The data-gathering capacity of spectrographic instruments can be improved by more efficient utilization of two-dimensional detectors. Multiple-object spectrographs employ optical fibers to capture the light simultaneously from several objects in the field. To use an area detector in the most efficient way, the fibers are assembled neatly to occupy one dimension while the spectrum is displayed in the other direction. Adaptive optics improves the coupling efficiency of the images to the fibers, as well as increasing the signal-to-background ratio.

Integral field spectroscopy is a method of optically integrating small extended fields to obtain their spectra. The use of adaptive optics improves the spatial resolution so that integral spectra may be obtained over smaller fields without overlap.

2.3.3 Interferometers and Telescope Arrays

Long-baseline interferometers and multitelescope arrays are playing an increasingly important role in ground-based astronomy. Interferometers operate with baselines far exceeding the diameter of the largest filled-aperture telescopes, giving milli-arc-second angular resolution that enables the diameters and other properties of stellar systems to be measured. Telescope arrays have great versatility: the instruments can be used individually, in pairs for Michelson interferometry, or in groups for interferometric imaging with large effective apertures. The full potential of these instruments can only be realized with adaptive optics.

All optical imaging systems depend on interference to convert wavefront phase variations into measurable intensity variations in the image plane. The image of an incoherent source can be considered as the superimposition of interference fringes produced by pairs of rays collected by the optical aperture. In the case of a conventional circular aperture, interference occurs between pairs of rays at all orientations and spacings up to the full diameter D. The limiting angular resolution is then proportional to λ/D, where λ is the wavelength of the light. Currently, the diameter of the largest filled-aperture optical telescopes (Keck I and II) is 10 m. Long-baseline interferometers and telescope arrays use two or more individual apertures, separated by distances of up to several hundred meters, to collect the light beams, which are then brought together to produce interference fringes.

The imaging process can be modeled using the Fourier transform. As shown in section 4.2, the intensity distribution in the image plane is the squared Fourier transform of the complex amplitude of the received wavefront. An image may be considered as the superposition of discrete Fourier components, each produced by a symmetrical pair of wavefront samples in the optical pupil, defined by the u,v plane. To obtain a fully detailed two-dimensional image, it is necessary to have complete coverage of the u,v plane. This is automatically obtained (with considerable redundancy) when using a conventional filled aperture.

The technique of interferometry is different from conventional imaging in that the Fourier components (or related quantities) are measured and recorded individually. The data may be used directly, or processed digitally to reconstruct an image. Techniques for image reconstruction from interferometer measurements were originally developed for radio astronomy, which employs baselines that are necessarily much larger than the telescope apertures. The baseline length and orientation with respect to the sky define the u,v coordinates of each measurement. Coverage of the u,v plane is obtained by using multiple baselines, by moving the telescopes, and by using Earth's rotation.

The basic theory of interferometry is treated in Thompson et al. [1986]. The principles and practice of interferometric imaging in optical astronomy are described by Roddier [1988b]. Recent advances in long-baseline optical interferometry are reviewed by Shao and Colavita [1992].

The potential advantages of long-baseline interferometers have been known for more than a century, but the practical difficulties involved in their construction and operation have restricted their use in astronomy. Not until the late 1980s were they available for regular observations. The two main problems with ground-based optical interferometers are (1) atmospheric turbulence, and (2) instability of the optical paths between the collecting apertures and the detection device. Both of these cause uncertainty in the position or phase of the interference fringes. As a result, the quantity usually measured is the intensity (amplitude squared) of the fringes, as a function of the length and orientation of the baseline. If the structure of the object is known (e.g., a disk or binary star), then useful information, such as the diameter or angular separation, can be deduced from fringe intensity measurements alone.

To obtain interference fringes in a long-baseline interferometer, three conditions must be satisfied:

- The beams collected by the interferometer apertures must be superimposed and coaligned to within the diffraction limit of the individual apertures.
- The optical pathlengths between the object and image plane must be equal to within the coherence length of the radiation, $\lambda^2/\Delta\lambda$. In the case of visible white-light fringes, the coherence length is on the order of 1 μm.
- To obtain fringes of good contrast, the (differential) wavefront errors within each of the interfering beams should not exceed about 1 rad rms.

To meet these requirements, and to maximize the signal-to-noise ratio, a high degree of active control is required in ground-based interferometers. Fast-tracking mirrors are used to superimpose the beams, eliminating the tilts produced by atmospheric turbulence. Optical pathlengths are equalized with delay lines, automatically driven by fringe-tracking systems. Larger telescope apertures improve the limiting magnitude of the interferometer (or fringe signal-to-noise ratio), but the full benefit is obtained only when wavefront errors due to turbulence are compensated with adaptive optics.

Long-baseline interferometers can be used in several modes to exploit their inherently high angular resolution. The most important applications are:

- astrometry,
- parametric measurements,
- interferometric imaging.

Astrometry requires an extremely stable, fixed baseline, which provides the reference for stellar angular position measurements. Using broadband light, the interferometer acquires each star in turn and measures the delay line length required to bring the white-light fringe to a fixed position. Measurements made using two or more baselines at different orientations enable the two-dimensional position angle of the star to be determined. The Mark III Stellar Interferometer [Shao et al. 1986] was designed for astrometry, using two 12-m baselines. It employs fringe-tracking with a servo bandwidth of 20 Hz. The apertures are 7.5 cm, smaller than the normal value of r_0 at the site (Mount Wilson), so no adaptive optics is used, other than guiding mirrors to align the beams. Shao and Colavita [1992] report that average position errors of 9.2 and 5.6 milli-arc seconds were achieved on stars brighter than about visual magnitude $m_V \approx 6$, significantly better than is obtained with conventional wide-angle astrometric instruments.

Parametric measurements are made on simple objects, such as disks or binary stars, by measuring the fringe contrast or visibility function (ignoring phase) as the interferometer baseline is varied. The visibility drops to zero when the separation is $\theta_0 = B\lambda/\theta$, where λ is the mean wavelength and B depends on the intensity distribution of the object. For two point sources of angular separation θ, B is equal to 0.5, while for a uniform circular disk of angular diameter θ, B is equal to 1.22. The first practical success with this technique was achieved by Michelson [1890, 1891], who measured the angular diameter of Jupiter's Galilean satellites using the 12-inch refractor at Lick Observatory. Michelson and Pease [1921] later extended this technique to baselines larger than the telescope aperture and were able to determine that the diameter of Betelgeuse was 0.047 arc second. When measurements are made with a good signal-to-noise ratio, it is not necessary to seek the null in the visibility function, and the modern approach is to fit a series of visibility measurements to the appropriate visibility curve to determine the required parameter. This technique enables measurement of the angular diameter, separation, and magnitude differences of single and multiple objects that are unresolvable by conventional telescopes.

Interferometric imaging. True imaging is the most demanding of interferometric techniques as it requires the complex amplitude of the fringes to be determined at a number of different baselines and orientations. It is not necessary to obtain complete coverage of the u,v plane. Generally, the number of measurements should be comparable to the number of points required in the image. To build a two-dimensional image, fringe phases must be determined with respect to a common reference, defined by a point on or near the object. In principle, a natural star within the same isoplanatic patch as the object could be used as a common phase reference for all interferometer baselines. Unfortunately, the probability of finding a suitable star is extremely small. Laser beacons, even in the sodium layer at 90 km, are not useful for this purpose, because their positions cannot be controlled with sufficient accuracy. Even if a beacon could be accurately placed, parallax would limit its useful area to a few meters on the ground, totally insufficient to cover a telescope array. Because of the difficulty of making consistent optical phase measurements, it is not possible, with current technology, to form an image using an array of completely independent apertures.

One solution to interferometric imaging is to use the closure phase technique, pioneered in radio interferometry [Rogstad 1968, Pearson and Readhead 1984]. The closure phase is the sum of phases around a closed loop of telescope baselines. The atmospheric and telescope errors then cancel, but object phases add. Because of the time-varying nature of atmospheric turbulence, all phases must be measured simultaneously. An array of A' telescopes produces $N(N-1)/2$ independent fringe phases, one on each baseline, and the number of independent closure phases is $(N-1)(N-2)/2$. The minimum value of N is 3, which yields only one closure phase. In this case, although phase information is lost because there are three baselines but only one measurement of closure phase, it can be combined with the three measurements of fringe intensity to resolve ambiguities, enabling the reconstruction of simple images using iterative algorithms. The fraction of the object phases recovered by phase closure is $(N-2)/N$, so that an array of 10 telescopes would retrieve 80% of the object phase information. The closure phase technique is not without its drawbacks: because it is a differential measurement, information on the position of the object is lost, and the method involves considerable data processing. Closure phase has been used at optical wavelengths to obtain a two-dimensional image, using the Cambridge Optical Aperture Synthesis Telescope (COAST) interferometer [Baldwin et al. 1996].

There are other possible methods of obtaining images from long-baseline interferometer measurements. Spectral dispersion of the fringes in a two-telescope interferometer was suggested by Koechlin et al. [1979] as a method of obtaining two-dimensional information. Schloerb [1990] has shown by numerical simulation that this technique is capable of reconstructing two-dimensional images.

Another technique is to drive the delay line rapidly to scan the point of pathlength equality across the field of view, within the change time of the atmospheric turbulence [Dyck et al. 1995]. This "delay referencing" procedure generates fringes that are referenced to a common component, preserving the relative positional information within the isoplanatic patch. The object brightness distribution in each scan direction is obtained using the CLEAN algorithm [Hogbom 1974, Clark 1980]. To obtain a two-dimensional image, it is necessary to make delay-referenced measurements with at least two differently oriented baselines. Complex sources would require a more complete sampling in the u,v plane, using a tomographic approach with multiple baselines generated by Earth's rotation.

Telescope Arrays

Multiple telescopes can be combined incoherently (the light bucket mode), which increases the light-gathering power but not the angular resolution, or coherently (the phased mode), in which the angular resolution is determined by the overall dimensions of the array. Incoherent addition requires mutual alignment and scaling of the beams and has been demonstrated with the Multiple Mirror Telescope (MMT), employing, most recently, FASTTRAC II, an active optical package that corrects the overall wavefront tilt over each of the six primary mirrors independently, using a natural star as the wavefront reference [Gray et al. 1995]. Coherent phasing of separated telescopes is complicated not only by mechanical considerations but also by the effects of atmospheric turbulence, necessitating active control of optical pathlengths to correct systematic and random disturbances. Furthermore, the full gain of arrays using large telescopes is realized only when the individual apertures are compensated with adaptive optics.

Two different approaches are being followed for multiple telescope arrays:

1. To use a common mount for all elements, as in the MMT and Large Binocular Telescope (LBT). This arrangement simplifies the beam-combining optics because the optical paths are fixed relative to each element, but the mechanical structure must be capable of carrying the entire array and active control is required to eliminate gravitational strains and small random errors. The LBT consists of two 8.4 m primary mirrors on a common altitude-azimuth mount with a center-to-center separation of 14.4 m. One of the goals of this instrument is to achieve phase coherence across the full 22.8 m aperture at IR wavelengths. The length of the baseline is fixed and it rotates with respect to the sky as the object is tracked, giving the u,v plane coverage necessary for two-dimensional reconstruction of the image.
2. To use a separate mount for each element, as in the ESO VLT. In this case, the beam pathlengths and pupil geometry change with observing angle, requiring complex optical beam-combining systems that include continuously operating optical delay lines, pupil reimaging systems, and active beam stabilization [Merkle 1988].

In both cases, the individual telescopes are equipped with independent adaptive optical systems.

Adaptive Optics Requirements

When using adaptive optics with long baseline interferometers or telescope arrays, the large physical size of the telescope array and its effect on isoplanatism with various types of reference sources must be considered. Although the outer scale of turbulence may be comparable to the baselines used, most of the wavefront degradation occurring within separated apertures is uncorrelated. Each telescope in an array therefore requires an independent adaptive optics system. If the science object is bright enough to be used as a reference source, which generally requires a visual magnitude of < 10 at $0.7\ \mu$m and < 13 at $2.2\ \mu$m, or a natural guide star of this magnitude is within the isoplanatic angle of the science object, then the same reference source can be used by all telescopes in the array. (The wavefronts measured by the individual telescopes will, of course, be different.)

If no suitable natural star is available, then laser beacons may be used to measure the high-order wavefront errors (see section 2.9 and chapter 7). Each telescope in an array must have its own laser beacon, for the following reason. The highest altitude laser beacons are those generated in the sodium layer at about 90 km. The isoplanatic angle for average turbulence conditions is about 2 arc seconds at a wavelength of $0.5\ \mu$m, increasing to about 10 arc seconds at $2.2\ \mu$m. Because of parallax caused by its relatively low altitude, a sodium laser beacon will be within the isoplanatic angle over a radius on the ground of less than 1 m at $0.5\ \mu$m, and of less than 5 m at $2.2\ \mu$m. These dimensions are much smaller than the area covered by interferometer arrays, so it is necessary for each telescope to have its own laser beacon. The natural guide star required to provide the fixed reference for overall wavefront tilt correction can be shared.

Current Interferometer Projects

The modern era of optical interferometry was initiated by Labeyrie [1975], who succeeded in observing interference fringes produced by two independent telescopes of 25-cm aperture on a 12-m baseline, with the combining optics mounted on a moving table to equalize the pathlengths. The base-

line was later extended to 67 m [Koechlin and Rabbia 1985]. This instrument, the I2T (Interferometre a 2 Telescopes), has a limiting magnitude of about $m_V = 4$. A larger version, the GI2T, using 1.5-m telescopes, has since been constructed at the Observatoire de Calern, France, and it is planned to add a third telescope to allow image reconstruction.

The development of highly efficient photoelectric sensors, precision tracking mechanisms, and laser metrology have now solved many of the technical problems encountered in implementing amplitude interferometers at shorter wavelengths. Several versions of an interferometer dedicated to astrometry have been constructed at Mount Wilson Observatory [Shao and Staelin 1977, 1980]. The Mark III instrument has four baselines up to 20 m long with different orientations, and is also used for parametric measurements and the study of binary stars [Shao et al. 1986, Pan et al. 1994]. At the University of Sydney, a Michelson-type interferometer with an 11-m baseline was tested in 1985 [Davis and Tango 1985]. A much larger instrument, the Sydney University Stellar Interferometer (SUSI), using pairs of 14-cm apertures on baselines of up to 640 m, is now in operation [Davis 1993].

Most interferometers today are planned as multiple-aperture instruments, capable of producing images using the phase-closure technique, mentioned earlier. Measurement of closure phase at optical wavelengths was first achieved on a masked single-aperture telescope [Haniff et al. 1987], but has since been demonstrated with three elements of the COAST over a baseline of 6.1 m [Baldwin et al. 1996].

Many long-baseline optical interferometers are now being constructed. The Infrared-Optical Telescope Array (IOTA) is an imaging interferometer recently installed at the Smithsonian Astrophysical Observatory's F. L. Whipple Observatory at Mt Hopkins, Arizona [Carleton 1988, Carleton et al. 1994]. It covers the spectral bands 0.45-0.8 and 1.0–2.4 μm using a channeled spectrum detector array, which breaks the spectrum into narrow bands. The coherence length of the interferometer is determined by the spectral width of individual channels rather than by the full band. The baseline is 38 m and two optical delay lines are employed for pathlength compensation. The first has a range of 60 m to correct for the fixed observation geometry and is stationary during observations. The second delay line has a range of 4.8 m and is driven in real time to compensate for Earth's rotation; it will also be used for fringe tracking. The IOTA will ultimately use three 45-cm telescopes, to provide the closure phase necessary for image reconstruction.

The Navy Prototype Optical Interferometer (NPOI), located at Lowell Observatory, Arizona, consists of two subarrays, a four-element Astrometric Interferometer (AI), and a separate six-element Big Optical Array (BOA) for imaging. The AI employs a laser metrology system for monitoring the 20-m baselines [Hutter 1994]. The BOA imaging array has 250-m arms arranged in a Y and is intended for imaging stellar surfaces and close binary systems, as well as measuring stellar diameters and binary star orbits [Armstrong 1994].

The Astronomical Studies of Extrasolar Planetary Systems test-bed interferometer (ASEPS-0) at Mt Palomar is intended to measure star motions due to orbiting planets. It employs two 40-cm apertures, each of which simultaneously observes two stars. Fringes for both stars are detected at the same time, so atmospheric effects largely cancel when measuring the difference, enabling very small changes in their relative positions to be detected [Colavita et al. 1994].

The Center for High Angular Resolution Astronomy (CHARA) telescope array at Mount Wilson Observatory, built by Georgia State University, consists of five (eventually seven) 1-m apertures arranged in a Y-shaped configuration within a 400-m diameter circle. It operates at visible to near-IR wavelengths and is expected to achieve a limiting resolution of 0.2 milli-arc second in the visible [Mcalister et al. 1994]. Its stated purpose is the measurement of the diameters, distances, masses, and luminosities of stars, as well as image features such as spots and flares on their surfaces. Other applications include detection of other planetary systems, and imaging of the black-hole-driven engines of quasars and active galaxies.

Interferometric imaging is now an option in several large multiple telescope installations, enabling an angular resolution far greater than can be achieved with a single aperture. With adaptive optics, it is possible to compensate wavefront distortion over the largest telescope apertures, making the full aperture available for interferometric imaging and greatly increasing the limiting magnitude of the instruments. The 10-m Keck I and II telescopes on Mauna Kea will have the capability of operating in the interferometric mode with a fixed baseline of 85 m. The main telescopes will be supplemented by six 1.5-m apertures separated by up to 100 m.

The LBT, located on Mt Graham, Arizona, is a versatile instrument consisting of two 8.4-m F/1.14 mirrors on a common altitude-azimuth mount with a center-to-center separation of 14.4 m. When used in the phased combined mode, using interferometric imaging, the instrument has an effective aperture of 23 m. It is designed to operate over a wide spectral band, from 0.3 to 30 μm.

The most elaborate multitelescope installation is the ESO VLT, consisting of four fixed 8-m telescopes with different separations that can be used singly or in combination, together with eight movable 1.8-m telescopes. This instrument is described in more detail in chapter 10.

2.3.4 Active Control of Primary Mirrors

The primary mirror of an astronomical telescope is required to maintain its optical figure during the attitude changes resulting from tracking an object across the sky and during changes in the ambient temperature. For most of the twentieth century, primary mirrors have been made as stiff as possible, using a slab of low-expansion glass with a diameter-to-thickness (aspect) ratio of between 6:1 and 8:1. This approach is feasible only for mirrors up to about 5-m diameter. The flexure of a slab mirror due to its own weight increases as the square of the aperture, even for the same aspect ratio. For mirrors larger than 5 m, the weight and sag become excessive. Mechanical distortion can be compensated to some extent by adjustment of supporting pads, but such massive mirrors have another problem: thermal inertia. Although thermal distortion can be minimized by suitable choice of materials, it has been found that temperature differences of only 1 K between the mirror and the ambient air produce significant turbulence near the surface, degrading the image quality.

To make larger primary mirrors, it was necessary to solve both the structural problem and the thermal inertia problem. Three methods of fabrication have been developed to build mirrors of 6–10-m diameter: segmented mirrors, lightweight honeycomb mirrors, and meniscus mirrors. Segmented mirrors are composed of a large number of separate panels that fit together to form an almost continuous surface. The two 10-m primary mirrors of the Keck Observatory each consist of 36 hexagonal panels made of Zerodur, 1.8 m across and 75 mm thick. These panels are light and stiff enough to maintain their shape with a passive support, and their thermal inertia is relatively low.

Honeycomb mirrors consist of a thin faceplate and backplate, connected by a hexagonal structure of thin ribs [Angel, Davison et al. 1990]. They are cast in one piece, using a rotating mold to give the faceplate an approximately parabolic shape, which is then optically finished to the final figure. To minimize thermal stresses and surface convection, these honeycomb mirrors are maintained at ambient temperature by an internal air-conditioning system, which takes advantage of the low thermal time-constant of the lightweight structure. The largest mirrors of this type are the 8.4-m primaries for the LBT, made at the University of Arizona Mirror Laboratory.

Meniscus mirrors, as their name implies, are monolithic disks having a diameter-to-thickness ratio of about 40:1. They are supported by a large number of actuators, and rely entirely on active control to maintain their figure. The ESO 3.6-m telescope is of this type, serving as a prototype for the ESO VLT, which consists of four 8.2-m units that can be used separately or in combination. The VLT mirrors are cast of Zerodur, with a thickness of 175 mm. The active support system must accommodate three types of disturbances: systematic changes in the gravity vector as the telescope tracks across the sky, slow changes in the supporting structure due to temperature variations, and more rapid, random variations in loading due to wind buffeting. The meniscus is supported on passive axial and lateral hydraulic supports, together with a set of axial electromechanical actuators that provide the active control. The figure of the primary mirror is measured using a star as the reference source, and the necessary corrections are applied through these actuators.

The two Gemini 8-m telescopes use meniscus mirrors, made of Corning ULE™ (Ultra Low Expansion) quartz, with 1.2-m central holes. These mirrors have a novel active support system [Stepp and Huang 1994]. Eighty percent of the mirror weight is supported by uniform air pressure. The remaining 20% is taken by 120 axial supports, which provide both passive and active control of the mirror figure and position. Lateral support is provided by 72 passive hydraulic supports arranged around the circumference of the mirror. Two peripheral wavefront sensors monitor the incoming wavefront using reference stars just outside the science field. The wavefront measurements are analyzed to determine figure errors on the primary mirror and alignment errors between the primary and secondary mirrors. Corrections are made every few minutes.

Thermal control of the Gemini primary mirrors is provided by two interlinked systems [Gillett and Mountain 1996]. The bulk mirror temperature is controlled by a radiation plate located between the meniscus and the mirror cell. The bulk mirror temperature is preconditioned during the day and can be controlled slowly at night. The surface temperature of the mirror is adjusted to follow variations in the ambient temperature by ohmic heating of the reflective coating [Greenhalgh et al. 1994].

The Gemini telescopes also have provision for correcting rapid image motion and focus changes that are due to atmospheric turbulence and telescope effects by using small tilts and piston motions of the articulated secondary mirror.

2.3.5 Ground-Based versus Space Telescopes

Adaptive optics removes some, but not all, of the limitations of the atmosphere from ground-based telescopes. It is therefore of interest to compare the capabilities of space telescopes with those of ground-based telescopes equipped with adaptive optics. The two basic factors are cost and performance. Operation in space adds to the design requirements of a telescope in almost every category. Transportation into orbit produces constraints on the size, weight, and shock-resistance of the components. The space vehicle requires continuous attitude

control. Thermal control is more difficult than on Earth because of the more extreme environment, and because of the limit on power consumption. Human access to a space telescope for unexpected repairs or modifications is limited and expensive.

There is no question that with the advent of adaptive optics, ground-based telescopes of equivalent performance in the spectral bands accessible from Earth can be built and operated at lower cost, even including the laser beacons needed to obtain good sky coverage. They have better accessibility, and can be programmed or adapted quickly for special tasks or unexpected observations. The instrumentation can also be maintained and updated more readily. It is likely that ground-based telescopes will always be larger than space-based instruments, giving them an edge at IR wavelengths and in observations where the number of photons is important.

Space telescopes will always be needed for observations in spectral bands at which the atmosphere is opaque, mainly below 0.3 μm. At visible and near-IR wavelengths, space telescopes are free of the atmospheric airglow emission from OH molecules between 0.5 and 2.5 μm. A promising application for instruments in space is long-baseline interferometry, where high angular resolution can be obtained with relatively small optical apertures. In general, it is expected that the inherently higher cost of space telescopes will limit their application to specialized observations that are not possible from the ground.

The comparison can be summarized as follows:

- Atmospheric transmission cuts off at 0.3 μm, so that only space telescopes can work at shorter UV wavelengths.
- At visible wavelengths, ground-based and space observations are complementary. The Hubble Space Telescope (HST) is able to provide 0.05-arc-second resolution over a field of 2.6 arc minutes, with high contrast. Ground-based telescopes with adaptive optics are capable of higher resolution (0.015 arc second), but only over a much smaller field of view (~ 5 arc seconds), because of anisoplanatism. For seeing-limited observations not requiring adaptive optics, the field of view for ground-based telescopes is on the order of 1°, giving them the capability for covering large areas of the sky.
- In the near IR, large ground-based telescopes with adaptive optics have an advantage in both resolution and light-gathering power over the HST, with atmospheric radiation from water vapor being the only problem. The performance of ground-based telescopes at 10 μm and beyond is almost unaffected by the atmosphere.

2.4 Performance Cain with Adaptive Optics

The expected gain in performance of a ground-based telescope system equipped with adaptive optics will now be determined, and the parameters on which it depends will be identified. Two factors are of particular interest: the improvement in angular resolution and the increase in signal-to-noise ratio for faint sources in the background-limited case. The first of these defines the gain in spatial resolution for direct imaging and the gain in spectroscopic resolution, while the second enables the benefits of adaptive optics to be expressed in terms of an increase in limiting magnitude or a reduction in exposure time. The effect of adaptive optics on image structure and resolution is summarized in section 2.4.1, and the improvement in image signal-to-noise ratio is analyzed in section 2.4.2. Adaptive optics is also essential in long-baseline interferometers, especially those using large collecting apertures, to enable their full performance to be realized. The performance gains are reviewed in section 2.4.5.

2.4.1 Image Characteristics

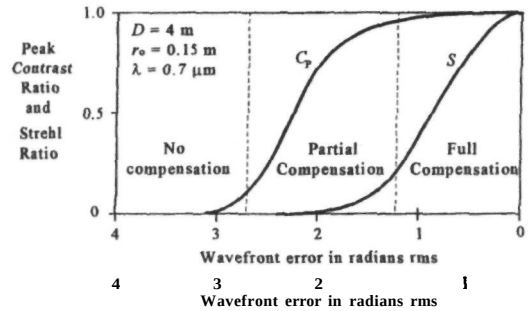
Full compensation of atmospheric turbulence with adaptive optics should improve the angular resolution of a telescope from its seeing-limited value determined by the turbulence parameter r_0, to its diffraction-limited value determined by the telescope aperture D. For astronomical applications, perfect image quality is not essential, and the amount of compensation is usually limited for economic reasons. Therefore, it is necessary to understand the effect of imperfect compensation on the structure of astronomical images. One might expect that the diameter of a compensated image would steadily decrease as the wavefront error is reduced. This is not the case: the shape of the image intensity distribution changes considerably as the turbulence is compensated. The following brief outline of optical image structure is based on the results of the detailed analysis given in chapter 4.

The image of a point source produced by a circular aperture consists of two basic components: a central core surrounded by a much larger halo. A perfect diffraction-limited system produces a central core of angular radius $1.22\lambda/D$, containing 84% of the light, surrounded by a series of diffraction rings. At the other extreme, a turbulence-degraded image consists of a large number of rapidly moving speckles, spread over a "seeing disk" with a radius of about λ/r_0. Each speckle has a diameter comparable to that of the diffraction-limited core. For exposures of more than a fraction of a second, the speckles blend into a continuous blur of light in which no central core is normally visible.

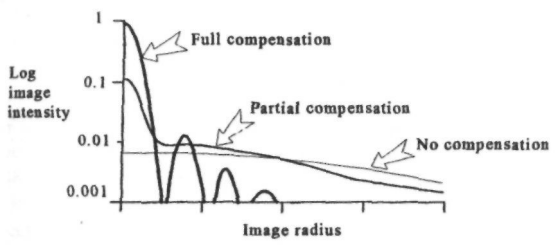
In the following discussion, it is assumed that the overall tilt component of atmospheric turbulence has

been compensated completely. A uniform wavefront tilt over the whole telescope aperture displaces the entire image. For very short exposures (less than about 1/50 s), it has little effect on image quality, but for long exposures, uncorrected random tilt excursions smear the image, with drastic effects on spatial resolution. Overall tilt is relatively easy to measure and correct with adaptive optics; the effects of residual tilt errors on image quality are treated in section 4.5.2.

When turbulence is compensated by adaptive optics, energy is transferred from the seeing disk or halo to the central core of the image, which starts to appear when the residual phase error in the wavefront has been reduced to about 2 rad rms. The process of wavefront compensation alters the relative intensities of core and halo, but does not significantly change their respective diameters. The peak intensity of the core, normalized to that of an unaberrated system, is known as the Strehl ratio. As shown in section 4.3.3, for random wavefront errors up to about 2 rad rms, the Strehl ratio is closely approximated by $S = \exp(-\sigma)$, where σ^2 is the mean-square wavefront error. Figure 2.5(a) shows how the Strehl ratio and the contrast ratio between core and halo vary as the turbulence is progressively compensated. The contrast ratio is defined as $C_p = S/(S + 2S_H)$, where S_H is the normalized peak intensity of the halo.



(a) Peak contrast and Strehl ratios of a turbulence-degraded point source as a function of the degree of compensation.



(b) Intensity profiles for three degrees of compensation.

Figure 2.5 Effects of wavefront compensation on the peak contrast ratio, Strehl ratio, and intensity profile of the turbulence-degraded image of a star.

There are three distinct zones of wavefront error:

- Uncompensated turbulence—both the contrast ratio C_P and the Strehl ratio 5 are near zero. The image is seeing-limited with no central core.
- Partial compensation—C_P rises sharply to a value near unity. The image has a distinct central core, with diameter close to the diffraction limit, but S remains below 0.1.
- Full compensation—S rises to near unity. The central diffraction-limited core is the dominant feature of the image.

Typical image profiles associated with each of these zones are shown in figure 2.5(b). The core retains its diffraction-limited diameter even when its normalized peak intensity (Strehl ratio) is less than 0.1, corresponding to a residual wavefront error of 1.6 rad(1/4 wave) rms.

Partial compensation of atmospheric turbulence can be achieved with much simpler (and lower cost) adaptive optics than is required for full compensation. It increases the contrast to the level at which image-processing algorithms can be used to restore the spatial information content of the picture. In principle, partial compensation allows the full diffraction limit of the telescope system to be achieved with indirect imaging. The main limitation is the presence of random noise in the data. Adaptive optics systems designed to provide full compensation at near-IR wavelengths usually give partial compensation in the visible band. Partial compensation is discussed further in sections 2.4.3 and 4.4.3.

2.4.2 Signal-to-Noise Ratio

The critical factor in detecting a faint star against the sky background is the signal-to-noise ratio, which depends on the number of photons received from the source and from the sky, within each measuring element of the detection system. The improvement in signal-to-noise ratio obtained by turbulence compensation depends on how this detector area is treated in the comparison. To simplify the analysis, wavelength effects and factors such as 1.22 and $\pi/4$ are ignored, as they do not change the basic results.

For a telescope of aperture D, the number of background photoelectrons detected within an angular area $a \times am$ time / is

$$p_B = N_B D^2 a^2 t\eta \qquad (2.1)$$

where

N_B = the sky radiance in photons per meter squared per second per steradian,
D = the aperture, m
a = the measuring element dimension in radians (the detector pixel size may be a submultiple of a)
t = the exposure time, s

η = the overall detection efficiency of the telescope

The number of detected signal photons in time t, from a point source of irradiance H_S photons per meter squared per second, is

$$p_S = H_S D^2 t\eta \tag{2.2}$$

In an uncompensated image, these photons are spread out into the "seeing disk" of angular diameter $2\lambda/r_0$, while for a compensated image they are contained mostly within the core angular diameter of 2A/*D*. The fraction of these photons detected within the a x a measuring element is defined as b.

The signal is equal to bp_S, the number of detected photons from the star, and the noise is the fluctuation in the value of the signal plus background $(bp_S + p_B)$. For large photon counts, the noise is equal to the square root of the signal plus background, giving

$$\begin{aligned} SNR &= \frac{bp_S}{(bp_S + p_B)^{1/2}} \\ &\approx \frac{bH_S D(t\eta)^{1/2}}{(bH_S + a^2 N_B)^{1/2}} \end{aligned} \tag{2.3}$$

For a bright star, $bH_S \gg a^2 N_B$, so that

$$(SNR)_{\text{bright}} \approx (bH_S)^{1/2} D(t\eta)^{1/2} \tag{2.4}$$

while for a faint star, $bH_S \ll a^2 N_B$, giving

$$(SNR)_{\text{faint}} \approx \frac{bH_S D(t\eta)^{1/2}}{a(N_B)^{1/2}} \tag{2.5}$$

For a bright star, $bH_S \gg a^2 N_B$, so that

$$(SNR)_{\text{bright}} \approx (bH_S)^{1/2} D(t\eta)^{1/2}$$

while for a faint star, $bH_S \ll a^2 N_B$, giving

$$(SNR)_{\text{faint}} \approx \frac{bH_S D(t\eta)^{1/2}}{a(N_B)^{1/2}}$$

The signal-to-noise ratio is proportional to $t^{1/2}$, so that, in principle, a star of any magnitude may be detected against any background, given a sufficiently long integration time. A signal-to-noise ratio of about 5 is required to detect the presence of an object in random noise. In practice, exposure times are limited and photon detectors have a finite storage capacity, so there is a de facto "limiting magnitude" for each system. This is not a hard number; one way to increase it is to make multiple exposures and combine the data, taking advantage of the fact that the coherent (signal) information in the image will add, while the random noise fluctuations will tend to cancel.

Most astronomical measurements involve faint sources for which background radiation sets the limit of detection. The signal-to-noise ratio for this situation is given in equation (2.5). To compare the signal-to-noise ratios of uncompensated and compensated images it is necessary to specify the relevant values of a and b. In a properly optimized system, the dimension a will match the size of the object to be detected and b will be near unity. Three cases of interest are depicted in figure 2.6.

In the first case, the size of the measuring element matches the uncompensated image and the same size is used for the compensated image, so that $a = 2\lambda/r_0$ and $b = 1$ for both conditions. This would occur if instrumentation that is optimized for normal seeing is used without modification when the telescope is compensated. In this case, adaptive optics does not improve the signal-to-noise ratio because all signal photons are collected, whether or not the image is compensated, and the background remains constant. This is a trivial case, but it draws attention to the fact that when a telescope is compensated with adaptive optics, the instruments must be upgraded also to take full advantage of the improved performance.

The second case represents a realistic situation in which the size of the measuring element is adjusted to match the image in each condition. For uncompensated operation, $a = 2\lambda/r_0$ and $b = 1$. Substituting these values in equation (2.5) gives

$$(SNR)_{\text{uncomp}} \approx H_S D \frac{r_0}{2\lambda} \left(\frac{t\eta}{N_B}\right)^{1/2} \tag{2.6}$$

To detect the faintest sources with a given exposure time, D and r_0 should be maximized and N_B minimized. This is the reason why astronomers use large telescopes situated on remote and arid mountain peaks, where they find good seeing (large r_0) and dark skies (small N_B).

For compensated operation, the value of a should match the diffraction-limited image diameter *2X/D*. The use of a smaller detector area for the compensated image reduces the background noise. If compensation were perfect, all the received photons would fall within the a x a area, but in practical adaptive optics systems the normalized peak intensity of the image (the Strehl ratio, *S)* falls between 0.1 and 0.8. The diameter of the core remains close to $2\lambda/D$ under these conditions, as shown in chapter 4. The value of b is then equal to the Strehl ratio 5, resulting in the following expression for compensated signal-to-noise ratio:

$$(SNR)_{\text{comp}} \approx H_S \frac{SD^2}{2\lambda} \left(\frac{t\eta}{N_B}\right)^{1/2} \tag{2.7}$$

The gain in signal-to-noise ratio due to adaptive optics for case 2 is then

$$G_{\text{ao}} = \frac{(SNR)_{\text{comp}}}{(SNR)_{\text{uncomp}}} = \frac{SD}{r_0} \tag{2.8}$$

This is an important result, the implications of which are discussed later.

In the third case, the size of the detection element is made equal to the diffraction limit for both uncompensated and compensated operation, so that *a = 2)./ D.* The value of b is then equal to $(r_0/D)^2$ for uncompensated operation, and is equal to *S* when the image is compensated. The ratio of the *SNRs* is then $S(D/r_0)^2$. This factor is the same as the increase in peak intensity of a point source due to turbulence compensation. This quadratic ratio does not represent the true improvement in signal-to-noise ratios, because it does not compare systems that are optimized individually for uncompensated and compensated operation. This is done in case 2, for which the correct ratio SD/r_0 is obtained.

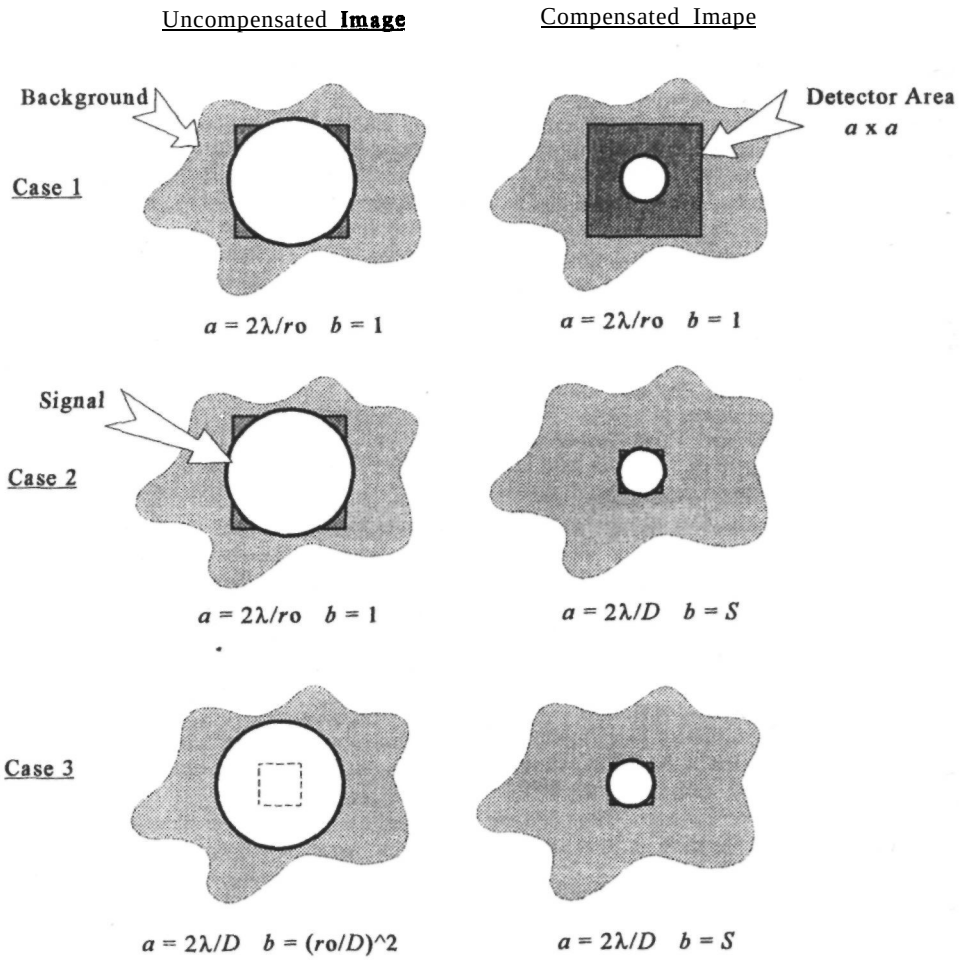


Figure 2.6 Calculation of signal-to-noise ratios for the detection of a star in the background-limited case. In case 1, the detector matches the (uncompensated) seeing disk. If the detector area is not reduced for the compensated image, there is no improvement in signal-to-noise ratio. In case 2, the detector area is optimized to match the compensated image. The signal-to-noise is improved by a factor of $S(D/r_0)$.In case 3, the detector area is constant at its diffraction-limited value. The signal-to-noise ratio increases by a factor of $S(D/r_0)^2$.

The gain factor $G_{ao} = SD/r_0$is the key quantity in evaluating performance improvements due to adaptive optics. For example, from equations (2.6) and (2.7) it is found that for the same signal-to-noise ratio, integration time, and background radiation, adaptive optics reduces the required star irradiance by $1/G_{ao}$, a stellar magnitude gain of 2.5 $\log_{10} G_{ao}$. The integration time required to achieve a given signal-to-noise ratio with a star of the same magnitude is reduced by $1/G_{ao}^2$.

Another way of assessing the benefit of adaptive optics is to determine the aperture of a compensated telescope that equals the performance of a larger uncompensated instrument. From equations (2.6) and (2.7), it is found that for equal signal-to-noise ratios,

$$\frac{D_{\text{comp}}}{D_{\text{uncomp}}} = \left(\frac{r_0}{SD_{\text{uncomp}}}\right)^{1/2} = \left(\frac{1}{G_{\text{ao}}}\right)^{1/2} \tag{2.9}$$

For example, if r_0 = 0.2 m and S = 0.4, then a compensated telescope of 2 m aperture will give the same signal-to-noise ratio (and have the same limiting magnitude) as an uncompensated 8-m telescope. The angular resolution of the smaller compensated telescope will be 10 times better than that of the larger uncompensated telescope (A/2 versus A/0.2). This is a convincing argument for the use of adaptive

optics in all ground-based astronomical telescopes, whether the aperture is large or small.

This gain directly improves the detectability of faint objects, as expressed by the limiting magnitude. Using a 19-actuator adaptive optics system on the ESO 3.5-m telescope, Peterson et al. [1991] obtained images with a limiting magnitude of 29 in the blue spectral band. This capability enables galaxies to be detected and counted at distances that have not been reached previously by ground-based telescopes.

2.4.3 Partial Compensation

Astronomical adaptive optics systems can be designed to compensate wavefront distortion to any required degree. As noted above, small amounts of phase distortion (up to about 2 rad rms) reduce the Strehl ratio but have little effect on the spatial resolution. The complexity of adaptive optics may be reduced by using partial compensation, in which the signal-to-noise ratio of the images is improved only to the level at which postdetection processing procedures, such as image deconvolution, can be used effectively. The advent of electronic imaging devices with a large dynamic range, together with digital processing, provides a great deal of flexibility in extracting the maximum amount of data from the raw image. Partial compensation is the result of using fewer subapertures than are normally employed, thereby reducing the number of wavefront sensor processing channels and deformable mirror actuators.

The effect of various degrees of wavefront compensation is illustrated in figure 2.7, which shows the basic structure of the modulation transfer functions (MTFs) and the corresponding image-plane point spread functions (PSFs). Wavefront distortion spreads the light that would normally be contained within an angular radius of $1.22\lambda/D$, the Airy disk, into a halo with radius on the order of λ/r_0, the seeing disk.

The effect of wavefront distortion on the MTF of an optical system is to reduce the magnitude of the diffraction-limited MTF, by a factor equal to the Strehl ratio, at all frequencies above r_0/λ, as shown in figure 2.7(a). If there is a residual response out to the cutoff of the full aperture, D/λ, then a core appears in the center of the image. The radius of this core is determined by the full telescope aperture.

The effect of partial correction is to restore some of the scattered light to the central core of the image. In zonal systems, partial compensation occurs when the correction elements in the deformable mirror are appreciably larger than r_0, as shown in figure 2.7(b). If d is the size of these elements, then the residual error after correction will be $\sigma_p^2 = a_F\ (d/r_0)^{5/3}$, where the parameter a_F depends on the influence function of the deformable mirror. The residual wavefront error variance σ_p^2 is reduced from its original uncorrected value, but is still significant. Again, the MTF consists of two parts, the low-frequency component cutting off at r_0/λ and the high-frequency component extending out to the telescope cutoff frequency, except that the latter component now has a higher amplitude because of the higher Strehl ratio. The corresponding point spread function has a central core, the intensity of which is proportional to the new Strehl ratio 5, superimposed on a halo with peak intensity proportional to $(1 - S)$. The result of partial compensation is to increase the intensity of the central core of the image in relation to that of the surrounding halo.

It is shown in section 4.6.2 that when $(D/r_0)^2 \gg 1$, which is the usual case for astronomical telescopes, the contrast ratio C_P between the core and the halo can be expressed as

$$C_P = \frac{S}{S + (r_0/D)^2} \tag{2.10}$$

This relation shows that large contrast ratios can be obtained even when the Strehl ratio of the image is low. For example, if $D/r_0 = 20$ and the image is partially compensated to give a Strehl ratio of only $S = 0.02$, the contrast ratio obtained is $C_P = 0.89$. This contrast ratio will support considerable image enhancement, providing that the photon count is reasonably high (> 100). The image can be deconvolved using the response of a nearby (unresolved) star, or blind deconvolution can be used based on the calculated system response.

Partial compensation provides a useful improvement in image signal-to-noise ratio when isolated point sources are being observed. In the case of closely spaced multiple objects, or extended objects, the haloes of adjacent image points may overlap, reducing the image contrast. Partial compensation is not appropriate for all astronomical instruments. In spectrographs, it is necessary to optimize the power passing through the slit, in which case a higher degree of wavefront compensation is required.

2.4.4 Wavelength Scaling

The observing wavelength has a profound effect on the image degradation produced by turbulence, and also on the design of the adaptive optics systems used to overcome it. The dispersion of air is very small, so the refractive-index changes produced by turbulence are almost independent of wavelength over the UV, visible, and IR spectral bands between 0.36 and $10\,\mu$m. As a result, the disturbances in optical pathlength are also essentially constant with wavelength. Images are formed by the interference of light waves, which depends on their relative optical phase. A given optical pathlength error due to turbulence therefore produces smaller phase shifts at longer wavelengths, reducing its effect on optical images.

Many optical effects of turbulence depend on the coherence length r_0, which varies as the 6/5 power of

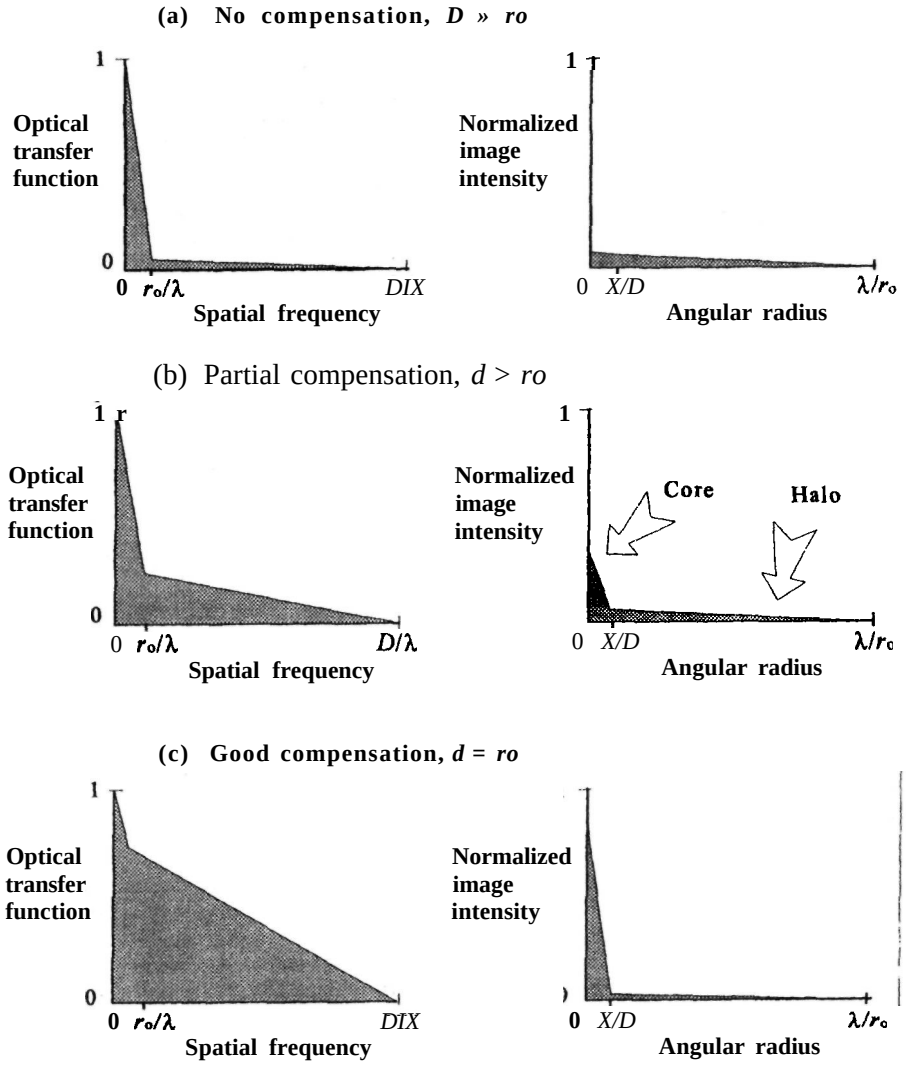


Figure 2.7 Partial wavefront compensation. This chart illustrates the effect of increasing the degree of wavefront compensation on the optical transfer function and point spread function of an astronomical telescope. Compensating the wavefront transfers energy from the halo (radius λ/r_0) to the core of the image (radius λ/D), changing the relative peak intensitites but having little effect on their radii.

wavelength. The angular diameter of an uncompensated image is proportional to λ/r_0, independent of the size of the telescope, so its wavelength depen-

image is proportional to λ/D and therefore increases directly with X. The result of these scaling factors is shown in figure 2.8, which compares the angular diameter of turbulence-degraded images with that of diffraction-limited images produced by 4-m and 8-m telescopes. At visible wavelengths, there is a large gap between uncompensated and compensated performance. At a wavelength of $10\,\mu$m, a 4-m telescope gives diffraction-limited performance in average turbulence, without any compensation. The results are summarized as follows: to obtain the best angular resolution without compensation, the longest observing wavelengths should be used. To obtain the best angular resolution with a perfectly compensated telescope, the shortest wavelengths should be used. Unfortunately, but not surprisingly, turbulence compensation is more difficult to implement at short wavelengths.

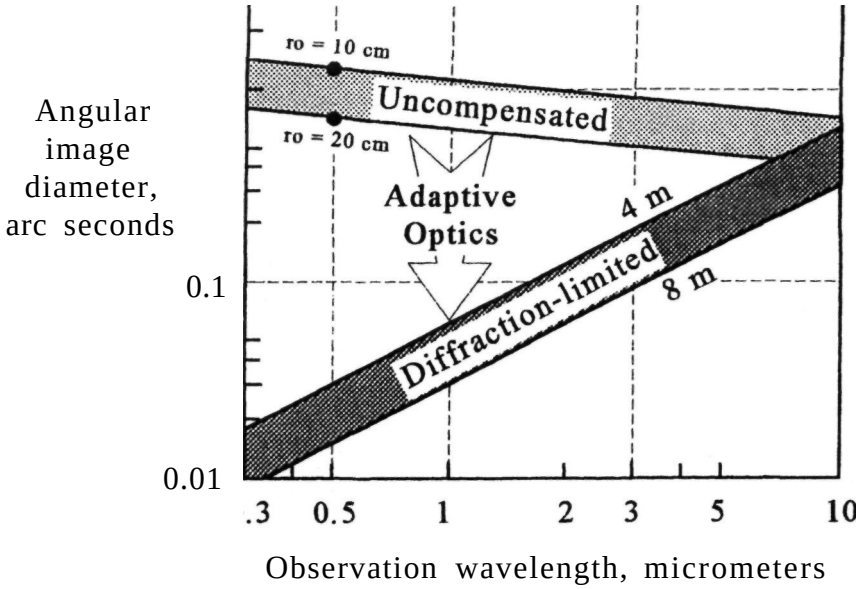


Figure 2.8 Wavelength scaling effects in optical images. This chart compares the image diameters of turbulence-degraded images with those produced by a diffraction-limited (compensated) aperture. The potential improvement due to adaptive optics is much greater at shorter wavelengths. At $10\mu m$, even large uncompensated apertures are close to diffraction-limited.

Considerable success has been obtained with adaptive optics of limited capability in the IR bands at 1.6 and 2.2 μm. At these wavelengths, the turbulence is relatively benign, with r_0 values of between 0.5 and 1 m. Compensation of a 4-m aperture can then be achieved with a corrector having only 16 actuators (degrees of freedom), which is relatively easy to implement. The angular resolution is improved by a factor of four to eight. At visible wavelengths, the potential improvement is far greater—a factor of 20 to 40.

2.4.5 Long-Baseline Interferometry

The performance gain achieved using adaptive optics with long-baseline interferometers is discussed in this section. Roddier and Lena [1984a] derived expressions for the signal-to-noise ratio of the fringe contrast in Michelson interferometers, operating at visible wavelengths (< $0.9\,\mu m$), in two operating modes: (1) visual detection of fringes and (2) photoelectric fringe-tracking. Their results also allow the performance to be evaluated in a third mode: (3) full wavefront compensation using adaptive optics.

1. This case represents the simplest type of interferometer, in which the average optical pathlengths are adjusted for equality. Short-term fluctuations between the apertures, due to turbulence, are not compensated in real time, resulting in relatively large pathlength errors. To maintain fringe contrast, a narrow spectral bandwidth is necessary. This bandwidth is proportional to $\lambda(r_0/L)^{5/6}$, where λ is the mean observation wavelength, r_0 is the coherence length of the turbulence, and L is the interferometer baseline. Without compensation of the collecting apertures, the resulting image is a speckle pattern, with interference fringes superimposed on each speckle. (The classic Michelson interferometer uses small apertures of size comparable to r_0, which produce an image containing a single "speckle".) The coherence area of the apertures is proportional to r_0^2. For small photon fluxes, the signal-to-noise ratio of the fringe contrast is

$$SNR_1 \propto \frac{\lambda}{v}\frac{r_0^{23/6}}{L^{5/6}} \tag{2.11}$$

The signal-to-noise ratio is extremely sensitive to the seeing conditions, varying almost as the fourth power of r_0. In this mode, there is no advantage in using an aperture larger than r_0. For the specific case of an interferometer with baseline L = 100 m, observing at a wavelength of $0.5\,\mu m$, with r_0 = 0.15 m, wind speed of $v = 5\,m\,s^{-1}$, overall optical transmission of 0.25, and a detector with quantum efficiency of 0.8, the limiting magnitude at which a signal-to-noise ratio of 5 can be achieved is about $m_V = 4$.

2. When photoelectric fringe-tracking is used, the optical pathlength variations over the baseline are compensated in real time, so the relevant coherence length becomes the aperture diameter D. Consequently, a larger spectral band can be used. For small photon fluxes, the fringe signal-to-noise ratio in this case is

$$SNR_2 \propto \frac{\lambda}{v} D^{1/6} r_0^{17/6} \tag{2.12}$$

Fringe-tracking greatly improves the signal-to-noise ratio, making it independent of the baseline length. The signal-to-noise ratio still has a high dependence on r_0, but is only weakly dependent on the aperture size D. The sensitivity of the interferometer is considerably improved. For the same conditions as the previous case, with $D \approx 1.5$ m, the limiting magnitude for a fringe signal-to-noise ratio of 5 is about $m_V = 11$, a gain of seven magnitudes.

3. The third case considered is that of an interferometer with fringe-tracking, in which the collecting apertures are fully compensated with adaptive optics. The spectral band is no longer limited by the coherence length of the turbulence, further improving the sensitivity. The signal-to-noise ratio for small photon fluxes is then

$$SNR_3 \propto \frac{\lambda}{v} D^2 r_0 \tag{2.13}$$

The factor r_0/v determines the exposure time, which must remain short in order to stabilize the fringes. For the standard set of conditions with $D = 1.5$ m and a signal-to-noise ratio of 5, the adaptive optics results in the limiting visual magnitude increasing to about 15, a gain of four magnitudes compared with an interferometer with fringe-tracking alone. The gains are even larger for bigger apertures. With interferometer apertures of 8 m, the adaptive optics gain over pure fringe-tracking for the same conditions is about six magnitudes.

2.4.6 Summary

Compensating atmospheric turbulence with adaptive optics achieves the following performance gains.

Direct Imaging

- The diameter of an image of a point source is reduced to approximately the diffraction limit of the telescope, improving (reducing) the angular resolution by a factor of r_0/D.
- The peak intensity of the image of a point source is increased by the factor $S(D/r_0)^2$.
- The signal-to-noise ratio of a point source with respect to the sky background is increased by the factor SD/r_0 compared with its uncompensated value, using an optimized detection system in each case.
- The peak contrast ratio of the image core to the halo is increased from its uncompensated value of near zero to approximately $S/[S + (r_0/D)^2]$.
- Adaptive optics systems produce higher Strehl ratios at longer wavelengths, because the turbulence coherence length r_0 increases as the 6/5 power of wavelength.
- The potential improvement in performance compared with the diffraction limit is much larger at short wavelengths.

Long-Baseline Interferometry

- Fringe-tracking and automatic pathlength compensation allow longer integration times, increasing the measurement precision and/or limiting magnitude. The improvement over visual fringe detection is about seven magnitudes, weakly dependent on aperture size.
- Wavefront compensation with adaptive optics allows apertures of any size to be used efficiently. With 1.5 m apertures at visible wavelengths, the combination of adaptive optics and fringe-tracking improves the interferometer sensitivity by about 11 magnitudes over a basic visual system. while for 8 m apertures, the improvement is about 13 magnitudes.
- The signal-to-noise ratio of interferometric imaging is extremely sensitive to seeing conditions, which suggests that partial compensation may produce a useful improvement in the performance of optical interferometers.

2.5 Optical Considerations

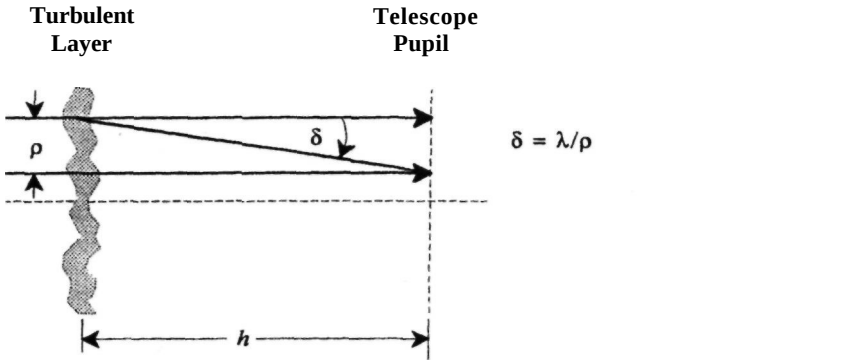
In this section, some basic issues regarding the optical systems employed in wavefront compensation are discussed. They include the validity of geometrical optics, the compensation of extended objects, the use of extended reference sources, and methods for compensating distributed turbulence. These topics have received scant treatment in the literature on adaptive optics, the last item being particularly neglected.

2.5.1 Validity of Geometric Optics

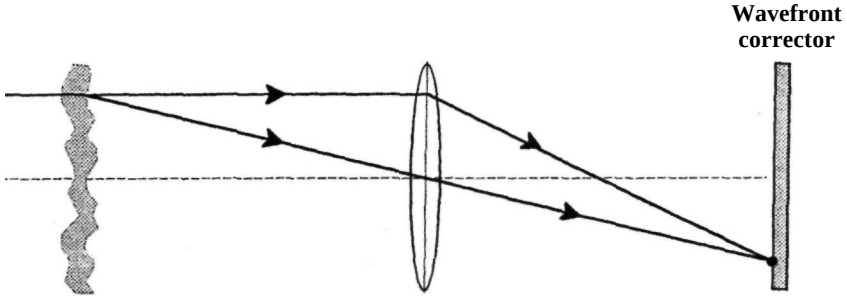
Conventional adaptive optics is based on the assumptions of geometrical optics, in which wavefront disturbances are summed along ray paths and diffraction effects are ignored. Only the optical pathlength disturbances (phase errors) are compensated. The geometric optics approach is generally valid when diffraction is insufficient to produce significant intensity changes or "scintillation" at the telescope,

which is often the case for the weak turbulence encountered at good astronomical sites. The conditions under which diffraction may be neglected can be derived in the following way. In figure 2.9(a), light incident on **a** region of turbulence is diffracted at an angle δ equal to λ/ρ, where X is the mean wavelength of the light and p is the characteristic size of the turbulent eddies. Diffracted rays will interfere when they have traveled **a** distance comparable to $h \approx p/S = \rho^2/\lambda$. If the propagation distance from a turbulent layer to the telescope is less than h, then interference effects may be neglected. Taking the specific example of a turbulent layer at the tropopause at 10-km altitude, with a scale size of 10 cm at a wavelength of 0.5 μm, the critical distance h is 20 km. This is much greater than the height of the layer above the telescope (especially for a mountain-top observatory), so the geometric optics approximation may safely be used. Under these conditions, the optical pathlength disturbances in light rays emanating from a single point source can be compensated by placing the wavefront corrector at any location in the ray path.

If the turbulence becomes more severe and the scale size decreases to 5 cm, the distance h drops to 5 km, in which case interference effects and scintillation become apparent at the telescope aperture. Diffracted rays are displaced by a distance comparable to p, producing wavefront phase and intensity variations that cannot be completely removed by inserting a simple corrector near the telescope aperture. The solution is to place the corrector at a conjugate image of the turbulent layer, as shown in figure 2.9(b). This arrangement restores the spatial relationship between the turbulence and the compensator, even when the turbulence is strong enough to produce scintillation. In other words, *conjugate imaging maintains the validity of geometric optics for wavefront compensation.* This principle also applies to the compensation of multiple layers of turbulence, providing that each layer is conjugated independently. Merely placing the correctors in the same (or reverse) order in which they occur in the atmospheric path does not fulfill this requirement. Optical systems necessary to implement multiple-layer compensation are discussed in sections 2.5.4 and 2.5.5.



(a) Diffracted rays interfere at the pupil when $h > p/S = \rho.\rho/\lambda$, causing scintillation.



(b) Diffracted rays are perfectly compensated when the wavefront corrector is located at a conjugate image of the turbulent layer.

Figure 2.9 Condition for compensation of a single turbulent layer. Placing the corrector at a conjugate image of the layer compensates turbulence strong enough to produce scintillation at the pupil.

Conjugate imaging is also necessary when compensating extended objects, as discussed in the following section.

2.5.2 Compensation of Extended Objects

It could be said that any object worth looking at must contain more than a single point, so the compensation of extended objects is fundamental to astronomical imaging. Although much information is obtained from unresolvable objects using photometry and spectroscopy, the increased angular resolution obtainable with large telescopes compensated with adaptive optics suggests that more and more celestial objects will become resolvable in the future. There are two main cases of interest:

1. Compensation of an extended object using a reference source that is separable from the science object itself. In this case, the reference may be either a point or an extended source, and the science object plays no role in the wavefront measurement process.
2. Compensation of an extended object using the radiation from the science object itself as the reference. In this case, the "wavefront" information must be separated from the "object" information.

The first case is addressed in this section and the second case in section 2.5.3.

The wavefront measured with a point reference source is valid only for an object in exactly the same direction as the reference. If the turbulence happened to be concentrated in a thin layer at the entrance pupil of the telescope, then the wavefront error could be compensated completely at all field angles by placing a corrector at the pupil. In reality, turbulence is distributed along the propagation path through the atmosphere, with the result that the wavefront error becomes decorrelated as the field angle increases. A single corrector compensates the wavefront measured at one field angle, so that only a small area of an extended object is compensated properly. This effect is known as angular anisoplanatism and is one of the major limitations to astronomical adaptive optics.

The wavefront error produced by angular anisoplanatism at a field angle θ is conveniently expressed as $\sigma^2 = (\theta/\theta_0)^{5/3}$, where θ_0 is the isoplanatic angle, a property of the turbulence distribution. This relation implicitly assumes that the wavefront corrector is located at the entrance pupil of the telescope (or at a conjugate image plane), which may not be the case. The anisoplanatic wavefront error that actually occurs in an adaptive optics system depends not only on the atmosphere but also on the configuration of the adaptive optics, specifically the location of the wavefront corrector(s) in relation to the turbulent layers in the atmosphere. The anisoplanatic error can be minimized by appropriate optical design.

The geometry for compensation of a single layer of turbulence is shown in figure 2.10. Light from the reference source *RS* passes through a turbulent layer T at distance h from the telescope entrance pupil P. Consider, first, a corrector C1 located at an arbitrary distance s from the pupil. To compensate the turbulence, the wavefront corrector inserts complementary pathlength corrections at each zone in the aperture. A ray from the reference source traversing the turbulent layer at J is compensated at zone K on the corrector. To compensate weak turbulence with a single source, the corrector may be placed at almost any location in the optical system, providing that it is properly scaled to the beam diameter. It cannot be placed near the prime focus ($s \approx f$), because in this position the scale size tends to zero and spatial correspondence with the pupil is lost.

Now consider an off-axis ray passing through the same point J in the turbulent layer, where the off-axis

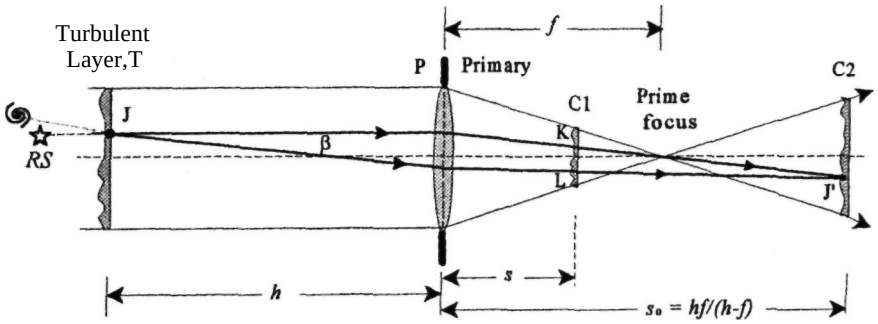


Figure 2.10 Geometry for compensation of a single layer of turbulence. If the corrector is located at C1, the isoplanatic angle is small and intensity variations (scintillation) due to diffraction are not compensated. By placing the corrector at the conjugate image plane C2, the compensated field is greatly enlarged and intensity variations are suppressed.

angle $\beta = \tan^{-} y/h \approx y/h$, where y is the offset at the telescope pupil. The ray passes through the corrector at point L, which is offset by distance y' from K. The off-axis ray will therefore be improperly corrected because it, too, should pass through point K. With the geometry shown, the position error of the ray at corrector C1 is

$$y' = \left(1 - \frac{s}{s_0}\right) = \beta h\left[1 - \frac{s}{h}\left(\frac{h-f}{f}\right)\right] \tag{2.14}$$

The layer T is imaged at distance $s_0 = hf/(h-f)$, where f is the primary focal length. When the corrector is located at position C2, at the conjugate image plane of T, then $s = s_0$ and the offset between the rays disappears. The turbulence is then compensated perfectly over a large field angle β. This condition is also shown in figure 2.10.

It is useful to define an effective isoplanatic angle θ_E that takes account of the configuration of the adaptive optics. As noted above, the canonical isoplanatic angle θ_0, which is a property of the turbulence, assumes that the corrector is located at the telescope pupil; that is, $s = 0$. For a single turbulent layer, the value of θ_E approaches infinity when $s = s_0$. It is of interest to determine how the effective isoplanatic angle θ_E varies with the corrector location; for instance, how accurately must the conjugate relation be satisfied in order to obtain a useful isoplanatic angle?

The effect of varying the corrector location on the isoplanatic angle may be determined as follows. Two rays of angular separation fi passing through one element of the turbulent layer produce a displacement of y' at the wavefront corrector. This displacement varies linearly with s, according to equation (2.14). The scale size of the wavefront corrector also changes with s, but in a different way. The diameter of the corrector is nominally equal to the telescope beam size, which (for an object at infinity) comes to a focus at distance / Near focus, the geometrical relation between the ray positions in the pupil and the beam break down because of diffraction, so this is not a suitable location for a wavefront corrector. The important point is that the diameter of the corrector varies linearly with its displacement from the prime focus, whereas the lateral displacement due to a ray separation of fi varies linearly with its displacement from the conjugate image s_0. The fractional displacement within the corrector is therefore a nonlinear function of s.

The fractional displacement of two rays at angle β, referred to the telescope pupil, is

$$y_P = y'\left(\frac{f}{f-s}\right) \tag{2.15}$$

The corresponding angle in object space is

$$\beta_P = \frac{y_P}{h} = \beta\left[1 + \frac{s}{h}\left(\frac{f}{f-s}\right)\right] \tag{2.16}$$

The effective isoplanatic angle θ_E, which takes into account the location of the wavefront corrector, is then

$$\theta_E = \theta_0\left(\frac{\beta}{\beta_P}\right) = \theta_0\left[1 + \frac{s}{h}\left(\frac{f}{f-s}\right)\right]^{-1} \tag{2.17}$$

When 5 = 0, the corrector is located in the telescope pupil and $\theta_E = \theta_0$, as expected.

If the normalized separation is defined as $s' = s/f$, then the ratio of the isoplanatic angles may be expressed as

$$\frac{\theta_E}{\theta_0} = \left[1 + \frac{f}{h}\left(\frac{s'}{1-s'}\right)\right]^{-1} \tag{2.18}$$

This ratio is plotted as a function of s' in figure 2.11 for a layer height h = 3000 m and a primary focal length f = 16 m. It is seen that the isoplanatic angle ratio drops to a sharp minimum when s' is near unity (at which value the corrector would be impossibly small) and rises to a sharp peak when $s' = h/(h-f)$. The latter value corresponds to the condition in which the turbulent layer at height h is imaged by the primary mirror onto the wavefront corrector, thereby fulfilling the conjugate imaging requirement. To take full advantage of this condition, the layer height must be known accurately. If the corrector is set up for a layer at height h, but the true height is h', then the isoplanatic angle ratio is

$$\frac{\theta_E}{\theta_0} = \frac{h'}{h'-h} = 1 + \frac{1}{\delta} \tag{2.19}$$

where S is the fractional error in determining the layer height. For example, to realize a ratio of $\theta_E/\theta_0 = 100$, the error in determining layer height must not exceed ±1%.

The isoplanatic angle ratio θ_E/θ_0 is independent of r_0 and wavelength, but the actual value of θ_0 is wavelength dependent, as it depends on r_0. A useful approximation to the isoplanatic angle for a single layer of turbulence is $\theta_0 = 0.31\ r_0/h$ (see section 3.7.2).

The above analysis shows that with a single turbulent layer, the effective isoplanatic angle θ_E is critically dependent on the placement of the corrector. This raises the question of whether the use of adaptive secondary mirrors materially affects the size of the isoplanatic angle. The use of adaptive secondary mirrors is an attractive method of simplifying adaptive optics for Cassegrain and Gregorian telescopes; the secondary mirror also functions as the wavefront corrector. The location of the corrector is therefore determined by the telescope configuration and is not a free parameter in the design of the adaptive optics.

Typical optical systems, unfolded for clarity, are shown in figure 2.12. In the generic Cassegrain configuration, figure 2.12(a), the secondary mirror is located inside the prime focus, with a diameter d_S typically between 10% and 25% of the primary diameter D. The Gregorian layout, with the secondary mirror outside the prime focus, is shown in figure 2.12(b).

The isoplanatic angle resulting from the use of an adaptive secondary mirror is easily determined by reference to figure 2.11(a), in which lines are drawn at s' values of 0.85 and 1.15, corresponding, respectively, to typical primary-to-secondary separations for Cassegrain and Gregorian telescopes having primary focal lengths of 16 m. It is seen that the Gregorian configuration gives a slightly larger isoplanatic angle, but in neither case is the ratio much different from unity. In a Gregorian telescope with primary focal length of 16 m and normalized separation $s' = 1.15$, the secondary is conjugate to a plane located 123 m in front of the telescope, considerably nearer than the mean turbulence height.

The conclusion is that the use of adaptive secondary mirrors forces the isoplanatic angle to a value close to θ_0, which is usually very small. Because of the fixed locations of the secondary mirrors, the angle θ_E cannot be expanded without the use of additional correction components, which, in this case, would be self-defeating.

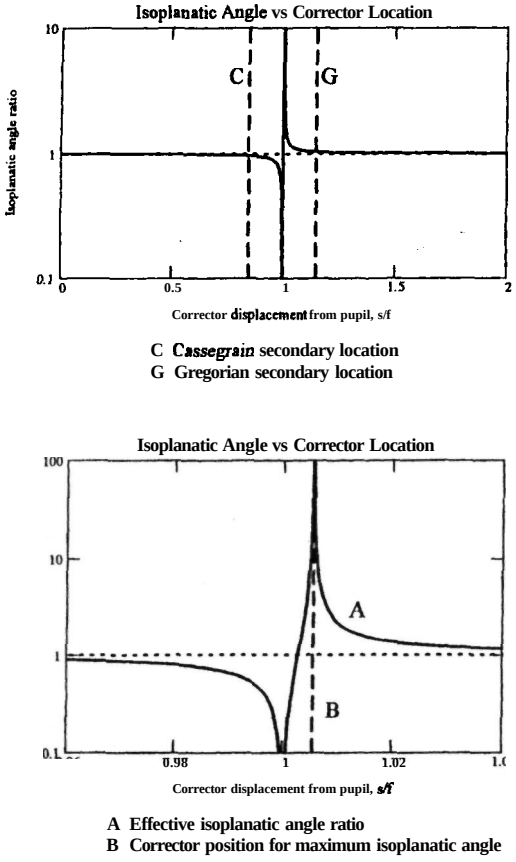


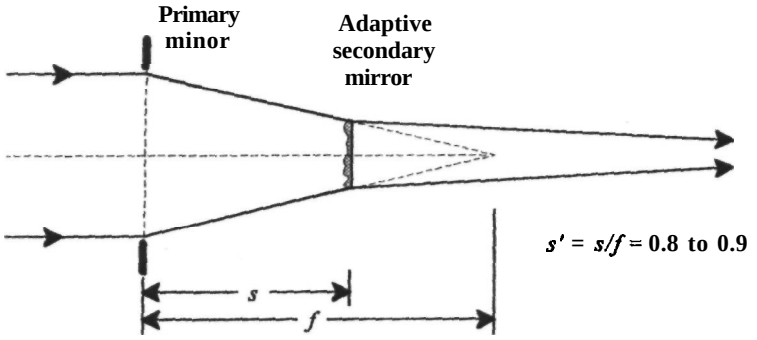
Figure 2.11 Effective isoplanatic angle of a compensated telescope, as a function of the corrector position, for a single layer of turbulence. The effective isoplanatic angle is normalized to that obtained with the compensator located at the entrance pupil of the telescope. The isoplanatic angle drops to zero when the corrector is located at the primary focus (in which case, there is no compensation) and is maximized when the corrector is located at the conjugate image of the turbulent layer. The focal length of the primary mirror is 16 m and the turbulent layer is 3000 m above the telescope.

2.5.3 Extended Reference Sources

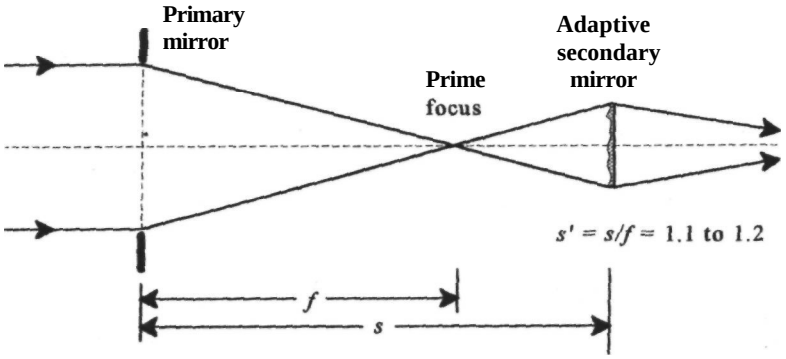
While it is clear that the wavefront from an object known to be a single point source can be compensated, it is not immediately obvious that the same technique can be used to compensate extended objects of arbitrary shape that are observed through a medium with randomly varying optical pathlength. In astronomical adaptive optics, few suitable point sources are available and reference sources are either laser beacons, which are typically 1 or 2 arc seconds in diameter, or small sections of the science object itself. To measure the wavefront, it is necessary to separate the intrinsic properties of the object from those of the intervening disturbance. The feasibility of this process and the limitations involved will now be discussed.

Astronomical objects, in general, may be modeled as arrays of incoherent point sources located at infinite distance from the telescope. The spatial information is contained in the two-dimensional intensity distribution of the object. The signal received from such objects consists of a number of superimposed, mutually incoherent wavefronts. In a perfect imaging system, each wavefront produces a point image; the intensities of these points add together to generate the two-dimensional image of the object. To form a faithful image of the source, the relative positions and intensities of each component must be preserved.

The wavefront sensors used in adaptive optics are designed to detect small variations in wavefront slope while being insensitive to the nature of the source itself. Two principles are used to attain this goal. The first principle is that the subaperture sizes employed for slope sensing are comparable to the turbulence coherence length r_0, for which the angular resolution λ/r_0 is about 1 arc second. Wavefront slope sensors are therefore blind to fine structure in the reference source. (Larger scale structure can influence the slope measurement capability, especially in Hartmann-type sensors.) The second principle is that variations in the overall brightness and size of the



(a) Cassegrain configuration, with elements unfolded for clarity



(b) Gregorian configuration, also with elements unfolded

Figure 2.12 Wavefront compensation with adaptive secondary mirrors. In the Cassegrain configuration, the effective isoplanatic angle is slightly smaller than that obtained with the corrector located conjugate to the primary mirror, whereas with the Gregorian configuration it is slightly larger. Because of the fixed location of the secondary mirrors, the compensated field cannot be widened without additional correction components.

reference source are minimized by "normalizing" the wavefront sensor output to eliminate offsets or bias. The effect of increasing the size or reducing the brightness of the reference source is mainly to increase the random noise on the wavefront measurements, rather than to produce systematic errors.

The first object type to be considered is one whose angular size is smaller than the isoplanatic angle of the turbulence. The radiation received from such an object is an ensemble of incoherent plane waves that all experience the same distortion. The correction of this distortion over the telescope aperture does not change the relative tilt angles or intensities associated with the structure of the object, so it is clear that the compensation process will not corrupt the image.

When the object size exceeds the isoplanatic angle, the situation is more complicated. This condition usually arises when turbulence is distributed along the propagation path, resulting in a small isoplanatic angle that cannot be expanded significantly with a single corrector. It is essential for the angular size of the reference source (or the area of the object used for reference) to be smaller than the isoplanatic angle, otherwise useful wavefront measurements cannot be made. This may require restricting the field of view of the wavefront sensor, especially with very large objects such as the Sun, the Moon or the planets. Within this field, the source object must have sufficient detail and contrast to allow precise slope measurements. For example, with solar adaptive optics, pores and small spots have been used successfully as lock points for the wavefront sensor [Acton and Smithson 1992].

A further requirement is that multiple reference sources are required to measure the wavefront errors over an object field greater than the isoplanatic angle,

with one source for each isoplanatic solid angle. In principle, the three-dimensional structure of distributed turbulence can be deduced from multiple source measurements, enabling multilayer wavefront correction to be implemented. Such systems have not yet been realized in practice.

A basic limitation arises with extended reference sources when a zonal wavefront slope sensor is used in high turbulence, in which case both the isoplanatic angle θ_0 and the turbulence coherence length r_0 have small values. It is necessary for the wavefront sensor subapertures to resolve the anisoplanatic angle, otherwise the slope measurements are meaningless because they involve more than one isoplanatic area. This requirement sets a lower limit on the subaperture size d, so that $d > \lambda/\theta_0$. On the other hand, accurate wavefront measurements require small subapertures. The wavefront fitting error of a zonal corrector, such as a deformable mirror, is given by $\sigma_F = a_F(d/r_0)^{5/6}$, where a_F is the fitting error coefficient. Substituting for d, it is found that there is a minimum wavefront measurement error given by ffmin $= a_F\ (\lambda/\theta_0 r_0)^{5/6}$. This error is imposed by the turbulence conditions, independent of the specific slope sensor employed. Under the benign turbulence conditions prevalent at good sites at night, it should not be a problem, but for solar observations at visible wavelengths during periods of high daytime turbulence, the minimum error may be on the order of 1 rad rms.

The general conditions for compensating an extended source may be summarized as follows:

1. The object must consist of incoherently radiating point sources.
2. The optical pathlength disturbance from each correctable point in the object to each independent zone of the telescope aperture must be known.

The first condition is satisfied by most of the objects observed in astronomy. The second condition appears to require a formidable number of optical path measurements, but, in practice, the number of independent measurements is drastically reduced by the following factors: (1) the finite size of the turbulence coherence length r_0 limits the number of independent zones in the aperture that must be compensated to about $(D/r_0)^2$; (2), the isoplanatic angle θ_0 limits the angular size of the object that can be compensated; (3) the angular resolution of the telescope λ/D sets an upper limit to the number of resolvable points in the object.

2.5.4 Distributed Turbulence

The combination of an extended object with distributed turbulence presents the most difficult compensation problem in adaptive optics. With distributed turbulence, a single wavefront corrector provides full compensation only over the isoplanatic angle θ_0, which may be only 2 or 3 arc seconds at visible wavelengths. To achieve full compensation over wide fields of view, it is necessary to employ three-dimensional wavefront compensation. In the following discussion, it is assumed that the turbulence distribution is known exactly, the problem being how to compensate it.

There are two approaches to wide-field compensation of distributed turbulence. One method is to separate the field of view into isoplanatic areas, using a separate compensator for each area. This approach requires a separate wavefront sensor and reference source for each isoplanatic area. Apart from its complexity, the system has the drawback that relative tilt errors between the multiple reference sources may cause mutual displacements between sections of the image. Such errors may be minimized by using a common tilt reference for the whole image, taking advantage of the fact that the tilt component of turbulence has a larger isoplanatic angle than the high-order components.

Another approach is to divide the continuous distribution of turbulence into slabs or layers, as shown in figure 2.13. If each layer could be compensated with a wavefront corrector in situ, as shown, then correction would be completely effective for any turbulence distribution and field angle. In practice, the turbulence can only be corrected after the light has been collected by the telescope. As a result, the light disturbed by layer 1 in figure 2.13 passes through layers 2 and 3 before it can be compensated. The sequence of correctors in the telescope optical path must therefore be 3, 2, 1. In addition, to achieve a useful compensated field of view, the conjugate rela-

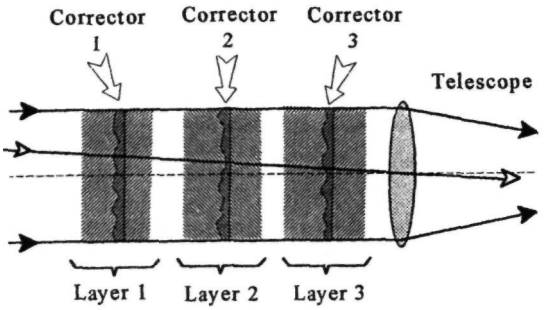


Figure 2.13 Compensation of multiple layers or distributed turbulence. If the correctors could be physically colocated with each layer as shown, compensation would be perfect, even for off-axis rays. In practice, the correctors must be located in the telescope, where layer 1 is seen through the distortion of layers 2 and 3. It is therefore necessary to compensate the layers in the reverse order; that is, 3, 2, 1.

tion must be maintained between the correctors and the layers.

Two separate requirements must therefore be satisfied to enlarge the compensated field of view using **multiconjugate** adaptive optics:

- Multiple layers must be corrected in an optical sequence determined by their distance from the telescope objective, the nearest layers being corrected first.
- Each corrector must be located at a conjugate image of the corresponding turbulent layer.

It might be expected that the required sequence of conjugate image planes would occur naturally in an optical imaging system, but, unfortunately, this is not the case. Optical images formed by a lens occur in the same order as the objects, which is the reverse of what is required. To satisfy all requirements, multilayer wavefront compensation consequently requires some ingenuity in the optical design. The payoff is that it can lead to a large increase in compensated field angle. Design principles for multiconjugate compensation systems are summarized in the following section.

2.5.5 Optical Configurations for Turbulence Compensation

In Earth's atmosphere, significant turbulence occurs at altitudes up to 20 km, often with large peaks in the vicinity of the tropopause, at around 10 km. This three-dimensional distribution of turbulence considerably restricts the angle over which adaptive compensation is effective. The angular extent of many objects of interest exceeds the isoplanatic angle, so the achievement of wide-field compensation is a major goal of astronomical adaptive optics. In this section, the optical configurations for multiple wavefront correctors are reviewed.

Because of the large aperture of most astronomical telescopes (2-10 m) and the relatively small size of practical wavefront compensators (0.01-0.3 m depending on the type), it is not feasible to place wavefront correctors in the telescope pupil; they must be located in a reduced beam, often near a conjugate image of the telescope entrance pupil. Most wavefront correctors consist of two parts: a two-axis tracking mirror to compensate overall tilt (image motion) and a device, such as a deformable mirror, to compensate the higher order distortions within the aperture.

Atmospheric turbulence generally has a three-dimensional distribution, although, at good sites, it may be confined to relatively few layers. Several workers have considered the possibility of using multiple compensators, such as deformable mirrors, each imaged on a specific atmospheric layer [Fried 1977, Beckers 1988, Johnston and Welsh 1994]. Exactly how this would be done has not been described. In fact, it is not a trivial problem; the optical systems needed to implement multiconjugate compensation are complicated.

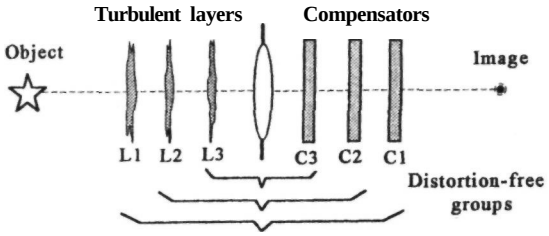


Figure 2.14 Symmetrical correction system for multiple turbulent layers. Although the layers are corrected in the right sequence, they are not imaged on the appropriate correctors, so that only axial rays are properly compensated. As a result, the compensated field angle is smaller than θ_0, making this approach useless for adaptive optics. A single compensator near the pupil would give better results.

The first matter to be discussed is the sequence of correction for multiple turbulent layers. Consider the arrangement shown in figure 2.14. Optical imaging is a symmetrical process in which a luminous object generates wavefronts that propagate in space, are collected by an aperture, and then have their curvature inverted by a lens or mirror to form an image of the source. The process is reciprocal in that the light follows the same path when propagating in either direction. Using a system of this type, the wavefront correctors C1, C2, and C3 could be located symmetrically with respect to the corresponding turbulent layers L1, L2, and L3, as shown. Each symmetrical group of layers, such as L3, P, and C3, would function as a distortion-free subsystem.

Because the phase disturbances are distributed randomly over the optical aperture, the correction is valid only with axial light. Off-axis rays are displaced in opposite directions on each side of the pupil, resulting in an isoplanatic angle smaller than the value of θ_0. The reason for compensating multiple turbulence planes is to enlarge the field of view, so this arrangement is useless. Better results would be obtained by combining the correctors into a single location near the pupil.

To enlarge the field of view with a single turbulent layer, the compensator must be placed at its conjugate image plane, as explained in section 2.5.4. With multiple layers, a telescope objective forms a sequence of images in the same order as the objects, as shown in figure 2.15. This system is also unsuitable for multiple-plane compensation because distortion in the intermediate layers spoils the conjugate imaging. For example, L1 would be imaged perfectly on C1 if no other layers were present, but, in fact, layers

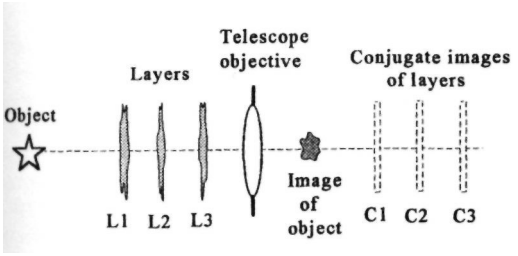


Figure 2.15 Location of images formed by a telescope objective. The conjugate images of turbulent layers naturally occur in the same order as the layers, which happens to be the wrong sequence for adaptive optics compensation.

L2 and L3 are intervened between L1 and the telescope, which invalidates the conjugate relationship.

How are these apparently conflicting requirements reconciled? The problem is resolved by separating the corrector planes with additional relay lenses, as shown in figure 2.16. In the first set of image planes, turbulent layer L3 (nearest to the telescope pupil) is imaged correctly and can be compensated directly by C3. Once this is done, the group L3, P, and C3, forms a perfect imaging system, enabling turbulent layer L2 to be compensated by C2 in the next relay section, and so on.

A further simplification is obtained from the fact that the reimaging requirements for turbulent layers are not very stringent and require only that the blur circle should be much smaller than the correlation distance r_0. As the effective f-number and field angles are very small, an appreciable out-of-focus distance is allowable. This enables the corrector for the lowest layer to be placed between the telescope primary and the prime focus, simplifying the optical system, as shown in figure 2.17.

When optical systems of this type are implemented, it is found that the physical size of the compensation elements varies greatly, because some are located near image planes where the beam diameter is very small while others are located near pupils where the beams are much larger. As a practical matter, it is desirable to use compensators of similar physical size. A. Walther [1974, private communication] has suggested adjusting the diameter of these components optically, by means of additional relay lenses. An example of the complete optical system for two-level correction using same-size compensators is shown in figure 2.18.

2.6 Implementation of Adaptive Optics

2.6.7 *Essential Functions*

Adaptive optics systems contain three essential functions: measurement of the wavefront error, computation of the corrections to be applied, and compensation of the wavefront itself. These functions can be arranged and implemented in many different ways, according to the specific application and system requirements. Most adaptive optics systems are built up using separate subsystems for wavefront sensing, data processing, and wavefront correction; these subsystems are connected together optically and electrically to form the complete wavefront compensation system. In addition to these basic functions, it is necessary to provide test and diagnostics capabilities for the adaptive optics, as well as an interface to communicate with the outside world. An efficient and user-friendly operating system is very desirable in such systems.

The functions of an astronomical adaptive optics system are typically arranged as shown in figure 2.19. Light waves from a distant reference source, initially undistorted, pass through the turbulent atmosphere and are collected by the telescope. The distorted optical beam is then reduced in diameter, passes through the wavefront corrector, and is sampled by a beam splitter, which sends a replica of the input wavefront to the wavefront sensor. The wavefront sensor measures the departure of the wavefront from a plane wave, usually by determining the local wavefront slope over an array of subapertures, thus producing electrical outputs corresponding to the measured optical errors. The data processor converts these error signals into electrical commands that are fed back to the wavefront corrector, thus closing the con-

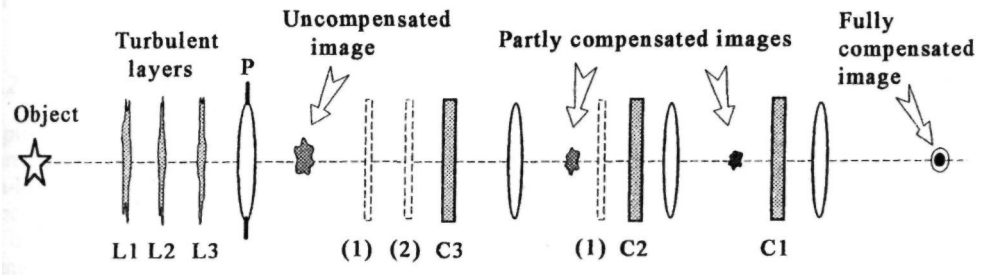


Figure 2.16 Optical reimaging system for compensation of multiple turbulent layers. This arrangement satisfies both the sequencing and the conjugate imaging requirements. Locations of intermediate images of layers L1 and L2 are shown as (1) and (2). Corrector locations are identified as C1, C2, and C3.

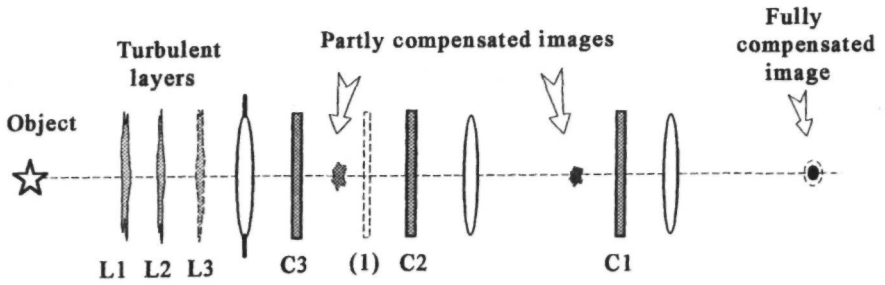


Figure 2.17 Simplified three-layer compensation system.

trol loop. The wavefront corrector is driven in such a way as to minimize the electrical errors at the output of the wavefront sensor; this is equivalent to nulling the optical wavefront error at the sampling point, the beam splitter. The light passing through the beam splitter forms the compensated optical output, which is focused onto a camera or other instrument.

The wavefront corrector usually consists of two components: a flat tracking mirror to remove image motion and a deformable mirror to compensate for defocus, astigmatism, and higher-order distortions. Tracking mirrors are usually several centimeters in diameter, with the capability of tilting in two orthogonal axes, at frequencies of up to about 100 Hz, using electromagnetic or piezoelectric actuators. Deformable mirrors employ a thin faceplate that ranges in size from about 5-cm diameter, for a small device with 21 actuators, up to about 40-cm diameter for the largest type with 2000 actuators. The actuators are arranged in a square or hexagonal array and are capable of displacing the mirror surface locally (within each zone) by a few micrometers up or down.

Wavefront sensors for astronomical adaptive optics generally measure local wavefront tilts or gradients within an array of contiguous zones or subapertures filling the telescope pupil. They operate with broadband light at visible or near-IR wavelengths and will accommodate natural stars, extended objects, or laser beacons as reference sources. Because of the limited brightness of the reference sources and the short time available for wavefront measurement, the efficiency of wavefront sensors in using the available photon flux from the reference source is of critical importance. The photodetector is usually the dominant component in this respect and low-noise CCD arrays are currently favored.

The computation required to convert the wavefront gradient data delivered by the wavefront sensor into commands for the wavefront corrector is known as wavefront reconstruction. The reconstruction process restores to each zone of the aperture the relative phase values that are lost when wavefront gradients are measured. This is a vital step in the operation of an adaptive optics system, because phase errors over the aperture have a far greater effect on the image quality than do local wavefront slopes. Reconstruction is a global process; that is, a change in any single slope measurement involves recalculation of all of the drive signals.

Wavefront gradient measurements are usually made simultaneously over the whole aperture and

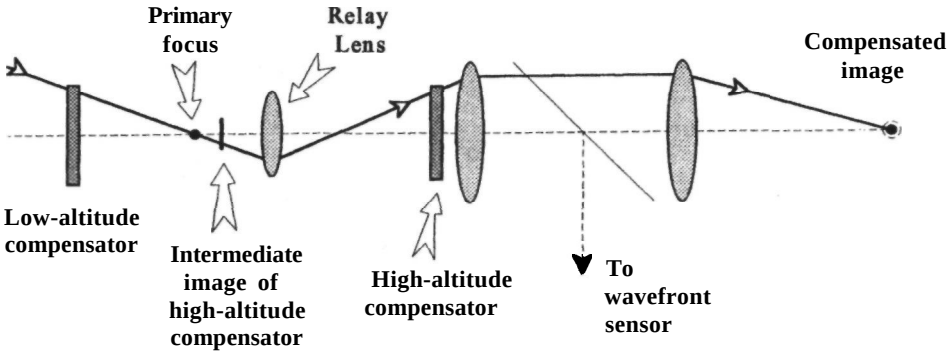


Figure 2.18 Practical compensation system for two layers using same-size correction devices. The intermediate image of the high-altitude layer is located near the primary focus and is consequently very small. An additional relay lens is inserted to magnify the image of this layer to a size similar to that of the low-altitude corrector.

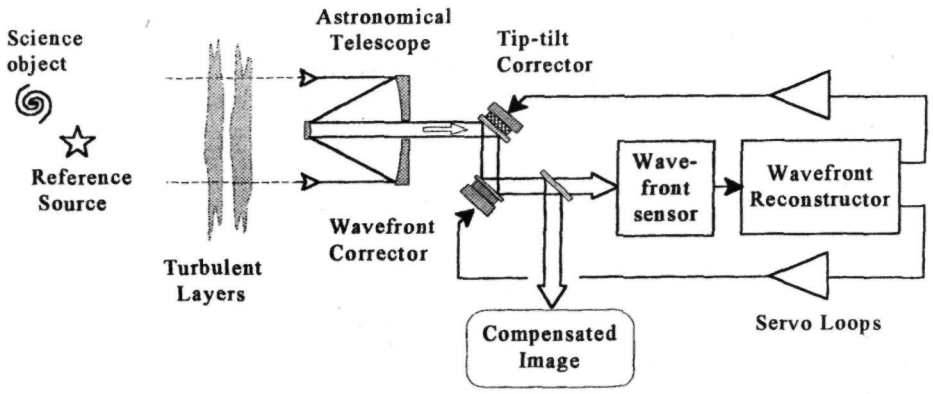


Figure 2.19 Components of an astronomical adaptive optics system.

are reconstructed and applied to the correction mirrors within a short time, termed the latency. The latency must be less than the change time of the wavefront being compensated; for atmospheric turbulence, this is usually only a few milliseconds. The amount of data that must be processed in each cycle of operation depends on the number of subapertures being compensated. In a zonal adaptive optics system with N correction channels, the number of multiply-add operations per cycle is $2N^2$, which produces a heavy data-processing load. Parallel processors are generally employed to solve this problem.

Atmospheric conditions are continually varying, resulting in changes in key parameters such as the turbulence coherence length r_0, the turbulence change time t_0, and the isoplanatic angle θ_0. Turbulence exhibits some fractal properties in that fluctuations occur at many different scale sizes, both in time and in amplitude. Adaptive optics systems are optimized to maximize performance for a set of specific operating conditions, and the performance depends critically on matching internal parameters, such as wavefront sensor integration time, to the external conditions. It is therefore important to design astronomical adaptive optics systems to accommodate these variations in turbulence characteristics, preferably on an automatic basis using adaptively controlled parameters within the adaptive optics system itself.

2.6.2 Parallel Operation

Adaptive optics systems may be configured in several different ways, using serial or parallel operation in open-loop or closed-loop configurations, with zonal or modal processing. The most successful architecture is the parallel closed-loop system using zonal sensing and compensation, a block diagram of which is shown in figure 2.20. This configuration was first used in the real-time atmospheric compensator (RTAC) system [Hardy et al. 1977]. The basic principle of whole-aperture wavefront measurement with multiple feedback paths to a zonal deformable mirror has stood the test of time and is used in most current systems. Parallel operation implies that the wavefront error over the whole pupil is measured during a single time interval, and that the wavefront correction is similarly updated over the whole aperture within a single interval. This mode of operation maximizes the time available for integration of light from the reference source, a vital factor in astronomical adaptive optics systems. Parallel operation can be employed with either zonal or modal sensing and compensation. The essence of parallel adaptive optics systems is the presence of multiple control paths between the wavefront sensor and the wavefront corrector. There is no limit to the number of parallel paths, so the largest telescopes can be compensated by adaptive optics without compromising the performance.

Most electronic devices used in adaptive optics, such as CCD detectors and electronic data processors, transfer data in the form of high-speed bit streams using pipelined operation. While this type of operation is not strictly speaking "parallel," its speed is such that the processing time for each data cycle is usually small compared with the time constant of the servo loop, so that the essence of parallel operation is maintained. In cases where the data processor is faster than the wavefront sensor due to the limited CCD pixel rate, the reconstructor can operate in the "delta" mode, in which the entire array of corrector drive signals is updated for each zonal measurement.

2.6.3 Serial Operation

The serial approach to adaptive optics is based on sequential measurement and correction of the wavefront errors in each zone or mode of the aperture. Serial operation simplifies the hardware required, but unavoidably restricts the time available during each

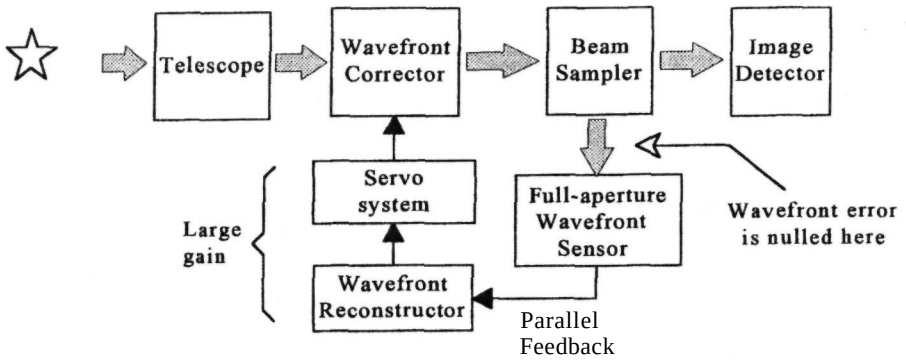


Figure 2.20 Closed-loop adaptive optics system. The parallel feedback loops have a large gain, minimizing the wavefront error at the wavefront sensor input. The null point of the wavefront sensor must be accurately defined, but variations in the gain of the feedback loop are relatively unimportant.

measurement for collecting photons from the reference source. Each cycle of operation, comprising wavefront measurement, computation, and correction over the whole telescope aperture, must be completed on a time scale much shorter than the change time of the turbulence. As a result, serial systems do not perform well with the low-brightness reference sources that are encountered in astronomical work. The time-sharing problem gets worse as the telescope aperture increases. In contrast, the parallel approach allocates a dedicated processing channel for each wavefront zone or mode so that the time available for wavefront measurement is independent of the number of actuators or telescope size.

Image sharpening is a type of serial operation that was used in the early days of adaptive optics. Its principle, shown in figure 2.21, is to divide the aperture into an array of zones, which are adjusted in sequence to minimize the overall wavefront error and thereby sharpen the image. An image-plane sensor is used, in which the reference source is focused onto a mask, which, in its simplest form, is a pinhole with a diameter of about half the Airy disk of the telescope imaging system. A photon detector located behind the mask measures the light flux passing through the pinhole, which is normalized to the total light received from the reference source. This flux is maximized when the wavefront is perfect. In operation, a small phase step is introduced sequentially into each zone of the aperture and its effect on the light flux is measured. If the flux increases, the phase step is retained; if not, it is removed and the system moves on to the next subaperture. The polarity of the trial steps is reversed for each cycle of operation. Three iterations are normally required to maintain diffraction-limited correction. This type of

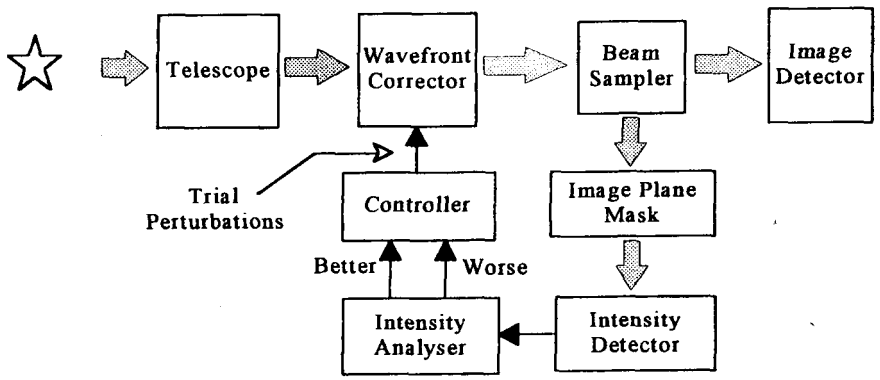


Figure 2.21 Image sharpening system. A small phase perturbation is applied sequentially to each zone of the deformable mirror. If the image intensity is improved, then the perturbation stays. If not, it is removed. A perturbation of opposite polarity is applied in the next cycle.

adaptive optics system requires a bright reference source and can support only a small number of subapertures.

The difference in performance between serial and parallel operation may be illustrated in the following way. For an adaptive optics system with N subapertures and a specified latency of t seconds, the time available for each wavefront measurement in a serial (image-sharpening) system is $t/3N$ seconds, allowing for three iterations per cycle. In a fully parallel system, approximately half of the latency is available for wavefront measurement, so the time available is $t/2$ seconds. To obtain the same signal-to-noise ratio, the parallel system needs a reference source with only 2/(3N) of the photon flux required by the image-sharpening system. If A' is 500, then the image-sharpening system would require a reference star about seven magnitudes brighter to achieve the same performance as a parallel adaptive optics system. This enormous difference in efficiency makes serial operation uncompetitive for astronomical adaptive optics.

2.6.4 Open-Loop and Closed-Loop Systems

Adaptive optics control systems may use either open-loop or closed-loop operation. In the open-loop or feed-forward configuration, shown in figure 2.22, the received wavefront is first measured in its raw or uncorrected state. The required corrections are then computed and fed to the wavefront compensation device. The time required for the measurement and correction process (the latency) must be less than the change time of the wavefront. Effective open-loop compensation depends on accurate wavefront measurement and correction; the wavefront sensor and deformable mirror must be calibrated accurately over the whole dynamic range of the (uncorrected) wavefront. Any error in the response of these components shows up directly as a wavefront error in the optical output. For this reason, open-loop systems are only used when the alternative closed-loop configuration is not feasible. Open-loop adaptive optics systems can operate either in the continuous mode or in the "one-shot" mode where only a single (whole-aperture) wavefront measurement can be made prior to wavefront compensation. This occurs, for example, in outgoing wave laser systems, where a point-ahead wavefront measurement must be made. Another application of open-loop adaptive optics, relevant to astronomical systems, is in laser beacon systems employing a pulsed laser, where the time separation between pulses is comparable to (or longer than) the change time of the atmosphere. In such cases, the adaptive optics must operate on a single-shot basis, using what is known as a "go-to" servo system, which is nulled after each cycle.

The closed-loop configuration, used in most adaptive optics systems, is shown in figure 2.20. In this case, the first component in the optical system is the wavefront corrector, which compensates the wavefront errors on the incoming beam before it is sampled and measured by the wavefront sensor. The wavefront sensor therefore sees only the residual error in the wavefront; that is, the difference between the current incoming wavefront and the last correction applied to the wavefront corrector. This error is processed to update the control signals applied to the wavefront corrector. A parallel system of this type contains a large number of feedback loops, one for each correction zone in the aperture. Feedback loops, which are used extensively in all kinds of automatic control systems, have several characteristics that make them especially useful for adaptive optics. Gain in the feedback loop drives the deformable mirror so as to force the residual error at the wavefront sensor output to zero. Because the wavefront sensor is only required to measure residual errors, its dynamic range and linearity are relatively unimportant; only the calibration of the null is critical. The wavefront sensor requirements for closed-loop adaptive optics systems, are therefore less exacting than

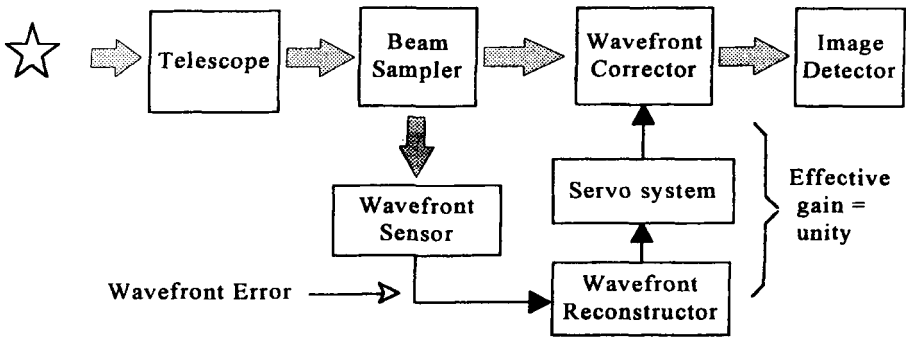


Figure 2.22 Open-loop adaptive optics system. The wavefront error must be measured exactly over its full range of variation. The accuracy of open-loop or "feed-forward" systems, such as this, depends on precise calibration of the correction loop.

for open-loop systems. In closed-loop systems, the residual error can be made arbitrarily small by increasing the loop gain. If the loop gain, a function of temporal frequency, is $G(\omega)$, then the errors detected by the wavefront sensor are reduced by the factor $1/[1 + G(\omega)]$. However, as the loop gain is increased, the phase shift around the loop also increases. The loop gain must be reduced to less than unity before the phase shift reaches 2π rad, otherwise oscillation sets in. This sets a limit to the amount of loop gain that can be used. Adaptive optics systems usually employ first-order servo loops containing a single integration, with a low-frequency gain of about 100. A model of the feedback loop and typical gain-versus-frequency plot are shown in figure 2.23.

The servo bandwidth, defined as the frequency at which the loop gain is reduced to unity, is typically on the order of 100 Hz for astronomical applications. To assure stability, a phase margin, typically 45°, is specified at the unity-gain bandwidth. The single integration produces a phase shift of 90°, so, with this phase margin, the allowable phase shift from all other sources at the unity-gain bandwidth is only 45°. To achieve stable operation, the temporal bandwidth of individual components, such as deformable mirrors, is generally specified as at least 10 times the required closed-loop bandwidth. Note that these stability considerations apply to any closed-loop control system. The actual values of gain and bandwidth required in an adaptive optics system depend on the specific performance requirements. This subject is discussed in detail in chapter 9.

2.6.5 Zonal and Modal Operation

There are two methods of specifying a random quantity such as wavefront error over a two-dimensional aperture: zonal and modal. In the zonal approach, the aperture is divided into an array of independent subapertures or zones. In each of these zones, the wavefront may be specified in terms of its optical pathlength (piston), its local gradient (tilt), or its local curvature. Most wavefront sensors and correctors used in adaptive optics work on a zonal basis, although modal algorithms may be used to compute the drive signals.

Modal analysis treats wavefront distortion as the sum of systematic functions. The wavefront is represented as a series of whole-aperture functions of increasing complexity. The most familiar modal functions in optics are the Zernike polynomials for a circular aperture. The first few Zernike terms correspond to familiar optical aberrations, such as overall wavefront tilt, defocus, astigmatism, and coma. Atmospheric turbulence may be decomposed into Zernike components, as described in section 3.5.1. The second and third modes correspond to overall tilt (angle of arrival fluctuations), which is usually compensated with a separate tip-tilt mirror. Each of these spatial components may have a characteristic temporal frequency associated with it, facilitating the processing of modal wavefront data.

While the Zernike modes form an orthogonal set for description of wavefront geometry, they are not statistically independent. To describe atmospheric wavefronts, a similar set of independent functions can be formed using the Karhunen-Loeve functions. It turns out that the low-order components of both the Zernike and the Karhunen-Loeve series are closely similar, so that little error is incurred in using the simpler Zernikes for the first 11 terms.

In principle, a turbulence-degraded wavefront may be compensated completely using either the zonal or modal wavefront representation. In practice, there are significant differences between these two approaches. The most obvious difference is that a zonal system may readily be expanded to arbitrarily high spatial resolution simply by increasing the number of subapertures, which are usually physically

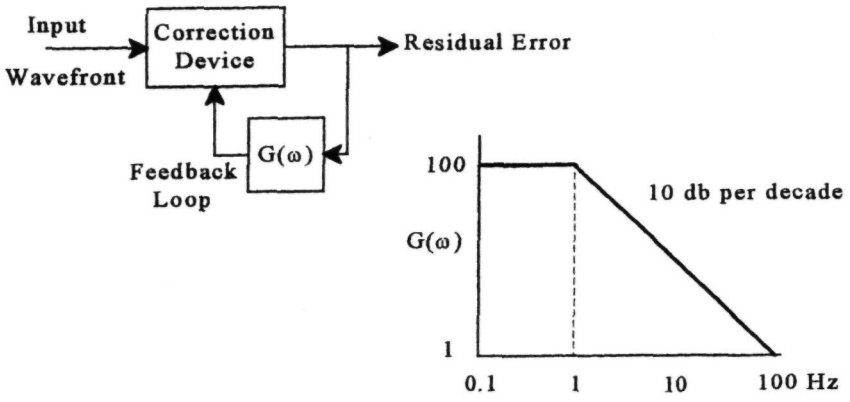


Figure 2.23 Feedback loop model and gain-versus-frequency plot for each channel of a closed-loop adaptive optics system.

identical. In contrast, while the first few Zernike terms represent easily implemented optical corrections, such as tilt and defocus, terms above Z_8 (coma) become increasingly difficult to measure and correct directly. Low-order aberrations can generally be implemented more smoothly using whole-aperture modes than with zonal correctors, which are liable to have small irregularities at the zonal period. For example, small amounts of defocus may be compensated by changing the radius of curvature of a spherical mirror; this can be done with a single actuator. A zonal deformable mirror requires six or more actuators to approximate defocus, and even then it produces an irregular shape.

The most important application of modal analysis is in adaptive optics data processing. For practical reasons, wavefront sensors and correctors are generally built using a zonal configuration, but the zonal data may be converted into modal form to perform the required computations. With atmospheric turbulence, there is a strong correlation between the spatial and temporal characteristics of the wavefront perturbations. The use of modal analysis enables adaptive optics performance to be maximized using optimal filtering in both the spatial and temporal domains.

2.7 Wavefront Sensing

2.7.1 Sensors for Adaptive Optics

The requirements for adaptive optics wavefront sensors are quite different from those developed for optical measurements in the laboratory, particularly in respect to the type of reference source and the speed of operation. Laboratory instruments used for testing precision optical components typically use laser interferometers to convert the optical phase information into intensity variations. The resulting interferograms are scanned to obtain the data from which the wavefront errors are computed. This process requires a local coherent reference, derived from the laser itself. Astronomical wavefront sensors, on the other hand, must operate with natural reference sources such as stars, as well as with laser-generated beacons. Most natural sources emit wideband thermal radiation that is temporally incoherent. In addition, stars that are potentially useful as reference sources have a wide range of surface temperatures, with peak radiation at wavelengths between blue and infrared. Astronomical wavefront sensors are usually photon starved, so it is vital that the available light from the reference source is used efficiently. Because of the need to use broadband light, direct measurements of optical phase are not feasible in astronomical adaptive optics.

The solution to the problem of wavefront sensing in astronomical adaptive optics is to measure the *direction of propagation* of the optical wavefront rather than its *optical phase.* This is done by measuring the wavefront gradients or curvature within an array of zones covering the telescope aperture. For a given value of turbulence strength r_0, the variations in wavefront gradient or tilt angle are invariant with wavelength, so this measurement can be made with incoherent (white) light.

Wavefront gradients are measured by producing an image of the reference source within each zone of the aperture and measuring the displacements of the spots from their mean position. Both the Hartmann sensor and the shearing interferometer are based on this principle, although their implementation is different. This principle is similar to that of the Foucault knife-edge test, which has been used since the 1860s to visualize the optical figure of large mirrors (see figure 1.5). A knife-edge was employed in the first image-stabilization devices developed for astronomy, described in chapter 1. Taking a larger perspective and stepping back further in time, one can see that the Foucault test is a brilliant elaboration of the basic process of establishing the direction of a distant object using a sighting device such as a pinhole and straight-edge. Such devices, depicted in figures 1.2 and 1.3, have been in use for millenia. The wavefront sensors in use today therefore have a pedigree stretching back to the earliest days of astronomy.

2.7.2 Wavefront Slope Sensors

The process of wavefront measurement using local gradients or slopes is shown in figure 2.24, which depicts the operation in one dimension. The wavefront is divided into separate subapertures, over each of which the average slope is determined. In this process, the relative phase or *piston* component of each subaperture is lost. The relative phase over the aperture is the essential information required; it is retrieved by the process known as *wavefront reconstruction.* In the simplest reconstructors, the piston components are recovered by two-dimensional spatial integration, in which the individual slopes are fitted together to form a surface having the smallest sum of discontinuities between elements. This process is made possible by the two-dimensional nature of the wavefront, which allows numerous redundant paths between elements. Mathematically, it involves minimizing the sum of the mean-square errors between all elements. More elaborate reconstruction algorithms have been proposed, weighting the slope inputs according to their signal-to-noise ratio and using operational and statistical information on the wavefront distortion [Wallner 1982].

The importance of reconstruction in wavefront slope sensors is illustrated in figure 2.25. If the wavefront slopes in each zone were measured and corrected without regard to their relative phase, the result would be an incoherent superimposition of the images formed by each zone. The result is shown as the upper curve in figure 2.25 (which is

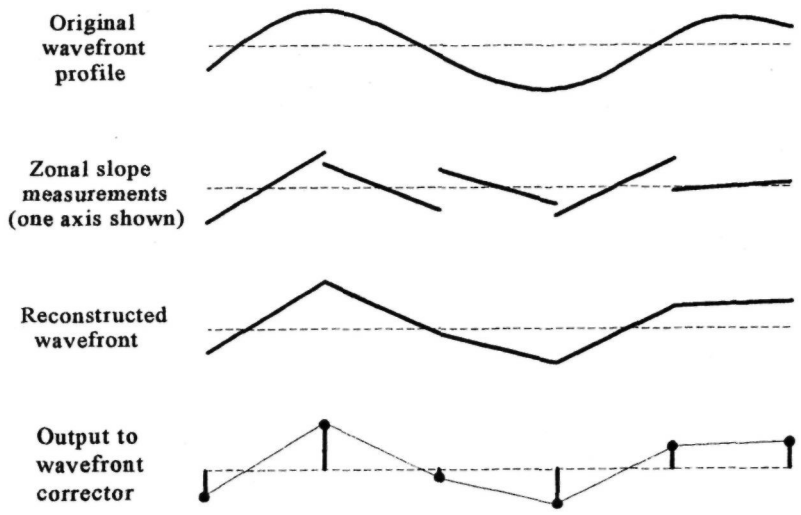


Figure 2.24 Wavefront measurement using zonal slope sensing. The wavefront is divided into contiguous zones and the average wavefront slope is measured in each zone. The relative phase of each subaperture is lost in this process. Individual slopes are then reconstructed into a continuous surface that best fits the measured data, retrieving the relative phase over the full aperture. Wavefront reconstruction is actually a two-dimensional process, only one of which is shown.

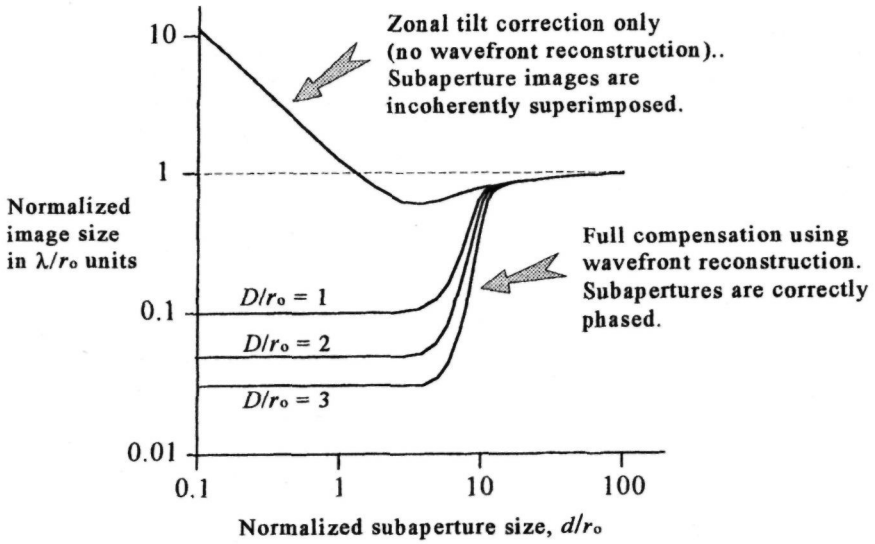


Figure **2.25** Performance improvement due to wavefront reconstruction. Reconstruction of the wavefront leads to a large gain in image quality compared with correction of local tilt only. The improvement factor is about D/d when the subaperture size d matches the turbulence parameter r_0.

the same as the short-exposure curve in figure 2.2). In the present context, tilt compensation removes the image motion.

When the zones are phased correctly, the zonal images add coherently, resulting in vastly improved angular resolution. It can be seen that for d/r_0 values around 1, the angular resolution with correct phasing is improved by a factor of D/r_0, which is often more than 10 for astronomical telescopes. In fact, a piston-type wavefront corrector that corrects the overall phase without correcting local tilt gives better compensation than that obtained by using zonal tilt correction alone. Wavefront slope sensing is therefore seen as only the first step in the process of measuring a wavefront, the reconstructor being an equally important part of the system.

The output of the reconstructor is a map of the original wavefront, at a sampling interval usually equal to the subaperture size. Wavefront variations at scales smaller than the subaperture size are lost in this process, the slope sensor acting as a low-pass filter. Obviously, the smaller the subapertures, the more accurately the wavefront will be measured. The error involved in wavefront slope measurements of this kind is known as *fitting error* and is always present in adaptive optics systems.

The reference source plays a major role in the design of adaptive optics systems. At visible wavelengths, it is rare for the science object itself, or even a nearby star, to be bright enough for full compensation, so that artificial sources, such as laser beacons, must be provided. The situation is more favorable at IR wavelengths, where compensation can be achieved with fainter reference stars. In all cases, the adaptive optics wavefront sensor must accommodate the peculiarities of whatever reference sources are planned for the system. Characteristics of reference sources used in astronomical adaptive optics systems are summarized in table 2.1.

Although most natural stars are not resolved by present-day telescopes and can therefore be considered as spatially coherent sources, the same is not true of laser beacons, which typically appear several arc seconds in diameter and may be spatially resolved by the telescope subapertures. It is therefore necessary for astronomical wavefront sensors to function with extended reference sources.

Adaptive optics wavefront sensors are required to measure complete wavefronts on time scales of a few milliseconds, with each measurement consisting of a large number of wavefront samples. The spatial resolution is determined by the number of subapertures corrected by the deformable mirror and ranges from a 5 x 5 array in small systems up to about 50 x 50 for the largest telescopes. The rate at which independent subaperture gradient measurements must be made in an adaptive optics sensor is therefore very high, over 10^6 per second for large systems. Each gradient or curvature measurement consists of at least two, and often four or more, pixel values (representing the number of detected photons), which must be read out of a CCD or similar detector array at megahertz rates. Noise added during the readout operation is minimized by using parallel output ports to reduce the temporal bandwidth.

Two major types of wavefront gradient sensors have been developed: the *Shack-Hartmann sensor* and the *shearing interferometer*. The Shack-Hartmann sensor is based on an optical test originally used for measuring the figure of large primary mirrors; in this test, a mask pierced with an array of holes is placed over the optical aperture, which is illuminated with light from a distant point source [Hartmann 1900]. Light reflected from the mirror through each subaperture produces a pattern of spots near the focal plane. Measurement of the displacement of the centroid of each spot from its expected position reveals the local gradient error in the mirror. In the late 1960s, Roland Shack of the University of Arizona was investigating methods of measuring wavefront distortion in order to enhance turbulence-degraded images. The small holes in con-

Table 2.1 Reference Sources for Astronomical Adaptive Optics

Type of Reference	Size (arc seconds)	Wavelength (μm)	Brightness (magnitude)
Visible stars (B-, V-, and R-bands)	< 0.1	0.4–0.8	m_V = -1 to +20
Infrared stars (J-, H-, and K-bands)	< 0.1	1.0-2.2	to m_V = +18
Small resolved disks (satellites)	0.25-6	0.4-0.8 (solar spectrum)	m_V = +5 to +10
Extended objects (solar or planetary surface)	1-1800	0.4-0.8 (solar spectrum)	m_V per arc second squared = -10 (Sun) to +9
Laser beacons, 8–15 km (Rayleigh scattering)	1-4	0.35-0.6	m_V = 0 to +8
Laser beacons, 90 km (Sodium layer fluorescence)	1-4	0.589	m_V = +5 to +12

ventional Hartmann masks used the available light very inefficiently, so Shack came up with the idea of replacing the holes with lenses, which could collect more light and produce an array of sharply focused spots. The centroids of the spots could then be measured precisely at high speeds, using, for example, quadrant cells. The main problem was to fabricate a suitable lens array consisting of 50 x 50 elements, with 1-mm center-to-center spacing. As no devices of this type were available commercially, the mask was cast in plastic in the form of an array of crossed cylinder lenses [Shack and Platt 1971]. The use of a lenslet array with contiguous subapertures, as shown in figure 2.26, enables essentially all of the light to be utilized and it produces a very compact design.

Shack-Hartmann sensors have reached a high degree of sophistication. Binary optics enables the lenslet arrays to be fabricated with high precision and stability. The preferred photon sensor is a CCD array; these have high quantum efficiency and good mechanical stability. The ability to work with either pulsed or continuous reference sources is a major advantage that enables these sensors to be used with pulsed laser beacons as well as stars.

The main disadvantage of Shack-Hartmann sensors is their inflexibility with respect to wavefront tilt sensitivity and dynamic range, which cannot be changed during operation. Quadrant detectors are normally employed for each subaperture, giving a truly linear range of only ±1/2 wave of tilt. When the local tilt approaches ±1 wave, which can occur randomly and suddenly, the output saturates. Quadrant detectors also have a nonlinear tilt response with extended reference sources. In spite of these drawbacks, the Shack-Hartmann sensor has become the standard wavefront sensor for adaptive optics systems.

There are many different types of shearing interferometer and only those capable of working with incoherent light are suitable for astronomical wavefront sensing. The principle of shearing interferometry is to generate two replicas of the wavefront to be measured; these are displaced by a small distance and then recombined. For use with natural sources that radiate temporally and spatially incoherent light, two requirements must be met: (1) the optical pathlengths of the two replicas must be equal to within the coherence length of the radiation, and (2) the interference fringes must coincide over the spectral band in use. These requirements are both met by using a diffraction grating to generate the two replicas. The shear or relative displacement of the two replicas depends on the diffraction angle, which is proportional to wavelength, so the interference fringes coincide at all wavelengths. Also, when symmetrical positive and negative diffraction orders are interfered, the optical pathlengths are automatically equal, resulting in a sensor that works with white light.

For wavefront sensing, lateral shear is most often used, resulting in an interference pattern having an intensity proportional to the wavefront gradient in the direction of shear. An array of photodetectors is used to measure the intensity of the fringe pattern, one detector for each subaperture. Two such sensors are required, with the shear directions at 90°, to measure the wavefront gradients in two orthogonal axes. For continuous reference sources such as stars (laser beacons are usually pulsed), the interference fringes can be modulated by using a rotating radial grating,

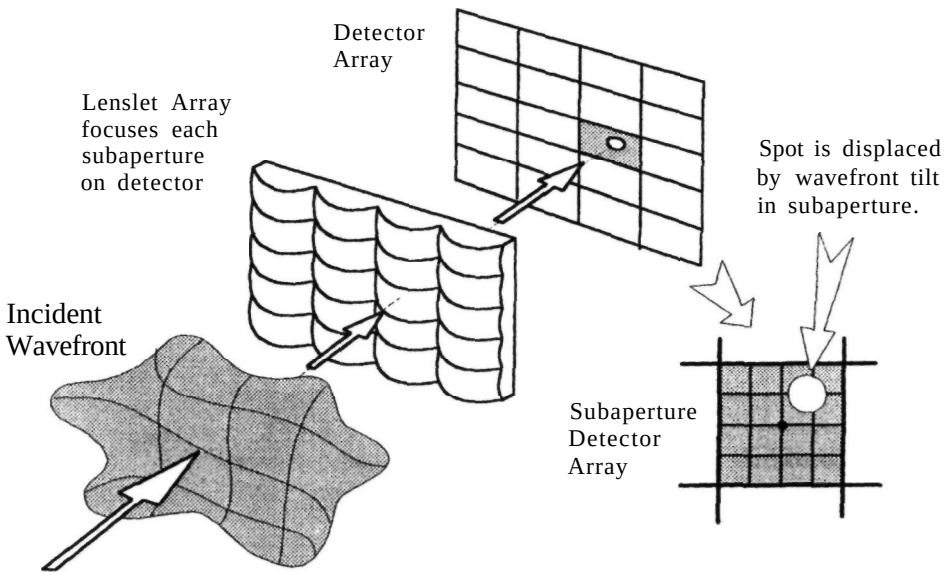


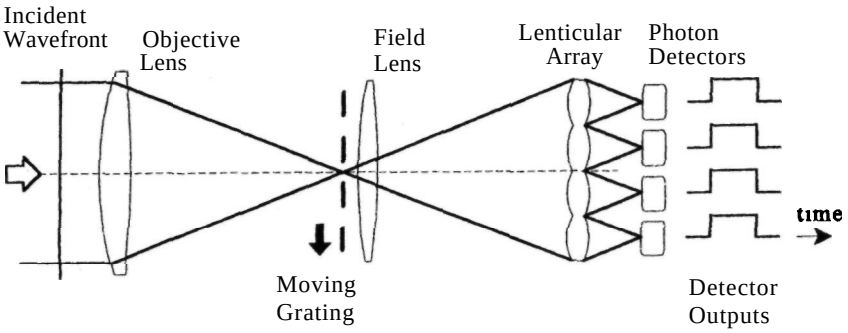
Figure 2.26 Principle of the Shack-Hartmann wavefront sensor.

located at an image of the reference source, to generate the sheared wavefronts. The operation of this type of shearing interferometer may be explained in terms of geometrical optics, as shown in figure 2.27. The detector array employs one pixel per axis to collect the light from each subaperture of the wavefront. With a plane wave input, the light from all subapertures is focused to a common point on the grating, which rotates at a constant rate, chopping the light reaching the detectors. In this case, the signals from all detectors are in phase.

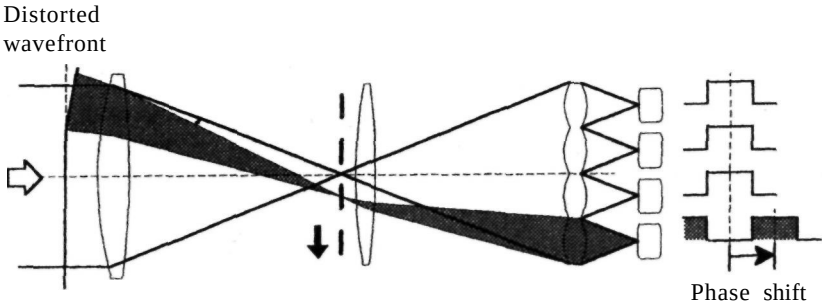
Distortion of the input wavefront produces local tilts across individual subapertures. The images from these subapertures are consequently displaced at the detector plane, causing them to be chopped at an earlier or later time by the grating. The individual detector outputs are therefore phase shifted in time by amounts proportional to the wavefront gradients within each subaperture. Electrical phase shifts are easily measured, giving the required wavefront slopes. This version of the shearing interferometer has the advantage that no calibration is needed because the phase values in all subapertures are generated by the same grating lines at a common focus, so any irregularities affect all channels equally and do not show up as measurement errors. While the principle of the shearing interferometer is not as obvious as that of the Hartmann sensor, the physical implementation is simple and it is a very stable and reliable device. An advantage of this sensor is that its sensitivity and dynamic range may be adjusted in real time by changing the grating spacing at the focus; this is easily achieved by moving the rotation axis of the radial grating.

2.7.3 Photon Detectors

In real-time wavefront compensation systems, it is necessary to perform a complete wavefront measurement (including photon integration, detector read-



(a) Plane wavefront: **all** detector outputs in phase.



(b) Distorted wavefront: output phase shift proportional to wavefront slope.

Figure 2.27 Shearing interferometer wavefront sensor with temporal modulation. A moving grating at the focal plane chops the light at a constant frequency. The light from each zone of the aperture is displaced according to the local slope, producing an electrical phase shift in the detector output that is directly proportional to the wavefront slope.

out, wavefront gradient estimation, and wavefront reconstruction) at several hundred locations in the telescope aperture, within a few milliseconds. To meet this requirement, wavefront sensors employ multielement photodetector arrays that cover the entire optical aperture, enabling all measurement zones to be sensed simultaneously. The overall performance of any type of wavefront sensor is determined largely by the photodetector employed.

Photon detectors suitable for wavefront sensing include photomultiplier arrays, intensified silicon diode arrays, silicon CCDs, and avalanche photodiodes (APDs). The two most important parameters are the quantum efficiency and the signal-to-noise ratio. The first wavefront sensors used for astronomical adaptive optics employed arrays of photomultiplier tubes (PMTs) for photon detection. Photomultiplier tubes optimized for photon-counting add practically no noise in the detection process, so they can be regarded as "photon-limited". In this case, the signal-to-noise ratio is equal to the square root of the number of photons detected. The main drawback of PMTs and intensifiers is the relatively low quantum efficiency of photocathodes, which peaks at about 20% at visible wavelengths, although Generation 3 devices using infrared InSb photocathodes may have quantum efficiencies as high as 50%. In comparison, the quantum efficiency of silicon devices is 80% to 90% at visible wavelengths. Another difficulty with photomultipliers is that they are relatively bulky devices and cannot easily be assembled into large arrays. Multianode devices with up to 100 channels have been made, but larger arrays are impractical.

Not only do CCD arrays have high quantum efficiency, but they can also be made in sizes compatible with wavefront sensors for astronomical adaptive optics: for example, 64 x 64 pixels. A major problem with CCDs has been the *read noise* added to the serial output when the charge packets are converted into electrical signals. The read noise is proportional to the pixel rate (bandwidth) of the serial output. Reading out a 64 x 64 array at 1000 frames per second produces a pixel rate of about 4 MHz. The readout rate may be reduced by using multiple ports. Low-noise devices are now available in which the readout noise is less than five electrons per pixel. High quantum efficiency is obtained in these devices by thinning the silicon substrate and using back illumination to avoid the obstructions of the electrode structure.

2.7.4 Image-Plane Sensing

In the wavefront sensors described above, the pupil is divided into an array of zones in which independent measurements of wavefront slope are made simultaneously. This is the preferred method, at least for adaptive optics systems containing a large number of sub-apertures. An alternative approach is to deduce the wavefront from measurements of the reference source intensity distribution in the image plane. The basic problem with image-plane sensing is that light from all zones of the aperture is superimposed at the image. There is a Fourier transform relationship between the distribution of light at the pupil and image planes, resulting in the loss of spatial information. Therefore, a change in intensity in the image cannot be associated directly with the wavefront in a specific part of the pupil. Deconvolution has been shown to be possible in some cases, but the process is image dependent and requires intensive computation. There are also practical difficulties in measuring the large intensity range within the small area associated with optical images.

Some of the problems of image-plane sensing can be overcome by making multiple intensity measurements near the focal plane. Even a small displacement from the focal plane allows the spatial relationship with the optical pupil to be established, eliminating the wavefront ambiguities that plague pure image-plane measurements. Wavefront distortion can be determined using phase diversity [Gonsalves 1982], in which two images are measured simultaneously, one in best focus and one with a known phase change, enabling image restoration with post-detection processing. In principle, it can also be used for real-time compensation, although considerable data processing is involved.

Neural networks also use intensity information from two planes near the image. This approach to wavefront sensing has been successful in detecting low-order distortion at low signal-to-noise ratios. The network must first be trained by exposure to the functions that it is required to recognize. The application of neural networks to adaptive optics is discussed in section 5.8.

Image sharpening, an adaptive optics technique in which all measurements are made in the image plane, has been described in section 2.6.3. Image sharpening may be regarded as a simplified version of the deconvolution process, in which a single intensity measurement is made at the image, usually at the central peak. To separate the contribution of each subaperture, trial phase corrections are inserted sequentially in each zone and their effect on the intensity of the image is measured. As shown above, serial adaptive optics systems do not perform well in the photon-starved environment of astronomy.

The output of a wavefront gradient sensor consists of an array of zonal measurements representing local wavefront gradients in two orthogonal axes. The necessity for reconstructing these measurements into a two-dimensional array that represents the optical path errors over the whole optical aperture was discussed in section 2.7.2. The reconstruction process was illustrated in one dimension in figure 2.24. In this case, the measured gradients were simply fitted

Adaptive Optics in Astronomy 67

together by matching the edges of adjoining pairs of subapertures to form the reconstructed wavefront profile. In the one-dimensional process, measurement errors or uncertainties due to noise on individual subapertures will propagate through the structure without attenuation, and will appear as accumulated errors in the output profile.

Wavefront reconstruction as used in adaptive optics is a two-dimensional process that has more desirable properties. A typical arrangement is shown in figure 2.28. The wavefront gradients in the x and y axes are measured by an array of sensing zones and the wavefront is evaluated at a corresponding square array of nodes. The total number of gradient measurements is approximately twice the number of evaluation points, so that each pair of points is linked by multiple measurement paths. This redundancy in the measurements has the beneficial effect of smoothing random errors, which do not accumulate as they do in the one-dimensional case.

It has been found that the *noise propagation* of two-dimensional reconstructors of this type varies as the logarithm of the number of nodes. For apertures containing on the order of 100 nodes, the noise propagation is approximately unity, so that the error at each point of the reconstructed wavefront is approximately the same as the average error on each input gradient measurement.

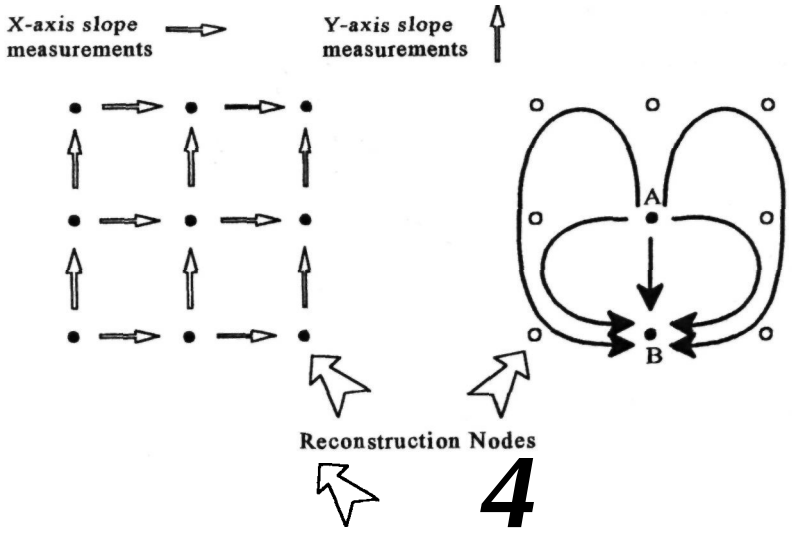
The first real-time wavefront reconstructor, used in the 21-actuator RTAC system [Hardy et al. 1977] was an analog device consisting of a two-dimensional electrical conductance network representing the wavefront nodes. Currents proportional to the measured wavefront slopes were injected into the network nodes, which then settled at wavefront values that minimized the mean-square error in the input data. A similar analog reconstructor was used in the 1982 Compensated Imaging System (CIS), which employed 168 actuators. The computation (settling) time of the network was less than $10\,\mu s$.

The current approach to wavefront reconstruction is to use parallel digital processing. Wavefront reconstruction involves a heavy data-processing load. For an aperture containing N zonal wavefront samples, there are approximately N x-gradient measurements and N y-gradient measurements. To obtain each reconstructed point, all $2N$ gradient values are weighted and summed. The reconstruction process therefore involves a total of $2N^2$ multiply–add operations, which must be performed in a small fraction of the turbulence change time, usually less than 1 ms. An adaptive optics system with 500 actuators produces a reconstructor data-processing load of 5×10^8 multiply-add operations per second. Wavefront reconstructors are usually implemented as a parallel array of multiplier-accumulator devices.

2.8 Wavefront Correction

2.8.7 Types of Wavefront Corrector

The wavefront corrector is the key element in determining the performance of all adaptive optics systems; the output wavefront can be no better than the uncorrectable errors on these devices, so a considerable effort has been expended on their design and fabrication over the last 25 years or so. Some



Reconstruction Nodes

Figure 2.28 Two-dimensional wavefront reconstruction. In the case shown, the x-axis and y-axis slope measurements are centered between adjacent reconstruction nodes. In the reconstructor, pairs of nodes, such as A and B, are connected by many different parallel paths, reducing the error in estimating the wavefront from noisy slope measurements.

early wavefront correctors were based on refractive index changes in electro-optical crystals, but these devices had a very limited range and required high voltages. The advantages of reflective devices soon became apparent and almost all current correctors are of this type. Deformable mirrors are also referred to as inertial compensators because they employ mechanical displacement of a reflecting surface. The basic forms of wavefront corrector are shown in figure 2.29. The main types in use, continuous facesheet deformable mirrors, bimorph mirrors, and segmented mirrors, are briefly described in sections 2.8.2-2.8.4.

Membrane mirrors appear to be an economical method of implementing a large number of actuators, but they have not yet achieved success in adaptive optics systems. A membrane mirror is being considered for this purpose in the Subaru Telescope of the National Astronomical Observatory of Japan [Takami and Iye 1994].

The only other promising technology for wavefront compensation in astronomical adaptive optics

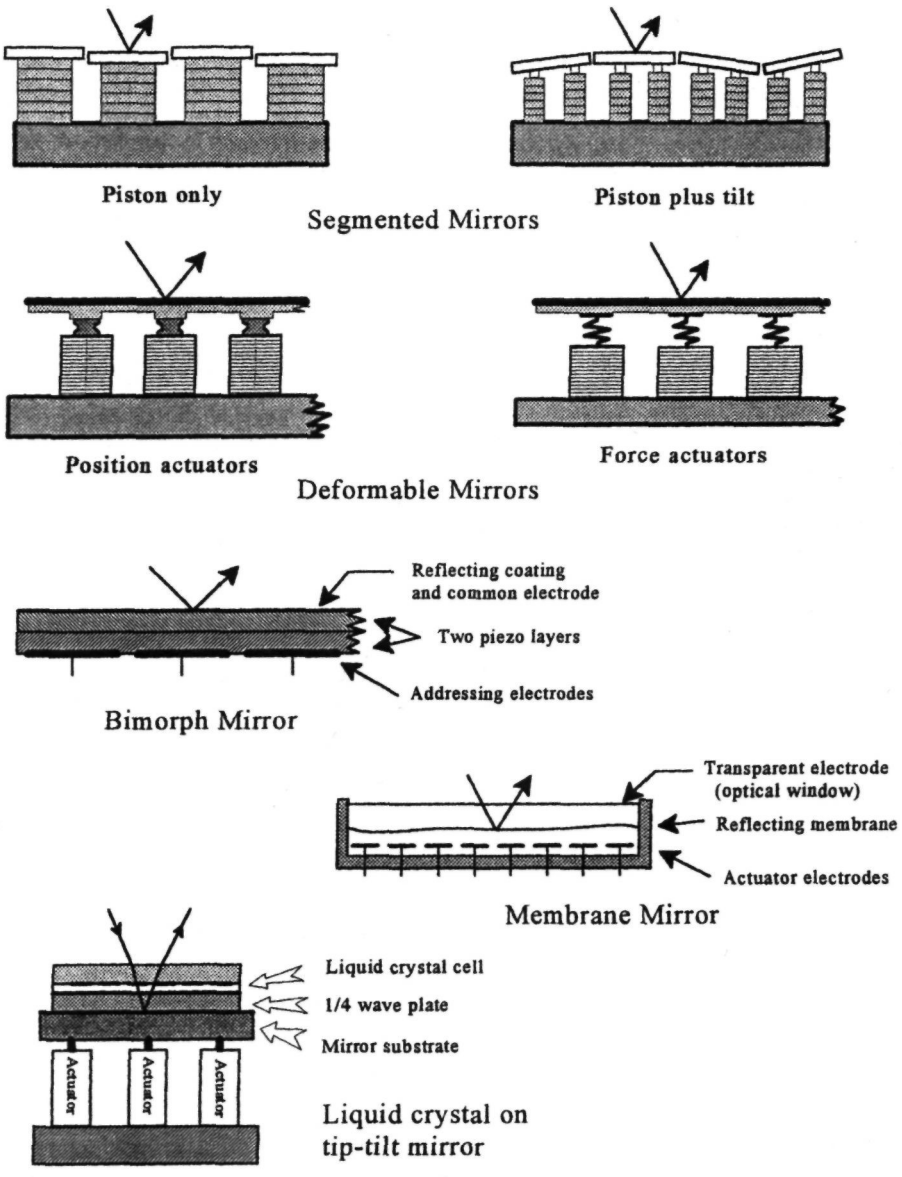


Figure 2.29 Wavefront correction devices.

appears to be refractive devices using liquid crystals, which have a potential advantage in being solid-state (or at least "liquid-state") components. At the present time, liquid crystal correctors operate best at IR wavelengths, as the dispersion rises significantly at visible wavelengths. Special techniques have been developed to improve the operating speed of liquid crystals: rise and fall times in the millisecond range are now feasible [Bonaccini et al. 1991]. These devices produce relatively large optical pathlength changes and can be made into arrays containing a large number of addressable pixels. A liquid crystal wavefront corrector, used in reflection, can be combined with a mechanical tip-tilt mirror as shown in figure 2.29.

2.8.2 Deformable Mirrors

The first successful deformable mirrors (developed at Itek in 1973) were of the monolithic piezoelectric type. They employed a thin aluminized glass facesheet bonded to a solid block of piezoelectric material on which an array of electrodes was mounted. Voltages applied between the electrodes and a common ground plane produced local deformations of the piezoelectric block that were imprinted in the facesheet. These devices could be made with several hundred actuators and had the required precision and stability, but their correction range was only 1.4 μm, restricting their application in large astronomical telescopes .

The present generation of deformable mirrors employs a thin continuous faceplate supported by an array of actuators attached to a massive and stable baseplate, often made of the same material as the faceplate (for example, ultralow-expansion quartz) in order to eliminate thermal distortion. The actuators are typically multilayer piezoelectric or electrostrictive ceramics. Deformable mirrors of this type have good stability over time and temperature changes, largely due to the use of a continuous faceplate, which requires the application of considerable force to produce distortion. The deflection or stroke obtainable with this type of mirror is limited by the stress induced in the faceplate by the actuator motion. Actuator spacing is typically around 7 mm. The greatest stress is produced between adjacent actuators, limiting the differential stroke to about $\pm 3\,\mu$m.

Continuous-plate deformable mirrors using low-voltage PMN (lead magnesium niobate) electrostrictive actuators were developed at Itek in the 1980s [Ealey and Washeba 1990], in sizes ranging from 69 to about 2000 actuators. The mechanical stroke of these mirrors was 5 μm peak to peak with a drive voltage of 150, and the minimum surface ripple was 0.02 μm rms. Low-cost deformable mirrors using replaceable PMN piezoelectric actuator cartridges have been developed by Xinetics [Ealey and Wellman 1994]. These mirrors have a stroke of 4 μm and are available in sizes of 37, 97, and 349 actuators.

Deformable stacked array mirrors (SAM) with 52, 88, and 249 actuators are manufactured by Laserdot, France [Gaffard et al. 1994]. Piezoelectric actuators are used, giving a maximum mechanical stroke of $\pm 5\,\mu$m, with a differential of $\pm 2.5\,\mu$m between adjacent actuators. The facesheets on these mirrors are single crystals of silicon.

2.8.3 Bimorph Mirrors

Bimorph mirrors also employ continuous faceplates, but these are supported and deformed in a totally different way from conventional deformable mirrors. The actuators in bimorph mirrors are flat disks of piezoelectric material that are bonded to the back of the faceplate. Voltages applied to the actuators cause their dimensions to change parallel to the plate, producing bending moments that curve the faceplate. Bimorph mirrors do not need a rigid baseplate because the bending forces are applied directly to the faceplate itself. Bimorph mirrors can be made at relatively low cost and are well suited to adaptive optics systems using curvature sensing and correction. A 13-actuator bimorph deformable mirror is being used for this purpose in the IR adaptive optics system developed by the University of Hawaii [Roddier, Northcott, and Graves 1991]. Bimorph mirrors with 13 and 19 actuators have been made by Laserdot [Gaffard et al. 1994].

2.8.4 Segmented Mirrors

Segmented mirrors consist of an array of triangular, square, or hexagonal elements, each supported on an actuator structure with three degrees of freedom, providing two axes of tilt plus piston motion. (Some large primary mirrors, such as those for the 10-m Keck telescopes, also employ segmented construction as a means for achieving large apertures. The segments of these primary mirrors are not intended to compensate atmospheric turbulence; they have a more elaborate supporting structure, as befits their size.) The segments of high-speed adaptive mirrors intended for turbulence compensation are typically about 1 cm in size. They have been made with as many as 500 segments, employing 1500 actuators [Hulburd and Sandler 1990]. Segmented mirrors have a good frequency response and large dynamic range because each element is unconstrained. The stroke obtainable with segmented mirrors is limited only by the capability of the actuators. They are also relatively easy to repair in case of actuator failure.

Segmented mirrors have two drawbacks: (1) Because the segments are free to move independently, they need frequent calibration to keep all elements phased properly. (2) The gaps between segments, however small, inevitably diffract some of the incident light, which shows up as a distinct pattern in the

image plane due to the periodicity of the segments. There are also abrupt changes in slope at the junctures between segments. These optical effects may not be important for some applications, but must be considered in low-scatter systems designed to work with very low contrast images. The gaps between segments could also be a problem with thermal IR systems.

The segments and actuators used in a segmented mirror are usually identical and can be mass-produced; as a consequence, the cost of segmented mirrors is often lower than that of continuous facesheet mirrors. Nevertheless, the need for a dedicated phasing sensor to keep the segments in proper alignment must be kept in mind; this adds to the cost of the adaptive optics system and complicates its operational use.

2.9 Laser Beacons

2.9.1 Sky Coverage

To maximize the value of image compensation in observational astronomy, adaptive optics must be usable over most (preferably all) of the sky. In reality, because of angular anisoplanatism, the area over which compensation is effective is restricted to a small radius surrounding the reference source. For full compensation at visible wavelengths, the isoplanatic angle is about 2 arc seconds for average conditions, and the required reference source magnitude is about $m_V = 10$. Using natural stars, the sky coverage is limited to about 1/100,000 of the hemisphere; this is of little value for deep-sky observations. The situation improves at longer wavelengths due to the larger values of r_0 and θ_0. To compensate images at 2.2 μm, a reference source of about $m_V = 14$ is required, and the isoplanatic angle increases to about 10 arc seconds, giving a sky coverage of about 1/1000.

Although sky coverage is an important issue in adaptive optics for compensating very faint objects, much useful work can be done using natural sources in the self-referencing mode. Many objects of current interest in astronomy, such as protoplanetary nebulae, contain stars bright enough to function as reference sources and are compact enough to lie within a single isoplanatic patch. For example, Roddier et al. [1995] have obtained compensated images of the bipolar nebula "Frosty Leo" with an angular resolution of 0.1 arc second in the J- and H- bands (1.25-1.65 μm). The wavefront sensor operated at 0.85 μm.

Laser beacons provide a partial solution to the problem of sky coverage. The artificial stars generated by laser backscatter within Earth's atmosphere provide reference sources for the measurement of most of the turbulence within the aperture of a telescope. The measurement is incomplete due to the proximity of the beacon to the telescope; this produces a residual error known *as focal anisoplanatism.* There is an additional problem: a beacon projected from the ground cannot provide information on the absolute position of the object viewed. To make a long exposure of an astronomical object, its position must be stabilized to within a fraction of the spatial resolution required. Even when using a laser beacon, a fixed reference star outside Earth's atmosphere is still required to stabilize the image. The sky coverage obtained with adaptive optics is therefore limited by the availability of natural stars bright enough to be used for image stabilization.

For the reasons discussed in section 2.9.4, tilt stabilization can be achieved with much fainter stars than are required for high-order wavefront compensation, thus resulting in useful coverage of the sky. For observations in the IR bands between 1.25 and 2.2 μm, stars as faint as visual magnitude m_V 16-18 can be used for image stabilization. It is shown in section 9.6 that it is possible to obtain a tilt stabilization Strehl ratio of 0.5 over 10% of the sky at 1.25 μm, and over 100% of the sky at 2.2 μm. By compensating the natural guide star (as well as the science object), the limiting magnitude can be extended to about 20, resulting in almost 100% sky coverage even for observations at 0.7 μm. Infrared guide stars may benefit directly from the primary adaptive optics because of the large isoplanatic angle, which eliminates the need for a separate compensation system.

2.9.2 Rayleigh and Sodium Beacons

The use of laser beacons for adaptive optics was pioneered for defense purposes during the 1980s and the technology was made public in 1992. Laser beacons consist of the backscattered light generated by ground-based laser beams passing through atmospheric constituents. There are two modes of scattering: (1) Rayleigh scattering, which is produced by air molecules and is useful at altitudes up to about 15 km; and (2) sodium resonance fluorescence, which occurs in a thin layer of sodium atoms deposited by meteors at an altitude of about 90 km.

Rayleigh scattering is most efficient at UV wavelengths and is useful up to about 0.6 μm. Many different types of laser have been used, including excimer, copper vapor, doubled neodymium - yttrium aluminum garnet (Nd:YAG), and tunable dye lasers. The strength of the Rayleigh scatter return depends on the atmospheric density, which drops exponentially with altitude, so Rayleigh beacons are used mostly at altitudes between 5 and 10 km. For continuous adaptive correction, the laser power needed is on the order of 100 W per beacon. The lasers are pulsed so that backscatter from low altitudes can be eliminated by range gating. The angular size of the spot is normally limited to 1 or 2 arc seconds by atmospheric turbulence. If the laser beam is launched through a telescope equipped with adaptive optics, the beam will be compensated

and a much smaller spot obtained. Rayleigh beacons have proved effective for small telescope apertures up to 2 m, but their relatively low altitude produces inadequate sampling of atmospheric turbulence with larger instruments. In principle, multiple beacons may be used to provide better wavefront measurements, but this process makes an already complicated system even more complex.

For large telescope apertures, sodium beacons offer improved sampling of atmospheric turbulence due to their much higher altitude. To excite the sodium resonance, the laser must be tuned accurately to the D_2 line at 0.589 μm.The process is much more efficient in generating backscattered light than is molecular scattering. The laser power required for each beacon is around 10 W. Tunable dye lasers have been used to generate the 0.589-μm radiation. The sodium wavelength is also produced as a sum-frequency by mixing the 1.064 and 1.319-μm outputs of Nd:YAG lasers in a nonlinear crystal.

In the process of resonance fluorescence, sodium atoms are pumped by the laser irradiation from the ground state into an excited state, from which their natural decay produces the desired backscattered radiation. If the population of excited atoms is large, then the incident light produces stimulated emission in the forward direction; such emission is lost to the telescope. This condition is known as *saturation* or *bleaching* and it limits the laser power that can be used efficiently. Saturation is controlled by limiting the peak pulse power of the laser and by tailoring its pulse width and repetition rate.

2.9.3 Focal Anisoplanatism

The artificial stars created by laser backscattering are necessarily within Earth's atmosphere and consequently they are relatively near to the telescope. These reference sources sample a cone-shaped volume of the turbulence in the optical path of the telescope, as shown in figure 2.30. Turbulence above the beacon is not measured, and that occurring below the beacon is only partially sampled. The resulting wavefront error is known as *focal anisoplanatism* or *cone effect.* It is characterized by the parameter d_0, the aperture diameter over which the mean-square wavefront error due to focal anisoplanatism is 1 rad^2. The larger the value of d_0, the smaller the error. The value of d_0 depends on the beacon height, the turbulence profile, and the observing wavelength. Because of the irregular nature of turbulence profiles, d_0 is a nonlinear function of height; it is defined and calculated in section 7.3.3.

Values of d_0 for Rayleigh beacons at 10-km altitude are typically about I m at a wavelength of 0.5 μm, 3 m at 1.25 μm, and 5 m at 2.2 μm. Most astronomical telescopes will therefore require arrays of such beacons. To provide compensation at visible wavelengths, a 4-m telescope would require at least a

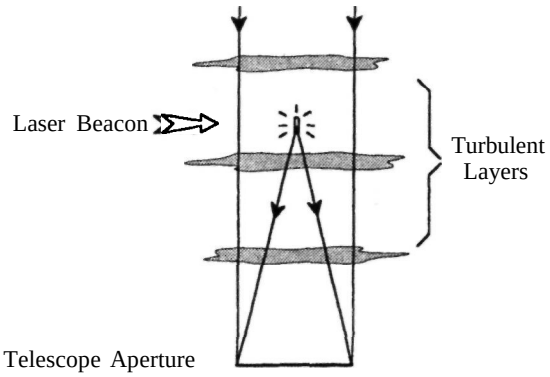


Figure 2.30 Focal anisoplanatism or cone effect due to incomplete sampling of atmospheric turbulence with a laser beacon.

4 x 4 array of Rayleigh beacons, while an 8-m telescope operating at 1.25 μm would require a 3 x 3 array. Laser beacon systems of this complexity have not been built.

For large telescopes, sodium beacons have a considerable advantage due to their much larger values of d_0. For a beacon at 90 km, the value of d_0 at a wavelength of 0.5μm is about 4 m, while at 1.25 μm it is close to 10 m. Thus, at IR wavelengths, even the largest telescopes would benefit from a single sodium beacon. Although the laser requirements for sodium are more stringent in terms of wavelength and pulse width (to avoid saturation), sodium beacons appear to be the preferred approach for astronomical adaptive optics.

2.9.4 Image Stabilization

The reference source provided by a laser beacon is suitable for measuring the wavefront errors within an optical aperture, but it cannot be used to determine the overall tilt, which specifies the absolute direction of the object to be compensated. The reason is that the position of a beam projected from the ground is randomly perturbed by atmospheric turbulence. When projected and viewed through the same telescope, a laser beam will always appear to be on axis no matter where the telescope is pointing, because the outgoing and incoming beams are deflected by the turbulence in exactly the same way. To stabilize an image, it is necessary to use a separate reference source that is not perturbed by the atmosphere. Natural stars are the obvious choice.

The sky coverage provided by natural stars is determined by two factors: (1) the star brightness required to provide a sufficient signal-to-noise ratio for overall tilt sensing; and (2) the isoplanatic angle over which tilt compensation is effective. Both of these factors become more favorable as the wavelength increases. Overall tilt is the lowest order wave-

front distortion component in single-aperture adaptive optics. The whole telescope aperture is used to collect the light and the integration time is relatively long, so the photon flux required from the fixed guide star is considerably less than that for high-order compensation. In addition, the isoplanatic angle for overall tilt is greater than that for high-order compensation.

For these reasons, the sky coverage obtained with laser beacons stabilized by natural guide stars is much greater than that of adaptive optics using natural stars alone.

2.10 Error Sources in Adaptive Optics

2.10.1 Fitting Error

Optical wavefronts degraded by atmospheric turbulence contain components at spatial scales ranging from several meters (outer scale) to about 1 mm (inner scale). According to Kolmogorov's model, these components are always present in fully developed turbulence. Exact compensation of such a wavefront would involve extremely high spatial frequencies, which are impractical to implement. The residual error due to the inability of the wavefront compensator to correct the wavefront exactly is known as *fitting error.* If the turbulent wavefront is $\phi_w(\mathbf{x})$ and the wavefront corrector function is $\phi_c(\mathbf{x})$ then the mean-square fitting error is defined as $\sigma_F^2 = \langle[\phi_w(\mathbf{x}) - \phi_c(\mathbf{x})]^2\rangle$, where x is the position vector and ( ) denotes ensemble average.

For a wavefront with a Kolmogorov spectrum (power a $\ell^{5/3}$), the fitting error of a zonal corrector may be expressed in the form $\sigma_F^2 = a_F(d/r_0)^{5/3}$ where a_F is a constant of proportionality dependent on the type of corrector. The fitting error is controlled by the ratio d/r_0, and may be reduced by using a small value of d; that is, by using more actuators. To obtain a high degree of wavefront compensation, the value of d/r_0 is usually between 1 and 1.5, but it may be larger with partially compensated systems.

The turbulence parameter r_0 is usually defined at a wavelength of 0.5 μm and varies as $A^{6/5}$. For constant turbulence, the value of r_0 increases with wavelength, and the fitting error is inversely proportional to wavelength.

2.10.2 Temporal Error

The temporal fluctuations of turbulence are caused mainly by transportation of perturbed layers by winds, so the temporal spectrum has a similar form to its spatial spectrum, scaled by the wind velocity. The inevitable time delay between the measurement and correction of turbulence in an adaptive optics system results in the temporal error, defined as $\sigma_t^2 = \langle[\phi_w(\mathbf{x}, 0 - \phi_w(\mathbf{x}, t + A/)]^2\rangle$, where *At* is the delay time. The mean-square temporal error has the same basic form as the fitting error, being proportional to $(\Delta t/t_0)^{5/3}$, where t_0 is the characteristic change time of the turbulence.

2.10.3 Anisoplanatic Error

When the reference source used for wavefront measurement is displaced angularly from the science object, the turbulence sampled by the wavefront sensor is different from that in the imaging path, as shown in figure 2.31. The wavefront error due to an angular separation *A0* between the reference source and the science object is known as the anisoplanatic error and is defined as $\sigma_\theta^2 = \langle[\phi_w(\mathbf{x},0) - \phi_w(\mathbf{x}, \theta + \Delta\theta)]^2\rangle$. The mean-square wavefront error due to angular anisoplanatism is expressed conveniently in terms of the isoplanatic angle θ_0, which is defined as the angle at which the anisoplanatic error is equal to 1 rad^2 at a wavelength of 0.5 μm. The anisoplanatic angle is determined by the vertical distribution of turbulence, with a weighting factor of $h^{5/3}$. For average conditions at visible wavelengths, its value is in the region of 2 arc seconds. The mean-square anisoplanatic error at an angle *6* is given by $\sigma_\theta^2 - (\theta/\theta_0)^{5/}$. This anisoplanatic error is encountered in adaptive optics systems using natural stars as the primary reference source. Its small value is the reason that the sky coverage with such systems is so small. The anisoplanatic error also varies as $A^{6/5}$.

When laser beacons are employed, the anisoplanatic errors have a different character, as described in

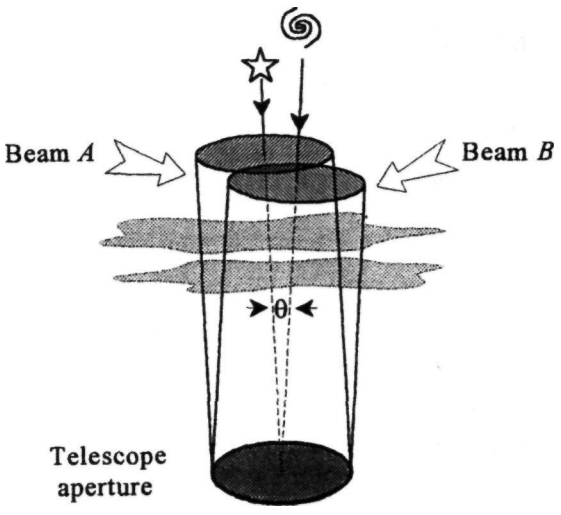


Figure 2.31 Angular anisoplanatism. The wavefront measured within beam A is not valid for beam B because the beams traverse different regions of turbulence. The error is dependent on the ratio of the beam separation angle *0* to the isoplanatic angle θ_0 of the turbulence structure.

section 2.9. Wavefront compensation within the aperture is degraded by focal anisoplanatism or cone effect, due to the proximity of the beacon to the telescope. The need for a fixed guide star to stabilize the image invokes an additional error due to tilt anisoplanatism. These errors have more complex dependencies and are treated in chapter 7.

2.10.4 Photon Error

The photon error is produced by random noise in the wavefront sensor, consisting of shot noise from the reference source and background, together with amplifier noise produced during detection of the reference signal. The shot noise component is inversely proportional to the square root of the number of photons counted, and it is dominant for high photon fluxes. Because of the small amount of light available from both natural and laser sources, wavefront sensors usually operate under photon-starved conditions, so the photon error is typically a large component of the error budget. The photon error may be reduced by increasing the integration time, but this increases the temporal error, so a tradeoff must be made to determine the optimum value.

2.10.5 Performance Estimation

The main sources of random error are summarized in figure 2.32, which shows their dependence on the external parameters of the atmosphere and on the parameters of the adaptive optics system. The overall performance of an adaptive optics system is estimated by summing the individual errors. If these sources are uncorrelated, the residual wavefront phase error is given by the sum of their variances:

$$\sigma^2_{\text{phase}} = \sum \sigma_i^2$$

In practice, there are correlations between some errors, so a straightforward sum of the variances may lead to an overestimate of the total residual error. This is not necessarily a disadvantage in system design because it tends to compensate for unrecognized errors, and the resulting performance predictions are usually more realistic.

Many of the parameters in adaptive optics systems are interrelated, so it is necessary to perform tradeoffs in order to optimize the overall performance. An important factor is the integration time, which controls both the temporal error and the photon error. This relationship is shown in figure 2.33. The optimum integration time minimizes the sum of the photon and temporal errors. The photon error varies with the reference source brightness and the turbulence strength, so it is necessary to update the integration time during operation as the observing conditions change.

From the total wavefront error variance, the Strehl ratio representing the normalized peak intensity of a compensated point source is computed from the relation

$$S = \exp(-\sigma^2_{\text{phase}})$$

The Strehl ratio is used conventionally as a performance criterion in adaptive optics systems. It does not represent the complete imaging capability of a system, for which the peak contrast ratio C_P may be more relevant. Images having Strehl ratios of 0.1 or less still contain important scientific information. The choice of a suitable performance criterion is a critical factor in the design of astronomical adaptive optics because it allows compromises in the hardware and software that may considerably reduce the complexity and cost of the system.

The issue of partial compensation was introduced in section 2.4.3. Partially compensated images consist of a diffraction-limited core surrounded by a much larger halo, the diameter of which is determined by the value of r_0 at the observing wavelength. The peak contrast ratio between the core and the halo may be the critical factor for some imaging tasks. The subject of suitable criteria for the evaluation and optimization of adaptive optics systems is discussed in detail in chapter 4.

2.11 Using Adaptive Optics

2.11.1 Integration with the Telescope

Adaptive optics may be added to an astronomical telescope in two ways: the adaptive system can be integrated into the basic telescope design, or it can be installed as a separate unit, attached to a focal station. With existing telescopes, there is usually little choice other than to install the adaptive optics at one of the focal positions, or in the coude path, as shown in figure 2.34. In such cases, additional optical components are unavoidably included in the imaging path.

As pointed out in section 2.4, adaptive optics enhances the performance of telescopes of all sizes, effectively giving them the capability of much larger instruments. For a telescope of aperture D operating in turbulence of coherence length r_0, the increase in limiting magnitude is potentially $2.5 \log(D/r_0)$, which for a 2-m telescope in average turbulence conditions of 0.1 m gives a magnitude gain of 3.25, equivalent to increasing the aperture to about 8 m. This suggests that a packaged adaptive optics unit, designed for use at the Cassegrain, Nasmyth, or coude foci, would be a valuable adjunct to most existing instruments. The package can be designed with the same input and output F-numbers, allowing the use of existing instrumentation, although the adaptive optics advantage is not realized unless the spatial resolution of the instruments is increased.

The advantage of designing the adaptive optics into the telescope itself is that the number of addi-

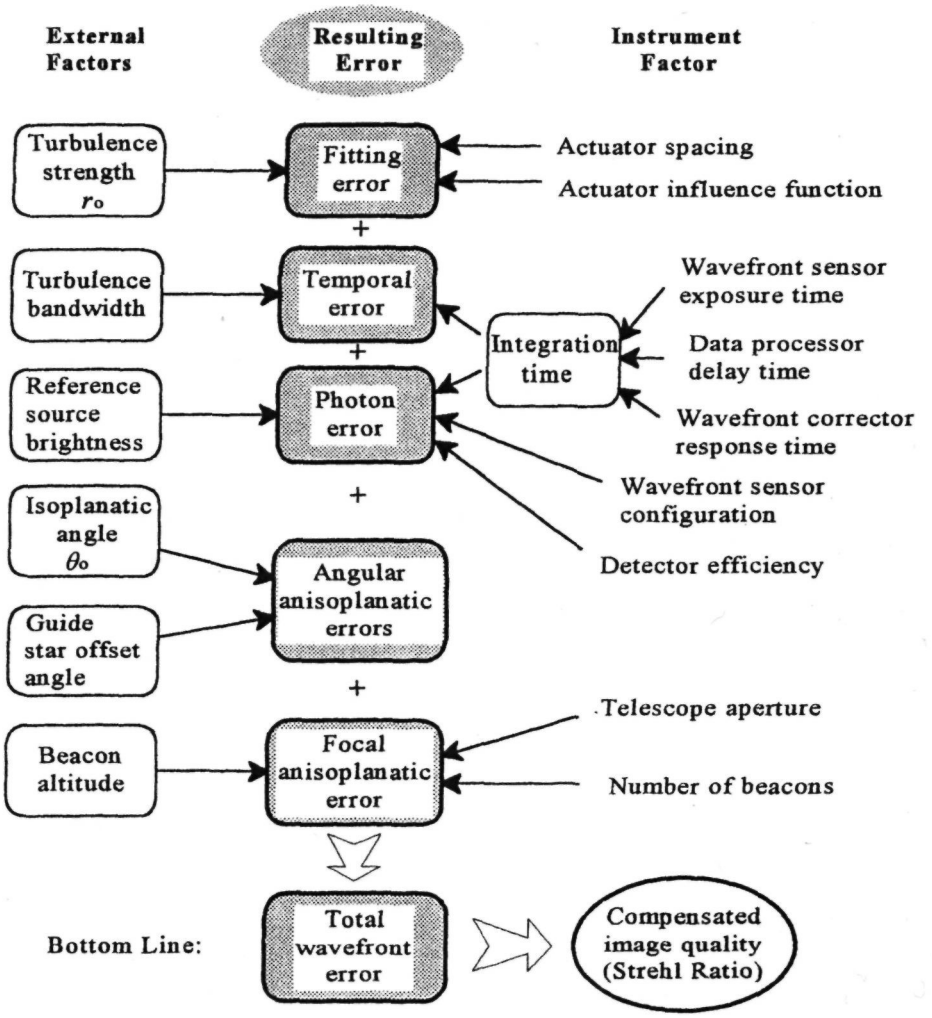


Figure 2.32 Main sources of wavefront error in adaptive optics.

tional optical components can be minimized; however, the resulting system may be less flexible and hardware modifications or upgrades more difficult and expensive. The use of an adaptive secondary mirror, as indicated in figure 2.35, is an elegant method of including adaptive optics into the design of Cassegrain or Gregorian telescopes. In some cases, it may be possible to replace an existing secondary with an adaptive component. This approach has the advantage of using an existing optical surface in the telescope, making it particularly attractive for IR operation. Adaptive secondary mirrors usually have the capability of compensating image motion, as well as higher order aberrations. A disadvantage of adaptive secondaries is that they are not located at a conjugate image of the turbulence, resulting in a small isoplanatic angle that cannot be enlarged without additional components.

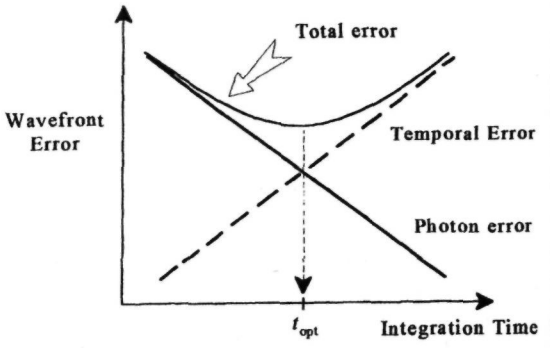


Figure 2.33 Optimization of integration time. The integration time of the adaptive optics control loop controls the photon (shot noise) error and the temporal (time delay) error, shown here on a log-log plot. The total error is minimized when the temporal and photon errors are equal.

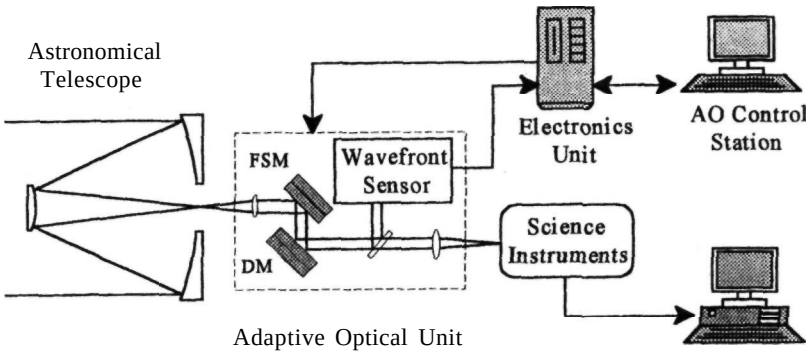


Figure 2.34 Adaptive optics installation at an existing focal station on an astronomical telescope. The adaptive optical unit contains the wavefront corrector, wavefront sensor, and interface optics. It is preferably mounted at the Cassegrain focus, but may also be placed at a Nasmyth or coude focus. The electronics unit and adaptive optics (AO) control station may be located some distance from the telescope. The laser beacon would be projected through a separate optical system. FSM = fast steering mirror; DM = deformable mirror.

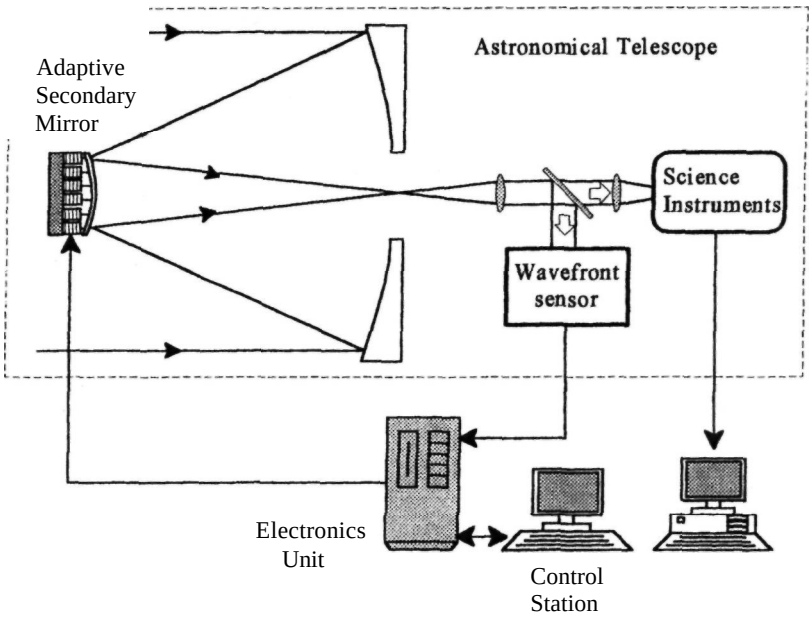


Figure 2.35 Adaptive optics system using a deformable secondary mirror. If the secondary mirror also has fast guiding capability, then no additional optics are required in the telescope imaging path other than a beam splitter. The optical path to the wavefront sensor may be designed as an intrinsic part of the telescope optical system.

In all adaptive optics systems, the location of pupil planes must be controlled carefully. The basic requirement for pupil imaging is for the wavefront corrector to be imaged at a plane conjugate to the layer having the largest turbulence moment, and for the wavefront sensor detectors to be imaged onto the wavefront corrector. This arrangement maximizes the isoplanatic angle and ensures independent measurement of the wavefront errors in each zone of the aperture. If a separate tilt corrector is employed, as is usually the case, it should be located at a conjugate image of the telescope entrance pupil, to eliminate vignetting due to angle-of-arrival variations.

2.11.2 The Operator Interface

It is recognized that the practical utility of adaptive optics depends greatly on the presence of a convenient and user-friendly interface with the telescope operator. The trend in design is to minimize the need for operator inputs by making the system autonomous. Setup and calibration procedures should be automatic, as should routine adjustments needed to optimize performance. The parameters most likely to vary during an observing session are the turbulence strength r_0, the isoplanatic angle θ_0, and the turbulence change time t_0. These variations are accommodated in the adaptive optics by optimizing the integration time and the control loop gain. The latter function is implemented by changing the coefficients or "weights" in the reconstruction matrix.

Although most parameters are optimized automatically, it is important for the operator to have a real-time presentation of all vital system parameters, as well as the capability for overriding commands and establishing manual control if necessary. This is especially necessary during the initial test and evaluation period, while the data required to establish a multiparameter automatic control system are being accumulated.

3 Optical Effects of Atmospheric Turbulence

"Everything should be made as simple as possible, but not simpler."
Albert Einstein

3.1 Introduction

The purpose of adaptive optics in astronomical telescopes is to improve the image quality by compensating the wavefront distortion due to atmospheric turbulence. A basic understanding of the effects of turbulence on optical image formation is necessary for the design and evaluation of adaptive systems; it is also the key to the development of new concepts and techniques. The characteristics of atmospheric turbulence and its effect on the propagation of light through Earth's atmosphere are reviewed in this chapter.

Adaptive optics systems designed to compensate atmospheric turbulence will also correct residual quasi-static aberrations in the optical components, which are usually dominated by figure errors in the primary mirror. Wavefront errors of this kind change slowly, on a time scale of greater than tens of seconds, compared with the millisecond time scale of atmospheric turbulence. Most modern telescopes with large apertures employ a separate low-bandwidth "active optics" control system to maintain the figure of the primary mirror in the presence of temperature and gravitational changes. Active control of the primary mirror is usually implemented separately from the adaptive optics, reducing the range of figure correction needed.

The requirements for real-time compensation of atmospheric turbulence are far more demanding than those of active optics in terms of spatial and temporal bandwidth. To begin, it is useful to get an overall picture of the turbulent atmosphere and its interaction with optical waves. The structure of Earth's atmosphere and a profile of turbulence strength are shown diagrammatically in figure 3.1. The density of the atmosphere falls exponentially with altitude above sea level. The temperature variations that give rise to optical effects of turbulence are only significant in the troposphere, below about 10 km. The vertical distribution of turbulence varies greatly with location and time. On average, turbulence is usually greatest near ground level and falls off exponentially with increasing altitude, except for a peak that often occurs at the tropopause, due to wind shear. As a result of the exponential distribution, the effects of turbulence are much less severe in near-vertical paths than in horizontal propagation paths of comparable length, for which the turbulence strength is roughly constant. For near-vertical paths, the variations in refractive index amount to a few parts in a million. This is significant because, although astronomical images appear to be degraded considerably by turbulence, they are, in fact, much more easily compensated than those obtained over long horizontal or slant ranges.

Analysis of the effects of turbulence on optical images is simplified by the large difference in scale, both spatial and temporal, between the turbulence and the radiation passing through it. A typical scale size for turbulence eddies at visible wavelengths is 10 cm, which is a factor of 200,000 larger than the wavelength of light. This means that diffraction angles due to turbulence are very small when looking up through the atmosphere, on the order of 5×10^{-6}

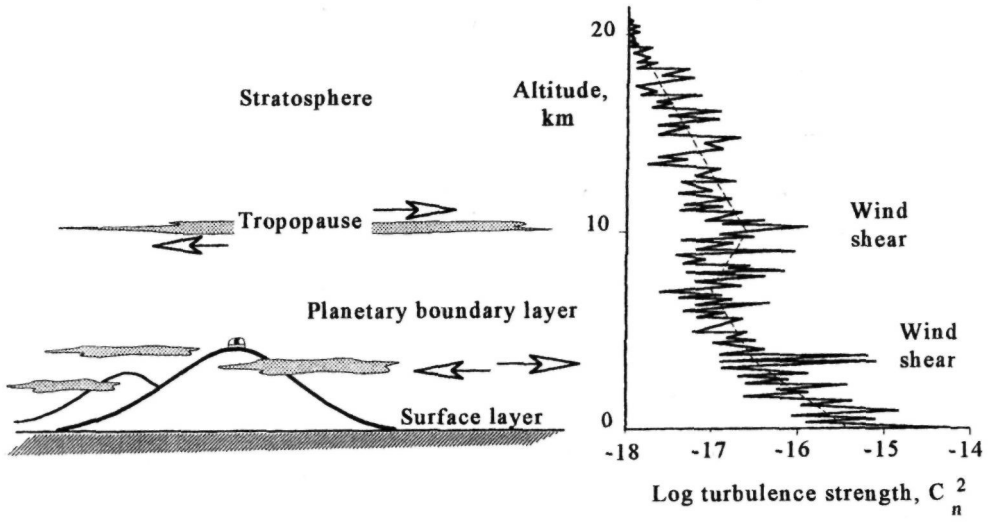


Figure 3.1 Structure of Earth's atmosphere and typical turbulence profile.

rad or about 1 arc second. In many cases, an adequate description of wavefront phase distortion can be obtained by considering only the refractive variations, which can be summed along straight ray paths in the atmosphere, using what is known as the geometrical approach.

With regard to temporal variations, there is a difference of about 12 orders of magnitude between the time period of one cycle of the optical radiation and the relatively slow changes in the structure of turbulence, which occur on a time scale of milliseconds. These large differences in scale facilitate analysis of optical effects, because not only can diffraction or scattering effects be neglected, at least in near-vertical paths, but the atmospheric structure can be considered as frozen in time as far as the radiation is concerned.

Light rays entering the atmosphere at angles away from the zenith are bent downward by the gradual change in refractivity, which is proportional to air density. This is a systematic effect that is well known to astronomers. The random effects of air turbulence are superimposed on this refraction. The dispersion of air over optical wavelengths is small but not negligible: its effect is to elongate broadband images of stars in the vertical direction. When high-resolution imaging techniques are employed with broadband sources, and especially with adaptive optics, it is necessary to compensate the atmospheric dispersion with a correcting prism in the telescope.

Atmospheric dispersion produces additional troublesome effects in multispectral adaptive optics in which the wavefront sensing and observing wavelengths are different. Dispersion effects are discussed in detail in section 8.6.

The sections of this chapter are organized as follows. In section 3.2, the model of mechanical turbulence proposed by Kolmogorov and the derived model of refractive index variations that produce the optical effects of atmospheric turbulence are described. The structure of Earth's atmosphere and its dynamics as a function of time and altitude are summarized in section 3.3. This section also includes a description of the analytical models for turbulence strength and wind velocity that have been developed for use in system analysis.

Section 3.4 reviews the various approaches that have been used by Tatarski [1961], Hufnagel and Stanley [1964], Fried [1966a], and others to analyze the optical effects of turbulence. The parameters of importance to adaptive optical systems are derived using the phase-screen approach developed by Lee and Harp [1969]. The effects of turbulence on long- and short-exposure images are reviewed. Decomposition of turbulence-distorted images into systematic modes such as Zernike polynomials is described in section 3.5.

A powerful method of analyzing the optical effects of turbulence using transverse spatial filters [Sasiela 1988] is introduced in section 3.6. This method is particularly useful in problems involving focal and tilt anisoplanatism, which are discussed further in chapter 7. In section 3.7, the important subject of isoplanatism and its dependence on the three-dimensional structure of the atmosphere is summarized.

3.2 The Mechanics of Turbulence

3.2.1 Kolmogorov Model

To analyze a complex, random phenomenon, such as atmospheric turbulence, it is necessary to develop a

model of the basic physical processes involved. The mechanical structure of turbulence was investigated by Kolmogorov [1941], who proposed a model for the velocity of motion in a fluid medium. This simple model explains most of the observed phenomena and has achieved wide acceptance. Kolmogorov's model assumes that energy is added to the fluid medium in the form of large-scale disturbances (the "outer scale"), which then break down into smaller and smaller structures. This process is characteristic of turbulent flow, which occurs when the Reynolds number exceeds a critical value that depends on the geometrical structure of the flow. The Reynolds number is defined as $Re = V_0 L_0/\nu_0$, where V_0 is the characteristic velocity, L_0 is the characteristic size of the flow, and ν_0 is the kinematic viscosity of the fluid. For air, $\nu_0 = 15 \times 10^{-6}$ m^2 s^{-1}. In the case of a moderate-sized atmospheric disturbance, with a scale size of 15 m and velocity of 1 m s^{-1}, the Reynolds number is 1×10^6, which greatly exceeds the critical value. As a result, airflow in the atmosphere is nearly always turbulent.

The ultimate source of energy in Earth's atmosphere is solar heating, which produces disturbances over a large range of scale sizes. During the day, the energy input occurs in the form of local convection cells generated directly by solar radiation on Earth's surface, while at night, the major energy source is the mixing of air masses of different temperatures and altitudes by the wind. As the turbulent flow breaks up, the kinetic energy is continually transferred to motions of smaller and smaller scale, forming an "energy cascade," until the Reynolds number drops below its critical value. At this stage, termed the "inner scale," the kinetic energy is dissipated into heat by molecular (viscous) friction and the turbulence dies away.

For the process to continue in a stable state, the rate of viscous dissipation must equal the rate of input turbulent energy. This condition implies that at any physical scale within the inertial range, the velocity fluctuations V are governed only by the scale size ℓ and the rate of energy input and dissipation ε. Dimensional considerations then lead to the relation

$$V \propto \varepsilon^{1/3} \ell^{1/3} \tag{3.1}$$

which implies that the nuctuational energy in perturbations of size ℓ is proportional to $\ell^{2/3}$. This is true for all scales within the inertial range. The size of the fluctuations depends only on the rate of energy input and dissipation ε. The relation between the inner scale ℓ_0 and the outer scale $\mathcal{L}_0$ is

$$\ell_0 = \frac{\mathcal{L}_0}{(Re)^{3/4}} \tag{3.2}$$

Thus, the greater the velocity of the flow, the larger is the Reynolds number and the smaller the inner scale of the turbulence.

3.2.2 Power Spectrum of Turbulence

The power spectrum of turbulence may be deduced from equation (3.1). The spatial wave-number for an eddy of scale size ℓ is defined as $\kappa = 2\pi/\ell$. If the spectral density of the fluctuations is $\Phi(\kappa)$, then the energy in an increment $d\kappa$ is proportional to V^2, or

$$\Phi(\kappa)d\kappa \propto V^2 \propto \kappa^{-2/3} \quad \text{or} \quad \Phi(\kappa) \propto \kappa^{-5/3} \tag{3.3}$$

This relationship, which defines the Kolmogorov spectrum in one dimension, is valid within the inertial range; that is, for $\mathcal{L}_0 > \ell > \ell_0$.

3.2.3 Structure Functions

Structure functions were introduced by Kolmogorov in 1941 to describe non stationary random functions, such as those encountered in turbulence theory. The reasons for using these functions are discussed by Tatarski [1961] in his seminal book on wave propagation; they are summarized as follows. The mean values of many meteorological variables, such as temperature, pressure, and humidity, are continually changing over time scales of minutes to hours. These changes are relatively smooth but may represent large differences in the mean value of a variable. When these variables are analyzed, a problem arises in distinguishing changes in mean value from those changes due to slow fluctuations.

To avoid this difficulty, in the case of a non stationary random function *fit)* where the mean value varies over time, the difference function $F_t(\tau) = f(t + \tau) - f(t)$ is used. Thus, slow changes in *fit)* do not affect the difference r, provided it is not too large, so that $F_t(\tau)$ may be considered a stationary random function of time, even though *fit)* is not. The structure function is then defined as

$$D_f(\tau) = \langle [F_t(\tau)]^2 \rangle = \langle [f(t + \tau) - f(t)]^2 \rangle \tag{3.4}$$

where ( ) represents the average value. The structure function D_f (r) is a measure of the intensity of the fluctuations in *fit)* over periods that are comparable to or smaller than r.

Structure functions are more general than correlation functions and may also be used to describe ordinary stationary functions. If *fit)* is a stationary random function with zero mean value, and the correlation function B_f (r) is defined as $\langle [f(t + \tau)f^*(t)] \rangle$, it follows that

$$D_f(\tau) = 2[B_f(0) - B_f(\tau)] \tag{3.5}$$

where B_f (0) is the mean-square value of the function. For most functions, the correlation at large separations is zero, so when $r = \infty$, $D_f(\infty) = 2B_f(0)$. Thus, for a random process with zero mean, such as atmospheric turbulence, the limiting value of the structure function at large separations is twice the mean-square value. This, of course, is just what would be expected, because the mean-square values of two independent quantities are being added. In these equations, the

arguments / and r can, of course, represent any physical quantity for which the function has appropriate statistics. In turbulence theory, the structure function is mainly used to describe the spatial characteristics of the medium, where (> represents a spatial average.

3.2.4 Temperature and Refractive Index Variations

Using the relationship in equation (3.1) and assuming that the medium is locally homogeneous and isotropic, it is possible to define a structure function between two components of velocity separated by distance r, along a coordinate x:

$$D_V(r) = |V(x) - V(x+r)|^2 = C_V^2 r^{2/3} \tag{3.6}$$

where the velocity structure constant C_V^2 is a parameter that depends on the energy of the process.

Kolmogorov's analysis describes the mechanical properties of turbulence. If the medium is in thermal equilibrium with no humidity variations, it is possible for physical turbulence to exist without producing any optical effects. These effects are produced only by variations in refractive index. In Earth's atmosphere, temperature and humidity vary with altitude. Mechanical turbulence produces mixing of air from different altitudes, causing variations in temperature and humidity of similar scale to the underlying turbulence. Obukhov [1949] and Yaglom [1949] showed that Kolmogorov's law also applies to additives in the turbulent medium, provided that they do not affect the dynamics or chemical composition of the medium. Such is the case with the mixing of air at different temperatures and the addition of water vapor, which consequently produce optical effects that have the same structure as mechanical turbulence. A temperature structure function based on Kolmogorov's law may therefore be defined as follows

$$D_T(r) = |T(x) - T(x+r)|^2 = C_T^2 r^{2/3} \tag{3.7}$$

where C_T is the structure parameter for temperature variations.

Temperature fluctuations affect the density of the air and therefore change its refractive index n. The refractive index of air depends on its temperature, its pressure, and the concentration of water vapor. Water vapor can be a significant factor in horizontal propagation, but has little effect on the refractive index in vertical propagation. The refractivity of air is closely approximated by

$$N = (n - 1)10^6 = 77.6P/T \tag{3.8}$$

where

P = pressure, millibars
T = temperature, K

Local pressure fluctuations are smoothed out at the speed of sound and are negligible compared with temperature fluctuations. Therefore, the refractive index fluctuations for vertical propagation are due mainly to temperature variations and can be be expressed as

$$\delta N = -77.6\left(\frac{P}{T}\right)\delta T \tag{3.9}$$

The structure function of refractive index variations is then

$$D_N(r) = C_N^2 r^{2/3} \tag{3.10}$$

where

$$C_N = \frac{\delta N}{\delta T} C_T \tag{3.11}$$

Atmospheric turbulence is often measured by using temperature probes that determine the value of C_T^2. In terms of this parameter, the refractive index structure function is given by

$$D_N(r) = \frac{78P}{T^2} C_T^2 r^{2/3} \tag{3.12}$$

Equation (3.3) indicates that the one-dimensional power spectrum of temperature fluctuations has the form

$$\Phi_T(\kappa) \propto \kappa^{-5/3}$$

To analyze propagation through turbulence, the three-dimensional power spectrum is needed. Tatarski [1961] has shown that this power spectrum is given by

$$\begin{aligned}\Phi_T(\kappa) &= \frac{\Gamma\left(\frac{8}{3}\right)\sin\left(\frac{\pi}{3}\right)}{4\pi^2} C_T^2 \kappa^{-11/3} \\ &= 0.033 C_T^2 \kappa^{-11/3}\end{aligned} \tag{3.13}$$

Similarly, the three-dimensional power spectrum of refractive index variations can be expressed as

$$\Phi_N(\kappa) = 0.033 C_N^2 \kappa^{-11/3} \tag{3.14}$$

The refractive index power spectrum of turbulence is shown in figure 3.2. Energy is input at the outer scale, where $\mathcal{L}_0$ is on the order of 10 m, and is dissipated at the inner scale ℓ_0, which is usually a few millimeters. The form of the spectrum outside these values is uncertain. At values of κ above $\kappa_m = 5.92/\ell_0$, Tatarski assumes that the spectrum is modified by an additional term $\exp(-\kappa^2/\kappa_m^2)$, producing the steep drop-off as shown. At values of κ below $2\pi/\mathcal{L}_0$, Strohbehn [1968] suggests a flattened spectrum that drops off slowly with increasing values of

3.2.5 Limitations of Kolmogorov Turbulence Model

The Kolmogorov turbulence model giving the "two-thirds law" [equation (3.3)] assumes an incompressible medium and it is only valid within the inertial range between the inner and outer scales of the turbulence. Roddier (1981) states that the inner scale

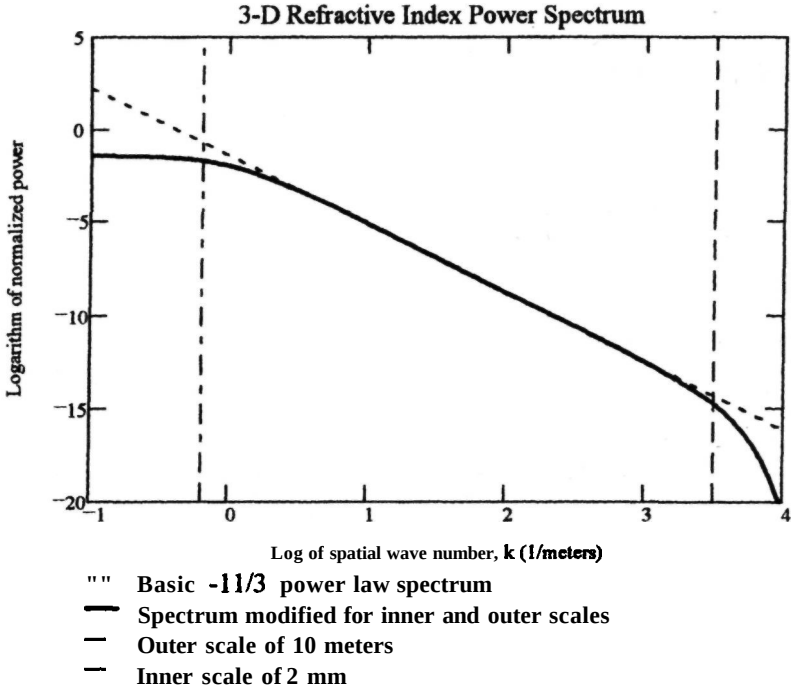


Figure 3.2 Three-dimensional power spectrum of refractive index variations. The vertical dotted lines represent spatial wave numbers $(2\pi/\lambda)$ for an outer scale of 10 m and an inner scale of 2 mm. The drop-off above the inner scale is due to dissipation of the turbulent energy. The flattening of the spectrum below the outer scale is necessary to represent a finite energy input.

varies from about 1 mm, near the ground, to about 1 cm, near the tropopause. The size of the outer scale has been the subject of much debate. There is no doubt that it varies considerably; estimates range from about 1 m to more than 100 m. At the lower end of this range, the outer scale is therefore comparable to the apertures of astronomical telescopes. For adaptive optics, the size of the outer scale determines the magnitude of the overall tilt component of atmospheric turbulence. In calculating overall tilt, it is usually assumed that $\mathcal{L}_0 \gg D$. When this is not the case, the overall tilt is smaller than predicted. In the analysis of adaptive optics performance, a small outer scale may be modeled by using the von Karman spectrum, as described in section 3.6.

Measurements made at several sites by Coulman et al. [1988] suggest that the outer scale varies strongly with altitude, reaching a maximum value of 5 m at an altitude of 7.5 km. Below 3 km, the outer scale observed was often less than 1 m. Based on their measurements at the European Southern Observatory at La Silla, Chile, Coulman et al. suggest the following model for the outer scale of turbulence at good sites:

$$\mathcal{L}_0(h) = \frac{5}{1 + \left(\dfrac{h - 7500}{2000}\right)^2} \quad \text{meters} \qquad (3.15)$$

Beckers [1993] has suggested that the outer scale of turbulence is not a unique quantity. Turbulent energy, due to various sources such as solar heating, wind shear, and the local environment of the telescope, is fed into the atmosphere at many different scales, and the outer scale at any time may depend on the dominant source.

A more fundamental issue concerns the validity of the basic mechanism by which atmospheric turbulence is generated. In the Kolmogorov model, changes in the strength of turbulence should occur smoothly, in response to variations in the energy input. In fact, there is evidence that turbulence is often intermittent, a condition in which small-scale structures of turbulence occur in bursts, with intervening quiescent periods. Kim and Jaggard [1988] have proposed a band-limited fractal model of atmospheric refractivity to account for intermittency. They show that the averaged spectrum of the one-dimensional intermittent fluctuation function for the average fractal dimension of 5/3 approaches the

Kolmogorov spectrum over a large range of intermittency. It is surmised that the Kolmogorov spectrum is recovered as a special case of this model, for fractal dimensions that are the negative of the exponents in the Kolmogorov spectrum. For long paths and long averaging times, intermittency effects tend to average out.

Intermittency is important in adaptive optics because turbulence strength is a key factor in real-time performance optimization. System parameters such as the servo bandwidth ideally should be adjusted continually in response to the observing conditions, to maximize the imaging performance. If sudden changes in turbulence level are ignored, then the adaptive optics performance will either drop drastically when the turbulence level rises significantly, or, if the system has been optimized for high turbulence, it will not deliver the best possible images during periods of good seeing.

3.3 Modeling Earth's Atmosphere

3.3.1 Physical Structure

Earth's atmosphere extends to an altitude of about 300 km. Its pressure decreases smoothly from the sea-level value of about 1000 millibars, following an exponential law with an initial gradient of -12.1 millibars per 100 m. At 30 km, the pressure has fallen to 10 millibars and there is little turbulence above this altitude. At 90 km, the altitude of the sodium layer, the pressure is only 10^{-6} of that at sea level. Turbulence has a different profile, occurring mainly in thin layers where wind shear is present. Considerable variations in turbulence strength have been observed over height intervals as small as 100 m. During the day, turbulence is usually strongest near the ground, in response to solar heating and the development of thermal currents. At night, especially at good observing sites, surface turbulence is low and most disturbances occur in layers at higher altitudes, especially at the tropopause, as a result of wind shear.

The best observing sites are located on mountain peaks near the ocean, which has a moderating effect on temperature variations. The wind structure is generally more consistent at latitudes within 35° of the equator, and isolated peaks on the windward side of a landmass avoid orographic disturbances.

3.3.2 Refractivity

The refractive index of the atmosphere is close to unity and it is convenient to describe it in terms of the refractivity, defined as the departure from unity in parts per million, $N = (n - 1) \times 10^6$. The refractivity of the atmosphere depends on its exact composition, including the presence of carbon dioxide and water vapor. Formulas for computing the refractivity to an accuracy on the order of 1×10^{-8} have been given by Owens [1967]. At optical wavelengths, the refractivity may be approximated with sufficient accuracy for the analysis of optical imaging systems by the simple expression given in equation (3.8):

$$N \approx \frac{77.6P}{T}$$

where

P = pressure, millibars
T = temperature, K

The value of N is about 280 at sea level, falling to about 95 at an altitude of 10 km.

The refractivity of dry air as a function of wavelength at standard temperature and pressure is given by the following formula [Edlen 1966]:

$$N_S(\lambda) = 8.34213 \times 10^{-5} + \frac{2.40603 \times 10^{-2}}{130 - \lambda^{-2}} + \frac{1.5997 \times 10^{-4}}{38.9 - \lambda^{-2}} \tag{3.16}$$

where λ is the wavelength in micrometers. The variation of refractivity of air over the visible and near-IR regions is shown in figure 3.3. The slope of the refractivity curve gives the spectral dispersion, which can produce errors in adaptive optics systems that use wideband or multispectral operation. At wavelengths above 1 μm the dispersion is negligible, but at visible wavelengths, especially below 0.5 μm, it increases rapidly. The effects of dispersion are discussed in section 9.3, and they must be taken into account in adaptive optics systems in which either the wavefront sensor or the observation wavelength are in the visible band.

For an atmosphere with a fixed composition, such as Earth's troposphere, the refractivity depends on the density, which is a function of altitude:

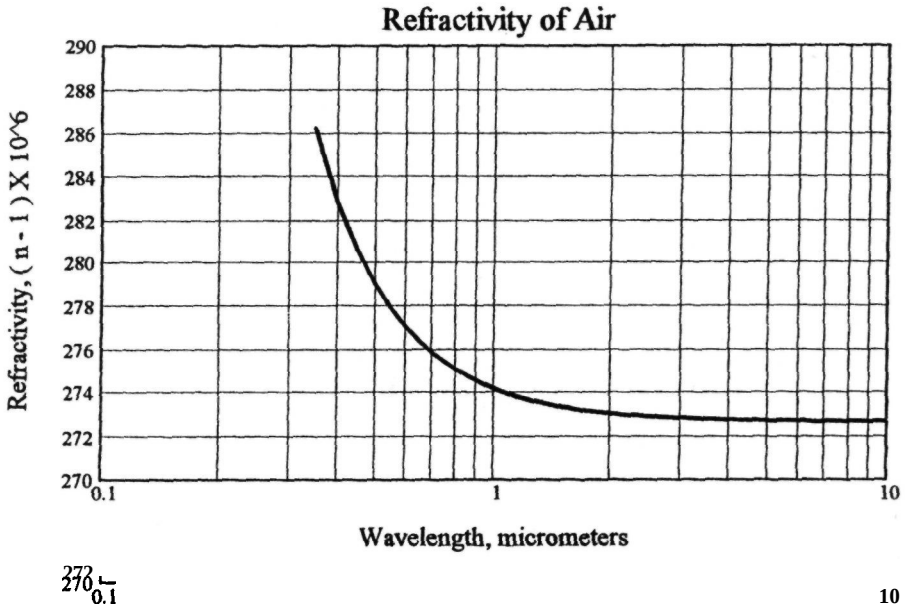
$$N(h, \lambda) = N_S(\lambda)\frac{\rho(h)}{\rho_S} \tag{3.17}$$

where $\rho(h)$ is the atmospheric density at altitude h and ρ_S is the standard density of the atmosphere, equal to 0.001225 g cm . In the equilibrium state, the density is proportional to P/T. The pressure P decreases exponentially with height, as shown in figure 3.4.

3.3.3 Atmospheric Turbulence Profile

The turbulence profile of the atmosphere is a critical factor in the design of adaptive optics systems. The profile consists of three main regimes:

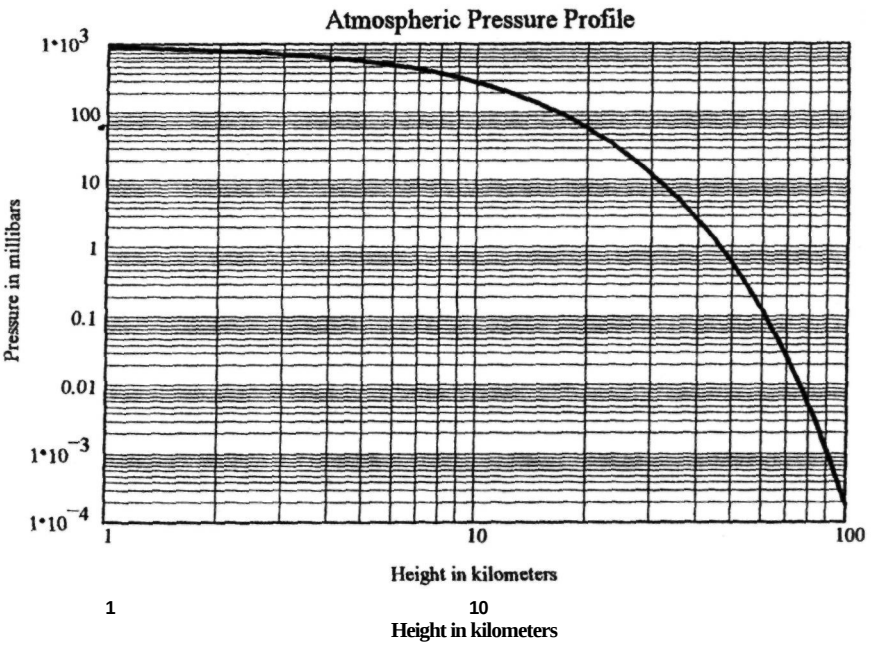
(1) Surface layer The surface layer extends to about 1 km above the ground and usually contributes most of the optical path fluctuations, except at mountaintop sites where it may be insignificant at night.



272
270
0.1
10
Wavelength, micrometers

Figure 3.3 Refractivity of dry air at standard temperature and pressure [Edlen, 1966].

Turbulence in this layer depends greatly on the local topography and climatic conditions. During the day, solar irradiation causes active convection, which may produce a factor-of-ten increase in the value of C_N^2 near the ground. Wyngaard et al. [1971] predicted that under these conditions the turbulence would follow an $h^{-4/}$ law, which has been confirmed by observations. Turbulence is lowest in this region just after sunrise and just after sunset, when the lower atmosphere is in thermal equilibrium. At night, radiative cooling of the ground often causes a small increase in turbulence. Turbulence near the ground is usually measured using high-speed thermal probes. Acoustic sounding has been used to measure the strength of turbulence up to about 1 km. In the surface layer, the average value of C_N^2 drops from about 10^{-15} to 10^{-14} $m^{-2/3}$ near the ground to about 10^{-16} $m^{-2/3}$ at 1 km.



1
10
Height in kilometers

Figure 3.4 Variation of atmospheric pressure with altitude.

(2) Planetary boundary layer The planetary boundary layer occupies the region above the surface layer up to about 8 or 10 km above sea level, in which the average turbulence level drops to around $10^{-17}\,\mathrm{m}^{-2/3}$. Within this region, turbulence may occur in several thin layers, 100 to 200 m thick, in which C_N^2 increases locally by more than a factor of 10. Measurements in this region have been made by Barletti et al. [1976] using balloon-borne thermal probes.

(3) Tropopause At the tropopause, around 10 km, there is often a peak in turbulence due to wind shear. Above this level, there is a rapid decrease, and turbulence effects disappear at altitudes around 25 km.

These three regimes of the turbulence profile affect the imaging properties of astronomical telescopes in different ways. All layers combine to produce the total wavefront phase disturbance that is characterized by the seeing parameter r_0. High-altitude turbulence layers are the dominant cause of intensity variations (scintillation) and are also the controlling influence on the isoplanatic angle that restricts the field of view over which adaptive optics compensation is possible. The integrated value of turbulence measured by the parameter r_0 is therefore an incomplete description of atmospheric turbulence. The distribution of turbulence with height must be known or specified in order to determine the performance of adaptive optical systems.

In addition to atmospheric turbulence, there are usually local sources of turbulence in and around the telescope enclosure. Observing sites having excellent seeing in their pristine state have often been degraded by the presence of man-made structures that create local turbulence and dissipate heat. To minimize these effects, such facilities are usually located downwind from the telescope.

Turbulence within the dome and telescope structure is caused by small temperature differentials, which can usually be reduced to acceptable levels by the use of passive and active thermal control. Radiation cooling can be controlled by the use of reflective surfaces. The diurnal temperature cycle can be largely suppressed by refrigerating the dome, as implemented in the Canada-France-Hawaii Telescope (CFHT) at Mauna Kea.

3.3.4 Turbulence Models

The optical effects of turbulence are produced by incomplete mixing of air at different temperatures. The key parameter is the refractive index structure constant C_N^2, which varies with altitude and time. Turbulence is a random process and the refractive index structure constant can vary by an order of magnitude around its median value. To produce the models that are needed for the design and evaluation of adaptive optics systems, measurements must be made over extended periods of time. The C_N^2 profiles seen in turbulence models represent data that are smoothed temporally and spatially; it must be kept in mind that the raw turbulence profiles are much more irregular.

A survey of optical turbulence models has been made by Good et al. [1988]. One of the most useful models is that of Hufnagel [1974], who proposed a heuristic model of the C_N^2 profile of the atmosphere for altitudes between 3 and 24 km, based on turbulence measurements made by many observers. Hufnagel found that stellar scintillation, which is due to high-altitude turbulence, is well correlated with the factor w^2 given by

$$w^2 = \left(\frac{1}{15\ \mathrm{km}}\right)\int_{5\ \mathrm{km}}^{20\ \mathrm{km}} v^2(h)\,dh$$

where $v(h)$ is the wind speed in meters per second at altitude h above sea level. This implies that the wind velocity, through factor w^2, controls the strength of the turbulence at high altitudes, as well as its temporal spectrum. The Hufnagel turbulence model has two components: (1) an exponentially decreasing value of C_N^2 through the troposphere, with a $1/e$ scale height of 1500 m; and (2) a peak at the tropopause around 10 km, scaled by the w^2 factor. The turbulence strength falls off rapidly at higher altitudes. The original model has the following form:

$$C_N^2(h) = A\left[2.2\times 10^{-53}h^{10}\left(\frac{w}{27}\right)^2\exp\left(-\frac{h}{1000}\right) + 1\times 10^{-16}\exp\left(-\frac{h}{1500}\right)\right]$$

The factor A models the fine structure of the turbulence as a function of height and time. It is defined as $A = \exp[r(h, t)]$, where r is a Gaussian random variable with zero mean, h is the height, and t is time. Atmospheric turbulence profiles measured directly by means of temperature probes show considerable local variations in turbulence strength, with a vertical scale size on the order of 100 m. Optical measurements of turbulence show variations over a time scale of minutes. This fine structure is generally ignored in the models used for the design of adaptive optics systems. The average turbulence profile is obtained by setting $A - e \approx 2.7$.

The Hufnagel model describes the turbulence structure above 3 km at mid-latitudes where tropospheric winds are high. Below 3 km, there are large local and diurnal variations in turbulence, produced by convection from solar heating. To satisfy the needs of the defense community for the design of adaptive optics systems that must operate under any conditions, day or night, Valley [1980] suggested the addition of a term for surface layer turbulence, which can be strong, especially at desert sites during

the day. The resulting **"Hufnagel-Valley"** (HV) turbulence model is given by

$$C_N^2(h) = A\left[2.2 \times 10^{-53} h^{10} \left(\frac{w}{27}\right)^2 \exp\left(-\frac{h}{1000}\right) + 1 \times 10^{-16} \exp\left(-\frac{h}{1500}\right)\right] + B\exp\left(-\frac{h}{100}\right) \quad (3.18)$$

The HV model can be adjusted for different seeing conditions, covering both night and day operation. The coefficient w controls the high-level turbulence strength, while the coefficient B, controlling the "Valley term," is the scale factor for surface turbulence.

Turbulence profiles measured by Ochs et al. [1976], Barletti et al. [1976], and Loos and Hogge [1979] generally confirm the basic structure of the Hufnagel model above 1 km. A comparison of these profiles with the Hufnagel model is shown in figure 3.5. Also shown is Barletti's "lucky observer" model, based on the lowest C_N^2 values measured at each altitude. This represents hypothetically the best seeing conditions that can be expected.

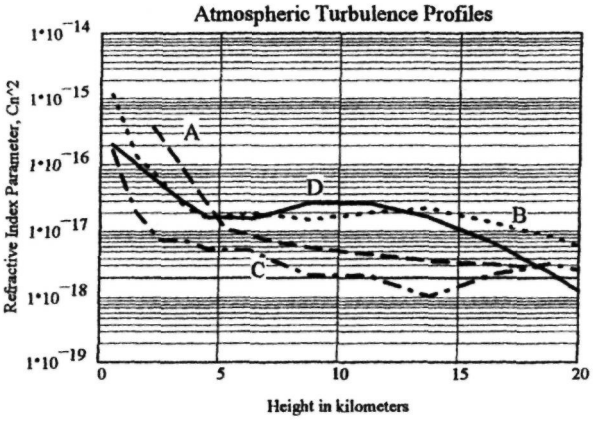
At high-altitude observing sites, such as Mauna Kea, Hawaii (4200 m), it has been found that the turbulence has a different character from that at lower elevations. There is little night-time turbulence in the surface layer, and turbulence in the troposphere tends to be concentrated in one or two thin layers. Turbulence profiles above Mauna Kea were measured by Roddier et al. [1990] using the scintillation detection and ranging (scidar) technique [Azouit and Vernin 1980] with the 2.2-m University of Hawaii telescope. The data consisted of 414 profiles representing 69 h of observing over 12 nights in November 1987 and 8 nights in June 1989. The profiles cover the range of 1-20 km with a vertical resolution of 300-500 m.

These data were further analyzed by Racine and Ellerbroek [1995], who found that the typical night-time turbulence profile consisted of an underlying background that had a C_N^2 value of less than 1×10^{-17}, on which were superimposed one or two thin layers of turbulence. These layers are produced by wind shear in the troposphere. One layer was usually present between 2 and 4 km, while a second, associated with the jet stream, occurred at altitudes between 6 and 8 km, sometimes persisting for days. The best seeing occurred when only the background turbulence was present, giving stellar image diameters with full-width at half-maximum (FWHM) of about 0.3 arc seconds at $\lambda = 0.5\ \mu m$. For 75% of the time, the seeing was dominated by strong layers. The seeing disk, as determined from the scidar, had a median FWHM of 0.5 arc seconds under these conditions, growing to 0.9 arc seconds during periods of bad seeing. Low-level turbulence (< 1 km) had little effect on the image spread.

The HV model can be generalized into the following form, capable of representing any C_N^2 profile as the sum of exponential terms:

$$C_N^2(h) = A\exp\left(-\frac{h}{H_A}\right) + B\exp\left(-\frac{h}{H_B}\right) + C h^{10} \exp\left(-\frac{h}{H_c}\right) + D\exp\left(-\frac{(h - H_D)^2}{2d^2}\right) \quad (3.19)$$

In this expression, A is the coefficient for the surface (boundary layer) turbulence strength and H_A is the height for its $1/e$ decay, B and H_B similarly define the turbulence in the troposphere (up to about 10 km), C and H_C define the turbulence peak at the tropopause, and D and H_D define one or more isolated layers of turbulence, with d being the layer thickness. Analytical models of this type can be adjusted to



A Loos and Hogge average
B Barletti et al. average
C Barletti "Lucky Observer"
D **Hufnagel** model

Figure 3.5 Comparison of measured turbulence profiles with Hufnagel model.

Table 3.1 Constants for Turbulence Models

Model	A (E-15)	H_A (m)	B (E-17)	H_B (m)	C (E-53)	H_C (m)	D (E-16)	H_D (m)	FWHM (arc seconds)	r_0 (m)	θ_0 (arc seconds)
Hufnagel	0		27	1500	5.94	1000	0		1.15	0.11	1.1
HV 5-7	17	100	27	1500	3.59	1000	0		2.5	0.05	1.4
HV 10-10	4.5	100	9	1500	2.0	1000	0		1.26	0.10	2.1
HV 15-12	2.0	100	7	1500	1.54	1000	0		0.84	0.15	2.5
Mauna Kea (1)	0		1	3000	1.63	1000	0		0.36	0.34	2.4
Mauna Kea (2)	0		1	3000	1.63	1000	1	6500	0.53	0.24	1.9

FWHM and θ_0 are specified at $\lambda = 0.5\ \mu$m.
Mauna Kea (1) models the background turbulence only (excellent seeing).
Mauna Kea (2) includes a dominant layer of turbulence (median seeing).

match measured C_N^2 values at specific altitudes, and are easily integrated to give the image diameter and values of r_0 and θ_0. Parameters used in typical turbulence models are listed in table 3.1.

Figure 3.6 shows a model of the Mauna Kea turbulence, generated using three terms from equation (3.19). The background turbulence is represented by an exponential decay in the troposphere with $B = 1 \times 10^{-17}$ and $H_B = 3000$ m, added to a shallow peak at the tropopause with $G = 1.63 \times 10^{-53}$ and $H_C = 1000$ m.. There is no surface layer, so $A = 0$. The isolated turbulent layer is represented by a peak at $H_D = 6500$ m, with $D = 1 \times 10^{-16}$ and d — 300 m. The background turbulence alone has a coherence length of $r_0 = 0.34$ m, which produces a seeing disk with FWHM of 0.36 arc second and isoplanatic angle of 2.4 arc seconds at $\lambda = 0.5\ \mu$m. Addition of the single layer reduces r_0 to 0.24 m, giving a seeing disk of 0.53 arc seconds and an isoplanatic angle of 1.9 arc second.

The presence of a single dominant layer of turbulence is very favorable for the use of adaptive optics. As shown in section 2.4, if a single wavefront corrector is placed at the conjugate image of a turbulent layer, then not only is the isoplanatic angle maximized but also the diffraction effects that produce scintillation are minimized. On the basis of the Mauna Kea data, Racine and Ellerbroek [1995] estimate that the effective isoplanatic angle can be increased by a factor of between 1.7 and 4.1, by conjugating a single corrector to the turbulent layer, rather than to the telescope pupil, with a median improvement factor of 2. Because the background turbulence is low, the (uncompensated) resolution achieved outside the isoplanatic angle is fairly good — about 0.3 arc second. In this case, a single adaptive corrector neutralizes the

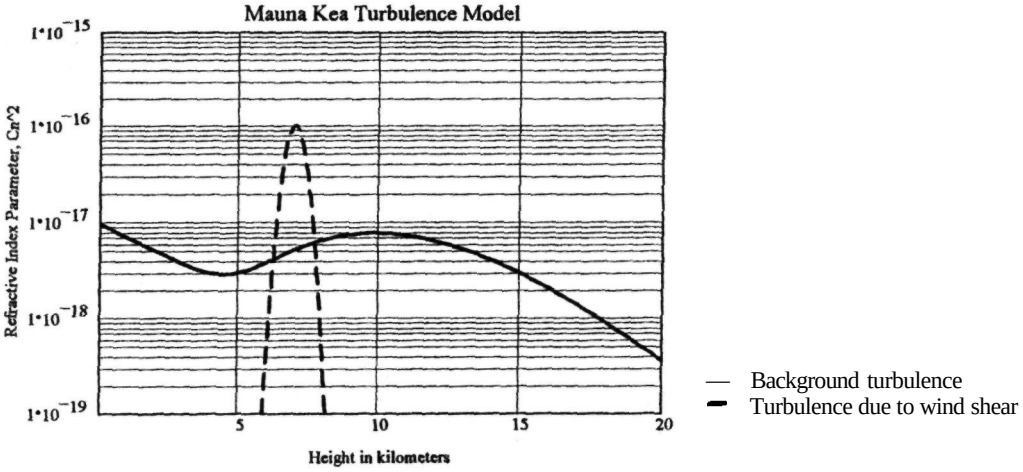


Figure 3.6 Turbulence model for a high-altitude observing site, based on measurements made at Mauna Kea by Roddier et al. [1990]. The background turbulence is always present, defining the seeing disk during the best periods in which the FWHM is about 0.3 arc seconds at A = 0.5 μm. Thin turbulent layers caused by wind shear in the troposphere are present 75% of the time, degrading the seeing to 0.5-0.9 arc seconds. Ground-level turbulence is very low.

turbulence produced in the dominant layer, maintaining the resolution due to background turbulence over a large field of view.

3.3.5 Wind Profiles

Wind has two effects on atmospheric turbulence: it controls both the intensity and the temporal frequency of the wavefront fluctuations. The intensity of the turbulence, measured by the refractive index structure parameter C_N^2, depends on the energy of the process and therefore increases with the square of wind velocity, as indicated by the factor w^2 in equation (3.18). The temporal bandwidth of the wavefront fluctuations increases linearly with wind velocity. The strongest winds normally occur between 9 and 12 km, the wind shear often generating a layer of turbulence at the tropopause. Evidence of a multilayer structure of turbulence produced by wind shear at altitudes of 5 and 10 km has been reported by Vernin and Roddier [1973]. The thickness of the turbulent layers was estimated to be on the order of 2.5 km.

Bufton [1973] reports simultaneous measurements of , wind velocity, and temperature at altitudes up to 25 km, using balloon-borne instrumentation. The turbulence measurements made with microthermal wire sensors correlated reasonably well with optical measurements of stellar scintillation. The vertical resolution achieved was 50 m. Turbulence occurred in thin layers, correlated with temperature inversions and wind shear. The peak wind velocities occurred near the tropopause with an approximately Gaussian distribution.

Greenwood [1977] suggested the use of a Gaussian model for the average wind profile transverse to the telescope. A general model for wind velocity is

$$v(z) = v_G + v_T \exp\left[-\left(\frac{z\cos(\zeta) - H_T}{L_T}\right)^2\right] \times \left[\sin^2\phi + \cos^2\phi\,\cos^2\zeta\right]^{1/2} \quad (3.20)$$

where

v_G = wind velocity at low altitude
v_T = wind velocity at tropopause
ζ = zenith angle of observation
H_T = height of tropopause
L_T = thickness of tropopause layer
ϕ = wind direction relative to telescope azimuth

Wind profiles for v_G = 5 and 8 m s^{-1} and v_T = 15 and 30 m s^{-1} are shown in figure 3.7.

The peak wind speed at the tropopause varies greatly with location and season. Wind profiles given in the *Handbook of Geophysics and Space Environments* [Valley 1961] indicate that the average (50% probability) wind speed at the tropopause in winter varies from 23 m s^{-1} in the calmest area of the United States (Northeast) to 46 m s^{-1} in the windiest area (Northwest).

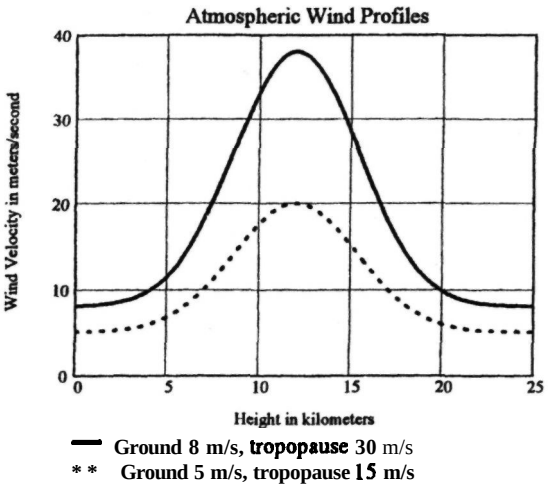


Figure 3.7 Models for average atmospheric wind speed.

The C_N^2 profile and wind speed models are used to generate the turbulence parameters required for adaptive optics system design and optimization, as described in chapter 9.

3.3.6 Turbulence Moments

In calculating the optical effects of turbulence, the refractive index structure parameter C_N^2 integrated over height, appears with several different weighting functions, depending on the particular error being evaluated. To simplify the calculations, these functions can be represented as full and partial turbulence moments, defined in the following way.

The full turbulence moment of order m is

$$\mu_m = \int_0^\infty dh\; C_N^2(h) h^m \quad (3.21a)$$

The partial turbulence moments at height H are

$$\text{Upper moment} \quad \mu_m^+(H) = \int_H^\infty dh\; C_N^2(h) h^m \quad (3.21b)$$

$$\text{Lower moment} \quad \mu_m^-(H) = \int_0^H dh\; C_N^2(h) h^m \quad (3.21c)$$

The turbulence-weighted velocity moment is

$$v_m = \int_0^\infty dh\; C_N^2(h) v^m(h) \quad (3.21d)$$

These moments are defined at the zenith; that is, at $\zeta = 0$. At other angles, they should be multiplied by

$\sec^{m+1}(\zeta)$. Turbulence moments define basic properties of the atmosphere and are independent of wavelength. The turbulence moments most frequently used are:

- zero-order, $m = 0$, for calculating r_0;
- five-thirds-order, $m = 5/3$, for calculating anisoplanatic and temporal errors;
- second-order, $m = 2$, for calculating overall tilt errors.

Higher order moments are used in the transverse filtering expressions (sections 3.6 and 7.3) to compute focal and tilt anisoplanatism.

3.4 Optical Effects of Turbulence

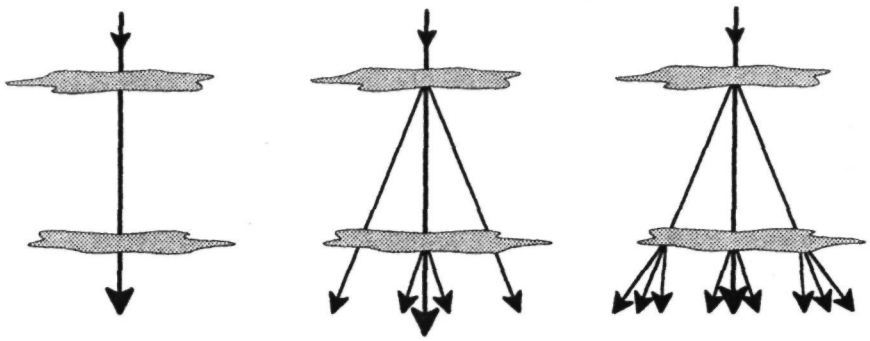
3.4.1 Analytical Approaches

The analysis of optical propagation through turbulence has progressed in several stages, from the earliest geometrical optics approach, accounting only for refraction, to more rigorous methods involving Maxwell's wave equation. Chernov [1960] and Tatarski [1961] developed an approximate solution to the wave equation, valid for weak scattering. This solution employs the so-called Rytov approximation, in which the amplitude of the wave is replaced by its logarithm. This approach was used subsequently by Fried [1965, 1966a] in his work on optical propagation, and it has more recently formed the basis of the transverse filtering technique developed by Sasiela [1988]. Another way of solving the wave equation was described by Hufnagel and Stanley [1964]. In this method, valid for small perturbations, the characteristics of a propagating wave are represented in its two lateral dimensions only, enabling an expression for the modulation transfer function (MTF) of a turbulent path to be found.

A different approach to solving optical propagation problems has been used by Lee and Harp [1969], in which the atmosphere is modeled by a series of phase screens. At each phase screen, the incident light generates two (first-order) diffracted beams that combine with the undiffracted light. In this way, single scattering is accounted for in a simple physical model. All of the solutions cited above produce essentially the same results in describing the optical effects of weak scattering at visible and IR wavelengths. These results generally agree with the observed properties of the atmosphere for near-vertical paths.

In horizontal paths near Earth's surface, strong turbulence is encountered. When the strength of the diffracted light becomes comparable to that of the undiffracted light, it is necessary to account for multiple scattering. These conditions have been treated by Uscinski [1977], Ishimaru [1978], and Strohbehn [1978]. The most successful approach is the "Markov approximation" [Tatarski 1971], which involves a parabolic approximation to the wave equation, followed by the use of delta functions to model the refractive-index fluctuations. These theories have been reviewed by Fante [1975].

The three main levels of complexity in calculating the optical effects of atmospheric turbulence are depicted in figure 3.8. The geometric optics approx-



(a) Geometric optics **(b) Single scattering** **(c) Multiple scattering**

Figure **3.8** Models for calculating the optical effects of turbulence. Using geometric optics (a), phase disturbances are summed along straight ray paths and scattering (diffraction) is ignored. This approach is valid for weak turbulence, such as that found at good astronomical sites. For moderate turbulence, the single-scattering approach (b) is appropriate. It accounts for scattering at each layer by the zero-order ray, but secondary scattering is ignored because it is much weaker. Single scattering is consistent with the Rytov approximation. With strong turbulence, such as that occurring on horizontal paths near the ground, a large fraction of the light is scattered, so that multiple scattering must be accounted for, as shown in (c). This situation is modeled by the Markov approximation.

imation involves summing the variations in refractive index along straight ray paths as shown in figure 3.8(a). It is valid when the wavelength A is much smaller than the scale of the disturbances, but only for weak turbulence and/or short propagation paths. The limitations to the geometric optics approach may be deduced from the simple model shown in figure 3.9. A plane wave of light is incident on a region of turbulence containing refractive-index variations. In the geometric optics approximation, the plane wave is represented by a bundle of parallel rays. The rays are advanced or retarded in phase by passing through regions of different refractivity, producing distortions in the wavefront.

In fact, some of the incident light will be diffracted (scattered) by the turbulence at an angle dependent on the characteristic size of the eddies *t,* which, in turn, depends on the turbulence strength. In weak turbulence, ℓ is large (around 1 m), while in strong turbulence, it is small (around 5 cm). Incident light is diffracted at an angle $0 = \lambda/\ell$. The diffracted rays will interfere when the pathlength L is comparable to $£/6 = \ell^2/\lambda$, producing intensity variations (scintillation). Note that $\ell = \sqrt{L\lambda}$ is the radius of the first Fresnel zone, which is also the correlation length of amplitude fluctuations.

If the pathlength is much smaller than L, then diffraction effects are negligible and the geometrical optics approach is valid. To take the specific case of a turbulent layer at the tropopause (altitude 10 km), if the eddy size ℓ is 20 cm and the optical wavelength is 0.5 μm, then L = 80 km, so the geometrical optics approximation may be used. If severe turbulence occurs in this layer, resulting in ℓ being only 5 cm, then L would be reduced to 5 km, resulting in considerable intensity fluctuations at ground level. For the relatively benign turbulence conditions at good observing sites, especially at 1R wavelengths where the eddy size ℓ may approach 1 m, the geometrical optics approximation is generally valid.

The next step in complexity is the single-scattering approximation, shown in figure 3.8(b). In this case, weak scattering of the main component of the wavefront is accounted for at each region of turbulence, but secondary scattering is ignored. The only interference effects accounted for are those between the zero-order (undiffracted) and first-order diffracted beams. It is assumed that the intensity of the zero-order light is undiminished by scattering.

Several methods of analysis are based on the single-scattering approximation. In the approach that Tatarski describes as the method of smooth variations, based on Maxwell's wave equation, the assumption is made that the refractive-index variations consist of small fluctuations superimposed on larger ones that vary smoothly. The Rytov approximation then allows the amplitude of the the electromagnetic field to be replaced by its logarithm in the exponent, greatly simplifying the analysis. While it is generally valid for near-vertical propagation paths, it has been found experimentally [Gracheva et al. 1970] that the Rytov approximation is invalid for propagation paths in which the parameter $\sigma^2 = 1.23k^{7/6}C_N^2x^{11/6}$ is greater than 0.3, corresponding to horizontal pathlengths of about 1 km, because of the neglect of multiple scattering.

This limitation led to the development of the third stage of complexity, addressing the case of strong turbulence in which a large fraction of the incident light is scattered. As indicated in figure 3.8(c), the scattered light is itself scattered by successive layers. This situation is modeled using the parabolic approximation to the wave equation, which is solved using

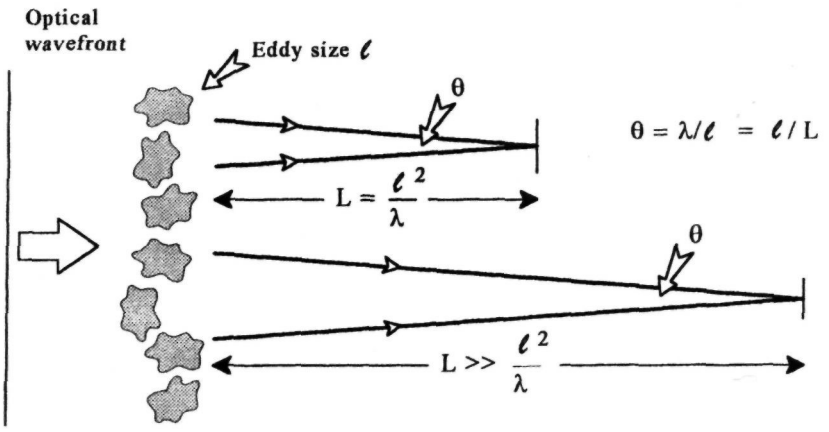


Figure 3.9 Limitations on the geometrical optics approximation for evaluating the optical effects of turbulence. It is valid for propagation over distances L, where $L < \ell/\lambda$. At greater distances, diffracted rays can interfere and scattering must be taken into account.

the Markov approximation as mentioned above. As this regime is rarely, if ever, encountered in astronomical adaptive optics, it is not discussed further. The reader is referred to the books cited above for more information.

For astronomical observations involving near-vertical propagation through Earth's atmosphere, analytical methods based on single scattering are adequate for the design and optimization of adaptive optics systems. In many cases, geometrical optics is sufficient. Only in special situations, such as observations at short wavelengths or at large zenith angles, may the more rigorous approach be necessary.

Even using simplified methods of analyzing atmospheric turbulence, the complexity of the mathematical models has usually necessitated the use of numerical solutions, which are inconvenient for performing the tradeoffs necessary to optimize a design. Sasiela [1988] has recently developed an analytical approach to solving the most difficult propagation problems involving anisoplanatism and multiple beams. This method, using transverse spatial filtering, allows wavefront errors to be expressed directly in terms of the basic parameters of the adaptive optics system. It is described in section 3.6.

3.4.2 Effects of Turbulent Layers

The following treatment is based on the approach of Roddier [1981]. The first step is to find the effect of a thin turbulent layer on an incident plane wave of unit magnitude, as shown in figure 3.10. The thickness of the layer *Sh* is large compared with the scale size of the turbulence cells, so that Gaussian statistics apply, but it is small enough for diffraction effects to be negligible. With a unit incident field, the scalar field emerging from the thin layer is

$$\Psi_h(\mathbf{x}) = \exp[i\phi(\mathbf{x})] \tag{3.22}$$

where the phase shift produced by the refractive index fluctuations is

$$\phi(\mathbf{x}) = k \int_h^{h+\delta h} dz\ \mathrm{n}(\mathbf{x}, z) \tag{3.23}$$

and k is the wave number, $2\pi/\lambda$. The coherence function of the field at the output of the layer is

$$B_h(\mathbf{r}) = \langle \exp\ i[\phi(\mathbf{x}) - \phi(\mathbf{x}+\mathbf{r})] \rangle \tag{3.24}$$

Because $\phi(\mathbf{x})$ has Gaussian statistics with zero mean, the coherence function can be expressed as

$$B_h(\mathbf{r}) = \exp -\tfrac{1}{2}\langle |\phi(\mathbf{x}) - \phi(\mathbf{x}+\mathbf{r})|^2 \rangle \tag{3.25}$$

or in terms of the structure function $D_\phi(\mathbf{r})$:

$$B_h(\mathbf{r}) = \exp -\tfrac{1}{2} D_\phi(\mathbf{r}) \tag{3.26}$$

The next step is to find $D_\phi(\mathbf{r})$ in terms of the fluctuations in refractive index. The covariance of $\phi(\mathbf{x})$ is

$$B_\phi(\mathbf{r}) = \langle \phi(\mathbf{x})\, \phi(\mathbf{x}+\mathbf{r}) \rangle \tag{3.27}$$

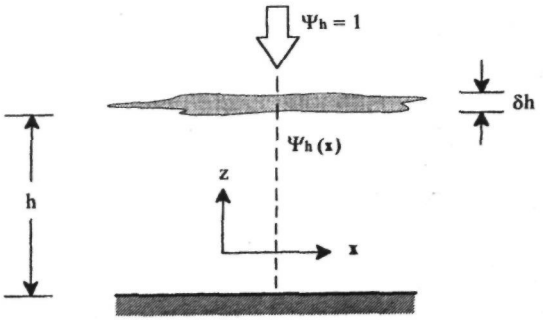


Figure 3.10 Thin turbulent layer model.

If $B_N(\mathbf{r}, z)$ is the three-dimensional covariance of the index fluctuations, then using equation (3.23)

$$B_\phi(\mathbf{r}) = k^2 \delta h \int_{-\infty}^{+\infty} dz \cdot B_N(\mathbf{r}, z) \tag{3.28}$$

Because *Sh* is much larger than the index fluctuations, the integral may now be taken from $-\infty$ to + oo.

The phase structure function is related to its covariance by

$$D_\phi(\mathbf{r}) = 2[B_\phi(\mathbf{0}) - B_\phi(\mathbf{r})]$$

Similarly,

$$D_N(\mathbf{r}) = 2[B_N(\mathbf{0}, \mathbf{0}) - B_N(\mathbf{r}, z)] \tag{3.29}$$

Using equation (3.28), it follows that

$$\begin{aligned} D_\phi(\mathbf{r}) &= 2k^2 \delta h \int_{-\infty}^{+\infty} dz \Big[B_N(\mathbf{0}, z) - B_N(\mathbf{r}, z) \Big] \\ &= k^2 \delta h \int_{-\infty}^{+\infty} dz \Big[D_N(\mathbf{r}, z) - D_N(\mathbf{0}, z) \Big] \end{aligned} \tag{3.30}$$

Using the expression for refractive-index structure function, equation (3.10),

$$D_N(r) = C_N^2 \left(r^2 + z^2\right)^{1/3}$$

$$D_\phi(r) = k^2 C_N^2\, \delta h \int_{-\infty}^{+\infty} dz \left[\left(r^2 + z^2\right)^{1/3} - z^{2/3} \right] \tag{3.31}$$

The integral term is

$$\begin{aligned} I &= \int_{-\infty}^{+\infty} dz \left[(r^2 + z^2)^{1/3} - z^{2/3} \right] \\ &= 2 \int_0^{+\infty} dz \left[(r^2 + z^2)^{1/3} - z^{2/3} \right] \end{aligned} \tag{3.32}$$

which is evaluated as

$$\begin{aligned} I &= \frac{1}{3}\,\frac{6}{5}\,\frac{\Gamma(\frac{1}{2})\Gamma(\frac{1}{6})}{\Gamma(\frac{2}{3})} r^{5/3} \\ &= 2.914 r^{5/3} \end{aligned} \tag{3.33}$$

The phase structure function at the output of a thin layer with Kolmogorov turbulence is then

$$D_\phi(r) = 2.914k^2C_N^2\ \delta h r^{5/3} \tag{3.34}$$

The coherence function at the layer output in terms of the refractive-index variations may then be expressed as

$$B_h(r) = \exp -\tfrac{1}{2}(2.914k^2C_N^2\ \delta h r^{5/3}) \tag{3.35}$$

For astronomical observations, the near-field approximation can be used, so that the coherence function at the ground is

$$B_0(r) = \exp -\tfrac{1}{2}D_\phi(r) \tag{3.36}$$

In propagating from the layer output at height h to the ground, the complex field will generally produce fluctuations both in amplitude (producing scintillation) and in phase. However, the amplitude fluctuations in the near field are small, so D_ϕ is a good representation of the phase structure function.

These results are extended readily to cover multiple layers and a continuous distribution of turbulence. For multiple layers, the contribution of each layer is $C_N^2(h_j)\,\delta h_j$. For a number of layers, the coherence function is therefore

$$B_0(r) = \exp -\tfrac{1}{2}\left[2.914k^2r^{5/3}\sum_j C_N^2(h_j)\delta h_j\right] \tag{3.37}$$

For a continuous distribution of turbulence, this becomes

$$B_0(r) = \exp -\tfrac{1}{2}\left[2.914k^2r^{5/3}\int_j dh\ C_N^2(h)\right]$$

The above expressions assume vertical propagation through the atmosphere. For observations at an angle of ζ from the zenith, the thickness of each layer is increased by the factor (sec ζ), so the coherence function at the telescope aperture is

$$B_0(r) = \exp -\tfrac{1}{2}\left[2.914k^2(\sec\zeta)r^{5/3}\int_h dh\, C_N^2(h)\right] \tag{3.38}$$

This expression is of fundamental importance in determining the effect of atmospheric turbulence on image structure.

3.4.3 Imaging Through Turbulence

The effect of the turbulence coherence function $B_0(r)$ on the images produced by an astronomical telescope will now be considered. The optical transfer function for the whole imaging system (telescope plus atmosphere) for long exposures is

$$S(f) = B(f)\cdot T(f) \tag{3.39}$$

where

$B(f)$ = atmospheric transfer function
$T(f)$ = telescope transfer function

In his study of the effects of atmospheric turbulence on optical images, Fried (1966a) introduced the resolving power of a telescope $\mathscr{R}$, which is defined as the integral of the optical transfer function. For a telescope looking through turbulence, it follows from equation (3.39) that

$$\mathscr{R} = \int df\cdot B(f)T(f) \tag{3.40}$$

In the case of a small aperture, turbulence effects are negligible, so for a diffraction-limited telescope with a clear circular aperture of diameter d'

$$\mathscr{R}_{SA} = \int df\cdot T_d(f) = \frac{\pi}{4}\left(\frac{d'}{\lambda}\right)^2 \tag{3.41}$$

For a large aperture, the resolving power $\mathscr{R}_{LA}$ depends only on the turbulence:

$$\mathscr{R}_{LA} = \int df\cdot B(f) \tag{3.42}$$

Fried also defined the parameter r_0, which is the diameter of a telescope (assumed to be diffraction limited) which would have the same resolving power as the atmospheric transfer function; that is, r_0 is the diameter of a perfect telescope having a transfer function $T(f)$, where

$$\int df\cdot T(f) = \int df\cdot B(f) \tag{3.43}$$

This leads to the relation

$$\mathscr{R}_{LA} = \int df\cdot B(f) = \frac{\pi}{4}\left(\frac{r_0}{\lambda}\right)^2 \tag{3.44}$$

From equation (3.38), the atmospheric transfer function may be written

$$B(f) = B_0(\lambda f) = \exp -Kf^{5/3} \tag{3.45}$$

where K is a constant that describes the seeing conditions. Then,

$$\begin{aligned}\mathscr{R}_{LA} &= \int df \exp -Kf^{5/3}\\ &= \frac{6\pi}{5}\Gamma\left(\tfrac{6}{5}\right)K^{-6/5}\end{aligned} \tag{3.46}$$

Equating equations (3.44) and (3.46)

$$\frac{6\pi}{5}\Gamma\left(\tfrac{6}{5}\right)K^{-6/5} = \frac{\pi}{4}\left(\frac{r_0}{\lambda}\right)^2 \tag{3.47}$$

which gives

$$K = 3.44\left(\frac{r_0}{\lambda}\right)^{-5/3} \tag{3.48}$$

Substituting in equation (3.45) gives

$$B(f) = \exp -3.44\left(\frac{\lambda f}{r_0}\right)^{5/3} \tag{3.49}$$

or, equivalently,

$$B(f) = \exp -3.44\left(\frac{r}{r_0}\right)^{5/3} \tag{3.50}$$

By equating this to equation (3.38), the definition of r_0 in terms of the integrated turbulence is obtained:

$$r_0 = \left[0.423k^2(\sec \zeta) \int dh\, C_N^2(h)\right]^{-3/5} \tag{3.51}$$

$$r_0 = \left[0.423k^2(\sec \mathrm{f}) \int dh C_N^2(h)\right] \tag{3.51}$$

The phase structure function can then be expressed in terms of r_0 as

$$D(r) = 6.88\left(\frac{r}{r_0}\right)^{5/3} \tag{3.52}$$

This expression is valid over the inertial range of turbulence, that is, between the inner and outer scales.

The parameter r_0 is a convenient measure of the strength of turbulence, as it represents in a single number the integrated effect of the refractive-index fluctuations for the entire atmosphere. Small values of r_0 correspond to strong turbulence and poor seeing, while large values mean weak turbulence and good seeing.

Equation (3.51) shows that r_0 is wavelength dependent, varying as the 6/5 power of wavelength. It is usually specified at the visible wavelength of 0.5 μm. Typical values for astronomical observations range from less than 5 cm in strong daytime turbulence to over 20 cm in good sites at night. Average values are generally in the range of 7-12 cm. Because r_0 has the dimension of length, it is often regarded as defining the "cell size" of atmospheric turbulence. This idea is misleading, because well-developed turbulence contains the complete Kolmogorov spectrum, ranging from an outer scale (several meters) to the inner scale (millimeters). The significance of r_0 is that it defines an aperture size over which the mean-square wavefront error is 1 rad^2. The "seeing angle" or image spread due to atmospheric turbulence for long exposures is given by λ/r_0 and varies as $\lambda^{-1/}$. This shows that the seeing angle gradually improves (that is, it gets smaller) at longer wavelengths, up to the point where r_0 becomes comparable with the aperture size. In other words, the angular resolving power of large ground-based telescopes improves as the observation wavelength increases. At a wavelength of 10 μm, r_0 is typically between 3 and 7 m, and atmospheric turbulence has little effect, even on the largest telescopes. It is also important to note that r_0 is certainly not a "constant" and is continually changing, often by a factor of 2, over periods of seconds to minutes. Measurements of r_0 are usually averaged over periods of several seconds to minutes.

Because of the very small dispersion of air, the optical disturbances produced by turbulence are practically independent of wavelength. This is an important fact for adaptive optics, because it means that turbulence is compensated at all wavelengths by a deformable mirror that corrects the optical pathlength directly.

3.4.4 Long- and Short-Exposure Images

The resolution of the imaging system $\mathscr{R}$ (the integral of the optical transfer function) corresponds to the performance of a telescope system averaged over long exposures. Long-exposure images are degraded by image motion caused by overall wavefront tilt variations across the telescope aperture, as well as by blurring of the image due to higher-order wavefront distortions. The effects of image motion can be eliminated by using short exposures, or, in an adaptive optics system, by using a tilt-compensating device such as a tracking mirror. It is of interest to determine the relative size of long- and short-exposure images as a function of the aperture diameter. These and similar questions were addressed by Fried [1965, 1966a].

From equations (3.39) and (3.49), the long-exposure MTF of the telescope and atmosphere is

$$\langle S(f)\rangle_{LE} = T(f) \exp\left[-3.44\left(\frac{\lambda f}{r_0}\right)^{5/3}\right] \tag{3.53}$$

where

$T(f)$ = transfer function of the telescope and
f = spatial frequency in cycles per radian.

Note that the exponent term is independent of telescope parameters, so it may be taken as the MTF of the atmosphere.

To determine the MTF for short-exposure images, it is necessary to eliminate the effects of overall wavefront tilt, which produces image motion. This is equivalent to eliminating the time-averaging of the atmospheric transfer function, which, in the case of long-exposure images, determines the average image position or phase. The phase terms are eliminated by setting them to zero and adding only the amplitudes of the Fourier components of the image. When this is done, Fried shows that the overall MTF for short exposure images has two forms, depending on whether the telescope is in the near field or far field of the atmospheric phase disturbances:

For $z \ll \dfrac{D^2}{\lambda}$ (near field)

$$\langle S(f)_{SE}\rangle = T(f) \exp\left\{-3.44\left(\frac{\lambda f}{r_0}\right)^{5/3}\left[1-\left(\frac{\lambda f}{D}\right)^{1/3}\right]\right\} \tag{3.54}$$

For $z \gg \dfrac{D^2}{\lambda}$ (far field)

$$\langle S(f)_{SE}\rangle = T(f) \exp\left\{-3.44\left(\frac{\lambda f}{r_0}\right)^{5/3}\left[1-\tfrac{1}{2}\left(\frac{\lambda f}{D}\right)^{1/3}\right]\right\} \tag{3.55}$$

The difference between long and short exposures is seen in the last term of the exponential. With short exposures, as f approaches D/λ, the cutoff frequency

of $T(f)$, the value of the negative exponent drops, so the short-exposure MTF is improved at high frequencies, especially in the near-field case. Note that the exponent depends on D and is therefore telescope dependent. The reason that the short-exposure MTF depends on the telescope aperture is simply that the amount of tilt removed depends on D.

The effect of the telescope diameter D on the performance of turbulence-degraded imaging systems may be determined using the normalized resolution $\mathscr{R}/\mathscr{R}_{LA}$. The quantity $\mathscr{R}_{LA}$, defined in equation (3.44), is the resolution that would be obtained through the atmosphere with an arbitrarily large telescope, the performance of which is limited entirely by turbulence. The turbulence strength is defined by r_0.

The normalized values of resolution are then

$$\mathscr{R}_{LE} = \frac{1}{\mathscr{R}_{LA}} \int df \langle S(f) \rangle_{LE} \tag{3.56}$$

$$\mathscr{R}_{SE} = \frac{1}{\mathscr{R}_{LA}} \int df \langle S(f) \rangle_{SE} \tag{3.57}$$

These integrals were evaluated by Fried [1966a] for a range of values of the parameter D/r_0, giving the resolution in units of cycles squared per steradian field of view.

The resolution may also be expressed as an angular image size, as shown in figure 3.11, which plots the long-exposure and short-exposure image sizes in λ/r_0 units against the normalized telescope aperture D/r_0. Also shown is the standard deviation of the image motion due to turbulence, and the resolution of a diffraction-limited aperture. These curves illustrate several basic properties of imaging through the turbulent atmosphere in the absence of adaptive compensation:

1. For long exposures, atmospheric turbulence sets a limit on the resolution achievable, no matter how long the exposure or how large the aperture. This resolution is given by equation (3.44), in units of cycles squared per steradian field of view. In Figure 3.11, it corresponds to a normalized image size of $\sqrt{4/\pi} = 1.13$. The resolution improves very little for values of D/r_0 greater than 10. Because of the wavelength dependence of r_0, the angular size of an atmospherically-degraded image improves as the 1/5 power of wavelength.
2. For short exposures in the near field ($Z << D^2/\lambda$), the maximum angular resolution is obtained with an aperture of $3.8r_0$. For this condition, the ratio of short-exposure to long-exposure resolutions $\mathscr{R}_{SE}/\mathscr{R}_{LE}$ is 4.18, corresponding to an image size reduction of 2.045. As the aperture size increases, the short-exposure resolution becomes asymptotic to the limiting value for long exposures.
3. The maximum ratio between short-exposure and long-exposure resolutions occurs when the telescope aperture is $3.5r_0$. If the image motion due to angle-of-arrival fluctuations is compensated actively, the maximum possible improvement in one-dimensional resolution is a factor of 2.054. In practice, small differences in these ratios are not significant because of the random variations in atmospheric parameters.

3.4.5 Measures of Imaging Performance

At this point, it is useful to discuss the relation between two different performance measures, $\mathscr{R}/\mathscr{R}_{LA}$ and the Strehl ratio, that are used in adaptive optics. The Strehl ratio is defined as the ratio of the actual peak intensity of the image of a point source to that produced by a perfect (diffraction-limited) telescope of the same aperture. It is shown in section 4.3.3 that for wavefront errors up to about 2 rad rms, the Strehl ratio may be expressed as

$$S = \exp(-\sigma^2) \tag{3.58}$$

where σ^2 is the mean-square wavefront error in the imaging beam. For the large wavefront errors typical of an uncompensated image, the Strehl ratio is not a useful measure.

A related measure of performance in the frequency domain is the resolving power $\mathscr{R}$, defined as the integral of the optical transfer function (OTF). The parameter $\mathscr{R}$ is analogous to the bandwidth of an electronic signal processing system. Although the resolution ratio $\mathscr{R}/\mathscr{R}_{LA}$ and the Strehl ratio are both derived from the integral of the OTF, they are normalized quite differently and represent totally different aspects of system performance. The Strehl ratio is normalized to an error-free OTF and is used to describe the performance of well-corrected optical imaging systems. The $\mathscr{R}/\mathscr{R}_{LA}$ ratio is normalized to the uncompensated OTF, which has an extremely small Strehl ratio. The ratio between the normalizing factors in these two measures is approximately $(D/r_0)^2$, which is on the order of 1000 for astronomical telescopes.

Each of these performance measures has its uses in adaptive optics, depending on the application. For energy projection systems, in which the goal is to maximize the peak energy on a small target, the Strehl ratio is the appropriate performance measure because even small wavefront aberrations will reduce the peak intensity by a significant amount. This may be called the "top-down" approach, for which the goal is a perfect optical system, making it necessary to determine the effect of small imperfections.

If the design is approached from the "bottom-up" direction, with the goal of improving the quality of turbulence-degraded images, then the normalized resolution $\mathscr{R}/\mathscr{R}_{LA}$ is appropriate because it measures the improvement over an uncompensated image. The Strehl ratio gives a pessimistic view of system performance, while the normalized resolution gives a (very) optimistic view. This is not a trivial matter, because

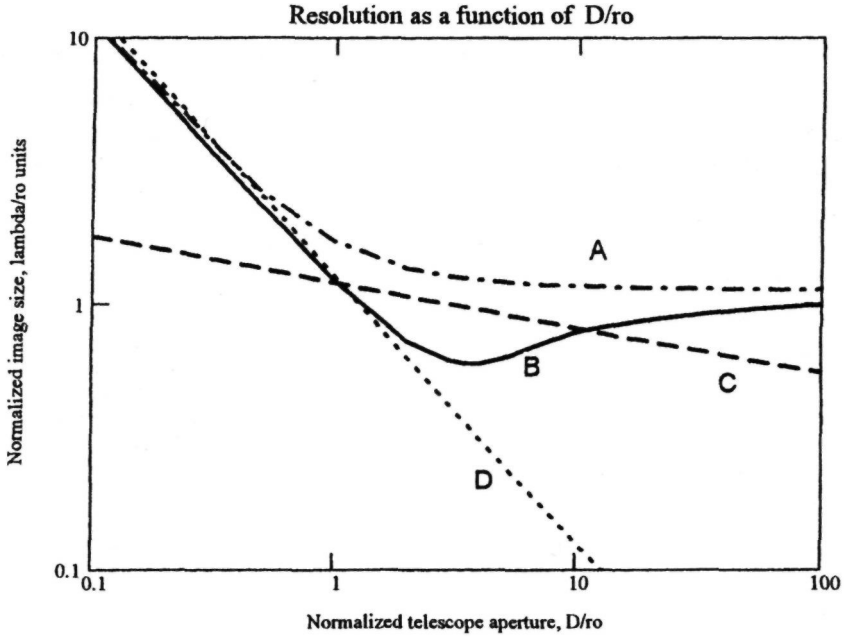


Figure 3.11 Normalized image size as a function of D/r_0, for an uncompensated telescope.

the selection of performance criteria is a central issue in the design and optimization of adaptive optics systems. It is discussed further in chapter 4.

3.4.6 Angle-of-Arrival Fluctuations

Turbulence-degraded images of a star generally consist of a blurred disk that moves randomly from its mean position. The image motion is due to the overall wavefront tilt component of the turbulence over the telescope aperture, which produces variations in the angle of arrival of the wavefront. Wavefront distortion components of second and higher order, such as defocus, astigmatism, and coma, enlarge the seeing disk from its diffraction limited diameter. The relative importance of the blurring of the disk and the motion of its centroid depends on the D/r_0 ratio. For small apertures where D/r_0 is on the order of unity, the short-exposure image is close to diffraction-limited and almost all of the long-exposure degradation is caused by image motion. As the ratio D/r_0 increases, the image motion decreases and the diameter of the long-exposure image approaches that of the short-exposure image.

The angle of arrival has a Gaussian distribution; its mean-square value in one axis has been determined by Fried [1965] to be

$$\sigma_\alpha^2 = 0.18\lambda^2 D^{-1/3} r_0^{-5/3} \tag{3.59}$$

Note that because r_0 varies as $\lambda^{6/5}$, the value of σ_α^2 for a given turbulence strength is independent of wavelength. This produces some interesting scaling effects, in comparison to the size of the image disk, as the aperture and wavelength are varied. The angular motion becomes a larger component of the uncompensated seeing disk as the wavelength increases, because the overall size of the seeing disk is proportional to $\lambda^{-1/5}$. However, for a constant telescope aperture, the image motion becomes a smaller fraction of the Airy disk as the observing wavelength increases, because the diameter of the diffraction-limited disk is proportional to λ. At a constant wavelength, the image motion becomes a larger fraction of the Airy disk as the telescope aperture is increased.

The effect of angle-of-arrival variations on astronomical images may be eliminated either by using short exposures to freeze the turbulence or by using an image motion corrector, such as a tip-tilt mirror

driven by an image position sensor. The first type of adaptive optics system to be implemented was of this type, as described in chapter 1.

The temporal power spectrum of angle-of-arrival variations depends on the telescope aperture, outer scale of turbulence, and the wind transport velocity, and is discussed further in section 9.2.5.

3.5 Modal Representation and Correction of Turbulence

3.5.1 *Zernike Polynomials*

The Kolmogorov model of atmospheric turbulence generates smoothly varying optical wavefronts with spatial components spanning a frequency range of over 1000 to 1 from the inner to the outer scale. Functions of this kind can be represented by a series of orthogonal modes of ascending order. The Fourier series of harmonically related sine waves used for one-dimensional waveform analysis is a familiar example of this approach. In optical systems, the two-dimensional Zernike polynomials perform a similar role, enabling complex wavefront shapes to be decomposed into a set of basis functions of ascending order. The lowest order Zernike polynomials correspond to the familiar and physically controllable wavefront aberrations, such as wavefront tilt, defocus, and astigmatism. Zernike polynomials and a related set based on Karhunen-Loeve functions can* be used to represent the instantaneous shapes of wavefronts generated by atmospheric turbulence. For low-order distortions, they are able to represent the wavefront more accurately than does the zonal approach (an array of flat subapertures) with the same number of degrees of freedom. Hence, the incentive for using Zernike polynomials is not just as a method for analyzing wavefronts, but also as a possible basis for compensating atmospheric turbulence effects.

Zernike polynomials are defined in polar coordinates on a unit circle as functions of azimuthal frequency m and radial degree n, where $m < n$, and $n - m$ is even. For statistical analysis, it is convenient to order the polynomials in a sequence j in such a way that even values of j correspond to the symmetric modes in $\cos(m\theta)$, while odd values of j correspond to the antisymmetric modes in $\sin(m\theta)$. This set of polynomials is defined as:

$$\begin{aligned} Z_{\text{even}} &= \sqrt{n+1}R_n^m(r)\sqrt{2}\,\cos(m\theta), \qquad m \neq 0 \\ Z_{\text{odd}} &= \sqrt{n+1}R_n^m(r)\sqrt{2}\,\sin(m\theta), \qquad m \neq 0 \\ Z &= \sqrt{n+1}R_n^0(r), \qquad m = 0 \end{aligned} \tag{3.60}$$

where

$$R_n^m(r) = \sum_{S=0}^{(n-m)/2} \frac{(-1)^S (n-S)!\; r^{n-2S}}{S!\;[(n+m)/2-S]!\;[(n-m)/2-S]!} \tag{3.61}$$

The normalization is such that each polynomial has a root mean-square value of 1 over the unit disk. The first 15 Zernike polynomials defined in this way are shown in table 3.2, together with the names of the classic optical aberrations associated with them.

An arbitrary phase function $\phi(r,\theta)$ over a unit circular aperture can be expanded as

Table 3.2 Zernike Polynomials

	Azimuthal Frequency, m				
Radial Degree n	0	1	2	3	4
0	$Z_1 = 1$ Constant				
1		$Z_2 = 2r\cos\theta$ $Z_3 = 2r\sin\theta$ x and y tilts			
2	$Z_4 = \sqrt{3}\,(2r^2 - 1)$ Defocus		$Z_5 = \sqrt{6}\,r^2\sin 2\theta$ $Z_6 = \sqrt{6}\,r^2\cos 2\theta$ Astigmatism		
3		$Z_7 = \sqrt{8}\,(3r^3 - 2r)\sin\theta$ $Z_8 = \sqrt{8}\,(3r^3 - 2r)\cos\theta$ Coma		$Z_9 = \sqrt{8}r^3\sin 3\theta$ $Z_{10} = \sqrt{8}\,r^3\cos 3\theta$ Trifoil	
4	$Z_{11} = \sqrt{5}$ $(6r^4 - 6r^2 + 1)$ 3rd order spherical		$Z_{12} = \sqrt{10}\,(4r^4 - 3r^2)$ $\cos 2\theta$ $Z_{13} = \sqrt{10}\,(4r^4 - 3r^2)$ $\sin 2\theta$		$Z_{14} = \sqrt{10}\,r^4\cos 4\theta$ $Z_{15} = \sqrt{10}\,r^4\sin 4\theta$

$$\phi(r,\theta) = \sum_{0}^{\infty} a_j Z_j(r,\theta) \tag{3.62}$$

where the amplitudes a_j of the Zernike components are given by

$$a_j = \int d^2r\, \phi(r,\theta) Z_j(r,\theta) \tag{3.63}$$

A wavefront distorted by Kolmogorov turbulence may be described by its structure function, defined in terms of the turbulence coherence length r_0, as

$$D_\phi(r) = 6.88\left(\frac{r}{r_0}\right)^{5/3}$$

The Wiener spectrum $\Phi(k)$ of the phase fluctuations is related to the structure function by

$$D_\phi(r) = 2\int dk \Phi(k)[1 - \cos(2\pi k \cdot r)] \tag{3.64}$$

from which the Wiener spectrum is found to be

$$\Phi(k) = \frac{0.023}{r_0^{5/3} k^{11/3}} \tag{3.65}$$

Noll [1976] showed that a Zernike representation of this spectrum may be obtained from equation (3.62) by evaluating the covariance of the expansion coefficients. When used to represent an atmospherically distorted wavefront, the Zernike components can be considered as Gaussian random variables with zero mean. The covariance between two Zernike polynomials Z_j and $Z_{j'}$ having amplitudes a_j and $a_{j'}$ may be expressed in the form [Wang and Markey 1978]:

$$\langle a_j a_{j'} \rangle = \frac{k_{ZZ'}\delta_{mm'}\Gamma[\frac{1}{2}(n+n'-5/3)](D/r_0)}{\Gamma[\frac{1}{2}(n-n'+17/3)]\Gamma[\frac{1}{2}(n'-n+17/3)]\Gamma[\frac{1}{2}(n+n'+23/3)]} \quad \text{for } j-j' \text{ even}$$
$$= 0 \quad \text{for } j-j' \text{ odd} \tag{3.66}$$

where $k_{ZZ'}$ depends on the frequency characteristics of z_j and $z_{j'}$ and is given by

$$k_{ZZ'} = 2.2698(-1)^{1/2(n+n'-2m)}\sqrt{(n+1)(n'+1)}$$

and $\delta_{mm'}$ is the Kronecker delta function, defined as

$$\delta_{mm'} = 1 \quad \text{when } m = m'$$
$$= 0 \quad \text{when } m \neq m'$$

Note that only those polynomial terms having the same values of azimuthal frequency m have nonzero covariance and that the covariance between sine and cosine terms is zero.

3.5.2 Orders of Modal Compensation

A characteristic of Zernike or similar polynomials is that the modes are ordered according to their radial and azimuthal frequencies. Consequently, when they are used to represent turbulence-induced wavefronts, the amplitudes of the modes fall steadily as the mode number increases, due to the $f^{-8/3}$ shape of the Kolmogorov spectrum. When Zernike modes are used as the basis for compensating turbulence, the lowest modes, such as tilt, defocus, and astigmatism, are compensated first. It is of interest to know how much distortion remains after each order has been applied. The answer is obtained readily with Zernike polynomials. If the first N modes are corrected, the phase correction, may be represented as

$$\phi_c = \sum_{j=1}^{N} a_j Z_j \tag{3.68}$$

The mean-square residual error after compensation is

$$\sigma_N^2 = \langle \phi^2 \rangle - \sum_{j=1}^{N} \langle |a_j|^2 \rangle \tag{3.69}$$

where $< \phi \quad >$ is the phase variance of the wavefront to be corrected. The first Zernike polynomial represents constant phase and therefore has no effect on the wavefront in a single aperture. Noll [1976] has shown that the residual error may be expressed as

$$\sigma_N^2 = A_N\left(\frac{D}{r_0}\right)^{5/3} \tag{3.70}$$

The first few values of the parameter A_N for a circular aperture are given in table 3.3.

When $N = 1$, only the overall (piston) phase is corrected and there is no compensation within the aperture. In this case, the value of A_1 represents the error due to uncompensated turbulence. Compensation of overall tilt in two axes is represented by A_3. When $N > 10$, the value of σ_N^2 is closely approximated by

$$\sigma_N^2 \approx 0.2944 N^{-\sqrt{3}/2}\left(\frac{D}{r_0}\right)^{5/3} \tag{3.71}$$

Table 3.3 Residual Error Coefficients for Compensation of Kolmogorov Turbulence Using Zernike Polynomials

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
1.030	0.582	0.134	0.111	0.088	0.065	0.059	0.053	0.046	0.040

3.5.3 Modulation Transfer Function

The plane-wave MTF for a compensated system using Zernike polynomials up to a specified order is calculated as follows. If $\phi(\mathbf{r})$ is the turbulence phase distribution as defined in (3.62) and $\phi_c(\mathbf{r})$ is the compensating phase distribution up to the *Nth* Zernike polynomial, then

$$\phi_c(\mathbf{r}) = \sum_{j=2}^{N} a_j Z_j(\mathbf{r}) \tag{3.72}$$

The plane-wave MTF of a system consisting of the atmosphere and a telescope may be written

$$\langle T(\rho)\rangle = \int d^2\mathbf{r}\langle G(\mathbf{r}+\rho)G_*(\mathbf{r}-\rho/2)\rangle \times W(\mathbf{r}+\rho/2)W(\mathbf{r}-\rho/2) \tag{3.73}$$

where $G(\mathbf{r})$ is the perturbation function at the aperture that is due to a point source, p is the separation between two points in the aperture, and $W(\mathbf{r})$ is the pupil function of the telescope:

$$\begin{aligned} W(\mathbf{r}) &= 1 \quad \text{for} \quad |\mathbf{r}| \le D/2 \\ &= 0 \quad \text{for} \quad |\mathbf{r}| > D/2 \end{aligned} \tag{3.74}$$

Following Wang and Markey [1978], the perturbation term after compensation of the phase errors can be expressed as

$$\begin{aligned} \langle G(\mathbf{r}+\rho/2)G^*(\mathbf{r}-\rho/2)\rangle = \exp\{&-\tfrac{1}{2}D(\rho) - \tfrac{1}{2}\langle[\phi_c(\mathbf{r}+\rho/2) \\ &-\phi_c(\mathbf{r}-\rho/2)]^2\rangle \\ &+ \langle[\phi(\mathbf{r}+\rho/2) - \phi(\mathbf{r}-\rho/2)] \\ &\times [\phi_c(\mathbf{r}+\rho/2) - \phi_c(\mathbf{r}-\rho/2)]\rangle\} \end{aligned} \tag{3.75}$$

where $D(\rho)$ is the wave structure function, defined in equation (3.52), and ϕ is the phase error due to turbulence. Evaluation of this expression gets prohibitively difficult for the higher order polynomials. To make the problem more tractable, Wang and Markey [1978] show that the exponent term in equation (3.75) may be expressed as

$$\begin{aligned} E(\mathbf{r},\rho) = &-\tfrac{1}{2}D(\rho) + \tfrac{1}{2}\langle[\phi_c(\mathbf{r}+\rho/2) - \phi_c(\mathbf{r}-\rho/2)]^2\rangle \\ &+ \sum_{j=2}^{N}\sum_{j'=N+1}^{\infty}\langle a_j a_{j'}\rangle[Z_j(\mathbf{r}+\rho/2) - Z_j(\mathbf{r}-\rho/2)] \\ &\times [Z_{j'}(\mathbf{r}+\rho/2) - Z_{j'}(\mathbf{r}-\rho/2)] \end{aligned} \tag{3.76}$$

This expression shows that $E(\mathbf{r},p)$ is the sum of three terms: the wave structure function, the autocorrelation of the compensating phase distribution, and the cross-correlations of the Zernike expansion coefficients. The magnitude of these cross-correlations decreases rapidly as the difference in order n increases, enabling finite sums of these terms to be used without loss of accuracy. For example, when compensating with Zernike polynomials up to Z_{10} (zero curvature coma, or trifoil), the infinite summation may be truncated a t / = 91. The advantage of this approach is that the MTF computation can be extended to arbitrarily high-order modal phase corrections without incurring a large computational burden. The MTF of the complete system comprising the turbulent atmosphere and telescope, with modal compensation through Zernike order N, may then be expressed as

$$\begin{aligned} \langle T(\rho)\rangle = 4D^2 \exp[-\tfrac{1}{2}D(\rho)] &\int_0^{\pi/2} d\theta \int_0^{L(\theta)} \mathbf{r}\, d\mathbf{r} \\ &\times \exp\left\{\tfrac{1}{2}Q(\mathbf{r},\rho) + \sum_{j=2}^{N}\sum_{j=N+1}^{\infty}\langle a_j a_{j'}\rangle\right. \\ &\times [Z_j(\mathbf{r}+\rho/2) - Z_j(\mathbf{r}-\rho/2)] \\ &\left.\times [Z_{j'}(\mathbf{r}+\rho/2) - Z_{j'}(\mathbf{r}-\rho/2)]\right\} \end{aligned} \tag{3.77}$$

where

$$\begin{aligned} Q(\mathbf{r},\rho) &= \langle[\phi_c(\mathbf{r}+\rho/2) - \phi_c(\mathbf{r}-\rho/2)]^2\rangle \\ &= \sum_{j=2}^{N}\sum_{j'=2}^{N}\langle a_j a_{j'}\rangle[Z_j(\mathbf{r}+\rho/2) - Z_j(\mathbf{r}-\rho/2)] \\ &\quad\times [Z_{j'}(\mathbf{r}+\rho/2) - Z_{j'}(\mathbf{r}-\rho/2)] \end{aligned} \tag{3.78}$$

and

$$\begin{aligned} L(\theta) = &-\tfrac{1}{2}\left(\frac{\rho}{D}\right)\cos(\theta-\phi) \\ &+\tfrac{1}{2}\left(\frac{\rho}{D}\right)\left[\left(\frac{\rho}{D}\right)^{-2} - \sin^2(\theta-\phi)\right]^{1/2} \end{aligned} \tag{3.79}$$

The plane-wave MTF produced by correcting a wavefront distorted by Kolmogorov turbulence with Zernike polynomials Z_2 through Z_{10} is shown for several values of D/r_0 in figure 3.12.

Zernike polynomials are not statistically independent, so wavefronts generated directly from the sum of Zernike coefficients do not have zero mean values over time. A similar set of orthonormal functions with completely uncorrelated elements may be generated using the **Karhunen-Loeve** expansion. Fried [1978] has shown that for a given number of modal corrections, the Karhunen-Loeve expansion is optimal in that it gives a better fit to atmospherically distorted wavefronts than any other orthogonal set, in terms of minimizing the mean-square residual error. However, the difference in performance between Zernike and Karhunen-Loeve functions is not large, producing at most a 10% improvement in the peak value of normalized resolution when 21 correction terms are used [Wang and Markey 1978]. For low-order correction of up to 10 terms (through coma), the Zernike polynomials are close to optimum.

3.5.4 Partial Compensation with Zernike Modes

The improvement in image quality produced by partial compensation with Zernike modes may be

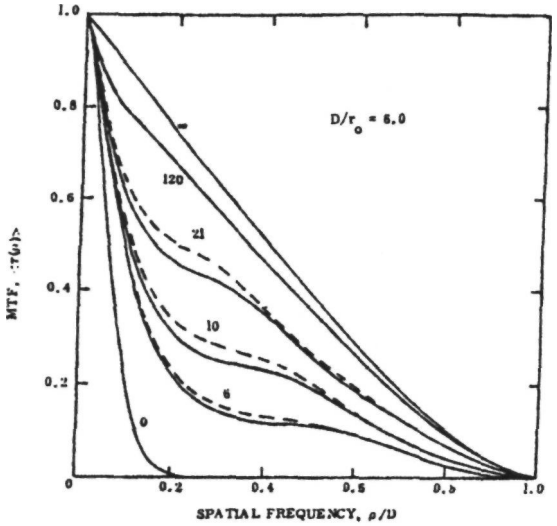


Figure 3.12 Modal compensation of an astronomical telescope [Wang and Markey 1978]. This chart shows the modulation transfer function (MTF) of a telescope with $D/r_0 = 6$, plotted against normalized spatial frequency, for several orders of modal compensation. The solid curves represent compensation by Zernike polynomials and the dotted curves represent compensation by Karhunen-Loeve functions.

determined using the normalized resolution $\mathscr{R}/\mathscr{R}_{LA}$, as shown in figure 3.13. The peak values of $\mathscr{R}/\mathscr{R}_{LA}$ and the optimum values for D/r_0, as determined by N. Roddier [1990], are shown in table 3.4 for Zernike corrections of up to 13 terms.

For example, with the optimum aperture size, the correction of the first 11 Zernike terms (through spherical aberration) improves the ratio $\mathscr{R}/\mathscr{R}_{LA}$ by a factor of 20.4, equivalent to reducing the image diameter to $1/\sqrt{20.4} \simeq 0.22$ of its uncompensated size. While this is a useful improvement, it only applies to apertures with a diameter of $8.3r_0$; the improvement drops very rapidly when D/r_0 is larger. At visible wavelengths, the optimum aperture would be in the region of 1.5 m, and the improvement factor for a 4-m telescope would be marginal. At IR wavelengths, it is a different story, because of the much larger values of r_0. At 1.6 μm or more, with moderately good seeing (r_0 = 15 cm at λ = 0.5 μm), the optimum aperture is over 4 m.

It appears to be very difficult to implement high-order Zernike corrections directly with a mechanical wavefront compensation device, such as a deformable mirror, so this approach is probably limited to low-order compensation with about 10 terms. However, it is possible that Zernike or Karhunen-Loeve modes could be closely approximated in a high-resolution zonal compensation device such as a liquid crystal.

3.6 Transverse Filtering Techniques

3.6.1 Outline of Method

Analysis of the optical effects of atmospheric turbulence generally involves two steps: (1) an integral expression for the quantity of interest is obtained; and (2) the integral is evaluated. Both of these steps involve simplifications and assumptions in order to make the process tractable. The wave equation can be simplified when the radiation wavelength is much smaller than the spatial wavelength of the refractive-index variations, which is true at optical wavelengths. Fortunately, the turbulence levels encountered in astronomical adaptive optics are relatively small, so the variance of the received field is much smaller than its mean value. These conditions justify the Rytov approximation, in which the normalized amplitude of the field is represented by its logarithm. The resulting theory gives accurate results for phase, the main concern in adaptive optics systems, but leads to inaccuracies in predicting scintillation effects when the log-amplitude variance exceeds about 0.3.

Even when these simplifications are made, evaluation of the resulting integrals is still extremely difficult. The integrals involved are three dimensional, consisting of two dimensions of the spatial spectrum transverse to the optical beam, plus the axial dimension in the direction of propagation. In the literature on optical propagation through turbulence, most problems have been solved individually. Because of the complexity of these calculations, especially those involving anisoplanatism, the results were usually presented in tabular or graphic form for a specific set of parameters.

The design and optimization of adaptive optics systems is facilitated greatly when analytic expressions are available for the error sources. A systematic approach that can be applied generally to the evaluation of optical propagation problems has great potential value. These desirable attributes have been realized in a method of analysis developed by Sasiela [1988]. The approach leads to analytic expressions for the effects of atmospheric turbulence on most situations involving astronomical adaptive optics, using basic system parameters such as aperture size, beam convergence, turbulence moments, and wavelength. The method can handle multiple-beam problems that involve the differences between focused and collimated beams, such as focal anisoplanatism.

The analysis is based on Tatarski's [1961] results for wave propagation, which have been expanded to obtain a general formula that contains the product of filter functions that operate on the transverse spatial components of the refractive-index spectrum. Filter functions for a large number of propagation problems have been derived. The formula contains a triple integral, with two covering the refractive index

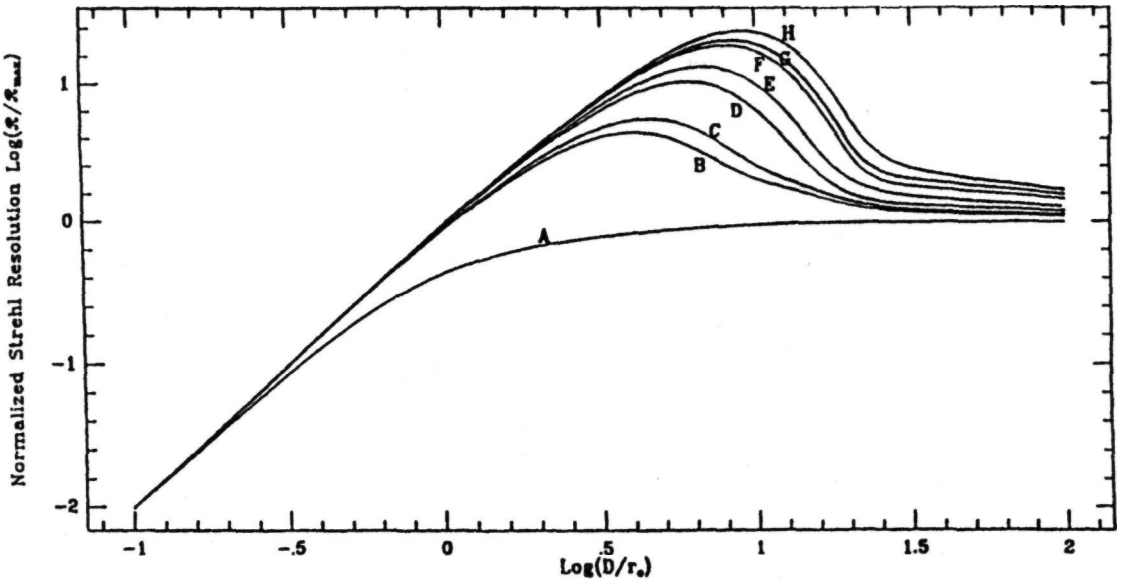


Figure 3.13 Normalized Strehl resolution as a function of the degree of modal compensation [N. Roddier 1990]. (A) Complete turbulence spectrum, no compensation; (B) overall tilt removed; (C) tilt and defocus removed; (D) tilt, defocus and astigmatism removed; (E) all terms through pure coma (Z_8) removed; (F) all terms through trifoil (Z_{10}) removed; (G) all terms through spherical aberration (Z_{11}) removed; and (H) all terms through fifth order astigmatism (Z_{13}) removed.

variations in the transverse plane and one for the axial variations. Two of these integrations are straightforward, but the third requires use of Mellin transforms [Sasiela and Shelton 1993b, Sasiela 1994]. In most cases, the results are obtained in the form of rapidly converging power series or asymptotic series involving relatively few terms.

The basis and results of this method will now be outlined. A complete description is beyond the scope of this book; Sasiela's monograph should be consulted for more detailed information. The starting point of the analysis is the vector wave equation for the electric field in a medium of constant permeability:

$$\nabla^2\mathbf{E} + k^2\mathrm{n}^2\mathbf{E} + 2\nabla[\mathbf{E}\cdot\nabla\log\ \mathrm{n}] = 0 \qquad (3.80)$$

where

∇^2 = transverse Laplacian operator
k = free space wave number $2\pi/\lambda$
n = the refractive index of the medium

If $X \ll \ell_0$, where ℓ_0 is the inner scale of turbulence, then the last term, which contains the interaction between orthogonal components of the field (depolarizing effects), can be neglected. Either component of the field can then be represented as a scalar wave equation:

$$\nabla^2 E + k^2\mathrm{n}^2 E = 0 \qquad (3.81)$$

The Rytov method represents the scalar field as

$$E = A\ \exp[iS] \qquad (3.82)$$

Table 3.4 Peak Values of Normalized Resolution versus Number of Zernike Modes Corrected

Zernike Modes Corrected	Optimal Value of D/r_0	Peak Value of $\mathscr{R}/\mathscr{R}_{LA}$
3	4.07	4.47
4	4.47	5.62
6	6.03	10.23
8	6.76	13.18
10	7.94	18.62
II	8.32	20.42
13	8.91	23.99

where A is the amplitude and S is the phase. If ψ is defined as $x + i\phi$ where $x \equiv \log A$, then the field can be expressed as

$$E = \exp(\psi) \tag{3.83}$$

The wave equation may then be written in the form

$$\nabla^2\psi + (\nabla\psi)^2 + k^2 n^2 = 0 \tag{3.84}$$

This equation must be solved for the fluctuations in log amplitude and phase. In the Rytov method, the variations in the refractive index that are due to turbulence are assumed to be very small, so it may be represented as $n = 1 + n_1$. The quantity ψ may be represented as a large steady part ψ_0 and a small fluctuating part ψ_1, so that

$$\begin{aligned} \psi &= \psi_0 + \psi_1 \\ \text{where} \quad \psi_1 &= x + i\phi \end{aligned} \tag{3.85}$$

If it is assumed that $|\nabla\psi_1| \ll |\nabla\psi_0|$, and neglecting n_1^2 in comparison with $2n_1$, then ψ_1 can be expressed in terms of the variations in refractive index n_1:

$$\nabla^2\psi_1 + 2\nabla\psi_0\nabla\psi_1 + 2k^2 n_1 = 0 \tag{3.86}$$

This equation is solved using a perturbation approach.

The refractive index may be written in the form of a random function with stationary increments as

$$n_1(r, z) = n_1(0, z) + \int d\nu(\kappa, z)[1 - \exp(i\kappa \cdot r)] \tag{3.87}$$

where r is the transverse coordinate. The transverse structure of the turbulence is transformed into Fourier space so that it may be expressed as a spectral expansion in terms of the spatial wave number κ. The field may be written in the same form, as

$$\psi_1(r, z) = \psi_1(0, z) + \int d\phi(\kappa, z)[1 - \exp(i\kappa \cdot r)]$$

When these equations are inserted into equation (3.86), an equation relating the phase and refractive index is obtained. Sasiela [1988] gives the solution of this equation for a wave propagating from 0 to z as

$$d\phi(\kappa, z) = ik \int_0^z d\nu(\kappa, z') \exp\left[\frac{-ik^2(z - z')}{2k}\right] dz' \tag{3.89}$$

where $d\nu(\kappa, z')$ is the transverse spatial spectrum of the refractive-index variations.

This transverse spectrum may be modified by multiplying it by a filter function $G(\kappa, Z)$. The two-dimensional transverse spatial Fourier transform of the phase is obtained from the imaginary part of equation (3.89)

$$\mathcal{F}[d\phi(L)] = k \int_0^L d\nu(\kappa, z') \sin\left[\frac{k^2(L - z')}{2k}\right] G(\kappa, z') \tag{3.90}$$

Similarly, the Fourier transform of the log-amplitude variations (scintillation) is obtained from the real part of equation (3.89)

$$\mathcal{F}[dx(L)] = k \int_0^L d\nu(\kappa, z') \cos\left[\frac{k^2(L - z')}{2k}\right] G(\kappa, z') \tag{3.91}$$

These equations are valid for the relatively low turbulence encountered in the atmospheric paths employed in astronomy. The last expression loses its validity when the log-amplitude variance exceeds 0.3, which may occur for long horizontal paths.

To obtain the mean-square values of phase and log-amplitude, these expressions are integrated over κ-space, multiplied by their complex conjugates, and ensemble averaged. The filter functions $G(\kappa, z')$ can be a cascade of individual filters that are tailored to the specific propagation problem to be solved. Sasiela [1988] derives the following formulas:

Phase and scintillation variance

$$\begin{bmatrix} \phi^2 \\ \chi^2 \end{bmatrix} = 0.2073k^2 \int_0^L dz\, C_N^2(z) \int d\kappa f(\kappa) \begin{bmatrix} C_p \\ C_a \end{bmatrix} \prod_i F_i(\kappa, z) \tag{3.92}$$

Phase and log-amplitude structure function

$$\begin{bmatrix} D_\phi(\alpha) \\ D_\chi(\alpha) \end{bmatrix} = 0.4146k^2 \int_0^L dz\, C_N^2(z) \int d\kappa f(\kappa) \begin{bmatrix} C_p \\ C_a \end{bmatrix} \times [1 - \cos(\kappa \cdot \alpha D)] \prod_i F_i(\kappa, z) \tag{3.93}$$

Temporal power spectral density

$$F(\omega) = 1.303k^2 \int_0^L dz \frac{C_N^2(z)}{v^2(z)} \int_0^\infty \frac{c\, dc U(c - 1)}{\sqrt{c^2 - 1}} \times f\left(\frac{\omega c}{v(z)}\right) \begin{bmatrix} C_p \\ C_a \end{bmatrix} \prod_i F_i\left(\frac{\omega c}{v(z)}, z\right) \tag{3.94}$$

where

k = optical wave number,
$C_N^2(Z)$ = turbulence profile,
$f(\kappa)$ = transverse turbulence spectrum,
$v(z)$ = wind velocity profile.

The functions C_p and C_a specify the beam convergence; that is, whether it is collimated, focused at the source, or focused at the object. For a collimated beam (plane wave), the values of these functions are

$$\begin{aligned} C_p &= \cos^2\left(\frac{\kappa^2(z - L}{2k}\right) \\ C_a &= \sin^2\left(\frac{\kappa^2(z - L)}{2k}\right) \end{aligned} \tag{3.95}$$

For spectral density (equation 3.94) use $K = \omega c / v(z)$ in equation 3.95.

The transverse turbulence spectrum $f(\kappa)$ is basically the normalized **Kolmogorov** spectrum, which is proportional to $\kappa^{-11/3}$ in three-dimensional space, if the inner and outer scales are neglected. To take account of finite inner and outer scales, the inner scale may be modified by an exponential term [Strohbehn **1968**] and the outer scale may be accommodated by using a von **Karman** spectrum, as follows:

$$f(\kappa) \simeq \left(\kappa^2 + \kappa_0^2\right)^{-11/6} \exp(-\kappa^2/\kappa_i^2) \qquad (3.96)$$

The turbulence scales are defined by $\kappa_0 = 2\pi/\mathscr{L}_0$ and $\kappa_i \simeq 5.92/\ell_0$, where $\mathscr{L}_0$ is the outer scale and ℓ_0 is the inner scale. The multiplying factor of 0.033 has been included in the initial coefficients of the main equations.

3.6.2 Transverse Filter Functions

The final term $\prod_i F_i(\kappa, z)$ is the product of transverse spatial filters that are specified according to the problem being solved. A comprehensive list of filter functions is given in Sasiela [1988]. A common requirement is to extract or subtract one or more of the Zernike modes, such as average phase or tilt, from the wavefront. Zernike polynomials are defined in polar coordinates by two parameters, azimuthal frequency m and radial degree n, as described in section 3.5.1. The necessary filter functions are derived from the spectral representation of Zernike polynomials given by Noll **[1976]**. The Fourier transforms of the Zernike modes are complex, and, for use as spectral filters, it is necessary to determine the square of the absolute value of each component. These are given by the general equation

$$\left.\begin{matrix} F_{\text{even } m,n}(\kappa) \\ F_{\text{odd } m,n}(\kappa) \\ F_{m,n}(\kappa) \end{matrix}\right\} = (n+1)\left[\frac{2J_{n+1}(\kappa D/2)}{\kappa D/2}\right]^2 \left\{\begin{matrix} 2\cos^2(m\phi) \\ 2\sin^2(m\phi) \\ 1\ (m=0) \end{matrix}\right. \qquad (3.97)$$

where

J_{n+1} = a Bessel function of the first kind
D = optical aperture diameter.

For example, a filter to extract the average phase of the wavefront is produced by setting $n = m = 0$. The resulting filter function is then

$$F_0(\kappa, z) = \left[\frac{2J_1(\kappa D/2)}{\kappa D/2}\right]^2 \qquad (3.98)$$

The overall tilt components of the wavefront are found by setting $m = n = 1$, giving the result

$$\left.\begin{matrix} F_x(\kappa, z) \\ F_y(\kappa, z \end{matrix}\right\} = \left[\frac{4J_2(\kappa D/2}{\kappa D/2}\right]^2 \left\{\begin{matrix} \cos^2(\phi) \\ \sin^2(\phi) \end{matrix}\right. \qquad (3.99)$$

To obtain the two-axis tilt phase variance, the two components are summed. It is often useful to obtain the tilt components as angles rather than phase, thereby corresponding to the angle-of-arrival of a wavefront. This is done by multiplying the phase variance by $(4/kD)^2$.

In some cases, it is necessary to remove components such as average phase and tilt from a wavefront. The filter function required to perform this task is

$$F_{\text{PTR}}(\kappa, z) = 1 - F_0(\kappa, z) - F_x(\kappa, z) - F_y(\kappa, z) \quad (3.100)$$

3.6.3 Computation of Anisoplanatic Errors

The real power of the transverse filtering approach is demonstrated when computing the phase differences between beams traveling along different paths in the atmosphere. The approach enables analytical expressions to be developed for most of the configurations encountered in optical propagation, including those using multiple beams, beams focused at different distances, and extended reference sources. If the paths are completely separated, the individual wavefront variances are uncorrelated and the variances may be added. However, in most cases encountered in adaptive optics, two or more propagation paths partially overlap, resulting in partial correlation of their wavefront phases. This scenario, which has previously required numerical solution, has now been solved analytically using transverse filters.

The main application of this analysis in astronomical adaptive optics is to evaluate the errors associated with laser beacons, which is discussed further in chapter 7.

3.7 Anisoplanatism

3.7.1 Anisoplanatism in Adaptive Optics

The concept of isoplanatism was originally introduced into the design of optical systems in connection with the imaging of extended objects; the isoplanatic region or isoplanatic patch refers to the angular field over which a well-corrected image is obtained. This concept has been generalized to cover the wavefront errors, between two beams, that are due to displacements of any type, including lateral displacements, focusing of the beams, and chromatic effects, in addition to angular separation. The forms of anisoplanatism of interest in adaptive optics are depicted in figure 3.14. The dependency of the beam separations on the displacement parameters for each of these cases is as follows:

1. Aperture displacement, s = constant;
2. Angular displacement, $s(z) = \theta z$;
3. Time delay displacement, $s(z) = \int dz\ v(z)t$;

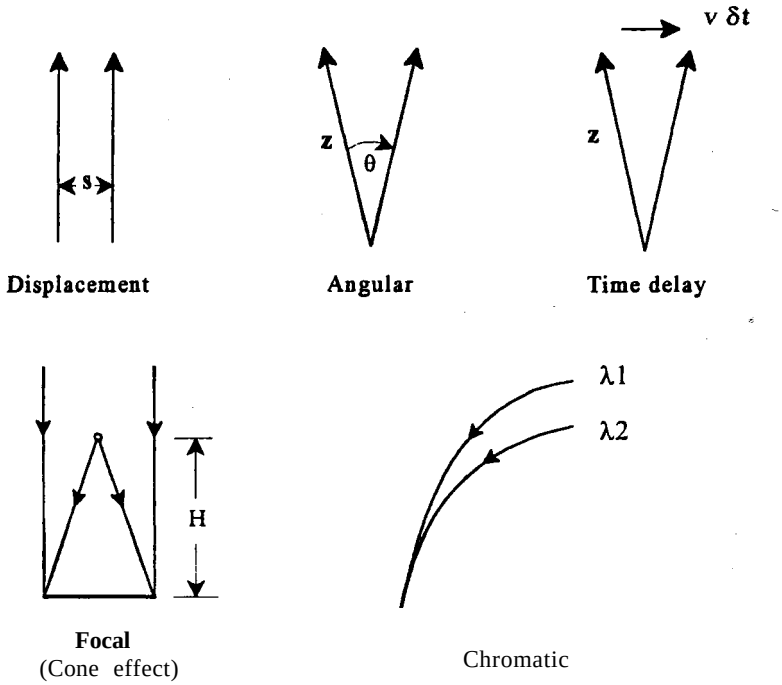


Figure 3.14 Forms of anisoplanatism in adaptive optics.

4. Focal displacement (axial), $s\{z, r) = rz/H$, where r = radius in aperture, and H = height of source;
5. Spectral displacement (color), s(z, A) = J *dzAn(z)* tan ζ, where An = refractive index difference between λ_1 and λ_2, and ζ = zenith angle.

Angular anisoplanatism degrades the performance of all optical systems that have finite fields of view. In adaptive optics, it is caused primarily by the three-dimensional distribution of atmospheric turbulence along the propagation path. Angular anisoplanatism is most sensitive to disturbances located at a distance from the receiving aperture, where they produce a wavefront error that varies with field angle. For ground-based telescopes, the isoplanatic angle is determined mainly by turbulent layers at higher altitudes.

The Achilles heel of adaptive optics is the need for a distant reference source to probe the turbulent atmosphere so that wavefront distortion can be measured and corrected. Available reference sources are rarely coincident with the science object, with the result that the optical path is not measured accurately. Even when the science object functions as the reference source, anisoplanatic errors may be produced if its size exceeds the isoplanatic angle. Using a single wavefront corrector, it is not possible to get real-time compensated images of extended areas on the Moon, the planets, or the Sun. In principle, it should be possible to compensate turbulence over much wider fields of view by using multiple wavefront correctors. The varieties of anisoplanatism of interest in astronomical adaptive optics are described below.

3.7.2 Angular Anisoplanatism

Angular anisoplanatism is produced by the gradual separation of two optical paths in the atmosphere as a function of their distance from the telescope. The mean-square wavefront error between two points in the wavefront separated by a distance r is defined by the phase structure function $D_\phi(r)$. From equations (3.30) and (3.32),

$$D_\phi(r) = 2.914k^2(\sec\zeta)r^{5/3}\int_h dh\, C_N^2(h) \qquad (3.101)$$

Angular anisoplanatism is modeled by two beams that are coincident at the telescope pupil and then separate at an angle 9. The separation distance at range z is therefore $r(z) = 6z$. Setting $z = h\sec(\zeta)$, where h is the height of the beam above the telescope and ζ is the zenith angle, the mean-square anisoplanatic error at angle 0 is

$$\begin{aligned}\langle\sigma_\theta^2\rangle &= 2.914k^2(\sec\zeta)\int_h dh\, C_N^2(h)(\theta h\sec\zeta)^{5/3}\\ &= 2.914k^2(\sec\zeta)^{8/3}\theta^{5/3}\int_h dh\, C_N^2(h)h^{5/3}\end{aligned} \qquad (3.102)$$

This equation is valid when $D \gg r_0$, which is generally the case with astronomical telescopes. If an angle θ_0 is defined such that

$$\theta_0 = \left[2.914k^2(\sec \zeta)^{8/3} \int_h dh\, C_N^2(h) h^{5/3} \right]^{-3/5} \quad (3.103)$$

then the anisoplanatic error for any angle θ may be expressed as

$$\langle \sigma_\theta^2 \rangle = \left(\frac{\theta}{\theta_0} \right)^{5/3} \quad (3.104)$$

The angle θ_0 is known as the "isoplanatic angle" and is a property of the turbulence distribution. It may also be stated in the form

$$\theta_0 = \left[2.914k^2(\sec \zeta)^{8/3} \mu_{5/3} \right]^{-3/5} \quad (3.105)$$

where $\mu_{5/3}$ is the five-thirds moment of the turbulence, as defined in section 3.3.6.

It is often convenient to express the anisoplanatic angle in terms of derived parameters, such as the turbulence coherence length r_0 and the mean turbulence height h. From equation (3.51),

$$r_0 = \left[0.423k^2 \sec(\zeta) \int dh\, C_N^2(h) \right]^{-3/5} \quad (3.106)$$

and combining this with equation (3.105) gives

$$\frac{\theta_0}{r_0} = \left[6.88(\sec \zeta)^{5/3} \frac{\int dh\, C_N^2(h) h^{5/3}}{\int dh\, C_N^2(h)} \right]^{-3/5} = \left[6.88(\sec \zeta)^{5/3} \frac{\mu_{5/3}}{\mu_0} \right]^{-3/5} \quad (3.107)$$

The mean turbulence height is defined as

$$\bar{h} = \left(\frac{\mu_{5/3}}{\mu_0} \right)^{3/5} \quad (3.108)$$

which leads to the relation

$$\frac{\theta_0}{r_0} = 6.88^{-3/5}(\sec \zeta)^{-1} (\bar{h})^{-1}$$

$$\theta_0 = 0.314(\cos \zeta) \frac{r_0}{\bar{h}} \quad (3.109)$$

When there is a single dominant layer of turbulence, h is the height of that layer. Otherwise h can be determined from the C_N^2 distribution or from the turbulence moments using equation (3.108). For example, if r_0 = 0.16 m and the mean turbulence height $\bar{h}$ is 5 km, then θ_0 = 2 arc seconds.

3.7.3 Other Forms of Anisoplanatism

Focal anisoplanatism and tilt anisoplanatism are of main concern in laser beacon systems and are discussed in detail in sections 7.3 and 7.4, respectively. Chromatic anisoplanatism, the displacement of optical beams by atmospheric dispersion, becomes significant for multispectral adaptive optics systems in which the observation wavelength is different from that of the wavefront sensor. The effects of atmospheric dispersion are discussed in section 9.3.

4 Optical Image Structure

4.1 Introduction

4.1.1 Image Structure in Adaptive Optics

The performance of an astronomical telescope is judged by its ability to concentrate the light from a distant star into the smallest possible radius. Laboratory tests reveal the intrinsic quality of the optical system, but when a telescope is operated in the dynamic environment of an observatory, external disturbances, such as atmospheric turbulence, become the limiting factor. The effect of atmospheric turbulence on telescope images is now well understood and many techniques are used to minimize its influence.

Adaptive optics adds another dimension to the imaging process. The wavefront compensation produced by adaptive optics is never perfect, so that compensated images have unusual structure. It is these images, formed at the focal plane of the telescope, that constitute the output of the system and by which the scientific results are judged. An awareness of the process of optical image formation, and of the effects of partial compensation on image structure, is essential for the design of adaptive optics systems for specific tasks. Conventional methods of specifying and evaluating image quality do not necessarily apply in such cases.

In this chapter, the factors that control the structure of compensated and partly compensated images are reviewed in order to determine the relation between the wavefront-correction performance of the adaptive optics and the quality of the resulting images. To establish performance criteria, it is necessary first to find the image characteristics required for specific scientific tasks in astronomy, and then to determine the type and degree of wavefront compensation necessary to produce those characteristics. In this way, the scientific objectives can be satisfied with minimum complications and expense. Without suitable performance criteria, arbitrary performance goals must be assumed, leading to the possibilities either that the overall performance will fall short of requirements, or that the adaptive optics will be overdesigned and consequently expensive to build and operate. For example, to ensure adequate image quality, it may be specified that the peak intensity of the image of a point source must be no less than 0.8 times its ideal (diffraction-limited) value. To achieve this peak intensity (Strehl ratio), the random phase errors due to turbulence must be reduced to about 1/14 wave rms at the observing wavelength. Producing this performance in a large ground-based telescope, especially at visible wavelengths, is an expensive proposition requiring a high-performance adaptive optics system.

For some imaging tasks, a high Strehl ratio may not be necessary. Useful scientific results have been obtained with simple adaptive optics that give much lower degrees of compensation. Even with a residual wavefront error of 1/4 wave rms, corresponding to a Strehl ratio of less than 0.1, the peak intensity of the core of a point image is more than 10 times greater than the halo of scattered light. For tasks involving the detection of point sources, the contrast ratio between the core and the halo is an acceptable

criterion. The important point is that each application of an astronomical telescope has its own set of requirements in regard to image quality. These image requirements, in turn, have a considerable impact on the design of the adaptive optics system.

4.1.2 Turbulence Effects

The wavefront distortion produced by atmospheric turbulence varies in both space and time. The characteristics of turbulence-degraded images are greatly different when examined over short rather than long periods of time. A short-exposure image is degraded by a single realization of the randomly varying wavefront. For a point source such as a star, the short-exposure image consists of a number of discrete speckles spread over an angular diameter of about λ/r_0. The speckles are produced by interference between r_0-sized sections of the aperture, separated by distances up to the full diameter, so their size is determined by the diffraction limit λ/D. The number of speckles is on the order of $(D/ro)^2$.

The speckle structure changes continuously as the turbulence evolves, with a time scale on the order of 1/100 s. Long-exposure images are blurred by two effects:

- The fine structure is averaged by turbulence within the aperture.
- The entire image is displaced randomly by the overall tilt component of turbulence.

Adaptive optics is capable of compensating both of these effects. Compensation of overall tilt removes the image motion that blurs the entire image, and is therefore mandatory in all adaptive optics systems. The fine speckle structure is only eliminated completely with perfect compensation; that is, when the system is diffraction limited. Practical adaptive optics systems always have some residual errors, depending on the observing wavelength and turbulence strength.

Atmospheric turbulence effects for short and long exposures have been described statistically in section 3.4, in which modulation transfer functions (MTFs) were derived in terms of atmospheric and telescope parameters. It is important to note that the short-exposure MTF, as well as the point spread function (PSF) derived from it, are statistical descriptions, representing the ensemble-average mean values of these functions; they do not represent the observable instantaneous image intensity distribution. For example, the short-exposure PSF of an uncompensated or partially compensated wavefront is represented statistically as a symmetrical function with a central peak, whereas the observable image consists of a number of constantly moving speckles that produce an instantaneous intensity profile which may be irregular and unsymmetrical. In the following discussion, the characteristics of turbulence-degraded images are described in statistical terms. For exposures of more than a few seconds, the averaging effect of turbulence should produce intensity profiles similar to the ensemble-averaged values. Only for very short exposures (less than a second or so) should there be any appreciable irregularity.

In the design of optical systems, it is customary to use normalized quantities to describe the properties of images. The peak image intensity (Strehl ratio) is usually normalized to that of a diffraction-limited system, while the angular diameter of images may be normalized either to the diffraction limit λ/D or to the turbulence limit λ/r_0. Because of the large range of apertures employed in astronomical telescopes and the wide spectrum of wavelengths used in any one instrument, normalized expressions must be used with great care, especially when performing tradeoffs in the initial design of a system. The scaling laws for atmospheric turbulence with regard to space, time, wavelength, and zenith angle are quite complicated and must be carefully accounted for.

When evaluating the effects of turbulence on images, two points are particularly important: (1) Atmospheric turbulence produces optical pathlength errors that are almost independent of wavelength. (2) Wavefront sensors usually operate at a fixed wavelength, independent of that used for the science observations. The consequence of these facts is that for a given set of input conditions, the residual wavefront error produced by an adaptive optics system is essentially constant in terms of optical pathlength for all observations within the spectral range of the telescope. Image structure, on the other hand, is determined by wavelength-dependent optical phase variations across the aperture. A given pathlength error therefore produces a different image structure at each observing wavelength. The structure of partially compensated images, in particular, varies greatly with wavelength and aperture size, because it depends on the turbulence parameters. When evaluating and comparing the performance *of* adaptive optics systems in different spectral bands, it is preferable to specify wavefront errors in terms of optical pathlength rather than optical phase. For the same reason, the errors due to overall wavefront tilt are best handled in terms of angular tilt, which is also independent of wavelength.

4.1.3 Summary

In the following sections of this chapter, the basic theory of optical imaging is reviewed and the effects of wavefront error, image motion, and photon noise on the structure of the images produced by an astronomical telescope are analyzed. The process of optical image formation is described from first principles, leading to a summary of the properties of aberration-free images. Next, the effect of small optical aberrations on the image intensity distribution is described using the classic Fresnel-Kirchoff diffraction integral,

from which the concepts of modal (Zernike) decomposition and Strehl ratio are derived.

To evaluate turbulence-degraded images, it is necessary to use statistical descriptions of the wavefront, using phase structure functions to determine the optical transfer function, followed by Fourier transformation to recover the image. The effect of partial wavefront compensation is handled by using structure functions of the residual wavefront errors. To facilitate system optimization, a shortcut procedure is described in which the image structure can be modeled with sufficient accuracy for first-order design using only four parameters: the telescope aperture, operating wavelength, turbulence parameter, and residual error. In the final section, imaging requirements for various astronomical tasks are reviewed and criteria for optimizing adaptive optics systems are developed.

4.2 Optical Image Formation

4.2.1 Huygens-Fresnel Principle

Huygens was the first proponent of the wave theory of light, originating the idea that every point in an optical wavefront may be considered as giving rise to spherical wavelets that combine to produce a continually expanding envelope. Fresnel later showed that interference between these wavelets explains the intensity variations that are apparent in diffraction patterns. Kirchhoff [1883] developed the concept mathematically to show that the Huygens-Fresnel principle may be regarded as an approximate solution to the scalar wave equation. These ideas form the basis of the theory of diffraction, a classic account of which is given in chapter 8 of Born and Wolf [1975].

The relevant properties of most optical imaging systems, including astronomical telescopes, are readily derived using a simple physical model. Consider a monochromatic point source P' illuminating an aperture A as shown in figure 4.1. When the source is at a great distance, the curvature over a small area is much less than the wavelength of light, so the wavefront reaching the aperture may be considered plane. Each point on the wave within an elemental area dS in the aperture gives rise to spherical wavelets. The size of dS is much smaller than the wavelength X, so these wavelets will be in phase when they reach point P, located a distance R from the aperture, and will combine constructively. If the strength of the plane wave has a constant value of E_A over the aperture A, then the optical disturbance at P at time r is given by

$$dU(P)\left(\frac{E_A}{z}\right)e^{i(\omega t-kz)}\,dS \qquad (4.1)$$

where *to* is the temporal frequency of the radiation, $k = 2\pi/\lambda$, and z is the distance between dS and P. As dS moves over the aperture, z will change by many wavelengths.

The total disturbance at P is found by integrating the contributions from each elemental area dS over the aperture. To do this, we must find the dependence of z on the positions of dS and P. Using a coordinate system with the origin at point O on the aperture, in which the coordinates of dS are $(x, y, z = 0)$ and the coordinates of P are (X, Y, Z), the distance between dS and P is

$$z = [Z^2(Y-y)^2 + (X-x)^2]^{1/2} \qquad (4.2)$$

This is the general case that applies when the distance R is comparable to the aperture size, producing the near-field condition known as *Fresnel diffraction.*

In the optical systems used in astronomical telescopes, the aperture is small compared with the distance R, and equation (4.1) may be simplified by replacing z, in the amplitude term, with R. However, small variations in the distance z represent large phase changes at optical wavelengths, so z must be calculated more exactly in the exponent term. Expanding equation (4.2) and using $R^2 = X^2 + Y^2 + Z^2$, the expression for z becomes

$$z = R\left[1 + \frac{y^2 + x^2}{R^2} - \frac{2(Yy + Xx)}{R^2}\right]^{1/2} \qquad (4.3)$$

Again, if the aperture is small compared with the distance R, then $(y^2 + x^2)/R^2$ may be neglected, reducing the right-hand side to two terms. Using the first two terms of the binomial expansion, the final approximation is obtained

$$z \approx R\left[1 - \frac{Yy + Xx}{R^2}\right] \qquad (4.4)$$

When the quadratic terms are dropped in this way, the expression reduces to the simplified case of *Fraunhofer diffraction,* which is valid for most practical optical imaging systems. Using equations (4.1) and (4.4), the expression for the disturbance at P is then

$$U(P) = \frac{E_A e^{i(\omega t - kR)}}{R} \iint_{\text{Aperture}} e^{ik(Yy - Xx)/R}\,dS \qquad (4.5)$$

Equation (4.5) describes the optical disturbance produced by a plane wave passing through an arbitrary aperture, at a point located at distance R from the aperture and displaced from the axis by a small distance Y, X. The disturbance consists of a time-varying component of amplitude E_A/R, angular frequency *co,* and phase angle kR, multiplied by a term proportional to the area of the aperture and the sum of the interferences between the spherical wavelets from each part of the aperture.

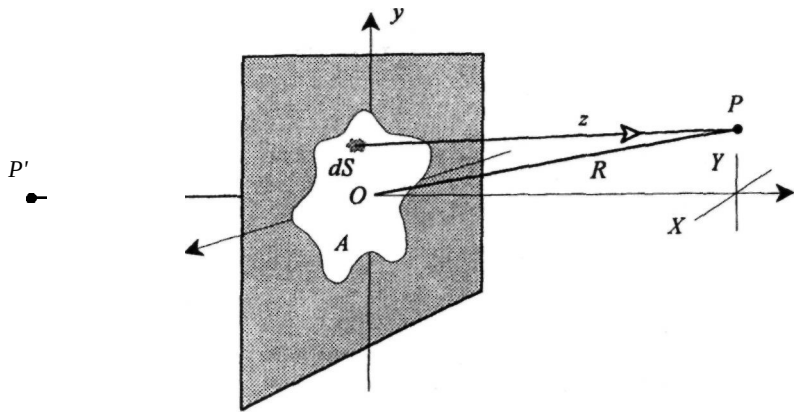


Figure 4.1 Notation for diffraction calculations.

4.2.2 Circular Aperture

The expression for $U(P)$ is considerably simplified when it is evaluated for specific aperture shapes. For the circular apertures used in most imaging systems, it is appropriate to use polar coordinates in the aperture and observation planes, as indicated in figure 4.2. The following quantities are defined

$$x = \rho \cos\theta \qquad y = \rho \sin\theta$$
$$X = r \cos\Psi \qquad Y = r \sin\Psi \tag{4.6}$$

The elemental area in the aperture is then

$$dS = \rho \, d\rho \, d\theta \tag{4.7}$$

Substituting in equation (4.5),

$$U(P) = \frac{E_A e^{i(\omega t - kR)}}{R} \int_0^a \rho \int_0^{2\pi} e^{i(k\rho r/R)\cos(\theta - \Psi)} d\rho \, d\theta \tag{4.8}$$

Because of the circular symmetry, the value of $U(P)$ is independent of the angle Ψ. The integral terms are considerably simplified by the use of Bessel functions. Using the definition

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{ix\cos\alpha} \, d\alpha$$

and the relation

$$J_1(x) = \frac{1}{x} \int_0^x x' J_0(x') \, dx'$$

equation (4.8) may be expressed as

$$U(t) = \frac{E_A e^{i(\omega t - kR)}}{R} 2\pi a^2 \left(\frac{R}{kar}\right) J_1\left(\frac{kar}{R}\right) \tag{4.9}$$

where a is the radius of the circular aperture.

For imaging purposes, the observable property of the light is its intensity, which is the time average of the energy falling on the unit area in unit time. The intensity at point P is equal to $\langle |U|^2 \rangle$, giving the result

$$I = \frac{2E_A^2 A^2}{R^2} \left[\frac{J_1(kar/R)}{kar/R}\right]^2 \tag{4.10}$$

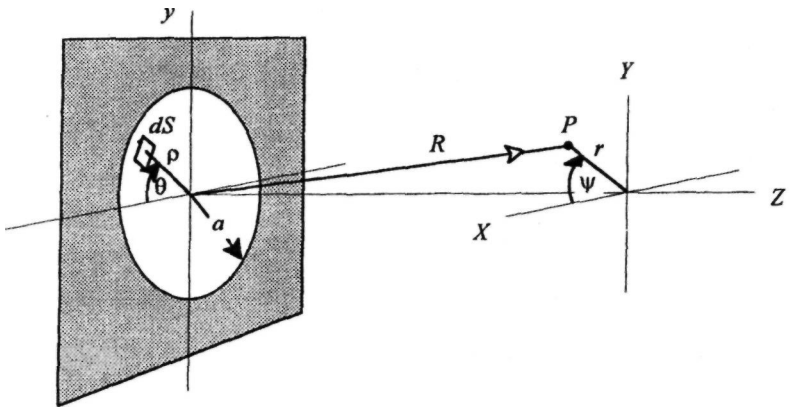


Figure 4.2 Model and notation for diffraction by a circular aperture.

where A is the area of the circular aperture. The irradiance at the center, where $r = 0$, is

$$I(0) = \frac{E_A^2 A^2}{2R^2} \tag{4.11}$$

The irradiance of the image in terms of the field angle β is found by setting $\sin \beta = r/R$ and assuming that R is constant over the pattern, the intensity is

$$I = I(0)\left[\frac{2J_1(ka \sin \beta)}{ka \sin \beta}\right]^2 \tag{4.12}$$

This expression, first derived by Airy [1835], describes the diffraction pattern formed at a large distance from a clear circular aperture. When a perfect lens (larger than the aperture, so as not to obstruct it) is placed close to the aperture, then the plane wave passing through the aperture is converted to a spherical wave converging to a point in the focal plane of the lens. The diffraction pattern is unaltered by this process, and its angular size remains the same, being a function only of k, a, and β.

Using both linear and logarithmic intensity scales, the profile of the Airy diffraction pattern is shown in figure 4.3 for an aperture of 4 m at wavelengths of 0.55 and 2.2 μm. The linear plot is dominated by the central peak, known as the Airy disk. The minima of the intensity function occur when $J_1(x) = 0$, except when $x = 0$. The first minimum gives a value of $ka \sin \beta = 1.22\pi$, so that the angular radius of the first dark ring is

$$\sin \beta = 1.22 \frac{\lambda}{D} \tag{4.13}$$

The secondary maxima occur when $J_2(x) = 0$; that is when $ka \sin\beta$ equals 5.14, 8.42, 11.6, and so on. The corresponding peak values of the first three rings are 0.0175, 0.0042, and 0.0016, normalized to the central peak of unity.

The fraction of the total incident energy contained within a circle of radius w_0 centered on the peak, is given by

$$L(w_0) = 1 - J_0^2(kaw_0) - J_1^2(kaw_0) \tag{4.14}$$

The fraction of energy within the first, second, and third dark rings is 0.838, 0.91, and 0.938, respectively. It should be noted that for a constant irradiance at the aperture, the peak intensity from equation (4.11) is proportional to A^2/R^2. For a constant focal length, the peak image intensity is proportional to D^4, which is explained by the fact that a larger aperture not only increases the amount of energy collected, but also reduces the size of the Airy disk.

The logarithmic plot shows the structure of the diffraction rings, which are inconspicuous in the linear plot. For visual use, the diffraction rings are relatively unimportant, but for astronomical imaging, especially when using charge-coupled device (CCD) detectors with a large dynamic range, the ring structure produced by an aperture with a sharp cutoff can be a serious limitation. This subject is discussed further in section 4.3.3.

The image of a point source is always of finite size, which implies that there is an uncertainty in the exact position of the source. This uncertainty is due to the fact that only a part (in the case of distant objects, such as stars, an extremely small part) of the radiation from the object is used to form the image. As the collecting aperture gets smaller, the fraction of radiation collected decreases, enlarging the image and increasing the uncertainty in the position of the source. In the limit of an infinitely small aperture, the diffraction pattern expands to form a spherical wave centered on the aperture, from which no information on the position of the original source can be obtained. Diffraction theory provides a consistent and intuitively satisfying explanation of these effects.

For a given photon flux (signal-to-noise ratio), the aperture size sets an absolute limit on the spatial information that can be obtained about the object. In practice, this limit is further reduced by disturbances in the optical path and the telescope system. The function of adaptive optics is to remove the effects of these disturbances and to restore the full spatial resolution determined by the aperture size.

4.2.3 Fourier Transforms

In general, both amplitude and phase variations will exist in the wavefront at an optical aperture. In equation (4.5), the amplitude term before the integral can be regarded as a single complex quantity

$$\mathscr{A}(y, z) = \mathscr{A}_0(y, z)e^{i\phi(y,z)} \tag{4.15}$$

where

$\mathscr{A}_0$ = amplitude variation across the pupil
$e^{i\phi(y,z)}$ = point-by-point phase variations

Equation (4.5) may then be rewritten as

$$E(Y, X) = \int_{-\infty}^{+\infty}\int \mathscr{A}(y, x)e^{ik(Yy+Xx)/R}\,dy\,dx \tag{4.16}$$

If spatial frequencies f_y and f_x are now introduced, defined as

$$f_y = k\frac{Y}{R} \qquad f_x = k\frac{X}{R} \tag{4.17}$$

the diffracted field may be written as

$$E(f_y, f_x) = \int_{-\infty}^{+\infty}\int \mathscr{A}(y, x)e^{i(f_y y + f_x x)}\,dy\,dx \tag{4.18}$$

Equation (4.18) is recognized as the Fourier transform relationship. In other words, the field dis-

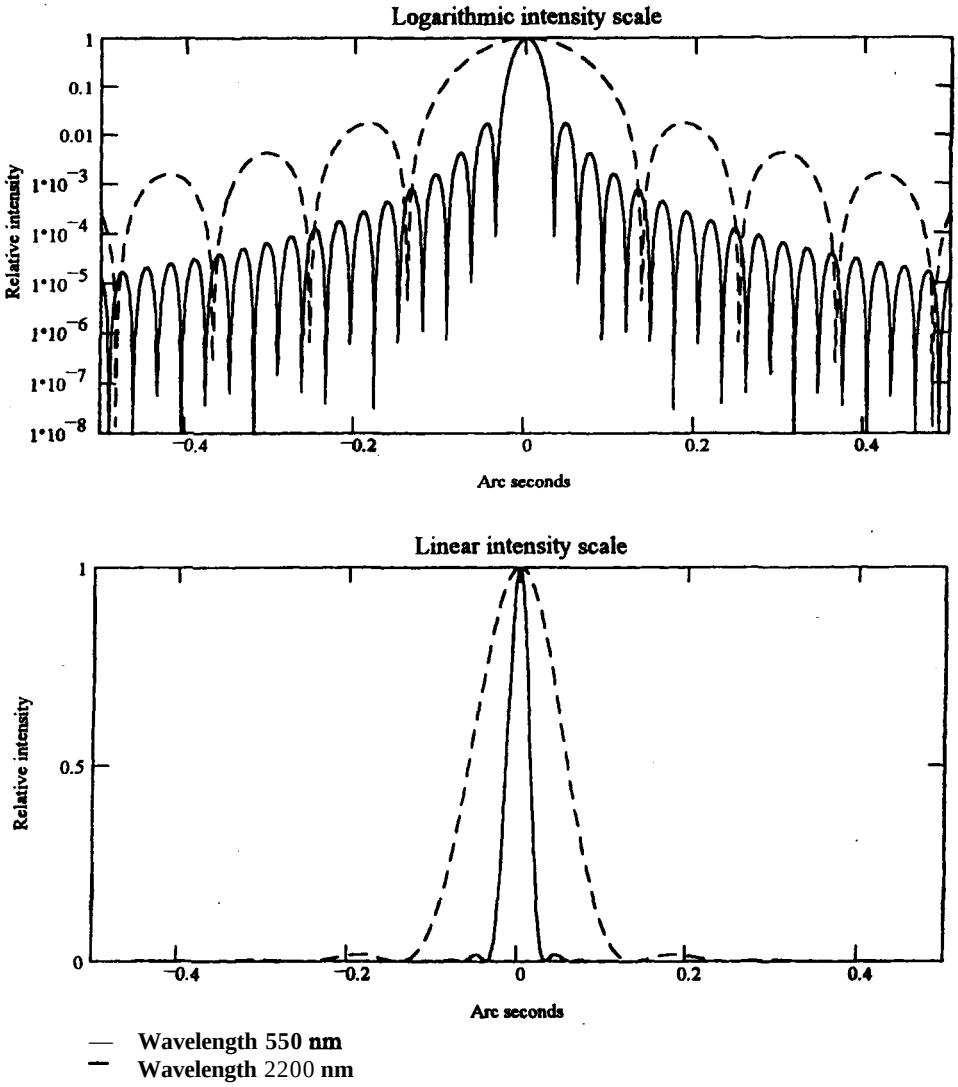


Figure 4.3 Intensity profiles of the image produced by a diffraction-limited circular aperture, shown with logarithmic and linear scales. The diameter is 4 m and the observing wavelengths are 550 and 2200 nm.

tribution $E(f_y f_x)$ in the image plane is the spatial frequency spectrum of the aperture function *ri{y, x).* The intensity distribution in the image plane for a given pupil function $\mathscr{A}(y, x)$ is given by $|E(f_y, f_x)|^2$.

The normalized optical transfer function of the imaging system may be expressed as

$$OTF(f_x, f_y) = \frac{\int_{-\infty}^{+\infty} \int \mathscr{A}(x, y)\mathscr{A}^*(x - f_x, y - f_y)\, dx\, dy}{\int_{-\infty}^{+\infty} \int |\mathscr{A}(x, y)|^2\, dx\, dy} \tag{4.19}$$

This equation relates the spatial frequency response of the imaging system directly to the wavefront deformation in the pupil.

For a perfect imaging system with a circular aperture, the optical transfer function for incoherent illumination is

$$OTF(\rho) = \frac{2}{\pi}\left\{\cos^{-1}\left(\frac{\rho}{2a}\right) - \frac{\rho}{2a} \times \left[1 - \left(\frac{\rho}{2a}\right)^2\right]^{1/2}\right\} \qquad 0 < \rho < 2a$$

$$= 0 \qquad \rho > 2a \tag{4.20}$$

where $p/2a$ is the normalized aperture radius. The resulting image intensity distribution or point spread function for a plane wave input is the Airy diffraction pattern given in equation (4.12).

4.2.4 Gaussian Beams

A circular aperture with a sharp cutoff produces the Airy diffraction pattern characterized by extensive diffraction rings, which contain about 16% of the total energy. If the transmission at the edge of the aperture is tapered off gradually, diffraction effects are considerably reduced. The Gaussian profile fulfills this requirement. The Fourier transform of a Gaussian function is also Gaussian, so that a Gaussian irradiance distribution with uniform phase in the aperture produces a Gaussian diffraction pattern, which decays smoothly with no sidelobes. The process of suppressing the subsidiary maxima surrounding the main response is known as apodization (literally "taking away the feet") and has been applied to astronomical telescopes. It is achieved by modifying the aperture function, which always involves a loss of energy, as well as an increase in the width of the image core [Jacquinot and Roizen-Dossier 1964].

A Gaussian distribution of energy in the focal plane may also be produced by random image motion during a time exposure, because of imperfect compensation of wavefront tilt. In this case, the Gaussian distribution is superimposed on the diffraction pattern because of higher order wavefront aberrations. Gaussian profiles have the form

$$A(r) = A_0 e^{-(r/\sigma)^2/2} \tag{4.21}$$

where A_0 is the peak value and σ is the standard deviation. The peak of a Gaussian profile has a similar shape to that of the Airy function, and by suitable choice of the parameter σ, the width of these functions can be made to match at any desired intensity. For example, to match the two functions at half of their peak values, the standard deviation of the Gaussian profile, as defined above, is $\sigma = 0.44\lambda/D$.

The Airy and Gaussian diffraction patterns are compared in figure 4.4, for peak amplitudes of unity, with $\sigma = 0.44\lambda/D$. The peaks of these two curves match very closely at values above half-maximum, but on the linear scale the Gaussian appears to have a wider base. It is apparent from the logarithmic plot that the Gaussian drops more rapidly at larger radii, due to the absence of sidelobes.

Many lasers generate a Gaussian irradiance profile. The TEM_{00} cavity mode produces a plane wavefront with Gaussian intensity distribution. A Gaussian beam must necessarily be truncated at some radius, and it is conventional to use the radius w at which the intensity is reduced to $1/e^2$. The irradiance distribution of a Gaussian TEM_{00} beam of total power P is then

$$I(r) = I_0 e^{-2(r^2/w^2)} = \frac{2P}{\pi w^2} e^{-2(r^2/w^2)} \tag{4.22}$$

In the absence of aberrations, the irradiance profile of a Gaussian beam remains intact as the beam propagates, irrespective of whether it is expanded or concentrated. This behavior is quite different from a beam of initially uniform intensity defined by a hard aperture; in this case, the profile changes considerably during propagation, as shown in section 4.2.2.

In accordance with the Huygens-Fresnel principle, a laser beam with Gaussian intensity distribution and initially flat phase will expand due to diffraction as it propagates. After travelling a distance z, the beam radius at the $1/e^2$ contour is given by

$$w(z) = w_0 \left[1 + \left(\frac{\lambda z}{\pi w_0^2}\right)\right]^{1/2} \tag{4.23}$$

where w_0 is the beam "waist"; that is, the radius of the $1/e^2$ contour at which the wavefront was flat. Further information on laser beam optics may be found in Siegman [1971].

4.2.5 Separability of Point Sources

The separation of closely spaced point sources, such as stars, has long been a measure of telescope performance. The ability to separate two images depends not only on the quality of the optical system (including the atmosphere) but also on the characteristics of the detector. The number of photons detected and the dynamic range of the detector determine the minimum intensity change that can be discerned, setting an ultimate limit to the spatial resolution of an optical imaging system. The well-known Rayleigh criterion was originally proposed as a means of comparing the resolution of spectroscopes having a $(\sin x/x)^2$ intensity distribution. It has since been adopted as a standard measure of the spatial resolution of optical imaging systems. The "Rayleigh resolution" is achieved with two point (or line) sources of equal brightness when each peak coincides with the first dark line or ring in the diffraction pattern of the other source. For a circular aperture, this occurs when the angle between the peaks is $1.22\lambda/D$ radians. This condition is shown in figure 4.5(a). Between the peaks, the intensity dips to 74% of the peak value.

The Rayleigh criterion does not represent a fundamental limit to resolving power. Around 1850, Dawes conducted experiments to separate double stars, finding that sixth-magnitude white stars as close as 0.8 arc second could be separated visually with a telescope of 150 mm aperture. This implies an angular separation of $1.05\lambda/D$ rad at a mean wavelength of $0.55\,\mu$m. The dip in this case is 3%, which is close to the minimum contrast that can be distinguished by the eye. The Dawes limit is depicted in figure 4.5(b).

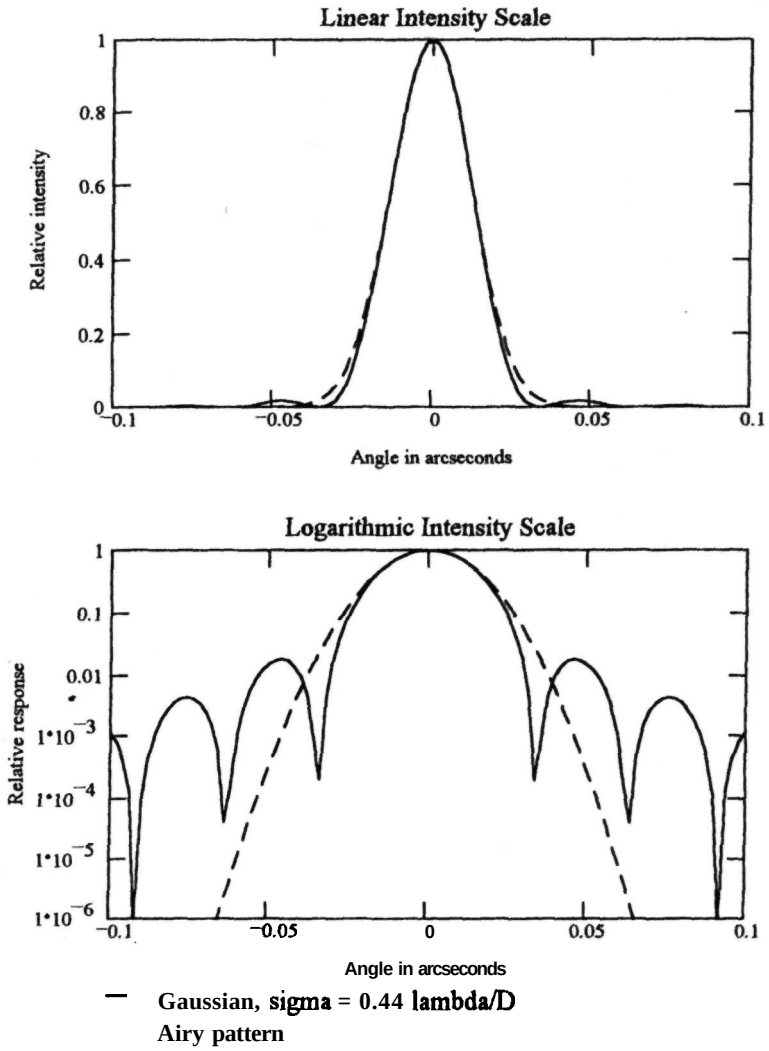


Figure 4.4 Comparison of Airy and Gaussian profiles. The angular scale is for a 4-m aperture at a wavelength of $0.55\,\mu$m.

The advent of photographic and, more recently, photoelectric detectors with a large dynamic range has dramatically improved the ability to distinguish small differences in brightness. Sparrow [1916] proposed a resolution limit corresponding to the distance between two point source images when the second derivative of the resulting envelope vanishes at the center of the pattern. The Sparrow criterion for two sources of equal intensity, which occurs at a separation of $0.95\lambda/D$, is shown in figure 4.5(c). With two stars, the resulting intensity pattern is significantly elongated in the direction of separation, making recognition more certain.

At closer separations, the peaks merge into a smooth profile, as shown in figure 4.5(d), which depicts a separation of $0.8\lambda/D$. In principle, the presence of multiple stars can still be detected by comparing the measured profile with the known profile of a single point. The overall width of the combined profile and the elongation of brightness contours provide the critical information.

With two stars of unequal brightness, the combined response becomes asymmetric, making discrimination more difficult for close separations. The main criteria outlined above are recognizable for small magnitude differences up to about 2.5, corresponding to an intensity ratio of 10:1. For larger magnitude differences, the presence of a faint companion is revealed by the elongation of intensity contours, until it becomes submerged in the diffraction rings or image halo.

The ultimate limit of detection is set by the signal-to-noise ratio of the measuring system. With perfect measurements, arbitrarily small differences in inten-

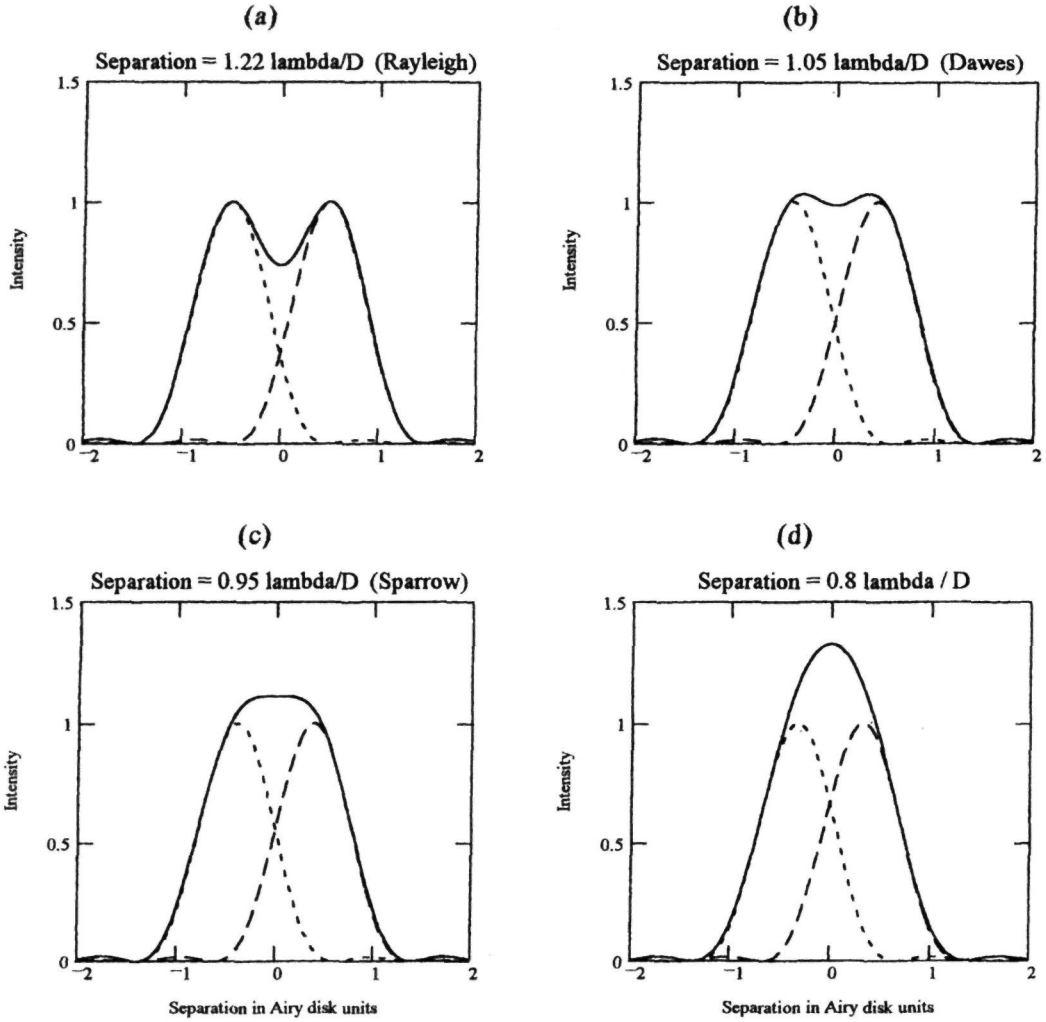


Figure 4.5 Separability of two point sources of equal intensity in a diffraction-limited optical imaging system.

sity can be detected, so there is no limit to the resolving power if the shape of the response is known. To quantify the effect of noise in optical images, Fried [1979] introduced the concept of "resolution scale" *80,* which is a basic property of an imaging system, defined as

$$\delta\theta = \frac{\lambda/D}{[2\pi^2 G(0)]^{1/2}}$$

where $G(0)$ is the optical transfer function of the system. The precision with which the location of a single point source can be measured is *S6/SNR,* where *SNR* is the signal-to-noise ratio of the measuring device. The resolution scale is optimized (that is, minimized) in two ways: (1) by making $G(0)$ as large as possible, which requires minimizing the wavefront aberrations; and (2) by maximizing the signal-to-noise ratio. Even small errors reduce (7(0), implying that adaptive optics potentially has a role even in optical systems that are close to diffraction limited.

4.3 Wavefront Distortion

4.3.1 Fresnel-Kirchhoff Diffraction Integral

Attention will now be turned to the effects of wavefront distortion on the structure of optical images. The starting point for this analysis is the Fresnel-Kirchhoff diffraction integral. The derivation of this formula is described in chapter 8 of Born and Wolf [1975] and will not be repeated here. The phy-

sical model and notation are shown in figure 4.6. A point source P' generates spherical waves that reach an opaque screen having an aperture that allows a section of the wavefront W to pass through. The resulting field is observed at a point P on the far side of the screen. The following boundary conditions were assumed by Kirchhoff: (1) The field distribution and its derivative within the aperture are unaffected by the presence of the screen. (2) The field and its derivative in the geometrical shadow of the screen are zero. These assumptions are not strictly true because they ignore diffraction at the edges of the aperture. However, in practical cases, where the dimensions of the aperture are large compared with the wavelength of the light, but small compared with the distances r and s of the points P' and P, respectively, the results agree closely with experiment. The disturbance at P may be expressed as

$$U(P) = -\frac{iA}{2\lambda}\iint_W \frac{e^{ik(r+s)}}{rs}(1+\cos\chi)\,dS \qquad (4.24)$$

where dS is an elemental area of the wavefront at position Q, as shown in figure 4.6.

This equation is a somewhat more general form of equation (4.5). The term (1 + cos x) is the inclination factor, χ being the angle between the propagation directions $P'Q$ and QP. If the aperture is much smaller than the distances r and s, or a lens is used in the aperture to reimage P' on P, then $(1 + \cos\chi)$ may be replaced by $2\cos 8$, where S is the angle between the direction $P'P$ and the normal to the aperture. When P' and P are close to the optical axis, as is usually the case in astronomical optics, then the simplification may be taken a step further by setting $\cos\delta$ equal to 1. In that case, the diffraction integral reduces to

$$U(P) = -\frac{iA}{\lambda}\iint_W \frac{e^{ik(r+s)}}{rs}\,dS \qquad (4.25)$$

This equation enables the three-dimensional image structure in the vicinity of P to be determined. The values of r and s depend on the positions of Q and P.

To investigate the effects of wavefront aberrations on the image structure, it is convenient to rearrange the geometry as shown in figure 4.7. A spherical

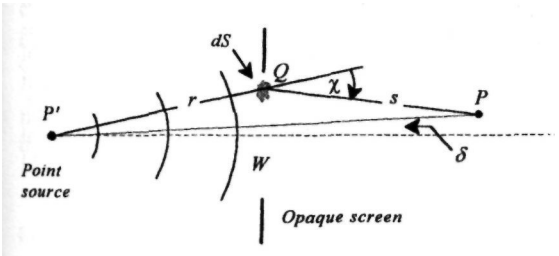


Figure 4.6 Geometry and notation for Fresnel-Kirchhoff diffraction integral.

reference wavefront of radius R is convergent on the point O in the image plane, which is the origin of the polar coordinate system. The wavefront is observed at point P, located at polar coordinates (σ, Ψ). The pupil radius is a. The normally spherical wavefront is deformed by small variations $\phi(\rho, 6)$ in its radius, where $(p, 6)$ are the polar coordinates in the pupil. The distance between P and an elemental area dS at point Q on the pupil is s. At points close to the axis, the distance s is approximately equal to R, so that s may be replaced with R in the amplitude term. Using equation (4.25), the resulting disturbance at P is

$$U(P) = -\frac{i}{\lambda}\frac{A}{R^2}\iint e^{ik(\phi+s-R)}\,dS \qquad (4.26)$$

Geometrical considerations show that $(s - R)$ is given by

$$(s-R) = \frac{a\rho r}{R}\cos(\theta-\Psi) - z\left[1 - \frac{1}{2}\frac{a^2\rho^2}{R^2} + \cdots\right] \qquad (4.27)$$

where z is the axial displacement of P from the image plane. Using

$$u = k\left(\frac{a}{R}\right)^2 z \quad \text{and} \quad v = k\left(\frac{a}{R}\right)\sigma \qquad (4.28)$$

it follows that

$$k(s-R) = v\rho\cos(\theta-\Psi) - u\left(\frac{R}{a}\right)^2 + \frac{1}{2}u\rho^2 \qquad (4.29)$$

The area of the element dS in the pupil is

$$dS = a^2\rho\,d\rho\,d\theta \qquad (4.30)$$

The disturbance at P may then be expressed as

$$U(P) = -\frac{i}{\lambda}\frac{Aa^2}{R^2}e^{iu(R/a)^2} \times \int_0^1\int_0^{2\pi} e^{i[\phi(\rho,\theta)-v\rho\cos(\theta-\Psi)-0.5u\rho^2]}\rho\,d\rho\,d\theta \qquad (4.31)$$

The intensity at P is

$$I(P) = |U(P)|^2 = \left(\frac{Aa^2}{\lambda R^2}\right)^2 \times \left|\int_0^1\int_0^{2\pi} e^{i[\phi(\rho,\theta)-v\rho\cos(\theta-\Psi)-0.5u\rho^2]}\rho\,d\rho\,d\theta\right|^2 \qquad (4.32)$$

There is no explicit solution to this equation, as there is for the diffraction-limited case that produced the Airy pattern, as treated in section 4.2.2. The aberration function $\phi(\rho, 8)$ can be handled in two ways. The first method is to expand the aberration function and find separately the diffraction pattern associated

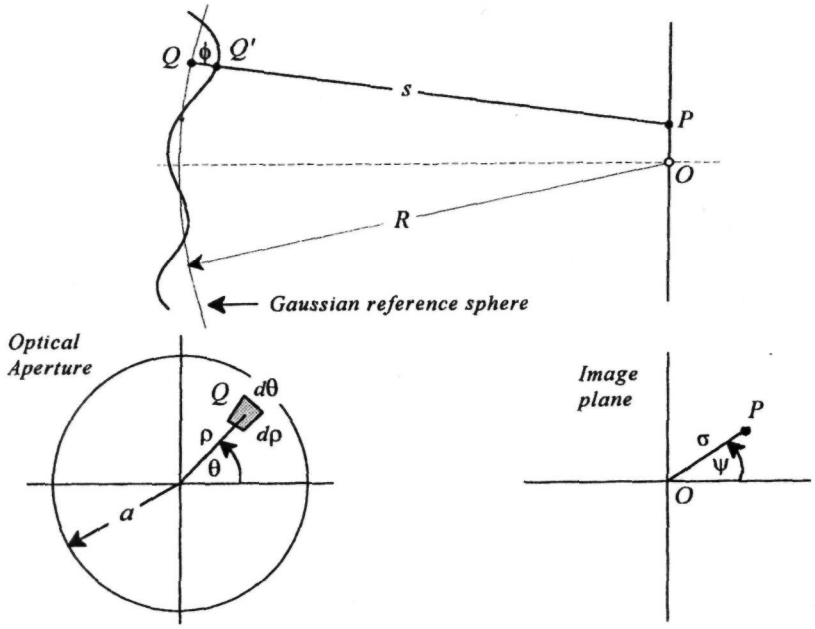


Figure 4.7 Geometry for diffraction integral with aberrations.

with each term. The aberration is then expressed as a sum of whole-aperture modes, as described in section 4.3.2.

The second method is to assume that the wavefront error is random, so that its effects can be determined statistically, in terms of a mean-square wavefront error. This approach may be used to determine the performance of adaptive optics systems, in which the errors are primarily random, and it is developed further in section 4.3.3. However, such an analysis is valid only for small errors.

4.3.2 Zernike Polynomials

The Zernike polynomials, which are orthogonal over a unit circle, are a useful basis for expanding the aberration function, because the terms represent the basic aberrations of conventional optical systems, such as defocus, astigmatism, and coma. This method is used in the design of "static" optical systems for balancing the various aberrations in order to obtain the best image.

Zernike polynomials may also be used to represent (and to compensate) atmospheric turbulence, as described in section 3.5. Zernike polynomials are not strictly orthogonal for compensating turbulence because they are not statistically independent. A more general set, using Karhunen-Loeve functions, has been configured for this application, although the difference is small for the low-order terms [Wang and Markey, 1978].

For small values of D/r_0, the modal approach to wavefront compensation has the advantage that efficient compensation is obtained with a smaller number of degrees of freedom than with zonal compensation.

4.3.3 Strehl Ratio

It is to be expected that wavefront errors will diffract light away from the center of the image and thereby reduce the peak intensity. To quantize this reduction in peak intensity, the intensity /* obtained at the Gaussian image point is first computed with no aberrations present. Setting $<p = 0$, $u — 0$, and $v = 0$ in equation (4.32), then

$$I^* = \pi^2 \left(\frac{Aa^2}{\lambda R^2} \right)^2 \tag{4.33}$$

The ratio of the actual peak intensity $I(P)$ to /* is a useful measure of the performance of an optical system. When this ratio is evaluated at a specific plane of observation (normally the Gaussian image plane), it is known as the Strehl ratio, after K. Strehl [1895, 1902], who first described the concept. For an optical system with aberration function $\phi(\rho, 9)$, the Strehl ratio is given by

$$S = \frac{I(P)}{I*} = \frac{1}{\pi^2} \left| \int_0^1 \int_0^{2\pi} e^{ik\phi(\rho,\theta)} \rho \, d\rho \, d\theta \right|^2 \tag{4.34}$$

To evaluate the Strehl ratio using equation (4.34), it is necessary to know the aberration function ϕ explicitly in terms of the pupil coordinates *(p, 0).* This is rarely the case in adaptive optics, where the random errors due to turbulence are usually described statistically.

Several expressions have been developed to allow the Strehl ratio to be determined directly from the mean-square wavefront error. These approximations have limited applicability and must be used with care. Born and Wolf [1975] derive an expression by first expanding the phase term in equation (4.34) to obtain a series representation of the Strehl ratio:

$$S = \frac{1}{\pi^2}\left|\int_0^1\int_0^{2\pi}\left[1 + ik\phi(\rho,\theta) + \tfrac{1}{2}[ik\phi(\rho,\theta)]^2 + \cdots\right] \times \rho\, d\rho\, d\theta\right|^2 \tag{4.35}$$

The next step is to define "aperture average" values of the wavefront errors over the pupil, with respect to a reference sphere centered at the focal point *P.* These average wavefront errors are given by

$$\overline{\phi_p^n} = \frac{\int_0^1\int_0^{2\pi}\phi^n(\rho,\theta)\rho\, d\rho\, d\theta}{\int_0^1\int_0^{2\pi}\rho\, d\rho\, d\theta} = \frac{1}{\pi}\int_0^1\int_0^{2\pi}\phi^n(\rho,\theta)\rho\, d\rho\, d\theta \tag{4.36}$$

Neglecting third and higher order terms in equation (4.35), the Strehl ratio can then be expressed as

$$S' \approx \left|1 + ik\bar{\phi}_p - \tfrac{1}{2}k^2\overline{\phi_p^2}\right|^2 = 1 - k^2\left[\overline{\phi_p^2} - (\bar{\phi}_p)^2\right] \tag{4.37}$$

The mean-square wavefront error is defined as $(\Delta\phi_p)^2 = \phi_p^2 - (\bar{\phi}_p)^2$. Setting the standard deviation of the phase as $\sigma_p = k\Delta\phi_p$, the Strehl ratio is approximated as

$$S' \approx 1 - (\sigma_p)^2 \tag{4.38}$$

This expression is useful only for very small phase errors, up to about 0.6 rad (1/10 wave) rms. It falls to zero at 1 rad of phase error, which is clearly unrealistic. Simulations of turbulence-degraded wavefronts show that well-defined diffraction-limited image cores with normalized peak values of at least 0.2 are obtained with random wavefront phase errors of 1 rad rms.

Marechal [1947] derived a similar expression for the normalized intensity at the diffraction focus in the presence of small aberrations:

$$S'' \geq \left[1 - \tfrac{1}{2}(\sigma_p)^2\right]^2$$

For small values of σ_p, up to about 0.5 rad rms, this gives essentially the same result as equation (4.38). It falls to zero when σ_p = 1.41 rad rms. Marechal suggested as a criterion, that an optical system should be considered well corrected when *S"* equals or exceeds 0.8, corresponding to a wavefront error of $\lambda/14$ or less. This is an acceptable criterion for fixed optical components, such as lenses or mirrors, but is unnecessarily restrictive as a system specification for adaptive optics systems.

Equation 4.38 represents the first two terms in the exponential expansion, leading to an alternative approximation for the Strehl ratio:

$$S \approx e^{-(\sigma_p)^2} \tag{4.40}$$

This equation appears to be valid over a somewhat larger range of phase errors than the other two; that is, up to about 2 rad rms. It is sometimes called the "extended Marechal approximation" and it is the most commonly used expression for the Strehl ratio in the field of adaptive optics. Unless otherwise stated, this is the definition of the Strehl ratio S used in this book.

Larger wavefront errors, such as those caused by uncorrected atmospheric turbulence, produce more radical changes in the image structure and must be treated in a different way, as outlined in the next section.

4.4 Turbulence-Degraded Images

4.4.1 Optical Transfer Function and Image Intensity Distribution

Although the total residual error provides a measure of the performance of a compensated telescope, enabling the Strehl ratio to be determined, far more information can be obtained from the Optical Transfer Function (OTF), which describes the complex amplitude of the received wavefront as a function of spatial frequency in the optical aperture. The corresponding image intensity distribution or Point Spread Function (PSF) is derived from the OTF by a Fourier transformation.

The optical transfer function (OTF) can be computed from the phase structure function of the compensated wavefront. The structure function may represent any random process that can be modeled with Gaussian statistics and zero mean value, which includes most of the errors that occur in adaptive optics systems. It is possible to determine the structure of images with any degree of compensation, ranging from images produced by uncorrected turbulence to those that are diffraction-limited..

For an aperture of area *A,* having an incident wavefront given by $\mathscr{A}(\mathbf{x}) = e^{\phi(\mathbf{x})}$, the optical transfer function from equation (4.19) is

$$OTF(\rho) = \frac{\int_A\int_A \mathscr{A}(\mathbf{x}')\mathscr{A}^*(\mathbf{x})\, d\mathbf{x}\, d\mathbf{x}'}{\int_A\int_A |\mathscr{A}(\mathbf{x})|^2\, d\mathbf{x}^2} = \frac{1}{A^2}\int_A\int_A e^{i[\phi(\mathbf{x}')-\phi(\mathbf{x})]}\, d\mathbf{x}\, d\mathbf{x}' \tag{4.41}$$

where x is the position in the aperture, $\mathbf{x}' = x + p$, and p is the displacement in the aperture. The corresponding spatial frequency in the image plane, in the axis of displacement, is $f = p/XF$, where X is the mean wavelength and F is the focal distance to the image plane.

If $[\phi(\mathbf{x}') - \phi(\mathbf{x})]$ is an unbiased Gaussian random variable, then the time-averaged (long-exposure) OTF may be expressed as

$$\langle OTF(\rho)\rangle = \frac{1}{A^2}\int_A\int_A e^{-0.5\langle[\phi(\mathbf{x}')-\phi(\mathbf{x})]^2\rangle}\,d\mathbf{x}\,d\mathbf{x}' \quad (4.42)$$

In section 3.2, the phase structure function is defined as

$$D(\rho) = D(|\mathbf{x}' - \mathbf{x}|) = \langle[\phi(\mathbf{x}') - \phi(\mathbf{x})]^2\rangle$$

$$\begin{aligned}\langle OTF(\rho)\rangle &= \frac{1}{A^2}e^{-0.5D(\rho)}\int_A\int_A d\mathbf{x}\,d\mathbf{x}' \\ &= e^{-0.5D(\rho)}T(\rho)\end{aligned} \quad (4.43)$$

The long-exposure OTF is equal to the product of the atmospheric OTF and that of the telescope, $T(p)$. The time-averaged intensity distribution in the focal plane is found by taking the Fourier transform of the OTF:

$$I(y) = \mathcal{F}\left[e^{-0.5D(\rho)}T(\rho)\right] \quad (4.44)$$

where y is the spatial dimension in the focal plane.

This procedure enables the image structure to be found for any wavefront for which a phase structure function $D(p)$ can be defined.

4.4.2 Structure Functions

The rationale for using structure functions in turbulence theory has been discussed in chapter 3, which also contained a summary of their properties. Adaptive optics systems generally operate with imperfect compensation, in which the partially corrected wavefronts contain the sum of individual errors from various parts of the system. If these errors are random and uncorrelated, their individual structure functions may be summed to obtain the overall structure function $D(\rho)$, which is the key to determining basic performance parameters such as the Modulation Transfer Function (MTF), and the Point Spread Function (PSF) of the image.

For uncompensated atmospheric turbulence, the phase structure function from equation (3.52) is

$$D_U(\rho) = 6.88\left(\frac{\rho}{r_0}\right)^{5/3} \quad (4.45)$$

where

p = displacement in the optical aperture

r_0 — turbulence coherence length

The structure function is zero when $p = 0$ and rises to a saturation value of $2\sigma^2$ when p is greater than the outer scale of the turbulence, where $\sigma^2 = 1.85(\mathscr{L}_0/r_0)^{5/3}$ for plane waves. When an optical wavefront is compensated, a correction device is placed in the optical path to reduce the value of $D(\rho)$. When the correction is perfect, $D(\rho)$ is zero for all values of p and the optical transfer function of the system becomes that of the telescope alone, $T(p)$.

4.4.3 Partial Compensation

Practical adaptive optics systems operate somewhere between the two extremes of zero and perfect compensation. The main factor that decides the level of compensation is the number of degrees of freedom in the wavefront compensator, which is one of the most important design decisions to be made in an adaptive optics system. In the case of a zonal compensation system, this is equivalent to fixing the ratio D/d, which determines the number and size of the subapertures. It is instructive to review the effect of varying the subaperture size, or more generally the ratio d/r_0, on the shapes of the transfer function and point spread function. To illustrate the results clearly, only the fitting error will be considered at this time.

In zonal compensation systems, the wavefront is corrected by an array of subapertures that match the local gradient and/or phase errors, relative to the aperture-averaged values. Segmented mirrors have independent elements with either piston-only correction, using one actuator per element, or piston and tilt correction, using 3 degrees of freedom per element. Deformable mirrors employ arrays of actuators with influence functions that may extend into the adjacent zones. It has been found by computer simulation that such devices are accurately modeled by a piston and tilt mirror, in which the segment size is $\sqrt{3}$ times the actuator spacing in the deformable mirror.

The first step in determining the effect of fitting errors on the performance of a zonal compensation system is to find the structure function of a turbulence-degraded wavefront partially corrected over an array of zones. The phase structure function as stated by Tatarski [1971] is

$$D(\rho) = 4\pi\int_0^\infty [1 - J_0(k\rho)]F_\Psi(k)\,k\,dk \quad (4.46)$$

where F_Ψ is the two-dimensional spatial spectrum of the phase variations. To determine the structure function of a partially compensated wavefront, Greenwood [1979] shows that these spectra may be expressed in terms of high-pass filter functions that depend on the type of correction:

$$\begin{aligned}F_\Psi &= F_\phi H_P \\ F_{\Psi'} &= F_\phi H_{PT}\end{aligned} \quad (4.47)$$

where H_P is the filter for piston correction only, and H_{PT} is the filter for piston and tilt. These filter functions are given by

$$H_P = 1 - \left[\frac{2J_1(kd/2)}{kd/2}\right]^2$$
$$H_{PT} = 1 - \left[\frac{2J_1(kd/2)}{kd/2}\right]^2 - 16\left(\frac{2}{kd}\right)^2 J_2^2\left(\frac{kd}{2}\right) \quad (4.48)$$

The turbulence spectrum in the simplest case (infinite outer scale, zero inner scale, no amplitude effects) is defined by

$$F_\phi(k) = 0.489 r_0^{-5/3} k^{-11/3} \quad (4.49)$$

When amplitude effects are taken into account, the spectrum is given by

$$F_\phi(k) = 0.489 r_0^{-5/3} k^{-11/3} \frac{1}{2}\left[1 + \operatorname{sinc}\left(\frac{k^2 \lambda L}{2\pi}\right)\right] \quad (4.50)$$

where $\sqrt{\lambda L}$ is the correlation length of the amplitude fluctuations (see figure 3.9). The phase structure functions for zonal compensation are then found by substituting equations (4.48) and (4.49) or (4.50) in equation (4.46), to obtain

$$D(x) = 6.145(d/r_0)^{5/3} \int_0^\infty [1 - J_0(ux)]H(u/2)u^{-8/3}\, du \quad (4.51)$$

where $u = kd$

$x = \rho/d$

The structure functions for zonal compensation are dominated by two asymptotes. For $x \ll 1$, the structure function is given by $D(x) = 6.88(d/r_0)^{5/3} x^{5/3}$, while for $x \gg 1$ the value of $D(x)$ saturates at $2a^2$, where $\sigma^2 = 1.03(d/r_0)^{5/3}$ for piston-only correction, and $\sigma^2 = 0.134(d/r_0)^{5/3}$ for piston and tilt correction. These asymptotes also control the structure of the corresponding MTFs and PSFs, as can be seen from figures 4.8 and 4.9.

To illustrate the effect of atmospheric turbulence on optical images, figure 4.8 shows three-dimensional plots of the MTF and PSF for diffraction-limited, partially compensated, and uncompensated telescope systems. The images are depicted with constant linear intensity scales. The basic conditions are D = 4 m and X = 1.62 μm, with r_0 =0.15 m at the reference wavelength of 0.55 μm. Even under these relatively benign conditions, the uncompensated image has a peak intensity of less than 1/20 of the diffraction-limited response. Assuming that the overall tilt is perfectly corrected, a wavefront compensator with only six elements will raise the Strehl ratio to over 0.5 at this wavelength, assuming that the total error is about twice the fitting error.

Figure 4.9 shows the transfer functions and corresponding PSFs for a partially compensated telescope with several different ratios of d/r_0. The MTFs are plotted on a linear scale, while a logarithmic intensity scale has been used for the PSFs in order to show the structure of the halo more clearly. The three pairs of charts depict a 1-m aperture with d/r_0 ratios of 1.0, 2.0, and 3.0, producing Strehl ratios of 0.79, 0.48, and 0.24, respectively.

The compensated MTF curves clearly show two components: (1) a low-frequency core, which drops steeply from its initial value of unity toward an intercept on the horizontal axis at a value of about d/D; and (2) a mantle consisting of a reduced version of the diffraction-limited MTF with a maximum value equal to the Strehl ratio. This general shape is typical of partly compensated random wavefronts. Also shown on the charts, for comparison purposes, are the diffraction-limited MTF and the uncorrected turbulence-degraded MTFs for short exposures and long exposures.

The corresponding PSF curves at the right have normalized peak intensities equal to the Strehl ratio. The diffraction rings typical of the Airy pattern are evident only in the first two plots. This diffraction pattern, which forms the halo around the central disk, is produced by the energy within the core of the MTF. As the d/r_0 ratio increases, the halo becomes more prominent. The central disk of the image is formed by the high frequencies in the outer mantle of the MTF, which extends out to the diffraction limit. These charts show clearly that while incomplete or partial compensation of the wavefront reduces the central intensity of the image, it has little effect on the radius of the image disk, which remains close to λ/D.

Another factor to consider is the ratio of the intensities of the image core and the surrounding halo. Even with a d/r_0 ratio of 3, which produces a Strehl ratio of 0.24, the peak intensity of the core is more than 10 times greater than that of the halo. A large ratio is maintained even when the Strehl ratio has decreased to 0.1, providing adequate signal-to-noise ratio for the detection of individual point sources against a dark background. The contrast is reduced with extended objects because the halos overlap and their intensities add. The contrast between the image core and the halo is discussed further in Ssection 4.6.2.

To summarize, images with partial zonal compensation have the following properties:

- The overall size of the aperture (or, more strictly, the ratio λ/D) determines the radius of the image core, but not the shape of its intensity profile, which depends on the d/r_0 ratio.
- The radius of the image core stays close to its diffraction-limited size even when the Strehl ratio is reduced to 0.1.
- The image shape is determined by the d/r_0 ratio.

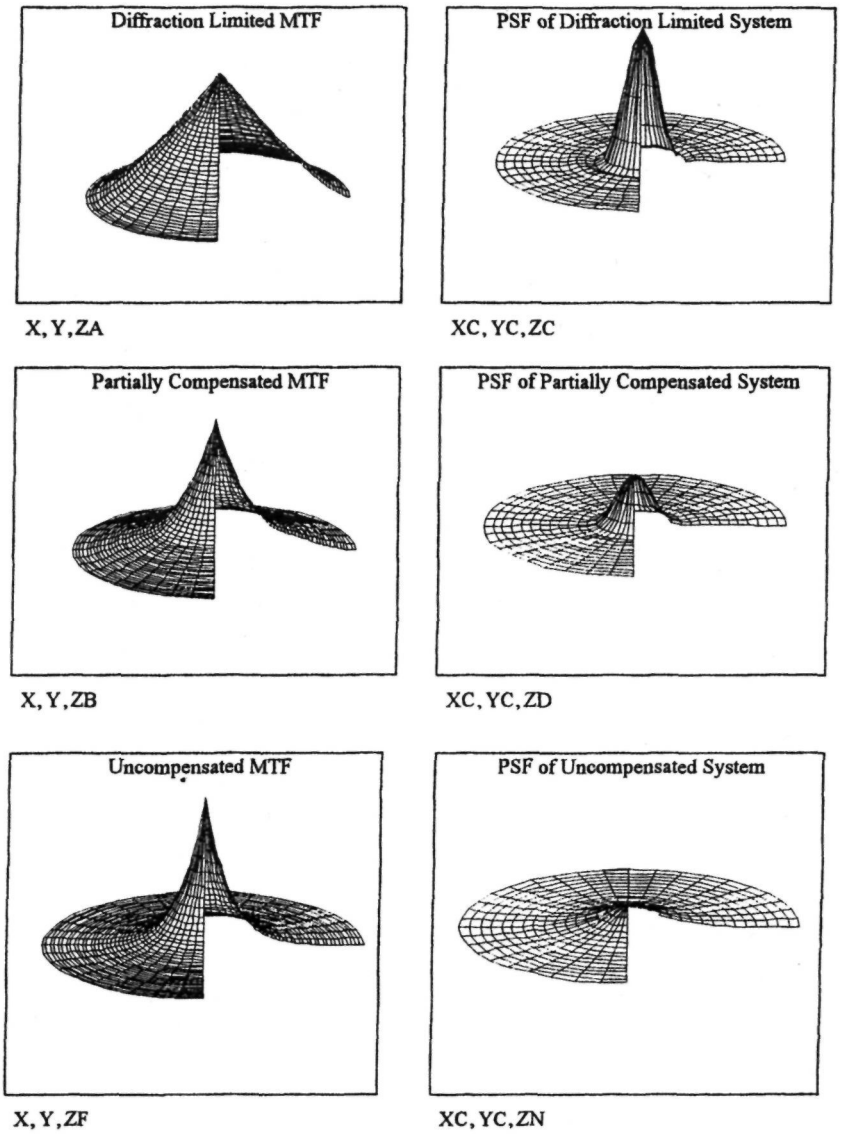


Figure 4.8 Three-dimensional plots of modulation transfer functions and image structure for diffraction-limited, partially compensated, and turbulence-degraded (uncompensated) images. The basic conditions are aperture 4 m, wavelength 1.62 μm, and turbulence parameter 0.15 m (at λ = 0.55 μm). The partially compensated image is obtained with only 6 degrees of freedom. The functions depicted are time-averaged with overall tilt removed.

4.4.4 Modeling Turbulence-Degraded Image Profiles

The effect of wavefront distortion on an optical image is to remove energy from the central peak, distributing it into a surrounding halo. This halo has a different origin from the diffraction ring structure produced by a hard-edged optical aperture and is superimposed upon it. The halo is absent for a perfect wavefront, but dominates the image structure in the presence of wavefront distortion typical of uncompensated turbulence. From equation (4.14), we know that in a diffraction-limited image, 84% of the light is contained within the first dark ring. When the distortion is small, the result is mainly to reduce the peak intensity, with little change in the basic image structure. As the wavefront error increases, more of this light is redistributed until,

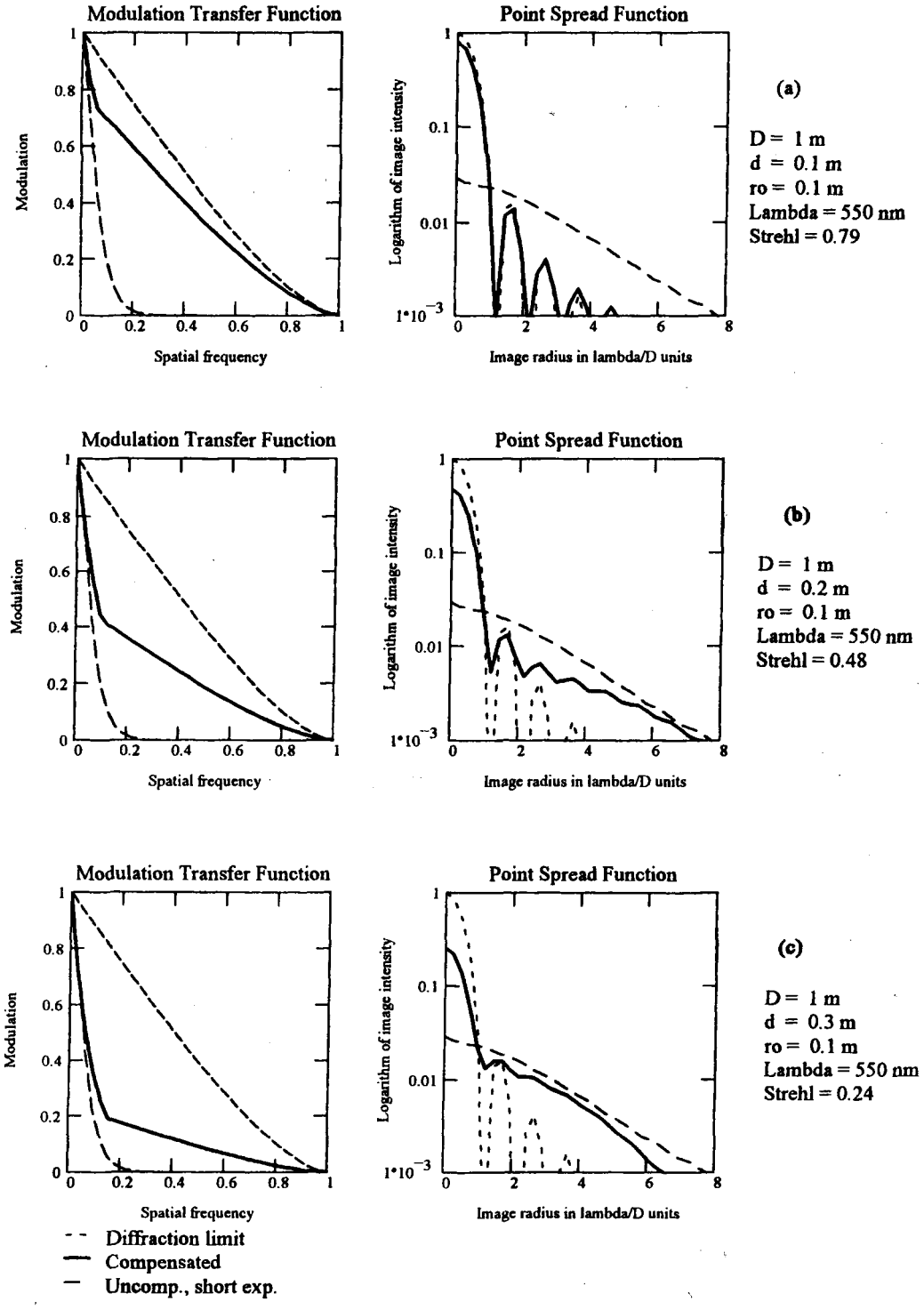


Figure **4.9** Modulation transfer functions and point spread functions for a compensated telescope at three values of d/r_0.

when the distortion reaches about 2 rad rms, the peak disappears. Where does this light go and what happens with larger wavefront errors?

An answer may be found by considering the structure of turbulence within an optical aperture. Such a wavefront has two key spatial parameters: the aperture diameter D and the coherence length of the turbulence r_0. No matter how strong the turbulence, we can consider the telescope aperture D to be filled with a random array of ro-sized subapertures, over each of which the wavefront error is 1 rad rms. The image structure must also be determined by D and r_0. The telescope aperture of diameter D produces an image core of full diameter $2.44\lambda/D$ at the first dark ring. The subapertures of size r_0 produce a halo that can be considered as a multiplicity of superimposed disks of diameter $2.44\lambda/r_0$. The value of r_0 is typically in the range of 0.1-0.2 m at a wavelength of 0.5μm. It grows as the 6/5 power of wavelength, with the result that r_0 may be as large as 1 m at infrared wavelengths.

To complete this model of the structure of a turbulence-degraded image, the relative intensities of the core and halo must be established. It was shown in the previous section that the normalized intensity of the core (the Strehl ratio, S) can be expressed statistically as $\exp(-\sigma^2)$, where σ^2 is the mean-square wavefront error. The Strehl ratio is also a good measure of the fraction of the total energy in the core, because its diameter changes very little with wavefront error. The fraction of energy in the halo is therefore $(1 - S)$. The expression for Strehl ratio is valid up to mean-square errors of about 4 rad^2, at which point the Strehl ratio has dropped to less than 0.02 and 98% of the energy is in the halo. This represents the upper limit of residual error for an optical image; beyond this point it is essentially uncompensated.

To determine the relative intensity of the halo, a useful approach has been suggested by Yura [1973]. The reduction in the on-axis irradiance of a turbulence-degraded image, compared with its value with no turbulence, can be expressed as the Factor F, where

$$F \approx \frac{1}{1 + (D/\rho_0)^2} \tag{4.52}$$

where

D = aperture diameter
ρ_0 = coherence length of the turbulence

The coherence length has different values for long and short exposures. For long exposures, ρ_0 is approximately equal to r_0. For short exposures, which eliminate the image motion caused by overall tilt, it is approximated by

$$\rho_0^{ST} \approx \rho_0 \left[1 + 0.37 \left(\frac{\rho_0}{D} \right)^{1/3} \right] \tag{4.53}$$

For the case of no turbulence, ρ_0 is equal to infinity and F becomes unity as expected.

The general model for a turbulence-degraded image (valid also for partially compensated images) consists of two components, a core and a halo, as depicted in figure 4.10. The dimensions for a short-exposure image, with no tilt error are:

Core diameter [full-width at half-maximum (FWHM)]

$$a_C^{SE} = 1.22\lambda/D \tag{4.54}$$

Core peak intensity

$$S_C^{SE} = e^{-\sigma_p^2} \tag{4.55}$$

Halo diameter (FWHM)

$$a_H^{SE} = 1.22(\lambda/D)[1 + (D/\rho_0)^2]^{1/2} \tag{4.56}$$

Halo peak intensity

$$S_H^{SE} = \frac{1 - e^{-\sigma_p^2}}{1 + (D/\rho_0)^2} \tag{4.57}$$

The total peak intensity is the sum of the core and the halo:

Total peak intensity

$$S_T = S_C + S_H$$

The total peak intensity S_T (which is always less than unity) may be regarded as a composite Strehl ratio, defining the normalized peak value of the core plus halo.

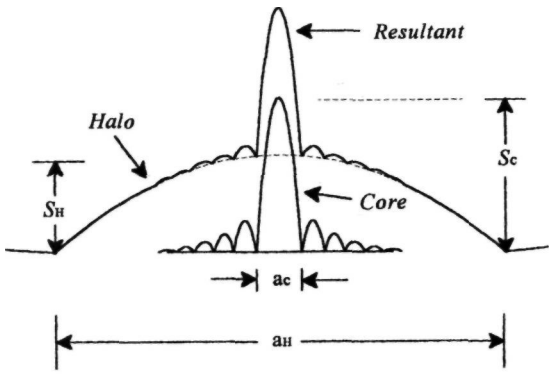


Figure 4.10 Model for partially compensated turbulence-degraded images. The image consists of two components: a diffraction-limited core, the diameter of which depends on the full telescope aperture, superimposed on a halo with a diameter dependent on the turbulence parameter r_0. As the degree of correction increases, a greater proportion of the light is concentrated in the central core.

4.5 Image Motion

4.5.1 Sources of Image Motion

The main sources of image motion in astronomical telescopes are atmospheric turbulence and wind buffeting, which can induce vibration of the telescope structure. Adaptive optics will ideally remove these disturbances, but residual errors remain because of imperfections in the compensation system. A major source of error is the tilt isoplanatism produced when the reference star is angularly separated from the science object.

Image motion is produced by the overall tilt component of atmospheric turbulence, generated by disturbances of size greater than the telescope aperture. There is some debate concerning the magnitude of the outer scale of atmospheric turbulence and whether, in fact, it exceeds the aperture of the largest telescopes now in use (10 m). Assuming an outer scale larger than the telescope aperture, the single-axis mean-square angular tilt, in units of radians squared, is

$$\sigma_\alpha^2 = 0.182\left(\frac{D}{r_0}\right)^{5/3}\left(\frac{\lambda}{D}\right)^2 \quad (4.58)$$

Because r_0 is proportional to $\lambda^{6/5}$, the angular tilt is, in fact, independent of wavelength and its mean-square value varies as $1/D^{1/3}$. The effect of wavefront tilt on the image structure depends on the tilt angle relative to the angular resolution $X\backslash D$, giving the relation

The image motion due to atmospheric turbulence, relative to the Airy disk, is therefore inversely proportional to wavelength. The reason is simply that the image size increases linearly with wavelength, while the angular motion is independent of wavelength.

4.5.2 Modeling Image Motion

The effect of a random tilt error of σ_α rad rms is to spread each point of the image into a Gaussian profile with standard deviation σ_α. This process reduces the angular resolution and diminishes the peak intensity. In some cases, image motion may be different in the x and y axes. When the difference is not large, the effective mean-square tilt error may be represented by $\sigma_\alpha^2 = \frac{1}{2}\left[\sigma_{\alpha_x}^2 + \sigma_{\alpha_y}^2\right]$. Computation of the resulting Strehl ratio is facilitated if the diffraction image profile is also modeled as a Gaussian curve. It was seen in section 4.2.4 that a Gaussian profile with standard deviation of $\sigma_A = 0.44\lambda/D$ has the same width at half-maximum as an Airy function.

The volume of a circular Gaussian profile of peak amplitude A_0 and standard deviation σ_A is equal to $2\pi A_0\sigma_A^2$. Tilt errors redistribute the energy into a Gaussian profile with variance $\sigma_A^2 + \sigma_\alpha^2$ but do not change the total energy in the image. The peak intensity is therefore reduced in the ratio

$$F_T = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_\alpha^2} = \frac{1}{1 + (\sigma_\alpha/\sigma_A)^2} \quad (4.60)$$

Substituting for $\sigma_A = 0.44\lambda/D$, the intensity reduction of the core due to tilt error is

$$F_C = \frac{1}{1 + (D/0.44\lambda)^2\sigma_\alpha^2} = \frac{1}{1 + 5.17(D/\lambda)^2\sigma_\alpha^2} \quad (4.61)$$

The diameter of the core is increased in the ratio $[(\sigma_A^2 + \sigma_\alpha^2)/\sigma_A^2]^{1/2} = F_C^{-1/2}$.

Similar calculations for the halo show that the peak intensity reduction ratio is

$$F_H = \frac{1}{1 + 5.17(r_0/\lambda)^2\sigma_\alpha^2} \quad (4.62)$$

The halo diameter is then increased in the ratio $F_H^{-1/2}$. When $D > r_0$, (which is usually the case in astronomical imaging, otherwise adaptive optics would not be needed), the effect of angular tilt errors σ_α is relatively much smaller on the halo than on the core, so F_H is close to unity and the dimensions in equations (4.56) and (4.57) may be used.

The core dimensions for a long-exposure image, including the tilt error, are therefore:

Core diameter (FWHM)

$$a_C = 1.22(\lambda/D)[1 + 5.17(D/\lambda)^2\sigma_\alpha^2]^{1/2} \quad (4.63)$$

Core peak intensity

$$S_C = \frac{e^{-\sigma_P^2}}{1 + 5.17(D/\lambda)^2\sigma_\alpha^2} \quad (4.64)$$

When $\sigma_\alpha = 0$, these expressions reduce to equations (4.54) and (4.55), so they may be used in the general case to replace these equations.

4.5.3 Criteria for Image Motion

Equations (4.61) through (4.64) enable the effects of random image motion to be related directly to the basic properties of the image of a point source, such as normalized peak intensity (Strehl ratio), core diameter, or peak contrast ratio of the core and the halo. In the case of images that are well-corrected except for image motion, the peak reduction factor F_C due to image motion can be cascaded with the Strehl ratios due to other errors in the system. The overall peak intensity S_C, given by equation (4.64), defines the relation between σ_α^2 and σ_P^2.

If the tilt is expressed as a fraction of the Airy disk radius by dividing by $1.22\lambda/D$, the relation between tilt and phase errors may be expressed in dimensionless form as

$$\sigma'_\alpha = \frac{\sigma_\alpha}{1.22\lambda/D} = 0.36\left[\frac{e^{-\sigma_P^2}}{S_C} - 1\right]^{1/2} \tag{4.65}$$

which defines the tradeoff between tilt and phase errors for a given value of core peak intensity (Strehl ratio) S_C. For example, to obtain a Strehl ratio of 0.8 from angular tilt only ($\sigma_P^2 = 0$), the rms angular tilt error must not exceed 0.18 times the Airy disk radius of the telescope at the observation wavelength.

In the case of partially compensated images where the structure contains both core and halo, the peak intensity is the sum of the core and halo intensities with random tilt errors:

$$\begin{aligned} S_T &= \frac{e^{-\sigma_w^2}}{1 + 5.17(D/\lambda)^2\sigma_\alpha^2} + \frac{1 - e^{-\sigma_w^2}}{1 + (D/r_0)^2} \\ &= S_w F_C + \frac{1 - S_w}{1 + (D/r_0)^2} \end{aligned} \tag{4.66}$$

where S_w is the Strehl ratio resulting from the total wavefront phase error σ_w^2 of the adaptive optics system (excluding tilt), and S_T is the final Strehl ratio including tilt.

The peak contrast ratio C_P between core and halo is also reduced by image motion. This subject is pursued further in the next section.

4.6 Image Evaluation

4.6.7 *Image Profiles*

The equations developed earlier, defining the peak intensities and diameters of the image core and halo, enable the basic image structure to be determined for any set of conditions, ranging from uncompensated turbulence to a perfect wavefront. The gross effects are illustrated in figures 4.11 and 4.12, which show image profiles for several degrees of compensation. The effect of changing the observing wavelength is shown in figure 4.11, with the telescope aperture fixed at 2 m and with r_0 (specified at 0.5 μm) constant at 0.15 m. The effective values of r_0 at the observing wavelengths are 0.22, 0.45, and 0.89 m at 0.7, 1.25, and 2.2 μm, respectively.

Figure 4.12 compares the normalized image profiles produced by apertures of 2, 4, and 8 m, at a fixed wavelength of 2.2 μm, with r_0 constant at 0.15 m at a wavelength of 0.5 μm.

These profiles clearly show the halo produced by uncompensated turbulence, as well as the effect of compensation in transferring energy from the halo to the core. In figure 4.12, each profile is normalized to unity for the specific aperture used. It is worth noting that if these profiles were compared on an absolute basis, corresponding to the light collected from a star of given magnitude, then the peak intensities of the 4-m and 8-m profiles would be higher by factors of 16 and 256, respectively, compared with the peak of the 2-m aperture.

4.6.2 *Evaluation of Partially Compensated Images*

The partly compensated image profiles shown in figures 4.11 and 4.12 have two distinct components of different radii. The effect of changing the degree of compensation is to change the intensity ratio of these components, but not their radii. When evaluating the imaging performance of an optical imaging system, the critical parameter is the size of the point spread function that convolves the source intensity distribution. It may be asked whether there is an equivalent image size for partial compensation. Alternatively, is there some other property of the image suitable as a criterion for evaluating partly compensated images?

A composite system resolution has been postulated by Parenti and Sasiela [1994], based on the weighted average of the relative intensities and diameters of the core and halo, as defined above:

$$A = \frac{\left[(a_C S_C)^2 + (a_H S_H)^2\right]^{1/2}}{S_C + S_H} \tag{4.67}$$

Parenti and Sasiela report that simulations of partially compensated short-exposure images have shown good agreement with the above expression. With a 4-m aperture and r_0 equal to 0.16 m, the threshold of wavefront distortion at which the resolution started to depart from the diffraction limit of the aperture was found to be about 2 rad rms.

As astronomical telescopes get larger, the ratio of the halo and core diameters also increases: for an 8-m telescope at a good site, it may be as high as 40 or 50. Such large differences between the halo and core diameters bring into question the meaning and validity of composite resolution as defined in equation (4.67). An alternative approach, suitable for large apertures, is to consider the core and the halo separable, on the basis of their greatly different intensities. The justification for this idea is that the photoelectric detectors now used in astronomy have a very large dynamic range, and when images are digitized it is relatively easy to select data within a certain range of intensities. Given an image structure similar to the partially compensated cases shown in figures 4.11 and 4.12, the ability to discriminate between the diffraction-limited core and the halo depends on their relative intensities.

For long exposures, the high-resolution information in the image is contained in the core, the normalized intensity of which is S_C. The total intensity is $S_C + S_H$. The peak image contrast ratio is defined as

$$C_P = \frac{S_C}{S_C + S_H} = \frac{S_C\left[1 + (D/r_0)^2\right]}{1 + S_C(D/r_0)^2} \tag{4.68}$$

where S_C and S_H are defined in equations (4.64) and (4.57), respectively. The contrast ratio has a value

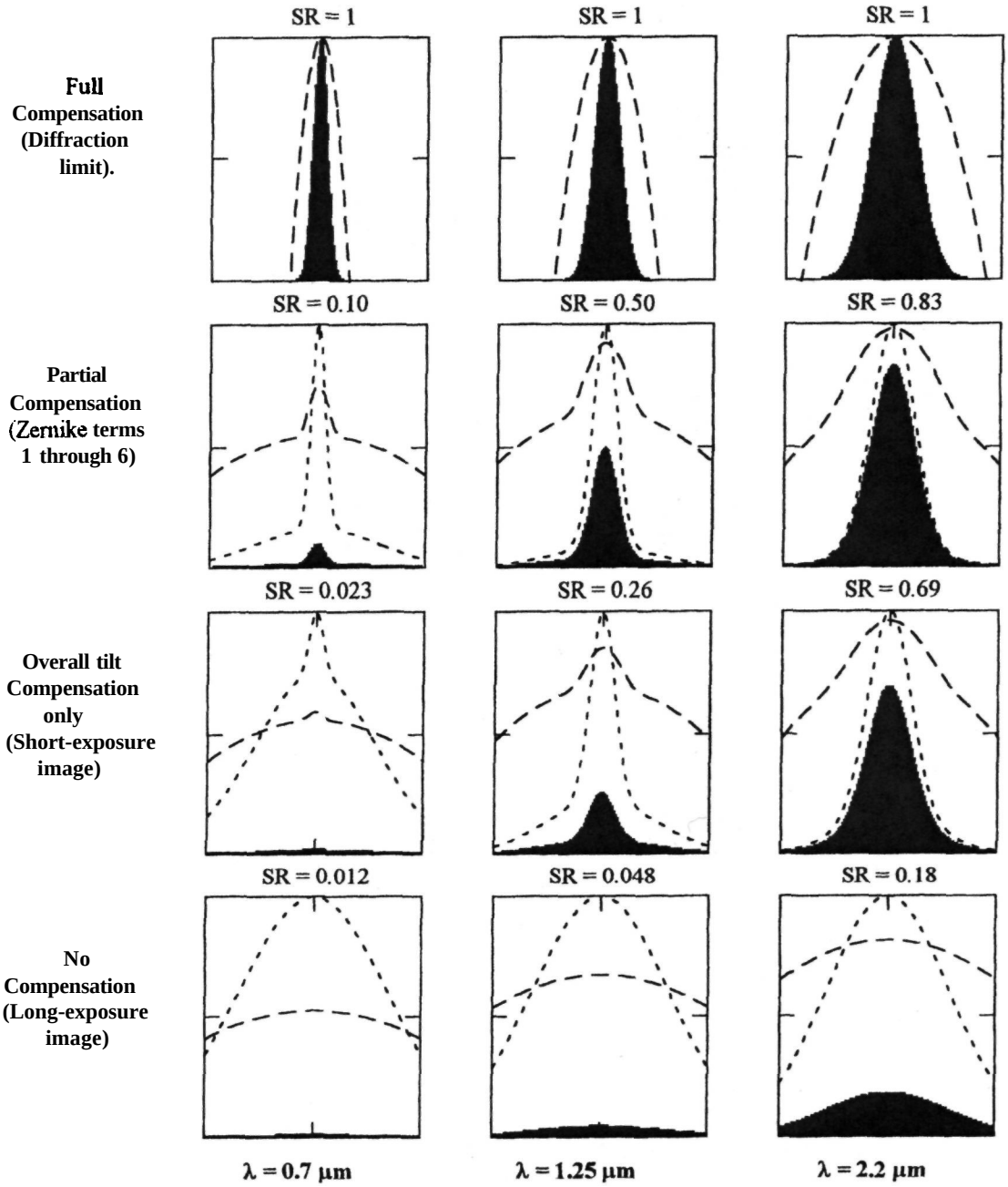


Figure 4.11 Image intensity profiles for various degrees of turbulence compensation, as a function of the observing wavelength. The peak image intensity (Strehl ratio) for each case is shown. The solid figure is the intensity profile with a linear intensity scale. The dotted line shows the shape of this profile, normalized to unity. The dashed curve is the intensity plotted on a logarithmic scale. The horizontal scale is ±0.5 arc seconds. Conditions: aperture = 2.0 m; r_0 = 0.15 m; λ = 0.7, 1.25, and 2.2 μm.

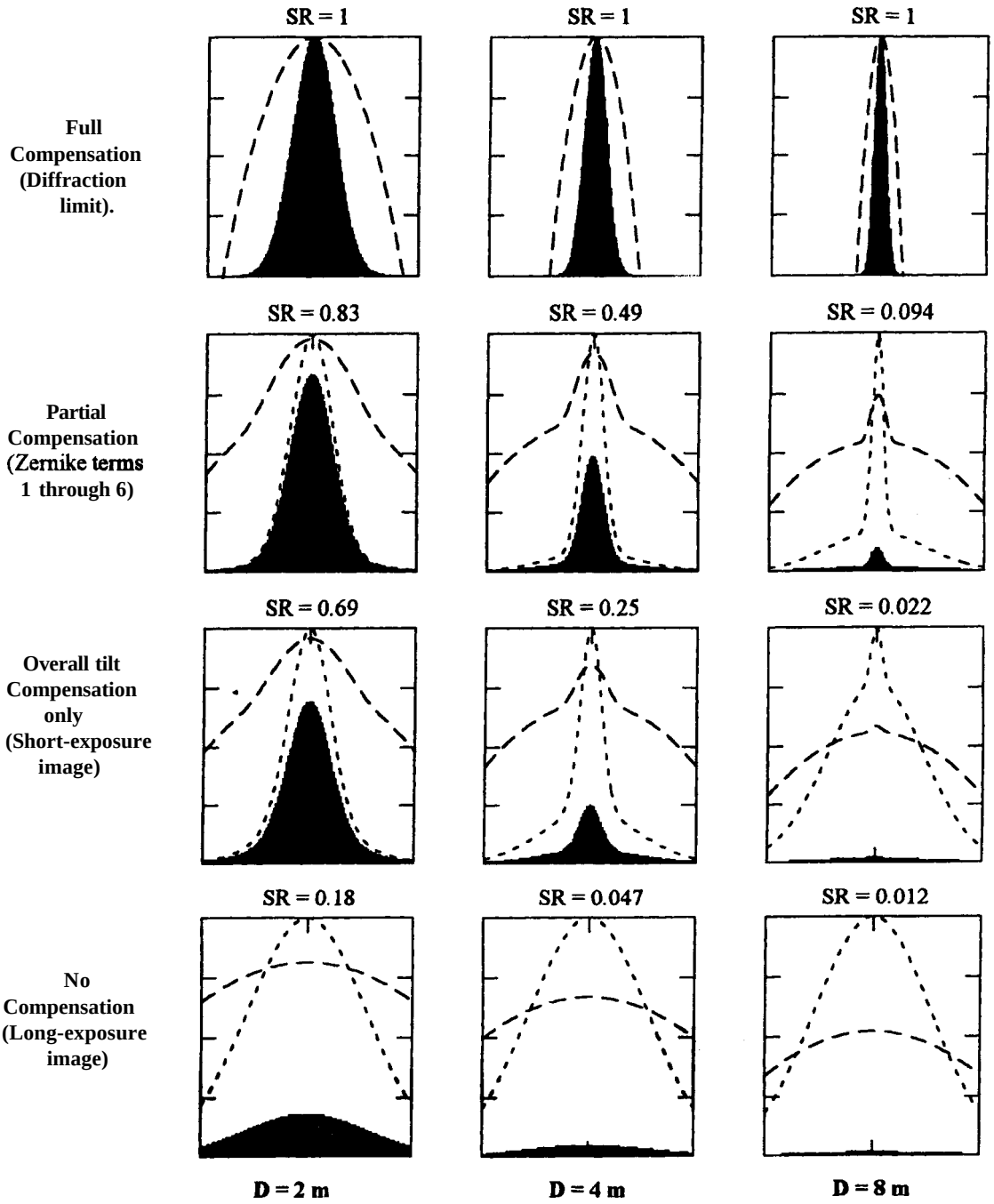


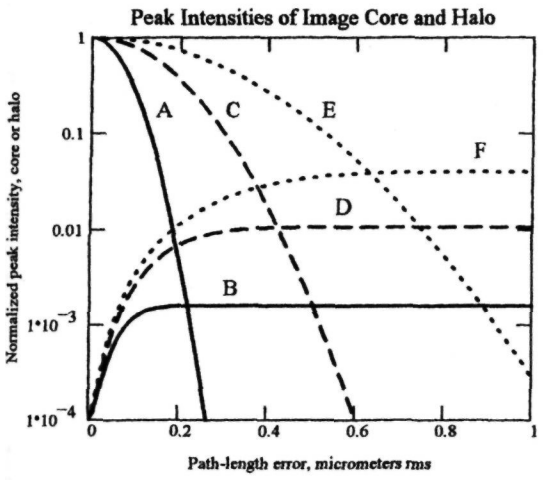
Figure 4.12 Image intensity profiles for various degrees of turbulence compensation, as a function of the aperture size. The solid figure on each chart is the intensity profile on a linear intensity scale, with the corresponding Strehl ratio. The dotted line shows each profile normalized to unity. The dashed curve shows the intensity profile on a four-decade log scale. Horizontal scale is ±0.5 arc seconds, r_0 = 0.15 m, λ = 2.2 urn.

between 0 and 1. When $(D/r_0) \gg 1$, it is closely approximated by

$$C_P = \frac{S_C}{S_C + (r_0/D)^2} \tag{4.69}$$

The variation of S_C and S_H with optical pathlength error is depicted in figure 4.13 for wavelengths of 0.55, 1.25, and 2.20 μm. Figure 4.14 compares the peak contrast ratio C_P with the composite Strehl ratio S_T for identical conditions. As the wavefront error increases, the contrast ratio initially remains near unity, while the Strehl ratio drops steadily. The contrast ratio remains high even when the Strehl ratio has fallen below 0.1. This region of low S_T and high C_P defines the regime *of partial compensation,* in which postdetection image-deconvolution techniques can often restore the full diffraction-limited quality to the image.

The peak contrast ratio C_P may be used as a measure of image quality for situations in which imaging sensors with a large dynamic range are employed. The ability to discriminate between different intensity levels ultimately depends on the signal-to-noise ratio of the image. As the contrast between core and halo decreases, a larger photon count is required to maintain the image signal-to-noise ratio, as discussed in section 4.7 below.



A core, 550 nm
B halo, 550 nm
C core, 1250 nm
D halo, 1250 nm
E core, 2200 nm
F halo, 2200 nm

Figure 4.13 Peak intensities of the image core and halo as a function of optical pathlength error for wavelengths of 550, 1250, and 2200 nm. The telescope aperture is 4 m and the turbulence parameter r_0 is 0.15m.

4.6.3 Effects of Image Motion

The effect of random image motion due to residual tilt errors on an otherwise well-corrected long-exposure image is to reduce the peak intensity of the core by the factor F_C and to increase the core diameter by the factor $(F_C)^{-1/2}$, where F_C is defined in equation (4.61). If the angular resolution of an imaging system is defined as the number of resolvable points per steradian of field angle, then it follows that the reduction in angular resolution due to tilt errors, compared with that of a diffraction-limited system, is given by the factor F_C.

The angular resolution is plotted against angular tilt error in figure 4.15 for a 4-m aperture at observing wavelengths of 0.7 and 2.2 μm. Also shown are the peak intensity curves for images with wavefront errors giving Strehl ratios of 0.5 and 0.2. (The angular resolution curve is the same as the peak intensity curve for an initial Strehl ratio of 1.)

The effect of image motion on the peak contrast ratio is found by multiplying the peak intensity of the core by F_C and multiplying the peak intensity of the halo by F_H, where F_C and F_H are defined in equations (4.61) and (4.62), respectively. The peak contrast ratio with image motion is then given by

$$C_P' = \frac{S_C F_C}{S_C F_C + \left[\frac{1 - S_C}{1 + (D/r_0)^2}\right] F_H} \tag{4.70}$$

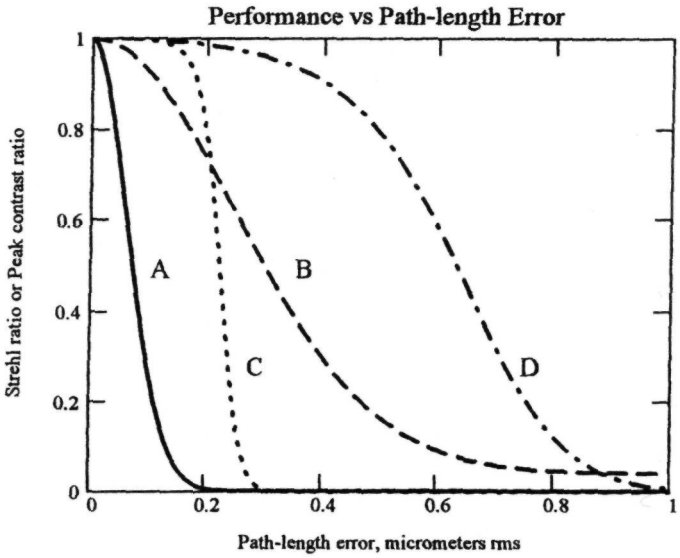
Values of the peak contrast ratio C_P' for wavefront Strehl ratios of 0.5 and 0.2 are also shown in figure 4.15. Overall tilt errors have a much smaller effect on the peak contrast ratio than on the angular resolution, which drops very sharply as tilt error is introduced, especially at short wavelengths.

4.7 Quantum Noise Effects

4.7.1 Photon Noise Limitations

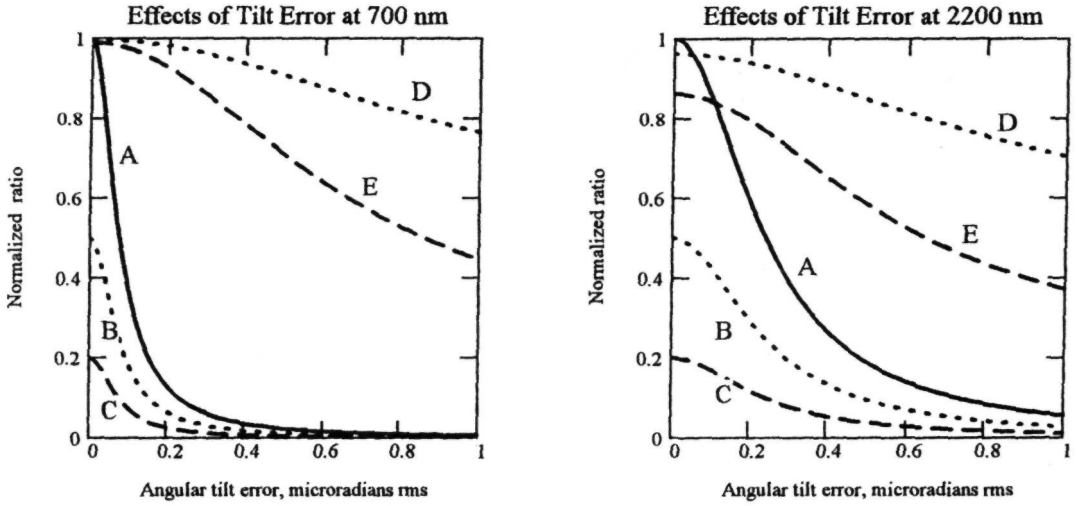
It has been implicitly assumed, up to now, that the image is formed by a large number of photons, so that quantum noise effects may be neglected. In practice, astronomical images are usually photon starved because of the faintness of the sources. The influence of photon noise on images is different from any of the wavefront effects discussed earlier. With the advent of adaptive optics using laser beacons, it is now possible for celestial objects at the limits of detectability to be well compensated.

The most basic problem is the detection of an object buried in background noise, typically because of radiation from the sky. Consider an array of $m + 1$ cells with a mean background photon count of N_b in each cell, as shown in figure 4.16. The object to be detected produces an additional mean count of N_s photons within a single cell. It is assumed that the



A composite Strehl ratio, 550 nm
B peak contrast ratio, 550 nm
C composite Strehl ratio, 2200 nm
D peak contrast ratio, 2200 nm

Figure **4.14** Comparison of peak contrast ratio C_P with composite Strehl ratio S_T at wavelengths of 550 and 2200 nm. The aperture is 4 m and r_0 is 0.15 m.



A angular resolution factor F_c
B peak intensity, initial Strehl = 0.5
C peak intensity, initial Strehl = 0.2
D peak contrast ratio, Strehl = 0.5
E peak contrast ratio, Strehl = 0.2

Figure **4.15** Effects of random tilt errors on the structure of compensated long-exposure point images for an aperture of 4 m with r_0 = 0.15 m and zenith angle 30°, at observation wavelengths of 700 and 2200 nm. These curves show the result of residual image motion on the angular resolution, peak intensity, and contrast ratio.

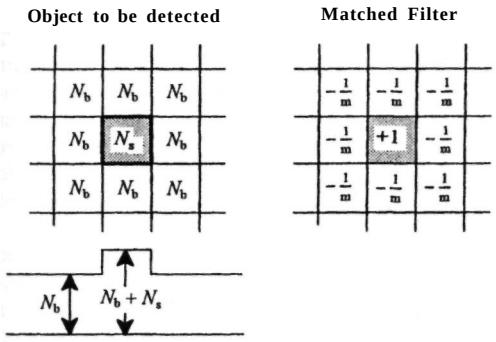


Figure 4.16 Detection of a source in a noisy background.

photon arrival times are random and independent so that Poisson statistics apply.

The signal s is the difference between the count in the cell containing the object and the average *of* the counts in the background cells. For a typical sample,

$$s = n_s + n_{b0} - \frac{1}{m}\sum_m n_{bm} \tag{4.71}$$

where

n_s = additional count from the object to be detected
n_{b0} = background count in the object cell
n_{bm} = background count in cell m

The mean value of the signal is

$$\begin{aligned} \bar{s} &= N_s + N_b - \frac{1}{m}(m N_b) \\ &= N_s \end{aligned} \tag{4.72}$$

Since all of the terms in equation (4.71) are independent, the variance of the signal is equal to the sum of the variances with the weights squared:

$$\begin{aligned} \sigma_s^2 &= \sigma_{ns}^2 + \sigma_{nb}^2 + \frac{1}{m^2}\sum_m \sigma_{nb}^2 \\ &= \sigma_{ns}^2 + \left(1 + \frac{1}{m}\right)\sigma_{nb}^2 \end{aligned} \tag{4.73}$$

For Poisson variables, the variances are equal to the mean values, so the variance of the signal is given by

$$\sigma_s^2 = N_s + \left(\frac{m+1}{m}\right)N_b \tag{4.74}$$

The signal-to-rms-noise ratio is then

$$SNR = \frac{\bar{s}}{\sigma_s} = \frac{N_s}{(N_s + k N_b)^{1/2}} \tag{4.75}$$

where $k = (m + 1)/m$. When m is large, the average background count can be estimated accurately and the value of k approaches unity. In this case, the signal-to-noise ratio is

$$SNR \approx \frac{N_s}{(N_s + N_b)^{1/2}} \tag{4.76}$$

For faint objects, the photon arrival rates are low, so the signal and background consist of discrete events with random arrival times. The exposure time necessary to achieve a given signal-to-noise ratio depends on the average photon rates of the signal and background. If the average signal and background rates are a and b detected photons per second, respectively, then the signal-to-noise ratio after t seconds is

$$SNR = \frac{at}{(at + bt)^{1/2}} = \frac{at^{1/2}}{(a + b)^{1/2}} \tag{4.77}$$

In principle, any source emitting photons can be detected, irrespective of the background noise, simply by accumulating the counts over a sufficiently long period of time. In practice, exposure times for ground-based astronomy are generally limited by saturation of the photon detector. This problem can be overcome by combining several separate exposures.

4.7.2 Contrast Ratios Required for Detection and Recognition

The preceding analysis can be broadened to include the detection of detail in extended objects. Work by Rose [1948] and Coltman and Anderson [1960] established that the capability of a human observer approximates that of an ideal matched filter for detecting visual signals buried in noise. They found that a minimum signal-to-noise ratio of about 5:1 is necessary to recognize simple shapes in a noisy environment. This information enables determination of the density of detected photons in the image plane required to detect or recognize simple objects. The required photon density depends on the contrast between the object and the background, which, in turn, depends on the degree of compensation. Consequently, there is a tradeoff between the photon density required to recognize an object and the amount of compensation applied to the image. The better the compensation, the smaller the photon density (or exposure time) required.

The contrast relevant to the detection of detail in an extended object consists of several components:

1. the contrast of the source, C_s;
2. the contrast reduction due to background radiation from the sky;
3. the contrast reduction due to imperfect compensation of atmospheric turbulence, C_p.

The simplest type of detail to be detected in an extended object may be modeled as a single spot on a pedestal, using two brightness levels. If the mean photon count of the spot is N_s, and the mean photon

count of the pedestal is N_p, then the contrast of the source is

$$C_s = \frac{\max - \min}{\max + \min} = \frac{N_s - N_p}{N_s + N_p} \tag{4.78}$$

The sky background adds photons to all detector cells. If the mean background count is N_b, then the contrast of the signal reaching the telescope is

$$C_t = \frac{N_s - N_p}{N_s + N_p + 2N_b} \tag{4.79}$$

There is a further contrast reduction in the telescope if the compensation is not perfect. This is due to the halo surrounding the image of a point source, as discussed in section 4.6.2. For a single point source, the contrast reduction is given by equation (4.69):

$$C_p = \frac{S_C}{S_C + (r_0/D)^2} \tag{4.80}$$

The overall image contrast is then C = C_t C_p, which can be written in the form

$$C = \frac{C_p N_1}{N_1 + N_2} \tag{4.81}$$

where

$C_p N_1$ = effective photon count of the signal to be detected (the spot)
$N_1 + N_2$ – total photon count of the background image (the pedestal)

From equation (4.76), the image signal-to-noise ratio can be expressed as

$$SNR = \frac{C_p N_1}{(N_1 + N_2)^{1/2}} \tag{4.82}$$

Using equation (4.81), the photon count in the extended image required to achieve a given *SNR* with overall contrast *C* is then

$$\begin{aligned} N_1 + N_2 &= \left(\frac{SNR}{C}\right)^2 \\ &= \left(\frac{SNR}{C_t^2 C_p^2}\right)^2 \end{aligned} \tag{4.83}$$

Equation (4.83) defines the tradeoff between contrast and photon count to achieve a specified signal-to-noise ratio for the detection of a single spot against a bright background. A minimum signal-to-noise ratio of 5:1 is required to discriminate a small spot from the background. For a contrast ratio of 0.5, the average background photon count per cell required to achieve this signal-to-noise ratio is 100. If the contrast drops to 0.2, the required photon count rises to 625.

A similar process can be used to determine the probable shape or orientation of an object. Figure 4.17 shows a matched filter for discriminating a short horizontal bar from a vertical one. Photons detected in the horizontal arms are weighted +1/2 and photons detected in the vertical arms are weighted —1/2. After a number of counts have accumulated, a positive output signifies a horizontal bar and a negative output signifies a vertical bar. If there is no background noise, then the signal-to-noise ratio is simply the square root of the number of detected photons.

In general, photons will be detected in all cells and the task is to determine whether the object is horizontal or vertical. If the mean number of counts in a detector cell due to the presence of the object is N_s and the mean number of background counts in every cell is *Nt*, the mean signal with a horizontal bar will be

$$\begin{aligned} \bar{s} &= \tfrac{1}{2}(2N_s + 2N_b) - \tfrac{1}{2}(2N_b) \\ &= N_s \end{aligned} \tag{4.84}$$

As above, the variance of the signal is the square weighted sum of the variances of the counts:

$$\begin{aligned} \sigma_s^2 &= \left(\tfrac{1}{2}\right)^2 \left(2\sigma_{N_s}^2 + 2\sigma_{N_b}^2\right) + \left(\tfrac{1}{2}\right)^2 \left(2\sigma_{N_b}^2\right) \\ &= \frac{1}{2} N_s + N_b \end{aligned} \tag{4.85}$$

The signal-to-noise ratio is then

$$\begin{aligned} SNR &= \frac{N_s}{\left(\frac{1}{2} N_s + N_b\right)^{1/2}} \\ &= C \frac{N_s + N_b}{\left(\frac{1}{2} N_s + N_b\right)^{1/2}} \end{aligned} \tag{4.86}$$

For low contrast objects, $N_b \gg N_s$ and the mean number of background photon counts required to achieve a given signal-to-noise ratio is

$$N_b = \left(\frac{SNR}{C}\right)^2 \tag{4.87}$$

This expression specifies the average number of detected background photons required per image element to discriminate between two simple objects having a contrast ratio of C. The required photon density, expressed as the average number of detected background events per unit solid angle, is

$$I_p = \frac{N_b}{\alpha^2} \tag{4.88}$$

where α is the angular size of the image elements.

In the context of image compensation, equation (4.87) can be used to determine the improvement, due to adaptive optics, in the detectability of photon-starved images. Image compensation concentrates more light into the core of each image element, improving the contrast *C*. This provides a better signal-to-noise ratio, or enables smaller detail to be discriminated. The approach is easily extended to more complex discrimination tasks by modeling the appropriate matched filters.

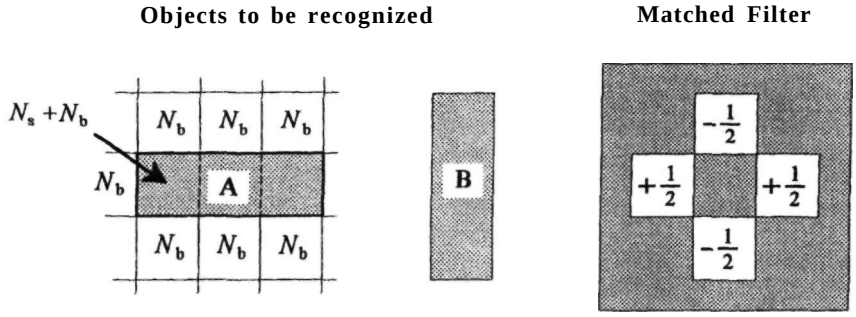


Figure 4.17 Recognition of an object in a noisy background.

4.7.3 Photon Noise in Partially Compensated Images

The partially compensated image of a point source is characterized by a bright halo that limits the contrast between the core and its surrounding area. The relative intensities of the core and halo are determined by the degree of compensation in the adaptive optics system. Large residual wavefront errors reduce the contrast between core and halo, necessitating an increased photon count to maintain the signal-to-noise ratio. For each level of wavefront compensation (defined by the Strehl ratio), there is a minimum image photon count that maintains the SNR above the threshold required to detect faint sources.

If I_0 is the photon count in the diffraction-limited core of the image of a point source, then the number of photons in the partially compensated core is $I_0 S_C$ and the number of photons in core-sized areas of the halo is $I_0 S_H$, where S_C and S_H are the intensities of the core and halo, as defined in equations (4.55) and (4.57), respectively. The signal-to-noise ratio between core and halo is

$$SNR = \frac{\sqrt{I_0} S_C}{\sqrt{S_C + S_H}} \tag{4.89}$$

Using equation (4.57),

$$SNR^2 = \frac{I_0 S_C^2 [1 + (D/r_0)^2]}{1 + S_C (D/r_0)^2} \tag{4.90}$$

If $(D/r_0)^2 \gg 1$, which is usually the case in astronomical telescopes, then the expression simplifies to

$$SNR^2 = \frac{I_0 S_C^2}{S_C + (r_0/D)^2} \tag{4.91}$$

$$S_C + (r_0/D)^2$$

Finally, if $S_C \gg (r_0/D)^2$, which is true except for very small values of S_C, then

$$I_0 \approx \frac{SNR^2}{S_C} \tag{4.92}$$

The number of photons in the image needed to maintain a given signal-to-noise ratio for partial compensation is therefore inversely proportional to the Strehl ratio S_C. For example, if $SNR = 6$ and $S_C = 0.1$, the minimum photon count required from a single point source is 360.

4.7.4 Detection of Other Planetary Systems

A subject of increasing interest in astronomy is the detection of planetary systems in nearby stars [Burke 1992, Angel 1994]. The main problem in planetary detection is the great difference in brightness between the two objects. Using the Sun-Jupiter system as a model, the light received from the planet is a factor of 10^9 less than that from the star. For a planetary system at a distance of 10 parsecs (pc), the angular separation is 0.5 arc seconds. This situation is depicted in figure 4.18, for the case of a 4-m telescope at a wavelength of 0.7μm. It is seen that the objects are easily separated angularly by a diffraction-limited telescope of this size, but the Airy diffraction rings of the star are a factor of about 2 x 10^4 brighter than the planet. The flux received from the planet is about 2 photons per second, while the background light from the rings is 4 x 10^4 photons per second. Even with a diffraction-limited optical system, exceedingly long integration times would be required to achieve a useful contrast between the planet and the background. In addition to the presence of diffracted light, small irregularities in the optical components scatter light from the bright parent star.

The classic solution to this problem is the Lyot coronagraph, [Lyot 1930, 1931], which employs an occulting mask at the image plane to suppress the light from the central star, together with a Lyot field stop, smaller than the telescope pupil, to reduce diffracted light. The use of a coronagraph combined with adaptive optics has been proposed by Malbet et al. [1994] for planetary detection using the Hubble Space Telescope (HST). In this case, the function of the adaptive optics is to correct for small figure errors

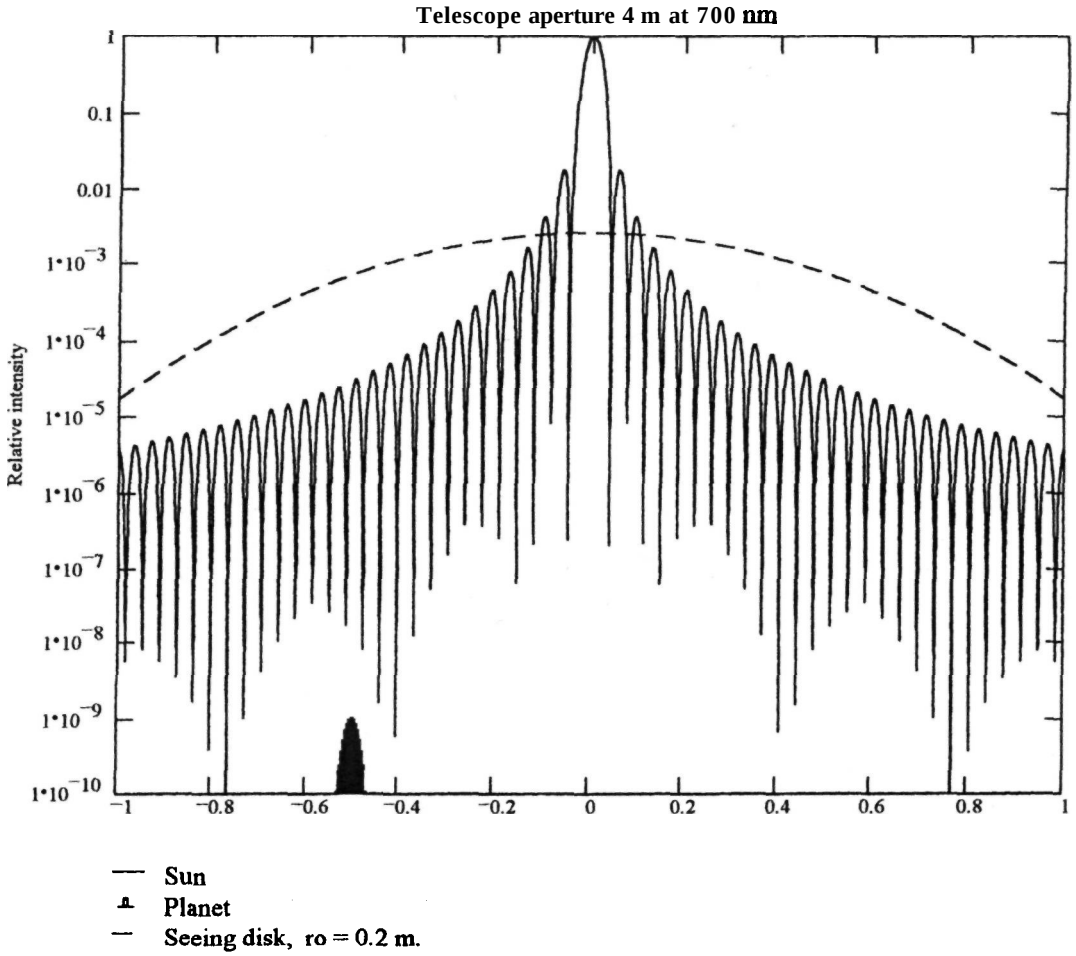


Figure 4.18 Detecting the planets of other suns. The chart shows the relative intensity distributions of a planetary system similar to that of the Sun and Jupiter, as seen at a distance of 10 pc by a diffraction-limited telescope of 4 m aperture at a wavelength of $0.7\,\mu$m.

on the primary mirror that would otherwise produce a background light level that is 10^4–10^5 times brighter than the planet. Simulations of the coronagraph using a Lyot stop with a transmission of 21% show that the background light is reduced by a factor of about 10^4. This enables the HST to achieve the signal-to-noise ratio of 5 that is required for reliable detection of a planet, in a reasonable integration time of about 1 h.

For ground-based astronomical telescopes, the detection of extrasolar planets is made more difficult by two factors: atmospheric turbulence and sky radiation. Planetary detection is probably the most difficult task for adaptive optics, because it involves a much higher degree of correction than is necessary just to improve the angular resolution: it is necessary also to reduce the light, scattered by turbulence, from the parent star. Diffracted light can be suppressed using coronagraphic techniques [Nakajima 1994], but partially corrected atmospheric turbulence acts like a random phase grating in front of the telescope to scatter light into a halo surrounding the central core of the image. In principle, scattered light can be suppressed if the wavefront corrector is conjugated to the source of the turbulence, but this strategy is difficult to implement in practice.

Angel [1994] has pointed out that the reduction of the stellar halo requires attention to both the spatial and the temporal aspects of adaptive optics. In addition to reducing the gross phase errors that control the peak intensity and shape of the image core, it is necessary to measure and correct the fine-grained phase errors that control the structure of the halo. This requires a large number of actuators and a fast control loop to ensure that the halo averages to a small value.

To make a planet detectable within a reasonable observation time, scattered light must be eliminated within a radius of about 1 arc second from the parent star. The spatial frequency of residual phase variations must therefore be high enough to diffract light at an angle greater than 1 arc second, or about 5 μrad. The diffraction angle produced by a periodic structure of spatial frequency f cycles per meter at a wavelength of A meters is $\alpha = \lambda f$ radians. For an observation wavelength of 0.7 μm and angle of 5/xrad, the minimum frequency required is 5/0.7 = 7 cycles per meter. Assuming two actuators per cycle, the actuator density at the telescope pupil must be at least 14 per meter. A 4-m telescope would therefore need 56 actuators across the diameter, for a total of about 2300, while an 8-m telescope would need over 9000 actuators.

One way of implementing such a dense wavefront corrector is to split it into two or more stages, as suggested by Sandler, Stahl and Angel [1995]. At good observing sites, the value of r_0 is about 0.15 m at 0.5 μm, equivalent to 0.22 m at 0.7 μm. The overall tilt and wavefront errors due to turbulence are compensated in the usual way, using a deformable mirror with actuator spacing of about 1.5 times the turbulence coherence length r_0 giving an actuator spacing of about 0.3 m. This results in a halo extending outward from an angular radius of about 1 μrad. The halo can be pushed out to the required angle of 5/xrad by using an additional high-spatial-frequency corrector, which requires an actuator spacing of 0.07 m, but only a small stroke. The wavefront must be measured at this spatial resolution at high precision, necessitating a relatively large photon flux. Bright reference sources will therefore be required to implement systems of this type. A laser beacon can be used if the parent star is too dim. The diffraction rings produced by the telescope pupil are suppressed in the usual way by means of a central mask and Lyot stop, following the adaptive optics correction.

4.8 Performance Criteria

4.8.1 Performance Specification for Adaptive Optics

The quality of conventional optical systems is usually specified as a maximum wavefront error for a given set of input conditions. The design process consists of modeling the errors generated by each component and then making the necessary trade-offs and compromises to achieve the required overall performance. In the case of astronomical telescopes, optical performance is often specified in terms of "power in the bucket"; that is, a specific fraction of the incident photons from an infinitely distant point source must fall within a given image radius, over the required angular field of view. This is a static specification, which can be verified by laboratory tests under controlled conditions.

In optical imaging systems, the shape (intensity distribution) of the image can be controlled, to some extent, by using apodization, in which an attenuation function is inserted in the optical pupil [Jacquinot and Roizen-Dossier 1964]. Apodization can improve certain properties of an image at the expense of others that may be of less importance for a particular application. For example, it can be used to reduce the diffraction rings surrounding the Airy disk, but in doing so it enlarges the diameter of the core. Apodization is only effective with real aperture functions and consequently always involves a loss of energy.

Performance specifications for adaptive optics systems have the added complication that they must include both the spatial and temporal effects of turbulence. A long-exposure image is degraded by overall wavefront tilt (image motion) as well as high-order distortion (figure errors). As these effects are usually sensed and corrected by separate components, it is necessary to determine specific values of the wavefront and tilt errors that will satisfy the performance objectives of the system, as well as the conditions under which it must be achieved.

How are the allowable wavefront and tilt errors determined? What performance criteria should be used? To answer these questions, it is necessary to review the basic requirements for scientific imaging tasks in astronomy. The Strehl ratio or the mean-square wavefront error on which it depends are commonly used as merit factors when evaluating the performance of imaging systems. There is no question that the Strehl ratio is the appropriate performance measure for power transfer in energy projection systems, where the concern is to maximize the peak energy on a target. It is also relevant in spectroscopy, where the maximum energy must be passed through a slit. However, there are several reasons for looking beyond the Strehl ratio as the primary merit factor for astronomical adaptive optics.

The first reason is that the Strehl ratio is not a linear measure of the usefulness or information content of an image. Although most, if not all, measures of imaging performance are maximized when the Strehl ratio is unity, a ratio of 0.5 does not mean that image quality has dropped to 50%. The analysis in section 4.4 shows that as the wavefront distortion increases, the diameter of the image core remains close to its diffraction-limited size, until the Strehl ratio has dropped to less than 0.1. The spatial resolution remains high until a critical wavefront error threshold is reached. The main effect of small random wavefront errors is to reduce the image contrast, not the resolution.

Another reason for not using the Strehl ratio as a sole criterion is that in astronomical imaging, the exposure time is not highly constrained; a sufficient number of photons may be collected to offset a

reduction in contrast caused by wavefront error. The availability of CCD detector arrays with a large dynamic range enables a large amount of information to be extracted from low-contrast images. Exposure times for ground-based telescopes are, of course, limited by sky brightness, but, within certain limits, the integration time is a useful parameter that can be traded off against wavefront error.

Each basic imaging task has an allowable threshold of wavefront error, which can be used to advantage in producing an economic design for the adaptive optics system, using partial rather than full wavefront compensation where appropriate. A hierarchy of basic tasks is considered in the next section.

4.8.2 Imaging Tasks

There are many basic imaging tasks in astronomy, each with its own special requirements and criteria. Some of these are listed below, in roughly increasing order of difficulty:

1. Detecting the presence of a point source.
2. Determining the position of a point source.
3. Detecting the presence of multiple sources.
4. Measuring the separation of two or more point sources.
5. Determining the centroid of a disk.
6. Measuring the size of a disk.
7. Determining the structure of low-contrast detail with a dark background (nebula or halo).
8. Determining the structure of low contrast detail on an extended bright object (solar granulation or planetary surfaces).
9. Detecting a dim object very close to a much brighter one (planet and star).

These tasks are concerned only with imaging and not with other tasks, such as spectroscopy, nor with operational issues concerning the adaptive optics, such as obtaining a suitable reference source.

These basic tasks fall into two categories: detection and measurement. In the first case, we are interested in the probability of detection, and in the second case, the precision of the measurement. In both cases, the critical factor is the signal-to-noise ratio of the image, which depends on many factors, such as the brightness, contrast, and spatial frequency content of the object; the sky brightness; the transfer functions of the atmosphere and telescope; and the characteristics of the image detector.

4.8.3 Measures of Imaging Performance

The performance of adaptive optics systems may be evaluated using formal criteria that have been proposed for optimizing the performance of optical imaging systems. Images of extended objects must be evaluated on a statistical basis, and Linfoot [1956] suggested the following quality factors:

(a) Relative structural content or sharpness of the image, compared with the object:

$$T = \frac{\int\int I^2\,dx\,dy}{\int\int O^2\,dx\,dy} \tag{4.93}$$

where $O(x,y)$ and $I(x,y)$ are the intensity distributions of the object and image, respectively. This figure of merit assesses the information content of an image, and is related to the equivalent bandwidth in a communication channel. If the image contains a similar amount of detail to the object, then the value of T will be near unity. However, the comparison is made statistically, so that T is insensitive to morphological distortions of the image; that is, the image does not have to look like the object. For detecting the presence of objects, this may not be important, but T appears to be an incomplete criterion for most astronomical imaging tasks.

(b) Correlation quality:

$$Q = \frac{\int\int O\,I\,dx\,dy}{\int\int O^2\,dx\,dy} \tag{4.94}$$

This factor is sensitive to the alignment or matching of details in the image and the object, penalizing small distortions in the image more than it does unsharpness. Therefore, it represents a completely different aspect of image quality from the factor T. A blurred image having good geometrical fidelity would give a high value of Q and a low value of T. Correlation quality appears to be relevant to astronomical imaging, but one objective of adaptive optics is to improve sharpness, to which this measure is somewhat insensitive.

(c) Image fidelity, which depends on the mean-square difference between image and object:

$$F = 1 - \frac{\int\int (O - I)^2\,dx\,dy}{\int\int O^2\,dx\,dy} \tag{4.95}$$

This factor seems to be nearer the mark, as it appears to measure, in the image plane, a quantity similar to that minimized by an adaptive optics system in the pupil, the mean-square wavefront error. In fact, F contains elements of both T and Q, as can be seen from the expansion of the term $(O - I)^2$, which produces both I^2 and OI terms. These three quality factors are related by the expression $F = 2Q - T$. This suggests that image fidelity F may be the most general criterion for image evaluation.

It is often more convenient to evaluate the performance of optical systems in frequency space rather than in image space. O'Neill [1963] showed that for the general case of an object composed of white Gaussian noise, these quality factors, nor-

malized to those of an ideal instrument, can be expressed in the form

$$t = \frac{T}{T_0} = \frac{\int\int |\tau(\omega)|^2 \, d\omega}{\int\int |\tau_0(\omega)|^2 \, d\omega} \tag{4.96}$$

$$q = \frac{Q}{Q_0} = \frac{\int\int \tau(\omega) \, d\omega}{\int\int \tau_0(\omega) \, d\omega} \tag{4.97}$$

$$f = \frac{F}{F_0} = \frac{q - \eta t}{1 - \eta} \tag{4.98}$$

where $\eta = T_0/2Q_0$. The quality factors t, q, and f are easily computed from a knowledge of the (degraded) system transfer function $\tau(\omega)$ and the ideal system transfer function $\tau_0(\omega)$, expressions for which are developed in section 4.4.

It can be seen that the factor q is the same as the Strehl ratio S, which measures the peak intensity of the image of a point source normalized to that produced by a perfect optical system. In the presence of a small amount of random wavefront distortion, the Strehl ratio may be approximated by $S = \exp(-\sigma^2)$. Maximizing the Strehl ratio is equivalent to minimizing the mean-square wavefront error σ^2.

It would be of interest to relate the factor f to specific types of image degradation, using models of the transfer functions and structure functions, in order to find the most important characteristics, which could then be used as optimization criterion. It is also of interest to include the effects of the signal-to-noise ratio on these criteria.

(d) A possible measure of image quality for partially compensated images is the peak image contrast ratio Cp, discussed in section 4.6.2, which is the ratio of the intensity of the central peak to that of the surrounding halo. In general, the image produced by an optical system with residual wavefront errors, when $D/r_0 \gg$ I, consists of a central diffraction-limited spike of radius λ/D, surrounded by a halo or pedestal of radius λ/r_0 that is produced by uncompensated turbulence. The value of C_P depends not only on the residual wavefront error, but also depends directly on the telescope aperture, wavelength, and turbulence parameter r_0. It is therefore a more comprehensive performance measure than the Strehl ratio, in which these factors only appear indirectly through their effect on the residual wavefront error.

4.8.4 Criteria for Image Quality

An appropriate quality factor can be assigned to optimize each of the basic imaging tasks listed in the preceding text. For tasks involving the detection of isolated detail against a dark background, such as tasks (1), (2), and (7), the signal-to-noise ratio is the governing factor, for which the peak image contrast ratio Cp is an appropriate measure. As noted above, Cp has useful values for partially compensated images when the Strehl ratio is low.

For the remaining tasks involving multiple or extended objects, in which there is considerable overlapping of adjacent point spread functions, the Strehl ratio, or factor Q, is the relevant criterion. The Strehl ratio of the image determines the minimum intensity modulation that can be discerned in the object. Maximizing the Strehl ratio optimizes the intensity ratio between the central disk and the surrounding diffraction rings or halo. This is the condition required for optimal imaging of extended objects containing low contrast detail, or for separating closely spaced point sources.

4.8.5 Information Content of a Partially Compensated Image

Continuing the more formal investigation of the quality of optical images, an expression is now derived for the information content of a partially compensated image, in the hope that it may shed some light on image-quality criteria, as well as suggest some parametric tradeoffs that may be of use in optimizing an adaptive optics system.

Consider an optical imaging system with pupil diameter D at wavelength λ, producing an array of resolution elements in the image plane. If the number of resolution cells is M, then the information content of a typical image can be expressed as

$$H = M \log_2(SNR) \quad \text{bits per image} \tag{4.99}$$

where SNR is the average signal-to-noise ratio per element. The quantity H may be used as a general image-quality criterion.

Each resolution element is the core of the image of a point source. When the wavefront errors are smaller than about 2 rad rms, the core diameter is essentially constant. In this case, the angular diameter of each resolution element is λ/D and the total number of elements per steradian field of view is $M = (D/\lambda)^2$. The signal-to-noise ratio depends on the number of photons detected per element n_p, which is given by

$$n_p = w D^2 t \eta S \frac{1}{M} = w t \eta S \lambda^2 \tag{4.100}$$

where

- w = irradiance at telescope aperture, photons $m^{-2}s^{-1}$
- t = exposure time, s
- η = photon detection efficiency of telescope and detector
- S = Strehl ratio

Note that n_p does not depend on D, because the number of elements is proportional to D^2, so the photon flux per element is independent of the telescope diameter.

The signal-to-noise ratio is then

$$SNR = \sqrt{n_p} = \sqrt{w\,t\,\eta\,S\,\lambda^2} \tag{4.101}$$

For small wavefront errors σ, the Strehl ratio may be expressed as

$$S \approx e^{-\sigma^2}$$

The image information content per unit angle is then

$$\begin{aligned} H &= \left(\frac{D}{\lambda}\right)^2 \log_2\left(w\,t\,\eta\,\lambda^2\,e^{-\sigma^2}\right)^{1/2} \\ &= \frac{1}{2}\left(\frac{D}{\lambda}\right)^2 \left[\log_2\left(w\,t\,\eta\,\lambda^2\right) - \sigma^2 \log_2 e\right] \end{aligned} \tag{4.102}$$

To maximize the information content, this equation specifies that the Strehl ratio should be maximized or the wavefront error minimized, which is entirely as expected. What is of more interest is the way in which H varies with wavefront error, and also the relation between D and σ. For example, given the number of correction subapertures V and the integration time t, is there an optimum value of D that maximizes the information content of the image?

If we assume that the fitting error of the wavefront correction device dominates the adaptive optics error budget, the wavefront error is

$$\sigma^2 = 0.3\left(\frac{d}{r_0}\right)^{5/3} \tag{4.103}$$

where d is subaperture size. Assuming a circular aperture filled with square subapertures, then $V = (\pi/4)(D/d)^2$. For a fixed number of subapertures, the wavefront error is

$$\sigma^2 = 0.25\left(\frac{D}{r_0\sqrt{V}}\right)^{5/3} \tag{4.104}$$

Defining the quantities

$$A = \frac{\log_2(w\,t\,\eta\,\lambda^2)}{2\lambda^2} \qquad B = \frac{0.25\,\log_2 e}{2\lambda^2 r_0^{5/3} V^{5/6}} \tag{4.105}$$

the following expression for information content is obtained

$$H = AD^2 - BD^{11/3} \tag{4.106}$$

The optimum value of aperture D is then

$$\begin{aligned} D_{\text{opt}} &= \left(\frac{6A}{11B}\right)^{3/5} \\ &= 1.6 r_0 \sqrt{V}\left[\log_2\left(w\,t\,\eta\,\lambda^2\right)\right]^{3/5} \end{aligned} \tag{4.107}$$

This relationship also defines the optimum subaperture size, which is given by

$$d_{\text{opt}} = 1.42 r_0\left[\log_2\left(w\,t\,\eta\,\lambda^2\right)\right]^{3/5} \tag{4.108}$$

It should be noted that the detected photon density and wavelength in this expression refer to the image and not to the wavefront sensor. In partially compensated images, the fitting error is dominant, so the photon error of the wavefront sensor may be ignored.

5 Optical Wavefront Sensors

5.1 Sensors for Astronomical Adaptive Optics

5.1.1 Background

The development of adaptive optics has fostered a new generation of wavefront sensors, instruments with speeds and sensitivities that are orders of magnitude greater than anything known before 1970. Compensation for atmospheric turbulence requires the measurement of randomly distorted wavefronts in real time, using only a faint star as the reference source. A totally new approach to optical wavefront measurement was needed to solve this problem. The wavefront sensors used in adaptive optics are different in almost all respects from the optical interferometers used to measure lenses and mirrors in the laboratory.

In the laboratory environment, lasers provide abundant coherent illumination and measurement times are not critical, so the optical efficiency of the instruments is irrelevant. The adaptive optics environment is the exact opposite, with very low photon flux available and measurement times counted in milliseconds. Together, these requirements dictate the need for high optical efficiency. In addition, adaptive optics wavefront sensors are required to operate with incoherent light and extended reference sources. On the other hand, the spatial resolution and accuracy requirements for adaptive optics are modest compared with those for laboratory instruments.

The special requirements for real-time wavefront sensing in the context of astronomical adaptive optics were reviewed in chapter 2. The operating principles and implementation of specific wavefront sensors are described in detail in this chapter, together with an evaluation of their performance. Some perspective on the characteristics of astronomical adaptive optics wavefront sensors for different tasks is provided in table 5.1. Typical characteristics of laboratory wavefront sensors used for testing optical components are included for comparison.

5.1.2 Wavefront Sensor Efficiency

The performance of astronomical wavefront sensors depends critically on their optical efficiency. The fundamental limit to the precision of all optical sensors is set by the quantum nature of light. In the case of wavefront sensors, the phase measurement error is inversely proportional to the signal-to-noise ratio of the detected radiation from the reference source, a star or laser beacon. In the photon-counting regime of astronomical wavefront sensors, the signal-to-noise ratio, *SNR*, is basically proportional to the square root of the number of photons counted, *N*. This relation may be expressed by

$$\sigma^2 \propto \frac{1}{SNR^2} \propto \frac{1}{N}$$

Increasing the number of photons counted by a factor of 2 should therefore halve the error variance of the measurement.

Table 5.1 Wavefront sensor characteristics

Requirement	Laboratory Measurement	Infrared AO	Visible AO	Laser Beacon AO
Spatial resolution, samples per meter (at primary mirror)	10-50	2-5	5-20	5-20
Measurement time, (ms):	» 1 0 0	10–50	2-10	2-10
Accuracy, waves rms:	0.01-0.05	0.05-0.1	0.05-0.1	0.05-0.1
Photon count, per sample:	» 1000	100	100	100
Reference source:				
Type:	Point	Extended	Extended	Extended
Spectral bandwidth:	Monochromatic	Wideband	Wideband	Monochromatic

AO, adaptive optics.

Ideally, the detection of 10 photons should enable the phase to be measured to a precision of 1/V10 radians, or about 1/20 wave. Using a subaperture area of 1 m^2 and an integration time of 0.02 s, with 100% detection efficiency, the photon flux required is about 500 photons $m^{-2} s^{-1}$, which is provided in the visible band by a star of $m_V = 20$. Current adaptive optics systems are capable of compensating images at near-IR wavelengths using reference stars of $m_V = 15$. While this is a significant achievement, there remains a potential improvement of about 5 stellar magnitudes (a factor of 100 in the number of photons) to be achieved in the performance of astronomical wavefront sensors.

The measurement efficiency of wavefront sensors is determined by three main factors:

1. The optical efficiency, including the optical phase-to-intensity conversion process and the fraction of the incident light that reaches the photon detectors;
2. The detector characteristics, such as quantum efficiency and signal-to noise ratio;
3. The effectiveness of the subsequent data processing, specifically the amount of wavefront information extracted from each detected photon.

The efficiency of the phase-to-intensity conversion process depends entirely on the optical configuration of the sensor; it is probably the most important factor and the least appreciated. Practical implementation of wavefront sensors has become focused on three closely related types: Shack-Hartmann, lateral shearing interferometer, and curvature sensor. The relative performance of these sensors depends a great deal on the specific optical configurations employed. Most performance comparisons are made on a fundamental basis that ignores engineering issues, which can, in fact, be the dominant factor in performance. For example, maximization of optical efficiency requires the avoidance of components with partial transmission, such as gratings, masks, and beam splitters.

With reference to detector characteristics, the availability of low-noise silicon charge-coupled devices (CCDs) with their high quantum efficiency (> 0.8) and low readout noise (< $10e^-$ per pixel) has been of great value for wavefront sensors operating at visible wavelengths. At longer wavelengths, the highest quantum efficiencies are presently obtainable with doped silicon and gallium arsenide detectors. The effects of turbulence at these longer wavelengths are smaller than at visible wavelengths, so that fewer subapertures and longer integration times are possible, resulting in a marked improvement in the performance of adaptive optics systems.

The significance of the efficiency of the data-processing algorithm in converting the number of detected photons into an estimate of wavefront error has been appreciated for many years. Wavefront reconstruction algorithms differ in their complexity according to the type of wavefront sensor, the optimization criterion used, and the degree to which collateral information, such as turbulence strength and signal-to-noise ratio, is taken into account. The principles of efficient wavefront estimation are well understood (see section 8.6), but these ideas have not yet been fully integrated into operating hardware.

5.7.3 *Wavefront Measurement Speed*

In astronomical wavefront sensors, the design of the detector and data-processing functions is dominated by the requirement for measurements of complete wavefronts at millisecond intervals. As a result, specialized detectors and data-processing systems have been developed for use in adaptive optics wavefront sensors. Even for a detector with relatively low spa-

tial resolution, such as 32 x 32, the detector pixel readout rate and the resulting data-processing rate are in the megahertz range. The detector readout noise is proportional to the pixel readout rate, so there is considerable incentive to reduce this rate to the absolute minimum. The use of CCDs with multiple readout ports is one solution to this problem.

In general, one wavefront measurement must be made per correction cycle for each degree of freedom in the adaptive system and at least two intensity samples are needed to generate each measurement. Wavefront slope sensors having two measurements (x and y tilt) per subaperture therefore require at least four detector samples per subaperture per complete wavefront. Shack-Hartmann sensors employ a spatial array of at least four pixels (a quad cell) per subaperture, and a larger array is necessary if the dynamic range is greater than ±1 wave. Shearing interferometers using alternating current (AC) modulation use three or four samples of the modulated signal for each measurement, and only one detector pixel is required for each subaperture. Curvature sensors need only one measurement per subaperture, together with gradient sensing at the edges of the full aperture.

The availability of low-noise 64 x 64 multiport CCD arrays has solved the adaptive optics wavefront sensor detector problem for telescope apertures up to about 2.5 m in the visible band, and for apertures up to about 10 m in the infrared (IR) at 2 μm. It is expected that even larger CCD arrays with parallel readouts will soon become available, and these will have pixel readout rates suitable for compensating even the largest telecopes at visible wavelengths.

The need for fast wavefront measurement in adaptive optics makes it preferable to use a deterministic technique for wavefront sensing, rather than an iterative process. The deterministic, or direct, approach has the advantage that a single set of measured data leads to an unambiguous realization of the wavefront. Wavefront sensors that make simultaneous measurements of wavefront properties in the telescope pupil (or at a reduced image of the pupil) are of this type, including all of the commonly used sensors. Iterative sensing schemes (also known as indirect wavefront sensors), employing image-sharpening and other "hill-climbing" algorithms, have been tested in adaptive optics systems using a small number of subapertures. Iterative wavefront sensors employ a simple photodetector, often only a single element located behind a mask at the optical focus, to sense image quality. The data processing is also relatively simple. Unfortunately, such systems are not competitive with direct wavefront sensing because the time required to make and evaluate trial wavefront corrections in each subaperture becomes an increasing burden with large apertures. Also, the uniqueness of the solution is often in question because it is not unknown for iterative sensors to converge on the wrong peak.

5.1.4 Reference Sources

To accommodate both natural and artificial reference sources, astronomical wavefront sensors must be capable of operating with spatially and temporally incoherent radiation. The radiation from most natural objects is of thermal origin, producing a broad spectrum. A broadband sensor is therefore needed to make the most efficient use of the radiation from such sources. Beacons produced by ground-based lasers are monochromatic, but the upgoing beams have less than perfect beam quality and are spread by atmospheric turbulence to produce spots with angular diameters of several arc seconds. These artificial "stars" are not point sources; they are small extended objects. Because of the varied nature of these reference sources, astronomical wavefront sensors do not measure optical phase, which has no meaning for incoherent radiation, but instead measure wavefront slope, from which optical pathlength variations are computed.

The reference source provides the direction from which the wavefront slope sensor determines the tilt angle of each zone of the wavefront. The simplest case to consider is that of an unresolved star. Slope sensors determine the center of intensity of the star's image produced by each subaperture. This measurement is valid over a radius equal to the isoplanatic angle of the atmospheric transmission path. It is not essential for the reference to be a point object, only that its angular size is smaller than the isoplanatic patch. Practical wavefront sensors must be capable of accommodating such objects. But, there are some even larger astronomical objects, such as the Sun, Moon, and the planets, the angular size of which considerably exceeds the isoplanatic angle, and the surfaces of which consist of low-contrast detail. For example, the solar granulation has a contrast (in the white-light continuum) of about 3%, with a characteristic cell size of 2-3 arc seconds. Provided that such surface detail has spatial variations within the isoplanatic angle, it can be used as a reference source.

To handle large extended objects, a fixed field stop with radius less than the isoplanatic angle is required to define the reference area. In this case, the angular tilt of the wavefront is determined by cross-correlation of the low-contrast image detail with a transmission function containing the same spatial frequency components as the image. The mask used for cross-correlation could be based on the image detail itself, with some means for configuring the mask function in real time. A system employing optical cross-correlation using a liquid crystal screen has been proposed by von der Luhe [1988]. It is also possible to use electrical cross-correlation employing a two-dimensional detector array; for example, using a CCD. One of the problems associated with the use of extended fields for wavefront sensing occurs when specific details in the random field straddle the edges of the field stop, causing transients in the wavefront sensor output. These effects can be minimized

by grading the edges of the field stop to make a soft transition.

5.2 Wavefront Sensing Techniques

5.2.1 Types of Wavefront Sensor

Requirements discussed above for wavefront sensing in astronomical adaptive optics eliminate most of the conventional methods of wavefront measurement used in the production and testing of optical components and systems. The techniques used in adaptive optics fall into two main classes:

1. Direct wavefront measurement, in which specific properties of the wavefront are measured, using zonal or modal decomposition of the pupil.
2. Indirect measurement, in which the wavefront properties are deduced from whole-aperture intensity measurements made at or near the image plane.

The wavefront sensors in each of these classes that have potential application to adaptive optics are listed in table 5.2. The relationship between these various approaches and their main characteristics will be discussed next.

5.2.2 Direct Wavefront Sensing

Direct wavefront sensors are those which determine the wavefront shape in the optical pupil, using either zonal or modal measurements. With zonal sensing, illustrated in figure 5.1(a), the wavefront slope, a vector, is measured within a number of contiguous zones in the pupil. The principle is to decompose a complex wavefront into simple elements that can be easily sensed. The measurements in each zone are independent and simultaneous. The zones are sized so that the dominant distortion within each zone is the average wavefront tilt. Large apertures are handled by replicating the same elements, the spatial resolution being determined entirely by the size and number of elements used. Direct sensing requires the light from each zone to be brought to a focus at which specific intensity measurements are made. This does not violate its status as a "direct" wavefront measurement because each zone is measured independently.

The principle of modal sensing is depicted in figure 5.1(b). The objective in this case is to decompose the wavefront into a number of distinct surface shapes or modes, each covering the whole aperture. These modes are also measured independently and simultaneously. In principle, modal sensing systems are capable of the same performance as zonal systems, but their implementation is difficult, except for the very lowest modes. Overall tilt and defocus can easily be detected, but beyond these first few orders the modal functions required to describe a turbulence-degraded wavefront become increasingly complex and difficult to distinguish. It is possible to convert zonal data to modal data and vice versa, so that a zonal wavefront sensor may be used with a modal wavefront corrector. There are advantages in using modal compensation when the wavefront errors are predominantly of low order (which can occur at IR wavelengths) because the local irregularities that are often present in zonal devices are avoided.

5.2.3 Indirect Wavefront Sensing

Indirect wavefront sensors are those in which the wavefront errors are deduced from their effect on a related parameter, usually the intensity distribution at or near the image plane. The wavefront at the pupil is not divided before measurement, so each measurement contains information averaged over the whole aperture. Wavefront sensors of this type are shown in figure 5.2. When the measurement is made at the image plane, as shown in figure 5.2(a),

Table 5.2 Wavefront-Sensing Techniques

Direct Measurements in Optical Pupil	Indirect Measurements at Image Plane
Zonal	**Aperture tagging**
Wavefront slope sensing	Sequential (time division)
Shack-Hartmann sensor	Image sharpening
Lateral shear interferometer	
Wavefront curvature sensing	Frequency division
Curvature sensor	Multidither systems
Modal	**Wavefront deconvolution**
Overall tilt sensing	Phase diversity
Focus sensing	

Specific wavefront sensors described in this chapter are underlined.

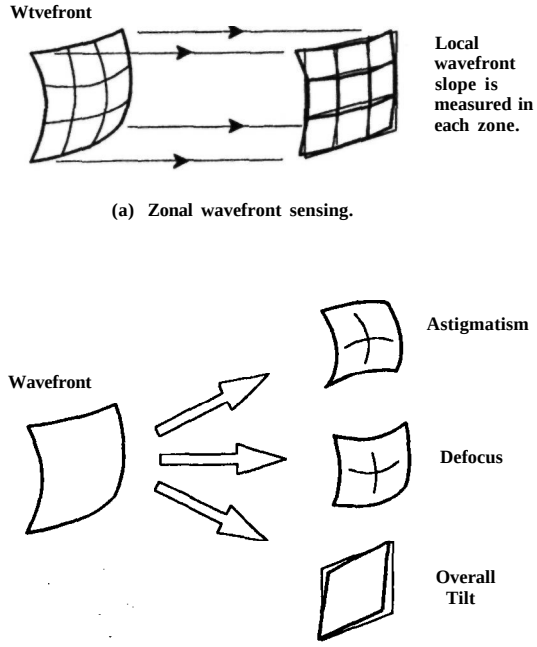
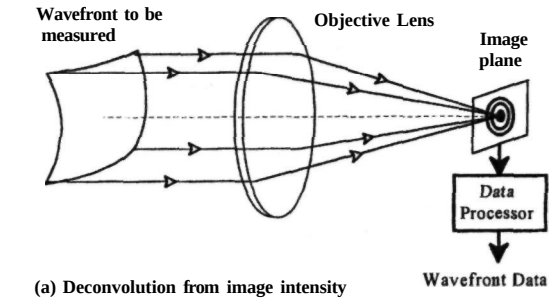


Figure 5.1 Direct wavefront sensing techniques. These methods measure the wavefront in the optical pupil. (a) In zonal sensing, the wavefront is divided into an array of equal size zones, in each of which the wavefront slope is measured. (b) In modal sensing, the wavefront is decomposed into whole-aperture modes, such as overall tilt, defocus, and astigmatism.

the intensity distribution is related to the wavefront through a Fourier transform. The process of deducing the wavefront from measurements at the image plane is known as deconvolution.

Optical image formation is a nonlinear process, which causes several difficulties with deconvolution, among them the fact that the process does not work well when the aberrations are large; other problems concern the uniqueness of the solution and the fact that the process is computation-intensive. The main difficulties stem from the spatial invariance produced by the Fourier transform; intensity variations in the image plane are not directly related to specific locations in the optical aperture.

Several methods have been devised to overcome the invariance problem. One approach is the use of a second out-of-focus detector array that restores the spatial data missing at the focal plane. Another approach is aperture tagging, shown schematically in figure 5.2(b), in which modulation is applied to specific zones or modes in the aperture to connect them with image-plane measurements. In one method known as image sharpening, small perturbations are applied to each actuator of the deformable mirror in sequence; the corresponding changes in image intensity are measured. Another technique, used in laser beam compensation systems, is known as multidither, in which small perturbations at different temporal frequencies are applied simultaneously to all zones of the optical aperture, using either the deformable mirror itself or an auxiliary device. Indirect wavefront-sensing techniques using aperture tagging have been shown to be useful for small apertures, but become increasingly difficult to apply to large apertures; they also suffer from a signal-to-noise disadvantage compared with direct wavefront-sensing methods. With sequential image-plane sensing, the available time for each wavefront measurement must be shared between all the zones in the aperture, as pointed out by Dyson [1975]. As a consequence, the signal-to-noise ratio varies inversely with the number of zones. But, with direct sensing in the pupil plane, whether zonal or modal, the zonal mea-



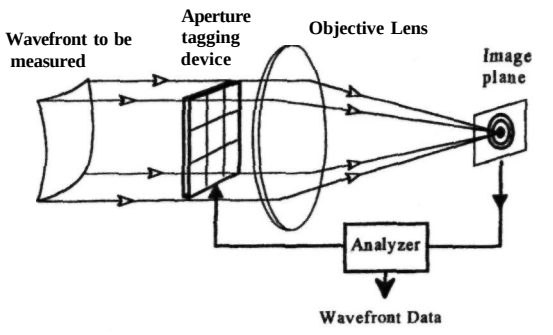


Figure 5.2 Indirect wavefront sensing techniques. These methods deduce the shape of the wavefront from the intensity distribution of the image of the reference source. (a) The deconvolution method is optically simple, but involves considerable data processing. This technique is more robust when augmented by an additional set of measurements displaced from the image plane. (b) The aperture-tagging system provides a means of relating the measured intensity variations in the image plane to the appropriate zones or modes of the disturbance in the pupil.

surements are independent and are made in parallel. The signal-to-noise ratio for direct sensing is therefore independent of the size of the telescope aperture.

5.2.4 *Principle of Wavefront Slope Sensing*

Because the reference sources used in astronomical adaptive optics emit incoherent radiation, it is not possible to make absolute measurements of optical phase. Instead, wavefront slope sensors measure the local wavefront gradient by sensing the direction of propagation of optical rays, which are perpendicular to the wavefront. Even if coherent reference sources were available, there are reasons for preferring gradient measurements to phase measurements, because gradient sensors are very robust and less susceptible to errors caused by temperature changes and vibration. In addition, the dynamic range of shearing interferometer sensors can be adjusted to match the expected wavefront excursions. This is particularly useful for measuring turbulence that produces large phase excursions of several wavelengths of light. Direct measurement of optical phase, even if it were feasible, would require tracking fringes through several rotations of the phase vector, involving much additional data processing.

The principle of wavefront slope sensing is to convert angular deviations of rays into intensity variations that can be sensed by a photodetector. The process is shown in figure 5.3. An optical wave, initially plane, propagates in direction z,and it is desired to find the wavefront surface $W(x,y,z)$ after it has passed through a region of varying refractive index. Wavefront distortion produces local variations in the gradient of the wavefront, and this causes interference after propagation for some distance. Intensity variations due to the interference are detected by measuring the intensity distributions at two planes, z_1 and z_2. These intensity variations reveal the shape of the wavefront surface.

The process of converting wavefront gradients into measurable intensity variations is formally treated using the irradiance transport equation, as described by Teague [1982]. This approach gives considerable insight into the relation between various wavefront slope measuring techniques. The irradiance transport equation describes the intensity variations in a beam with irradiance $I(x,y,z)$ as it propagates along the z axis of an optical system. Using the paraxial approximation, the complex amplitude is given by

$$A(x, y, z) = [I(x, y, z)]^{1/2} \exp[ik\, W(x, y, z)] \quad (5.1)$$

where $W(x,y,z)$ is the wavefront surface at distance z from the origin. The change in irradiance along the propagation path is then

$$\frac{\partial I}{\partial z} = -(\nabla I \cdot \nabla W + I\nabla^2 W) \quad (5.2)$$

where V = *d/dx* 4- *d/dy* is a gradient operator in the *x, y* plane.

The first term V/ • ∇W, the prism term, represents the irradiance variation caused by transverse shift of the beam due to local tilt of the wavefront. The second term / V^2 W, the lens term, can be interpreted as the irradiance variation caused by convergence or divergence of the beam, whose local curvature is proportional to V^2 W. These terms combine to give the variation in beam irradiance as it propagates along the z axis. The transport equation represents the law of light energy conservation. Although originally derived for coherent light, the transport equation has been shown to be valid for incoherent (extended) light sources when the source is uniform and symmetrical [Streibl 1984].

For wavefront sensing, the local wavefront slope *VW* or the local curvature $\nabla^2 W$ must be found. In general, this is done by measuring the distribution of intensity at two planes, *z*\ and Z2, as shown in figure 5.3. For small wavefront excursions such as those encountered in astronomical adaptive optics, the distance between *z*\ and z_2 would need to be very large to produce measurable intensity variations.

Fortunately, the intensity changes produced by small wavefront gradients can be greatly enhanced by placing a mask with a known transmission function at z_1, and then measuring the resulting intensity distribution at plane z_2. The mask is made to match the known characteristics of the wavefront. For example, its spatial resolution (and that of the intensity detector) should match the coherence length of the wavefront to be measured. In wavefront slope

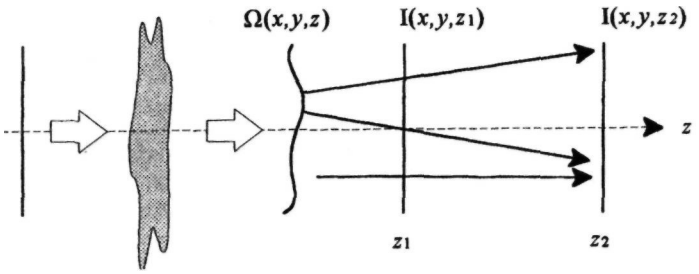


Figure 5.3 Principle of wavefront slope sensing.

sensors, the zones are sized so that the error involved in assuming a uniform gradient over each zone is small compared with the required measurement accuracy. The shape of the wavefront is then found by spatially integrating the individual zonal gradients over the whole aperture; this is the process known as wavefront reconstruction.

S.2.S Pupil-Plane Wavefront Sensors

Wavefront disturbances due to atmospheric turbulence are distributed over a considerable distance in the propagation path, causing two effects: (1), it restricts the isoplanatic region over which the wavefront is spatially correlated; and (2), it can produce intensity variations (due to scintillation) across the telescope pupil. These effects complicate the measurement of wavefront excursions. In this section, only a single isoplanatic region is considered.

Intensity variations in the telescope pupil, caused by scintillation, may be significant with unresolved reference sources, such as stars, but are negligible with laser beacons and other extended sources. In any case, pupil intensity variations have a small effect on the quality of the image, provided that the phase is correctly compensated. However, the presence of these input intensity variations may affect the accuracy of the wavefront phase measurement, because they will be superimposed on the intensity variations generated in the wavefront sensor to measure the phase. This situation is handled by encoding the wavefront phase information in a form that can be distinguished from input intensity variations. For example, wavefront gradients can be made to control the positions of spots or edges, which can be accurately measured over a wide range of intensities, or differential intensity measurements can be made that reject the input intensity variations.

In the case of a single isoplanatic region, and assuming that intensity variations due to scintillation are taken care of, the wavefront disturbances to be compensated can be considered as localized at the telescope aperture or entrance pupil. Using adaptive optics, the wavefront corrections are made at a plane conjugate to the telescope pupil. Therefore, it is logical to make the wavefront measurements at the same plane, because direct measurements of this kind avoid the need for transform operations on the data, which would be necessary if the measurements were made near the focal plane. This is the justification for preferring direct wavefront measurement at the pupil plane in adaptive optics systems.

The basic arrangement of figure 5.3 is the prototype for most wavefront sensors used in adaptive optics. It has been pointed out by F. Roddier [1990a] that many apparently different types of wavefront sensor can be modeled by placing an appropriate mask near the optical pupil. The canonical arrangement is shown in figure 5.4. The telescope pupil containing the wavefront disturbance to be measured is reimaged at reduced diameter at plane *P*, which is assumed to be uniformly illuminated. The detector array is at plane *D*, which is conjugate to *P*. In the absence of wavefront distortion and any other components in the optical path, the detector will also be uniformly illuminated, because it is imaged at plane *P*.

To convert wavefront phase deviations into intensity changes that can be measured by the detector array, a mask or modulation device $\boldsymbol{M}$ is inserted in the collimated beam near *P*. As the light travels from *M* to *D*, the interaction of the wavefront disturbance with the mask produces variations in the intensity, according to equation (5.2). These variations are measured by the detector, allowing the wavefront to be estimated. For example, if $\boldsymbol{M}$ is a periodic pattern consisting of equal-width clear and opaque bars, and is displaced from plane *P* by distance Az, then the image produced at the detector plane *D* becomes distorted by wavefront slopes perpendicular to the bars. This is known as the *Ronchi test.* From equation (5.2), it can be seen that the displacement of the bars in the direction of slope is

$$\Delta x = \left(\frac{\partial W}{\partial x}\right)\Delta z \tag{5.3}$$

When viewed as a diffraction grating, the system operates as a *spatial carrier shearing interferometer,* as described by Horwitz [1990]. In this case, it is easy

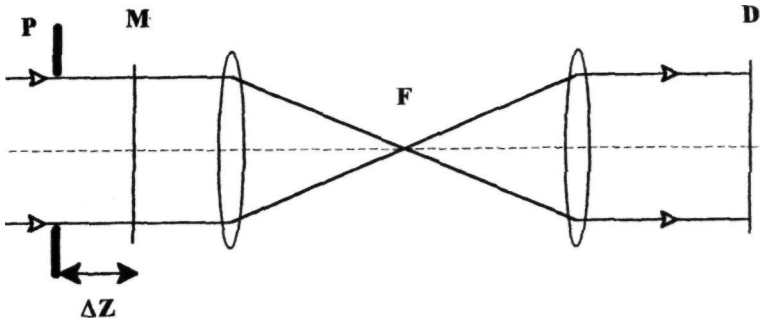


Figure 5.4 Canonical wavefront sensor.

to show that the diffraction fringes are displaced by an amount exactly proportional to the product of the wavefront slope and the distance Az, independent of wavelength; this is the same result as equation (5.3).

If a crossed grating is used, as shown in figure 5.5, then wavefront slopes can be detected in two dimensions simultaneously by measuring fringe displacements along the x and y axes. Carrying the process further, the mask can consist of a phase grating in the form of two sets of crossed cylinder lenses, as shown in figure 5.6, giving a high optical throughput. The lenses produce an array of spots, the displacement of which is proportional to wavefront slope. This arrangement is identical to Shack's modification of the *Hartmann sensor,* which uses an array of contiguous lenses in the aperture to concentrate the light in discrete spots at the image plane. The Shack-Hartmann sensor is probably the most-used wavefront sensor in adaptive optics and is described in detail in section 5.3.

Further configurations are possible by placing a modulation device at the focal plane F, at which the intensity distribution is the Fourier transform of the function at plane P. If a knife-edge is placed at F, excluding half of the frequency plane, then it produces the well-known *Schlieren system* for making phase objects visible. If the central order at plane F is retarded by 1/4 wave, then *Zernike's phase-contrast method* for viewing phase objects, used in microscopy, is obtained.

A grating located at plane F makes the optical system function as an *achromatic shearing interferometer;* this has different characteristics from those described above, where the grating was placed near the optical pupil. When located at the focal plane, the grating generates sheared replicas of the pupil at the detector plane D, as shown in figure 5.7. With no aberrations in the pupil, a fringe-free field is produced. Wavefront gradients in the pupil, in the direction of shear, produce intensity changes proportional to the gradient. The shear distance is proportional to the spatial frequency of the grating. When the grating is moved at velocity v in the focal plane, in a direction normal to the grating lines, the illumination at the detector plane is modulated at a temporal frequency v/a, where a is the grating period. The phase of the modulated signal at any point in the aperture is a measure of the wavefront gradient at that point. This device is the prototype for the *alternating-current shearing interferometer,* one of the basic wavefront sensors used in adaptive optics, which will be described in detail in section 5.4.

Another variety of wavefront sensor is created when a variable convergence device (such as a variable curvature mirror) is placed at the focal plane F, as shown in figure 5.8. The variable curvature mirror (VCM) produces longitudinal modulation, displacing the conjugate image of the detector away from its normal location at plane P. Intensity changes produced by wavefront slopes will therefore be scanned across the detector plane. Suppose, for example, that a Hartmann mask is placed at P and the VCM is driven with a sine wave deflection. If there are no aberrations in the system, then all rays are axial and the spots of light at the image plane will be stationary.

If aberrations are present in the pupil, then the ray directions will be deviated, causing the light spots to move cyclically at the detector plane, as shown in figure 5.9, and indicating the direction of the wave-

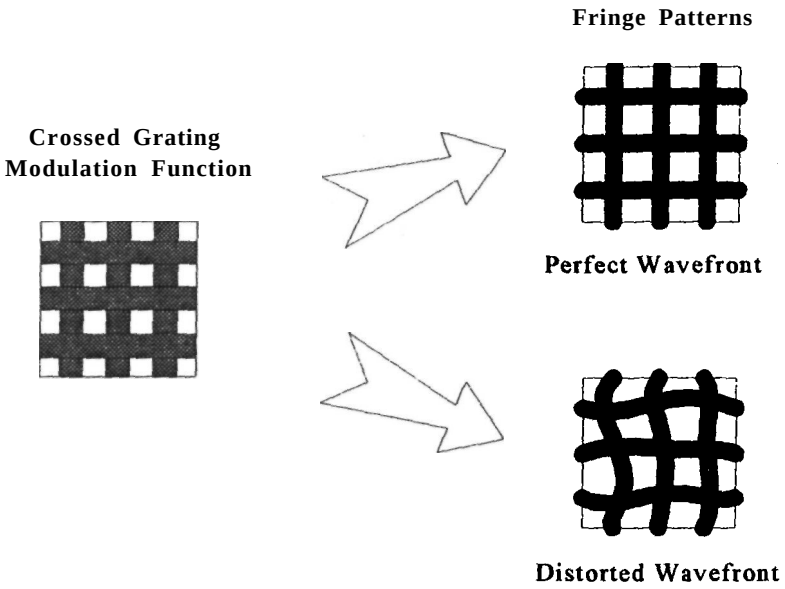


Figure 5.5 Crossed grating modulation function and resulting fringe patterns. This type of grating is used in shearing interferometer slope sensors.

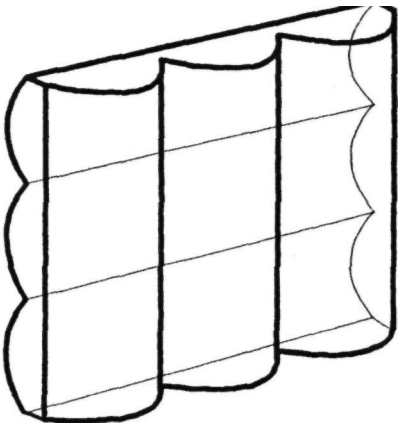


Figure 5.6 Phase grating in the form of two sets of cylinder lenses. This is functionally similar to the array of lenslets used in Shack-Hartmann sensors.

front slope. There are two advantages in using AC modulation of this type: (1) the position of the holes in the mask is noncritical, because it is the changes that are measured; and (2) the sensitivity of the device to wavefront slope can be varied by changing the amplitude of the modulation, which is a very useful capability in dealing with the wide variations in the strength of atmospheric turbulence.

A device of this kind can also be used to detect wavefront curvature directly [Roddier 1988a]. In this case, the mask defines the optical pupil, which has a sharp intensity gradient at its edges. Changes in wavefront curvature produce convergence or divergence of the normally parallel beam, resulting in changes in the beam size. These produce radial displacements in the edges of the beam and also changes in the average intensity of the beam, both of which can easily be measured. The simplest form of *curvature sensor* consists of two detector planes, one on each side of the focal plane. Local changes in wavefront curvature produce intensity changes of opposite polarity at the two planes, so that the ratio of the intensities in corresponding zones is a measure of the wavefront curvature. Methods of curvature sensing are described in more detail in section 5.5.

5.3 Shack-Hartmann Sensor

5.3.1 Principle of Operation

The principle of the Shack-Hartmann gradient sensor is extremely simple and intuitively appealing, which may explain its great popularity. An array of identical positive lenslets is placed in the pupil of the optical beam to be measured; one such lenslet is shown in figure 5.10. To obtain maximum efficiency, the lenslets are contiguous and cover the entire optical aperture. The single reference source, ideally an unresolved point, is brought to a separate focus by each lenslet, thus producing an array of spots in the focal plane. With a plane-wave input, each spot is located on the optical axis of its corresponding lenslet. Distortion of the input wavefront produces a local gradient $\alpha(x, y)$ over each lenslet, displacing each spot by a distance $s(x, y) = \alpha(x,y)Z$, where Z is the lenslet focal length. The lenslet array therefore converts wavefront gradients into measurable spot displacements.

In the usual case of square subapertures, the intensity distribution of each image is

$$I(x, y) = I_0\left(\frac{\sin ax}{ax}\right)^2\left(\frac{\sin ay}{ay}\right)^2 \qquad (5.4)$$

where

$a = \pi d/\lambda z$,
d = subaperture size,
λ = mean wavelength

The first minimum of the diffraction pattern occurs at $x' = y' = \pm kz/d$. A spot displacement of this value corresponds to a wavefront tilt of one wavelength over the subaperture. The distance between the minima, defining the effective width of each spot, is

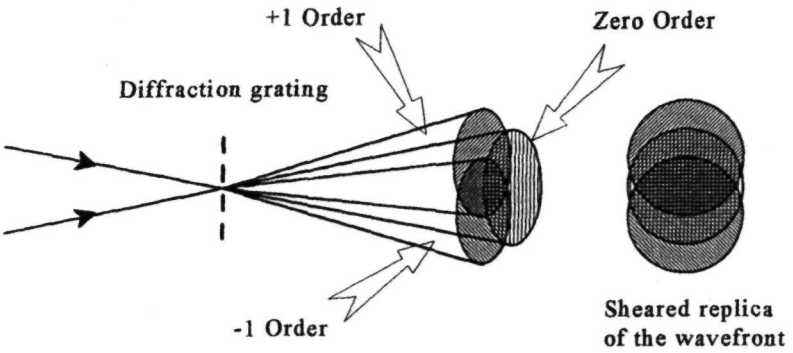


Figure 5.7 Shearing interferometer with temporal modulation using a moving grating. The diffraction grating at the focus of the reference source generates overlapping (sheared) replicas of the wavefront. As the grating moves at a constant velocity, the intensity pattern produced by the interfering orders is sinusoidally modulated, greatly simplifying the detection process.

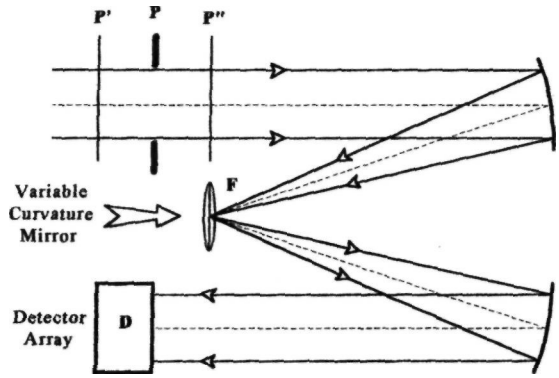


Figure 5.8 Wavefront sensor using a variable curvature mirror.

$$\Delta x = \frac{2\lambda z}{d} \tag{5.5}$$

The spot displacements, being vector quantities, are sensed by a two-dimensional detector array. The simplest arrangement is to use a quadrant detector, consisting of two orthogonal bicells, for each subaperture. To obtain better linearity and dynamic range, a larger array consisting of 4 *x* 4 pixels may be employed. The performance and constraints on the Hartmann sensor are different in these two cases, so they will be analyzed separately.

Bicell Detector

When using a bicell detector, the linear range of a Hartmann sensor is determined mainly by the width of the diffraction image. Transfer functions of a Shack-Hartmann sensor using a bicell detector are shown in figure 5.11. With a single point reference source focused to a diffraction-limited spot, figure 5.11(a), the linear range is limited to about 0.5 waves of tilt over each subaperture. One wave of subaperture tilt displaces the central core of the spot entirely into one cell of the detector array, so further motion produces no significant increase in electrical output. In terms of the tilt angle, the linear range of each subaperture of a Shack-Hartmann wavefront sensor using a bicell is limited to

$$\alpha_B = \pm\frac{\lambda_{ref}}{2d} \quad \text{rad} \tag{5.6}$$

No matter what the subaperture size or the detector size, the linear range of tilt measurement is restricted to about ±0.5 wavelength.

The question then arises as to what effect this limitation has on the measurement of atmospheric turbulence. When an adaptive optics control loop is initially closed, the full uncorrected wavefront errors due to turbulence are present and must be measured unambiguously in order to converge or "lock on." Once convergence has been achieved, then the residual errors appearing at the wavefront sensor (in a closed-loop system) are small. To answer this question, the wavefront tilt angle due to uncompensated turbulence is determined as a function of the subaperture size over which it is measured.

As shown in section 3.4.6, the mean-square tilt angle due to to atmospheric turbulence, over an aperture of size d, is

$$\sigma_\alpha^2 = 0.182\left(\frac{d}{r_0}\right)^{5/3}\left(\frac{\lambda}{d}\right)^2 \quad \text{rad}^2 \tag{5.7}$$

If the turbulence strength is specified as r_0 at wavelength A, then the value of r_0 at any other wavelength λ is $r_0 = \bar{r}_0(\lambda/\bar{\lambda})^{6/5}$. The standard deviation of the tilt angle due to turbulence may then be written

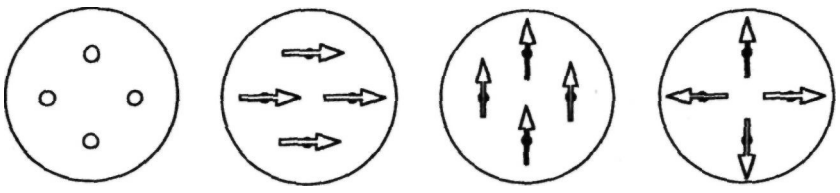
$$\sigma_\alpha = 0.427\frac{\bar{\lambda}}{d^{1/6}\bar{r}_0^{\,5/6}} \quad \text{rad rms} \tag{5.8}$$

If this atmospheric tilt angle is matched to the linear range of a bicell Shack-Hartmann sensor (that is, $\sigma_\alpha = \alpha_B = \lambda_{ref}/2d$), then

$$d' = 1.21\bar{r}_0\left(\frac{\lambda_{ref}}{\bar{\lambda}}\right)^{6/5} \tag{5.9}$$

where d' is the subaperture size (in object space) over which the uncorrected rms wavefront tilt is equal to 0.5 wavelength at the reference wavelength.

In astronomical adaptive optics, the reference source wavelength λ_{ref} is usually fixed by system considerations, such as the use of a sodium beacon at



No aberrations **X tilt error** **Y tilt error** **Focus error**

Figure 5.9 Effect of focus modulation on the spots produced by a Shack-Hartmann sensor.

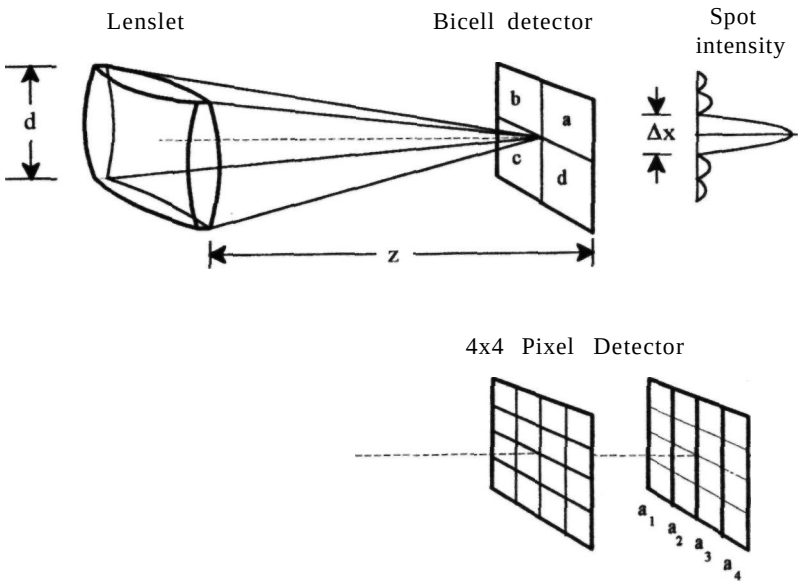


Figure 5.10 Geometry of one subaperture of a Shack-Hartmann slope sensor.

$0.589\mu m$, while the turbulence parameter r_0 is usually specified at a wavelength of $\lambda = 0.500\,\mu m$, giving the relation $d' = 1.47\bar{r}_0$. Typical values of $\bar{r}_0$ range from 0.1 to 0.2 m, so that d' generally falls between 0.15 and 0.3 m. If the subapertures used in the adaptive optics are smaller than d', the use of a bicell Shack-Hartmann sensor with limited linear range should not cause problems with initial convergence of the adaptive optics. Peak wavefront excursions with values of $2.5\sigma_\alpha$ or more will occur occasionally, but even if the linear range were exceeded and the bicell saturated, it would continue to give the correct polarity of control signal to the feedback loop, providing that the bicell size is large enough to accept the full tilt range. Once the control loops have locked on, the residual error is reduced to a small value and the dynamic range is not usually an issue. There are two scenarios in which the limited range of a Shack-Hartmann sensor using a bicell detector could cause problems. The first of these is when a short-wavelength reference source is used, such as a Rayleigh beacon using an excimer laser. The second occurs with partially compensated adaptive optics employing large subapertures.

Shack-Hartmann sensors using bicell (quadrant) detectors are very sensitive to the shape of the reference source, especially those containing multiple points. Double stars, for example, may produce a dead band in the response, as indicated in figure 5.11(b).

One way of improving the dynamic range is to enlarge the spots by defocusing, but this increases the measurement error due to the drop in sensitivity. A better way to improve the dynamic range is to use more detector elements for each subaperture. This may be done with CCD detectors by partitioning the array into 4 x 4 or larger sections for each subaperture, as described later.

The factors that determine the detector cell size and focal length of a Shack-Hartmann sensor using a bicell (quadrant) detector will now be considered. As explained previously, neither of these parameters changes the linear range of the sensor. The detector cell spacing usually matches the subaperture size d. For a fixed cell size, the focal length determines the total (saturated) tilt range of each subaperture of the sensor, and also the crosstalk threshold, which occurs when the image produced by one subaperture crosses over into an adjacent detector cell. This is most likely to happen during initial convergence, before the turbulence has been corrected by the system.

The maximum focal length is determined by the condition in which the diffraction image between the first nulls is entirely contained within one bicell. Assuming the width of each bicell is $d/2$, this condition can be stated as

$$\frac{2\lambda z}{d} \le \frac{d}{2} \quad \text{or} \quad z_{max} = \frac{d^2}{4\lambda} \tag{5.10}$$

In a typical case, with $d = 1$ mm and $X — 0.589\,\mu m$, the maximum focal length of the lenticular array is $z_{max} = 420$ mm.

The crosstalk threshold is then ±1 wave of tilt across each subaperture. The threshold may be made arbitrarily large, for a fixed subaperture size, by using a smaller focal length. A practical limit is set by the inevitable gap between the bicell components,

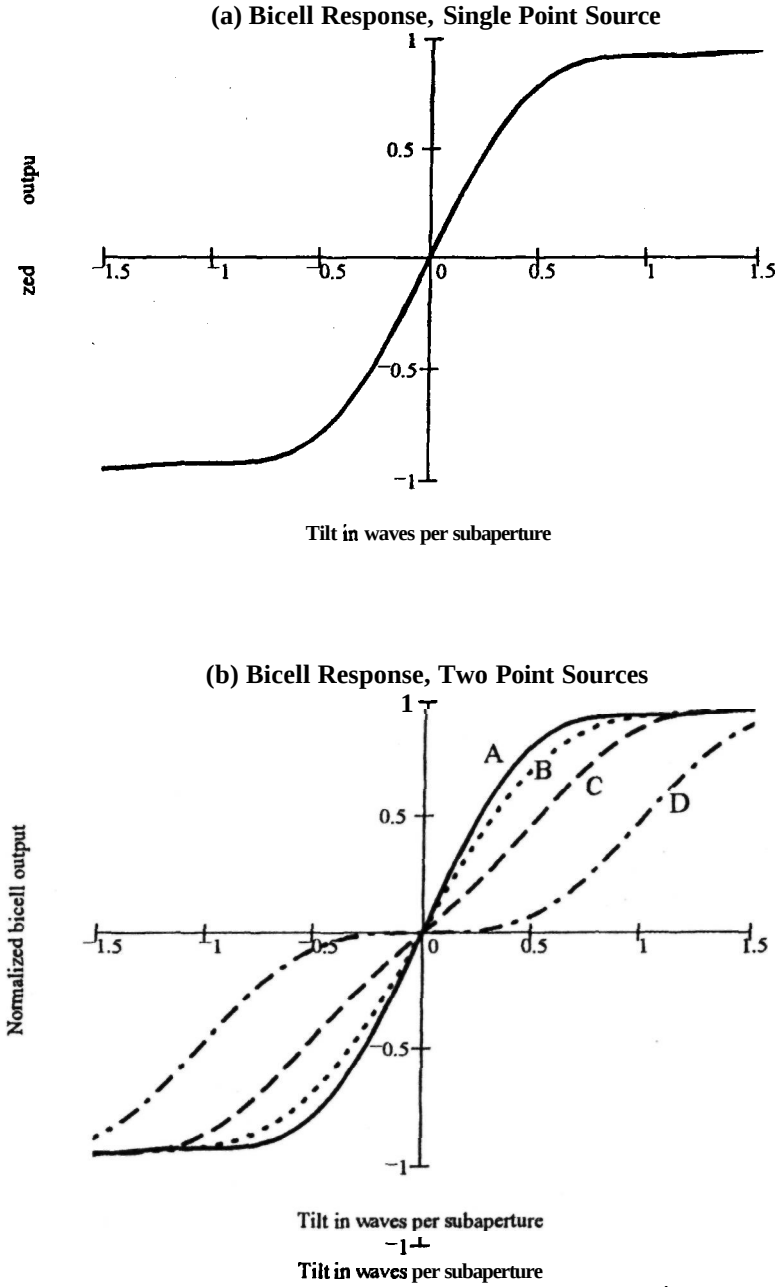


Figure 5.11 Transfer functions of a Shack-Hartmann sensor using a bicell detector with (a) a single point source, (b) two point sources of varying separation.

which increases the measurement error and consequently should be much smaller than the image core.

Multicell Detector Arrays

The design considerations for Shack-Hartmann sensors using multicell detector arrays, consisting of 4 x 4 or more elements, are considerably different from those using bicell detectors. With multicell detector arrays, the linear range of tilt measurement is determined by the geometry of the detector array and is no longer dependent on wavelength, as it was with a bicell detector. A 4 x 4 array has considerably greater flexibility than the simple bicell detector and provides a good compromise between performance and complexity. Because there are 16 pixels rather than 4, the electronic processing hardware is much more complex than for a quad cell. The discussion will be focused on 4 x 4 arrays as it is doubtful whether arrays larger than 4×4 can be justified.

The main parameters involved in the design of multicell Shack-Hartmann sensors are the number of subapertures *Ns,* the number of pixels per channel in the photon detector N_P , and the ratio of pixel size to spot width, *p.* The overall system design usually dictates the number of subapertures and the dynamic range over which slope measurements are required. The linearity depends on the number and weighting of the pixels, as well as on the ratio *p.*

The transfer function of a Shack-Hartmann sensor is easily calculated from a knowledge of the intensity distribution, as defined in equation (5.4), and the ratio of pixel size to spot width, *p.* The intensity within each pixel is computed as the optical axis is translated across the pixel array. As the responses are identical in the *x* and *y* directions, the pixels are summed across the array in the direction orthogonal to the scan to obtain the "four-bin" elements a_1 through a_4 for each axis. The basic four-bin slope-detection algorithm has the form

$$A = \frac{w_1 a_1 + w_2 a_2 + w_3 a_3 + w_4 a_4}{a_1 + a_2 + a_3 + a_4} \qquad (5.11)$$

where w_1 ... w_4 are scalar weighting functions applied to the outputs of the detector elements $a_1 \ldots a_4$

The effects of varying the parameters of a four-bin slope-detection system are shown graphically in figures 5.12 through 5.14. The results are summarized in table 5.3. The influence of pixel size on the dynamic range and linearity of the transfer function is shown in figure 5.12. Small pixels (p = 0.5) give good linearity over a range of about ±1.5 waves per subaperture. The dynamic range may be increased by using larger pixels, but at the expense of linearity. For closed-loop adaptive optics systems operating near the null, a pixel size of about 0.67 times the spot width is a good compromise, giving a dynamic range of about ±2 waves per subaperture. These results are obtained with weighting factors (w_1, w_2, w_3, w_4), equal to -3, -1, +1, +3.

The effect of varying the weighting factors is shown in figure 5.13. The pixel size in all cases is 0.67 times the spot width. For comparison, the first case shows the equivalent of a bi-cell, using weights of – 1, – 1, +1, +1. It is seen that increasing the weight of the outer elements increases both the output signal and the dynamic range. There is some degradation in linearity at large displacements, but all the responses have the same normalized slope in the region of the null, the normal operating point for closed-loop adaptive optics systems. It appears that weighting factors of $-3, -1, +1, +3$, represent a good compromise between dynamic range and linearity.

Figure 5.14 shows the effect of interference or "crosstalk" from adjacent channels on the response of a Shack-Hartmann sensor. The conditions are p = 0.67, with weighting factors of - 3 , - 1 , +1, +3. Curves are shown for an interfering spot located 2 pixels and 3 pixels from the center of the measured channel. Crosstalk displaces the null of the sensor by as much as 1 wave, which seriously compromises the performance. This is most likely to happen during initial convergence, before the wavefront errors have been compensated. Crosstalk is minimal when the adjacent spot is 3 pixels or more from the center of the measured channel, indicating that a guard band of 1 pixel between each 4×4 subaperture array should give sufficient protection.

5.3.2 Measurement Errors

There are two types of wavefront measurement error in a Shack-Hartmann sensor:

1. Random errors in determining the positions of the spots; these errors, caused by photon noise and electrical noise in the detectors, are unavoidable and determine the precision with which the wavefront can be measured.
2. Bias errors due to misalignment of the optics and variations in the responsivity of the detectors; these errors, which determine the accuracy of the measurements, can be reduced to acceptable levels by careful design and the use of calibration procedures.

The displacement signals from a quadrant detector are obtained in the form of two bi-cell measurements. If the outputs of the four quadrants are *a, b, c,* and *d,* as shown in figure 5.10, then

$$X \text{ displacement} = \frac{a - b - c + d}{a + b + c + d}$$
$$Y \text{ displacement} = \frac{a + b - c - d}{a + b + c + d} \qquad (5.12)$$

The random measurement errors depend on the size of the spots, which are images of the reference source produced by the subaperture lenslets. The spot size is determined by four factors: the subaperture

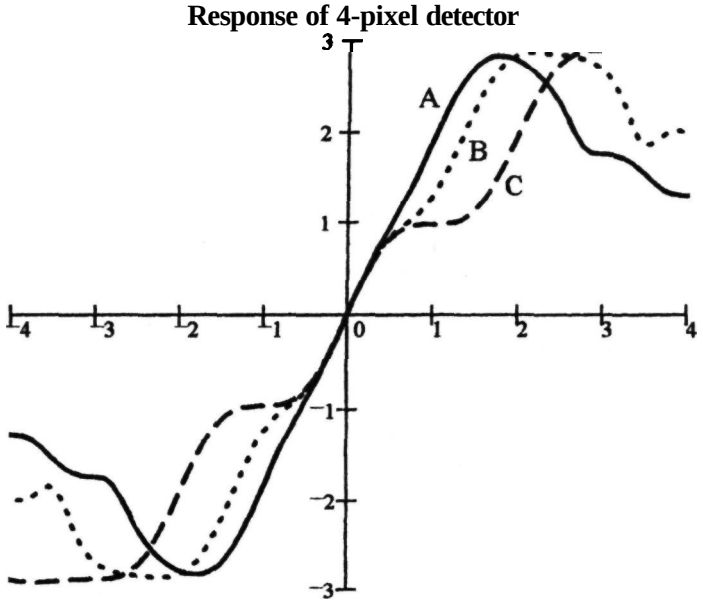


Figure 5.12 Effect of varying the pixel size on the linearity and dynamic range of a Shack-Hartmann sensor using 4×4 pixel detectors.

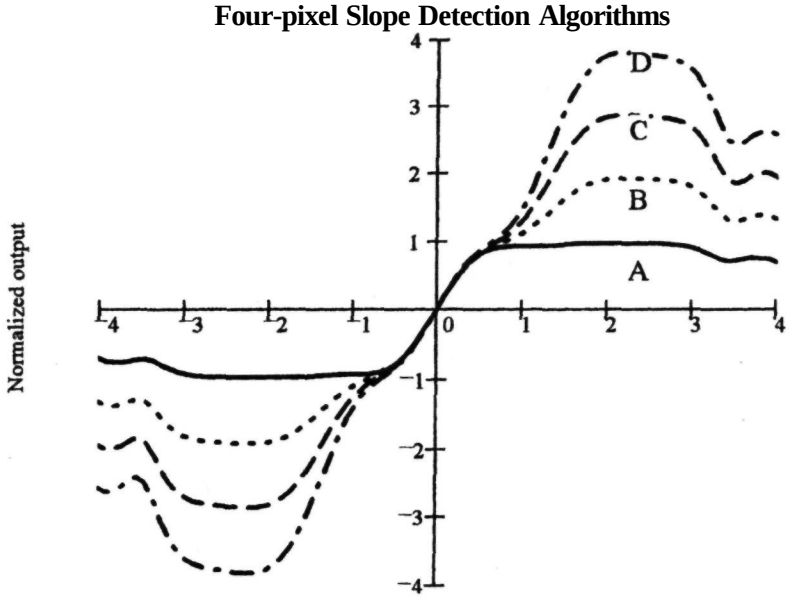


Figure 5.13 Comparison of four-pixel slope-detection algorithms for Shack-Hartmann wavefront sensors. The pixel size in all cases is 0.67 times the spot width. The number refers to the weighting of each of the four pixels. The first case, with uniform weighting, is identical to the bicell detector. Increasing the weighting of the outer pixels extends the dynamic range, at the expense of degraded linearity. The -3, −1, +1, +3 algorithm is a good compromise.

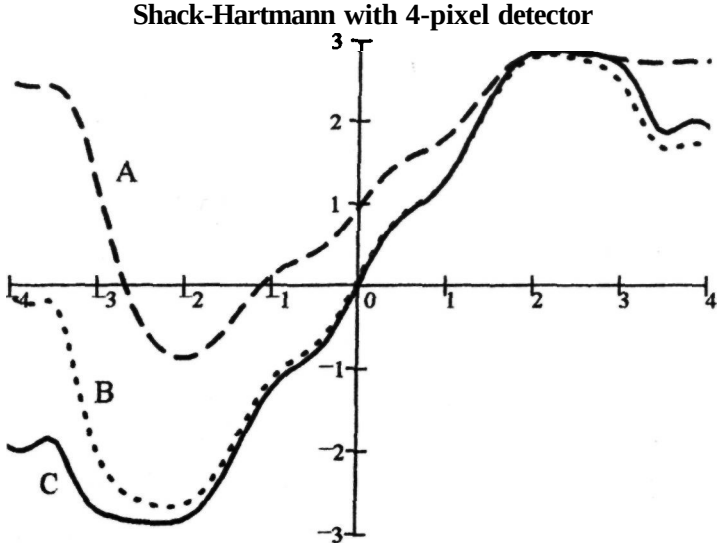


Figure 5.14 Effect of crosstalk on the response of a Shack-Hartmann sensor. Each channel has a 4 x 4 detector array using the —3, −1, +1, +3 algorithm. The pixel size is 0.67 times the spot width. Crosstalk is negligible when the adjacent spot is 3 or more pixels from the center of the measured channel, indicating that a guard band of 1 pixel between the 4 x 4 element detector arrays should be sufficient. Note that the effect of crosstalk is to displace the null of the sensor, in this case by as much as 1 wave; this seriously compromises the performance.

dimension *d*, the angular diameter of the source *9*, the turbulence strength r_0, and the sensor wavelength *X*.

It is assumed initially that the reference is a point source and that the wavefront is essentially flat over each subaperture; that is $r_0 > d$. For these conditions, the angular size of the spot is determined by the subaperture diameter and is given by λ/d. It is shown in appendix A that the standard deviation of the angular position error caused by random noise in one axis, when using a quadrant detector with a point source, is

$$\sigma_l = \frac{3\pi}{16}\frac{\lambda}{d}\frac{1}{SNR} \qquad (5.13)$$

where *SNR* is the signal-to-noise voltage ratio of the detected signal (all quadrants). The same result was obtained by Tyler and Fried [1982], using a different approach.

Table 5.3 Effect of Pixel Size on Dynamic Range and Linearity of Shack-Hartmann Wavefront Slope Sensors

Detector Size per Subaperture	Ratio of Pixel Size to Spot Size, *p*	Useful tilt range, ± waves	Departure from Linearity, waves rms
2 x 2	1.0-1.5	0.5	0.024
2 x 2		1.0[a]	0.13[a]
4 x 4	0.5	1.5	0.019
4 x 4	0.67	2.0	0.085
4 x 4	1.0	2.5	0.19

[a]Nonlinear response.

In the case of a large reference source, where $9 \gg \lambda/d$, the angular position error is determined mainly by the source size, and the single-axis error in this case is

$$\sigma_2 = \frac{\pi}{8}\frac{\theta}{SNR} \tag{5.14}$$

These equations may be combined to give a general expression for the angular position measurement error for a plane tilted wavefront:

$$\sigma_\alpha = \frac{\pi}{8\,SNR}\left[\left(\frac{3\lambda}{2d}\right)^2 + \theta^2\right]^{1/2} \tag{5.15}$$

When the turbulence parameter r_0 is smaller than d, the image size for an unresolved source is determined by the atmospheric r_0 rather than by the subaperture size, d. This is accommodated in equation (5.15) by replacing d with r_0.

The measurement error of practical sensors will be somewhat greater than that indicated by equation (5.15); this is because of the presence of spaces between the segments of the quadrant detector. The effect of segment gaps on the wavefront error of Shack-Hartmann sensors is evaluated in appendix A, section A.5. The additional error produced is a minimum at the null and increases as the spot is displaced, as shown in figure A.5. It is modeled in the error equations by a factor K_g, representing the increase in error at the null, with typical values between 1.2 and 1.5.

To evaluate the performance of adaptive optics systems, it is convenient to express the wavefront sensor error in terms of the phase difference over the subaperture. To convert the angular error to phase difference in radians, the angle must be multiplied by $2\pi d/\lambda$. The resulting expressions for the standard deviation of the one-axis measurement error in Shack-Hartmann sensors, in radians rms of phase difference per subaperture, are then

$$\sigma_\phi = \frac{\pi^2 K_g}{4(SNR)}\left[\left(\frac{3}{2}\right)^2 + \left(\frac{\theta d}{\lambda}\right)^2\right]^{1/2} \qquad r_0 > d$$

$$= \frac{\pi^2 K_g}{4(SNR)}\left[\left(\frac{3d}{2r_0}\right)^2 + \left(\frac{\theta d}{\lambda}\right)^2\right]^{1/2} \qquad r_0 < d \tag{5.16}$$

For random perturbations, the total (two-axis) error is $\sqrt{2}$ times this value.

The signal-to-noise ratio of the detector is determined by the photon (shot) noise associated with the detection of discrete events, together with any background or electrical noise added before the events are counted. In the case of a CCD array, the major additional source of noise consists of the electrons added to the charge of each pixel during transfer to the on-chip amplifier. If an intensifier is used before the detector, this noise is diluted by the gain of the intensifier. The signal-to-noise ratio for a generic detector is given by

$$SNR = \frac{n_p}{\left\{n_P + N_D\left[n_B^2 + (e/G)^2\right]\right\}^{1/2}} \tag{5.17}$$

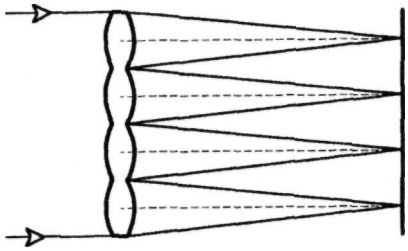
where

n_P = number of detected photoelectrons per subaperture (sum of all pixels)
N_D = number of detector pixels per subaperture
n_B = number of detected background electrons per pixel
e = read noise in electrons per pixel
G = intensifier gain (= 1 for a straight CCD)

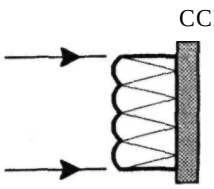
There are two approaches to obtaining a high detector signal-to-noise ratio. The first is to use a photomultiplier or image intensifier with a large gain G, which may be 1000 or more. The detector noise is then negligible in comparison with the photon noise, resulting in a signal-to-noise ratio proportional to $\sqrt{n_p}$. This assumes that the gain itself has no variance; when using a photomultiplier, it is equivalent to assuming that the pulse height distribution is very narrow, which is approximately true with pulse-counting detectors. The second approach is to use a nonintensified detector (G — 1) with high quantum efficiency and minimum read noise. At low photon counts and high readout rates, such detectors are usually limited by read noise, in which case the signal-to-noise ratio becomes approximately equal to $n_p/\sqrt{(N_D e^2)}$. Both types of detector have been used successfully in wavefront sensors. Characteristics of detectors for wavefront sensing are described in section 5.7.

5.3.3 Aperture Division

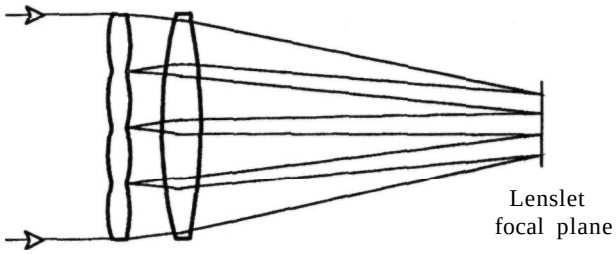
There are several different methods of subdividing the aperture in Shack-Hartmann sensors. For example, the aperture may be divided by an array of plates or prisms that diverge the light in each subaperture away from the optical axis; a single lens then focuses all subapertures at the image plane, where they form an array of spots. It is essential for the aperture division components to be optically stable and immune from temperature variations. To engineer even a modest-sized device with 16×16 elements, using discrete components, is a formidable and expensive task. Consequently, most successful Shack-Hartmann sensors employ simple lens arrays, as shown in figure 5.15(a), which are fabricated as monolithic devices. The element sizes are typically 0.4-1 mm^2, with focal lengths of 70-100 mm. Using this method of construction, there are only two basic components, the lens array and the detector, for which the mechanical tolerances for alignment and stability are at the micrometer level. The success of the Shack-Hartmann sensor is largely due to the availability of stable, inexpensive, monolithic arrays for these two critical components.



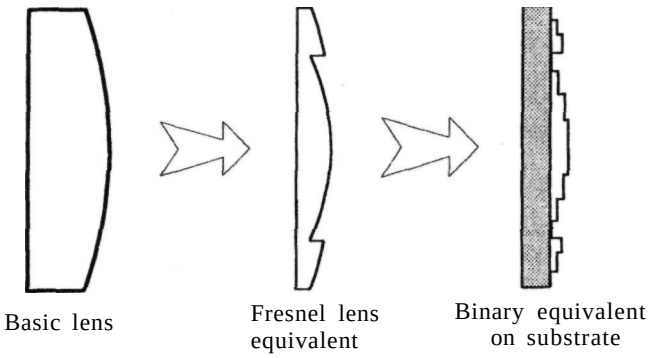
(a) Basic lenslet array



(b) Microlens array integrated with CCD chip



(c) Lenslet array with **demagnification**.



(d) Binary optics

Figure 5.15 Lens arrays for **Shack-Hartmann** sensors.

The microlithographic and plasma etching techniques developed for integrated circuit fabrication can also be used for producing microlens arrays. The focal length of the lenslets is proportional to the square of the pixel size, leading to some very tiny devices when the pixels are only $25\mu m^2$. Roland et al. [1994] describe a microminiature Shack-Hartmann sensor with microlenses only $69\,\mu m$ in diameter, etched into the cover plate of the CCD array, as shown in figure 5.15(b).

When using a lenticular array, it is not necessary for the lens spacing to match the pixel spacing of the detector. A reimaging lens may be used, following the lenticular array to reduce the spacing and size of the Hartmann spots to values appropriate for a small CCD detector array, as indicated in figure 5.15 (c).

The first microlens arrays were made of plastic, using a replication process from a master mold formed by stamping with a steel ball. More recently, Artzner [1992] has described microlithographic techniques for making microlens arrays in photoresist coatings on glass substrates. Binary optics technology has been used at Massachusetts Institute of Technology (MIT), Lincoln Laboratory to fabricate 16×16 lens arrays for Shack-Hartmann sensors [Barclay et al. 1992]. Binary optics employs plasma etching methods to approximate a Fresnel lens using a relatively small number of optical steps, as shown

in figure 5.15(d). Plasma etching was originally developed for Very Large Scale Integration (VLSI) integrated circuit fabrication [Veldkamp and Swanson 1983]. The maximum optical path difference required for a Fresnel lens is 1 wavelength of light. This was approximated with five binary masks, producing $2^5 = 32$ steps, yielding a theoretical diffraction efficiency of 99.7%. The individual lens elements were 0.4 x 0.4 mm square, having a 100% fill factor, with focal lengths of 72.5 mm. These devices were made to operate with a laser beacon operating at 512 nm.

Binary optics enables the fabrication of complex phase masks that could not be generated by conventional techniques [Neal et al. 1995]. This capability may be used to make multitiered Hartmann masks that contain several superimposed focusing elements covering the aperture at different scales. Neal et al. [1994] describe a hierarchical wavefront sensor of this type, for use in adaptive optics, using up to four tiers. The smallest elements function in the same way as a conventional Hartmann sensor, providing the slope information for individual subapertures or segments. Larger elements cover 2×2 subarrays, producing separate spots that provide the information needed to reconstruct the wavefronts over the smaller elements. The largest element, covering the whole aperture, provides the lowest order information, the overall tilt data. Typical parameters for Shack-Hartmann sensors using CCD detector arrays at visible wavelengths are listed in table 5.4.

5.3.4 Photon Sensing

Charge-coupled device imaging arrays are the ideal complement to microlens arrays in Shack-Hartmann sensors, as they are solid-state devices that are fabricated on a similar physical scale. Silicon CCDs have a useful spectral range between about 400 and 900 nm, and some achieve a quantum efficiency of over 90% at wavelengths of 600–700 nm. The devices used in wavefront sensors have different characteristics from those employed in imaging. The main differences are a relatively small number of pixels (64 x 64 is a commonly used size), a high readout rate of several thousand frames per second, and extremely low readout noise. Dark current is not important in CCDs used for wavefront sensing because of the very short exposure times. To achieve high pixel readout rates, multiple (parallel) output ports are employed. By using an on-chip summing process called "binning," in which several pixels can be added together before readout without incurring additional noise, spatial resolution can be traded-off to increase the frame rate. A CCD camera developed at MIT Lincoln Laboratory for use in wavefront sensors can operate at a readout rate of 7000 frames per second with 16 x 64 resolution and readout noise of only 10 electrons per pixel. Characteristics of photon detectors currently being developed for wavefront sensing are summarized in section 5.7

5.3.5 Sensor Calibration

An essential element of all adaptive optics systems is a built-in optical calibration capability to ensure that the wavefront sensor and corrector maintain proper alignment. Calibration is particularly important with Shack-Hartmann sensors because of their extreme sensitivity to physical misalignment. There are two basic requirements: (1) calibration of the null point of each sensor subaperture with a plane-wave input; (2) alignment of each sensor subaperture to the corresponding actuators of the wavefront corrector. With the millimeter-sized subapertures now used in Shack-Hartmann sensors, alignment accuracies on the order of 1 μm are required. The alignment system should be designed as an integral part of the sensor hardware and software, and automated as far as practicable.

The key element of the alignment system is a built-in reference source that takes the place of the external reference. It is injected into the optical train ahead of

Table 5.4 Shack-Hartmann Sensor Characteristics

	SWAT	Visible Band		IR
Characteristic	Sensor	Sensors		Sensor
Pixel size, μm	27	25	25	25
Wavelength, μm	0.5	0.5	0.75	2.0
Pixels per spot size	1.8	2	1	2
Tilt range, waves per subaperture	1.1	1	1	1
Pixels per subaperture	4 x 4	4 x 4	2 x 2	4 x 4
Lenslet f-number	181[b]	50	16.7	12.5
Lenslet dimension, μm	400	100	50	100
Lenslet focal length, mm	72.5*	5	0.83	1.25

[a]SWAT, Short Wavelength Adaptive Techniques.
[b]Optically reduced by a factor of 3.7.

the deformable mirror, as shown in figure 5.16. The built-in source provides a plane wavefront, enabling the null position of each Hartmann spot on the detector array to be established. Within the sensor itself, the source allows radiometric calibration of the detectors. Dark frame calibration is included in the case of CCDs. Using the adaptive optics tip-tilt mirror, the tilt transfer function of each subaperture may also be determined. These data may be saved and used to make individual channel corrections in the electronic processor.

Two levels of calibration are generally required in a Shack-Hartmann sensor. Small random offsets in the nulls of each subaperture, caused by mechanical or optical irregularities, may be compensated numerically during the wavefront computation. Larger systematic errors appearing as defocus, lateral displacements, or rotation, must be removed by physical adjustment of the detector if they impose a restriction on the dynamic range of the sensor.

The relative alignment of the wavefront sensor and corrector is usually verified by "poking" individual actuators and measuring the response at the sensor. The presence of a physical misalignment, such as translation or rotation between the components, is shown by a systematic pattern of offsets. The provision of automated test and diagnostic capabilities to carry out such tests ensures that the performance of the sensor is maintained at a high level.

5.3.6 Wavefront Slope Computation

The output from a wavefront sensor detector consists of a serial stream of digitized pixel values that is sent to the wavefront processor, where it is converted into *x*- and *y*-axis slope data. In the case of a 64 x 64 CCD array operating at a frame rate of 2 kHz, the data rate is between 8 and 12 megabytes per second. The wavefront processor performs three basic operations to convert the pixel data into wavefront slope measurements that are sent to the wavefront reconstructor:

1. radiometric calibration,
2. raw slope computation,
3. wavefront slope calibration.

These operations are shown in figure 5.17. Wavefront processors of this type usually employ the "pipeline" architecture that involves a minimum of buffer storage.

The radiometric calibration consists of removing the fixed pattern noise on the detector by subtraction of a stored "dark frame," followed by normalization of the quantum efficiency of each pixel by means of a gain correction. These are standard operations on detector arrays such as CCDs, but in this case they must be performed at a high data rate.

Calculation of the centroid for each subaperture is performed using equation (5.12) for a quadrant detector. A similar algorithm, equation (5.11), is used with larger detector arrays, but in this case there is more flexibility in weighting the individual pixel outputs. The total light received in each subaperture is computed by summing the outputs from all pixels in order to normalize the raw slope measurements, and also to provide a 'low-light flag' that provides warning when the photon count drops below a predetermined level. This flag may be used to set the slope output of any subaperture to zero if the signal-to-noise ratio becomes unacceptable.

The final step in processing the wavefront is the slope calibration, in which the gain or scale factor of the slope outputs is adjusted, and offsets are introduced to account for fixed aberrations in the optical system that are not part of the common path shared with the imaging sensor. The scale factor controls the loop gain of the servo-system driving the wavefront corrector and may, for example, be adjusted to optimize performance with different turbulence conditions. Fixed offsets in the gradients may be required

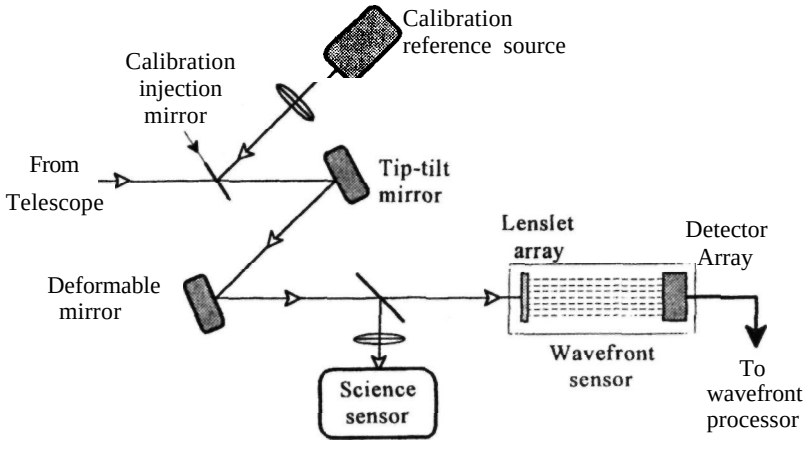


Figure 5.16 Adaptive optical system using a Shack-Hartman wavefront sensor. To calibrate the system, a local reference source is injected into the optical path by means of a retractable mirror. Two calibration procedures are required: (1) optical alignment of the wavefront sensor lenslet array to match the subapertures on the deformable mirror; (2) photometric calibration of each pixel of the detector array in the Shack-Hartmann sensor.

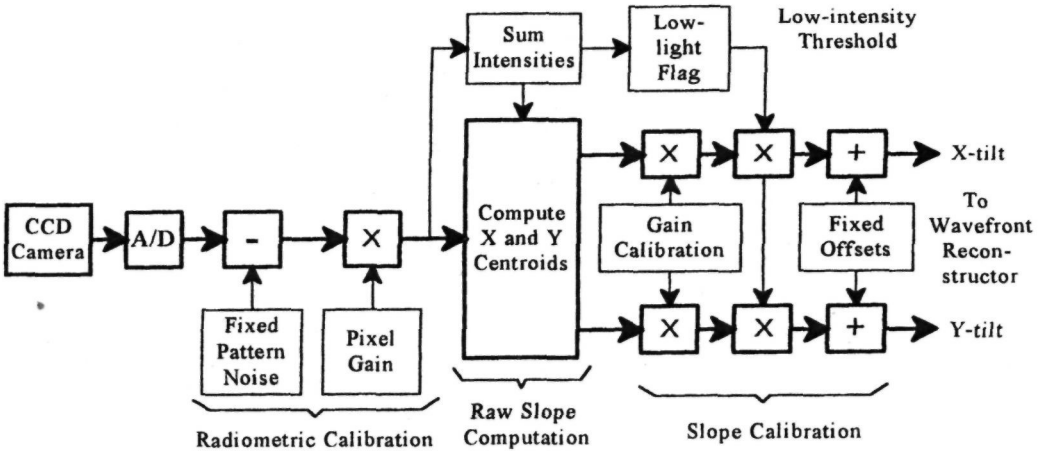


Figure 5.1 7 Block diagram of Shack-Hartmann wavefront slope processor. The main signal path is shown with heavy lines. A/D, analog to digital convertor.

to account for any differences in the optical paths to the wavefront sensor and imaging sensor, following the beam splitter, as shown in figure 5.16. These data-processing operations have been shown as examples and may not all be required in any particular adaptive optics system.

5.4 Shearing Interferometers

5.4.1 Principle of Operation

Shearing interferometry is a technique for measuring phase differences in optical wavefronts. The basic principle is to combine the original wavefront with a displaced or "sheared" replica, so that the optical waves interfere. In this way, optical phase differences are converted into intensity variations that can be sensed with a photoelectric detector. Shearing interferometry eliminates the need for the coherent plane-wave reference source that is required by conventional interferometers; with shearing interferometers, each point on the wavefront interferes with a copy of itself. In fact, common-path shearing interferometers will work with light sources that are temporally incoherent and of finite size. When the shear distance is small compared with the scale of the disturbances in the wavefront, an achromatic shearing interferometer effectively measures the gradient of the wavefront, independent of the wavelength of the light.

Wavefronts may be sheared in several different ways. Most commonly used is lateral shear, in which the wavefront is displaced linearly and at the same scale, by a distance s. The principle of a lateral shear interferometer is illustrated in figure 5.18. The wavefront to be measured is described by $W(x, y)$, with amplitude distribution $A(x,y)$. The complex amplitude is then

$$U(x, y) = A(x, y)e^{ikW(x,y)} \tag{5.18}$$

Two copies of the wavefront are generated, displaced by the shear distance s in the x axis:

$$\begin{aligned} U_1(x, y) &= A(x + s/2, y)e^{ik\,W(x+s/2,y)} \\ U_2(x, y) &= A(x - s/2, y)e^{ik\,W(x-s/2,y)} \end{aligned} \tag{5.19}$$

The intensity produced by the interference between these wavefronts is

$$\begin{aligned} I(x, y) = \; & 2\{1 + |A(x + s/2, y)||A(x - s/2, y)| \\ & \times \cos[kW(x + s/2, y) - kW(x - s/2, y)]\} \end{aligned} \tag{5.20}$$

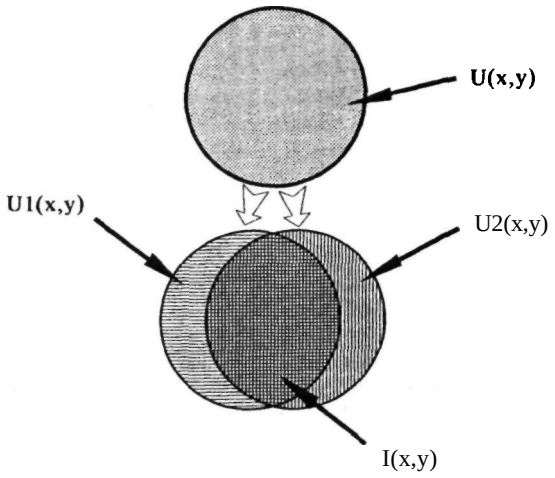


Figure 5.18 Principle of lateral shearing interferometry.

When the shear distance is small compared with the spatial period of $W(x,y)$, this can be expressed as

$$I(x, y) = 2\{1 + |A(x, y)|^2 \cos[k\, s\, dW(x, y)/dx]\} \quad (5.21)$$

The intensity of the interference pattern is proportional to the cosine of the wavefront slope multiplied by the shear distance. The sensitivity of a shearing interferometer may therefore be varied by adjusting the shear distance. By measuring the intensity of the interference pattern and knowing the shear distance, the wavefront slope in the shear direction at any point *(x,y)* may be determined. Wavefront slope is a vector quantity, so it is necessary to make two lateral shear measurements, in the *x* and *y* directions, in order to characterize a two-dimensional wavefront.

The simplest lateral shear interferometer consists of a tilted, plane parallel plate. Light reflected from the front and back surfaces is displaced (sheared) by a distance proportional to the tilt angle of the plate. In this case, the shear distance is approximately constant with wavelength (small variations can occur because of dispersion in the plate), so that for a given wavefront slope the optical phase angle between the interfering wavefronts varies as $k - 2\pi/\lambda$. This means that an interferometer with fixed shear is not achromatic. To achieve achromaticity, the shear must vary linearly with wavelength. One method of implementing this relationship is to use a diffraction grating to generate the shear, as discussed in section 5.4.2.

Radial shear is obtained by magnifying or demagnifying the replica and combining it with the original. Radial shear is useful for measuring rotationally symmetric aberrations, such as defocus and spherical aberration [Peters et al., 1974]. It is also employed in an extreme form in the point diffraction interferometer, in which a small section of wavefront, usually at the center, is magnified up to the full aperture and interfered with the original wavefront. Phase variations over the magnified wavefront are very small, causing the wavefront to act effectively as a constant phase reference, producing an interferogram referenced to the center of the aperture. Unfortunately, this technique is optically inefficient, and for that reason it is unsuitable for astronomical wavefront sensors.

Rotational shear is orthogonal to radial shear, and measures the tangential component of wavefront slope. An interferometer with 180° of rotational shear contains all possible shear distances within the aperture, from zero in the center to the full diameter at the edges. This feature has been used to measure the Fourier components of radiant objects [Roddier 1979].

Lateral shearing interferometers using diffraction gratings to shear the wavefront have achromatic characteristics and are therefore well suited for wavefront sensing in adaptive optics systems. They have high optical efficiency and will operate with broadband, extended sources. The theory and practice of this type will be considered in some detail in the following sections. Even within this category, there are numerous variations, including devices with temporal modulation, spatial modulation, and variable shear. Shearing interferometers with multiplexed outputs have recently been developed to reduce the number of detector elements required.

5.4.2 Alternating Current Shearing Interferometer

A shearing interferometer using a moving grating to provide temporal modulation is depicted in figure 5.19. A wavefront function $W(x,y)$ is located at the system pupil, which is defined by the pupil function $A(x,y)$. The complex amplitude following the pupil is

$$U(x, y) = A(x, y)e^{ik\,W(x,y)} \quad (5.22)$$

Lens L1 generates the Fourier transform of $U(x,y)$ at its focal plane, where it falls on a grating with amplitude transmittance $M(x_0,t)$. The amplitude distribution after the grating is then

$$U(x_0, y_0, t) = \tilde{U}(x_0, y_0)M(x_0, t) \quad (5.23)$$

where $\tilde{U}(x_0, y_0)$ is the Fourier transform of $U(x,y)$. The second lens L2 produces an image of the pupil

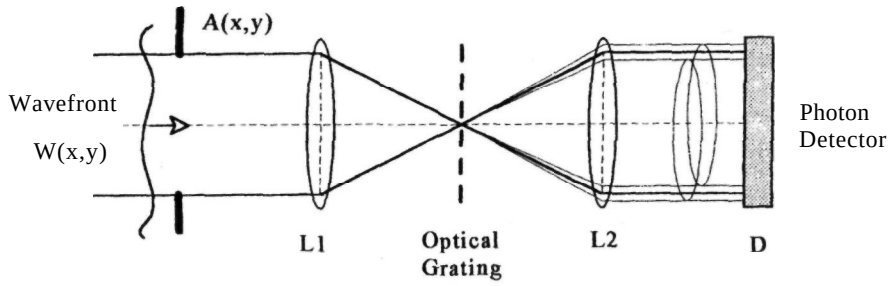


Figure 5.19 Shearing interferometer using a moving optical grating to generate temporal modulation.

containing the wavefront disturbance at the detector plane *D*. The intensity at the detector plane is

$$I\{x,y,t) = |U'(x, y, t)|^2 \tag{5.24}$$

where $U'\{x, y, t)$ is the convolution of $U(x_0, y_0)$ with the Fourier transform of $M(x_0,t)$.

In the general case, the grating function can be expressed as

$$M(x_0, t) = \sum_{n=-\infty}^{\infty} F_n e^{j2n\pi(x_0-vt)/g} \tag{5.25}$$

where

- g = grating period in the x direction
- *v* = grating velocity in the direction normal to the lines
- $F_n s$ = Fourier coefficients of the periodic grating amplitude transmittance.

Interference occurs between each pair of diffracted orders n and *m*, which have a fixed phase difference α_{nm}. The shear distance between adjacent orders at the detector plane is $s = kZ/g$ and the fundamental temporal modulation frequency produced by grating motion is $\omega = 2\pi v/g$.

The intensity at the detector plane may then be written as

$$I(x, y, t) = \sum_{n=-\infty}^{\infty} |F_n|^2 + \sum_{\substack{n=-\infty \\ n\neq m}}^{\infty} |F_n||F_m| \times \cos\{k[W(x - ns, y) - W(x - ms, y)] + (m - n)\omega t + \alpha_{nm}\} \tag{5.26}$$

In this general case, the intensity distribution at the detector plane is composed of a number of overlapping shear patterns, each modulated at a multiple $(m - n)$ of the modulation frequency ω. The wavefront data appear as temporal phase shifts proportional to the wavefront pathlength differences over the effective shear distances, $(m - n)s$. For small displacements, this temporal phase shift can be regarded as proportional to the wavefront slope.

The desired shear pattern may be extracted by spectral filtering of the temporally modulated detector outputs. But, in the general case, many different shear patterns will be modulated at the same frequency. To ensure that the desired shear pattern is represented by a unique frequency, it is necessary to use specific grating configurations that restrict or control the diffraction orders *n* and *m*. Two useful types of grating are the sine-wave amplitude grating and the Ronchi grating.

The sine-wave amplitude grating produces a transmittance given by

$$M(x_0, t) = \sin[2\pi(x_0 - vt)/g] \tag{5.27}$$

The resulting intensity at the detector plane is then

$$I(x, y, t) = \tfrac{1}{2} + \tfrac{1}{2}\cos\{k[W(x - s, y) - W(x + s, y)] + 2\omega t\} \tag{5.28}$$

In this case, there is only one shear pattern, produced by the interference of the +1 and —1 diffraction orders from the grating, with a shear value of *2s*. The phase difference between the interfering beams, which appears as the phase shift at the modulation frequency 2ω, is

$$\phi(x, y) = k[W(x - s, y) - W(x + s, y)] \tag{5.29}$$

Measurement of the electrical phase ϕ at detector location *x, y* therefore gives the wavefront difference between points $x + s$ and $x - s$. For small values of *s*, this is equivalent to the wavefront slope.

Sine-wave amplitude gratings contain both intensity and phase components, with a phase reversal each half cycle. As a result, they are difficult to make. An alternative and more easily manufactured type is the Ronchi grating, which has a square-wave transmission function:

$$M(x_0, t) = \tfrac{1}{2}\sum_{n=-\infty}^{\infty} \frac{2}{n\pi}\sin\left[\frac{n\pi}{2}\right] e^{j2n\pi(x_0-vt)/g} \tag{5.30}$$

The diffraction orders produced by this type of grating are shown in figure 5.20. Even orders are suppressed, leaving only the zero-order and the odd-order terms. The fundamental modulation frequency ω can be produced only by interference between the zero-order and the first-order terms. This frequency is selected by temporal filtering. The remaining interference products at frequencies of *3a), 5*ω, and so on, are rejected.

Koliopoulos [1980] has shown that the signal produced by a moving Ronchi grating can be written as

$$S(x, y, t) = \tfrac{1}{2} + \tfrac{2}{\pi}K_d K_i K_\theta \cos\left[\omega t + k s \frac{\partial}{\partial x} W(x, y)\right] \tag{5.31}$$

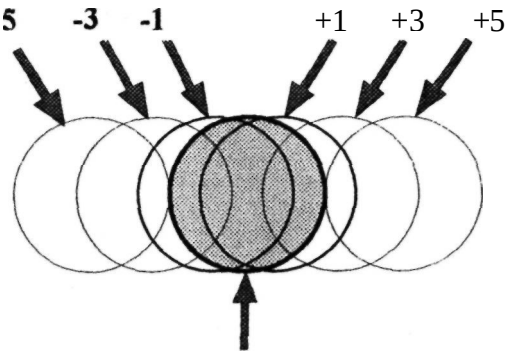


Figure 5.20 Diffraction orders produced by a Ronchi grating.

The detector output signal at frequency ω has a phase shift $ks\ \partial W(x, y)/\partial(x)$, proportional to the product of the shear distance s and the wavefront slope. Measurement of the electrical phase of this signal therefore gives the wavefront slope. The modulation of the detected signal, ideally unity, is reduced by the contrast reduction terms, K_d, K_λ, and K_θ. These terms do not directly produce errors in the wavefront measurement, but they reduce the signal-to-noise ratio of the wavefront sensor output, which may increase random errors at low photon counts.

The first contrast reduction term is due to the fact that a Ronchi grating generates two first-order shear components, produced by interference between the 0 and +1 orders and the 0 and −1 orders, respectively. These shear components are displaced by the shear distance, resulting in a phase difference that is proportional to the second derivative of the wavefront. Then,

$$K_d = \cos\left[\frac{k s^2}{2}\frac{\partial^2}{\partial x^2} W(x, y)\right] \tag{5.32}$$

With severe wavefront disturbances, the second derivative of the wavefront may become large, so that this term departs significantly from unity. In such cases, the shear can be adjusted to optimize the modulation. The displacement between the sheared interferograms may produce small errors in the measured slope when high-order wavefront distortions are present, but these errors are negligible in most practical cases.

The second contrast reduction term is due to the spectral bandwidth, if high-order distortions are present within the measurement subaperture. Although a lateral shearing interferometer using a diffraction grating to produce the shear is achromatic in making wavefront slope measurements, the presence of third- and higher order wavefront distortions produces local differences in the optical pathlengths between the interfering beams, reducing the fringe contrast. For a source of spectral bandwidth $\Delta\lambda$ at mean wavelength λ, the modulation is reduced by the factor

$$K_\lambda = \mathrm{sinc}\left[\frac{\pi}{2}\frac{\Delta\lambda s^2}{2\lambda^2}\frac{\partial^2}{\partial x^2} W(x, y)\right] \tag{5.33}$$

The third contrast reduction term takes the reference source characteristics into account. In a shearing interferometer, the signal modulation depends on the degree of coherence between points on the wavefront separated by the shear distance. The coherence function is given by the Fourier transform of the intensity distribution of the source. For a uniformly illuminated source of angular size θ in the direction of shear, the contrast reduction factor is

Circular source

$$K_\theta = \frac{2J_1(\pi s\theta/\lambda)}{\pi s\theta/\lambda} \tag{5.34a}$$

Rectangular source

$$K'_\theta = \frac{\sin(\pi s\theta/\lambda}{\pi s\theta/\lambda} \tag{5.34b}$$

The factor K_θ is maximized when the source size is zero (a point source); it falls to zero when the source size is about equal to the grating period and has local maxima for higher values of shear. When using extended objects as reference sources, it is necessary to adjust the shear (grating period) to maximize the product of shear and K_θ, as will be explained in the following section. The condition for zero contrast can be used to suppress the modulation caused by an interfering background such as the sky or a bright surface. In this case, a field stop having a width exactly equal to the grating period is placed close to the grating. This technique enables a shearing interferometer to make wavefront measurements using a bright star in full daylight, or using solar pores on the surface of the Sun to compensate solar observations.

5.4.3 Measurement Error

The output signal from an AC shearing interferometer has the form

$$I = I_{av}[1 + K\sin(\omega t + \phi)] \tag{5.35}$$

where

$K = K_d K_\lambda K_\theta$ - overall modulation or fringe contrast
ω = modulation frequency
ϕ = electrical phase shift, proportional to the wavefront slope being measured.

The output is generally contaminated with shot noise due to the detection of discrete photoelectron events and random noise in the signal amplifiers. The output signal is demodulated with respect to a fixed phase reference at frequency ω. The added noise causes perturbations in the electrical phase that translate into wavefront measurement errors. The magnitude of these errors will now be determined.

From basic considerations, it can be deduced that the error in determining the phase ϕ is proportional to $1/K(SNR)$, where SNR is the signal-to-noise ratio. In an AC shearing interferometer, the phase of the modulation signal represents the average wavefront slope over the shear distance s. The measurement error over one subaperture of dimension d is therefore

$$\Delta\phi = \frac{A}{K(SNR)}\frac{d}{s} \tag{5.36}$$

The value of the parameter A is ideally 1; its actual value will be greater than 1, depending on the method used to extract the phase information from the signal. One method of phase determination is to correlate the photon events with sine and cosine reference signals at frequency ω. The correlated values of the sine and cosine are accumulated over the integration time

T, and the phase angle is computed from the inverse tangent. If the arrival time of the jth photon event is t_j, the sine and cosine sums can be expressed as

$$S(\phi) = \sum_0^T \sin \omega t_j$$
$$C(\phi) = \sum_0^T \cos \omega t_j \tag{5.37}$$

The phase angle is then

$$\phi = \tan^{-1} \frac{S(\phi)}{C(\phi)} \tag{5.38}$$

Using this detection scheme, the wavefront measurement error is

$$\sigma_\phi = \frac{\sqrt{2}}{K(SNR)} \frac{d}{s} \quad \text{radians of phase, rms} \tag{5.39}$$

To implement this scheme, the reference signals sin ωt and cos ωt must be sampled at the exact time at which each photon is detected, to obtain the correct sine and cosine sums. Errors in establishing the arrival times result in phase estimation errors. As the arrival times are random, the detectors must be sampled at a high rate so that time errors are much smaller than the integration time T. The sampling problem may be eliminated by using a parallel detector array in which the output of each element is continuously monitored. An array of photomultiplier tubes has been employed for this purpose in the radial grating interferometer described in section 5.4.5.

An alternative scheme for measuring the phase of the modulation frequency, which is proportional to the slope of the wavefront, is to divide the period of the sine wave into three or more sectors or "bins" and to accumulate the detected events in each bin. In this case, the detector array needs to be sampled only three or four times per cycle of the modulation frequency. A four-bin algorithm of this type, described by Wyant [1975], is a robust method of phase determination. The sine-wave period t_S is divided into four bins, B_1, B_2, B_3 and B_4, in each of which the photon counts $\delta(t_j)$ are accumulated during the integration time:

$$B_1 = \sum_{-t_s/8}^{+t_s/8} \delta(t_j) \qquad B_2 = \sum_{t_s/8}^{3t_s/8} \delta(t_j)$$
$$B_3 = \sum_{3t_s/8}^{5t_s/8} \delta(t_j) \qquad B_4 = \sum_{5t_s/8}^{7t_s/8} \delta(t_j) \tag{5.40}$$

The phase angle is then

$$\phi = \tan^{-1} \frac{B_1 - B_3}{B_2 - B_4} \tag{5.41}$$

Using the four-bin algorithm, the wavefront measurement error of an AC interferometer, excluding the wavefront reconstruction process, is given by

$$\sigma_\phi = \frac{\pi}{2K(SNR)} \frac{d}{s} \quad \text{radians of phase, rms} \tag{5.42}$$

Although the error is greater than with sine/cosine correlation, the difference is not significant in practice, and the four-bin phase detection algorithm is widely used for phase detection in both temporal and spatial modulation systems.

In the three-bin phase-detection algorithm, each period is divided into three 120° sectors and the phase angle is computed from the relation

$$\tan \phi = \frac{\sqrt{3}(I_2 - I_1)}{2I_0 - I_2 - I_1} \tag{5.43}$$

where

I_0 = signal from 0° bin
I_1 = signal from +120° bin
I_2 = signal from −120° bin.

5.4.4 Optimization of Shear

It is evident from the foregoing analysis that three factors must be maximized in order to optimize the performance of a shearing interferometer: (1) the signal-to-noise ratio SNR; (2) the ratio of shear to subaperture size, s/d; and (3) the modulation K. These factors are interrelated; the conditions for minimizing the wavefront measurement error will now be considered in more detail.

The signal-to-noise ratio relevant to a wavefront sensor is given by equation (5.17). In the following analysis, it is assumed that the detector noise is small compared with the photon shot noise, in which case the signal-to-noise ratio is given by the square root of the number of photons counted; that is, $SNR = \sqrt{n_p}$. For a circular extended source of uniform brightness, the photon count in a subaperture of size d x d is given by

$$n_p(N_s \lambda/hc) d^2 t q To \theta^2 \tag{5.44}$$

where

N_s = source radiance, watts m^{-2} sr^{-1}
h = Planck's constant
c = velocity of light, m s~
t = integration time, s
q = detector quantum efficiency
To = optical transmission to wavefront sensor
6 = angular subtense of reference source, rad.

The signal-to-noise ratio may then be expressed as

$$SNR = F d \theta \tag{5.45}$$

where $F = \sqrt{(N_s \lambda t q To/hc)}$.

From equation (5.39), the wavefront error of a shearing interferometer using the sine/cosine detection system is

$$\sigma_\phi = \frac{\sqrt{2}}{K(SNR)} \frac{d}{s} \qquad \text{rad rms}$$

The wavefront error is minimized by maximizing the product of modulation and shear, Ks. However, K and s are related by the characteristics of the reference source: as the shear distance is increased, the modulation produced by a source of finite size decreases. This relationship is analyzed using the Van Cittert–Zernike theorem [Born and Wolf 1975, Ch. 10], which shows that the degree of coherence between two points separated by distance s and equidistant from a quasi-monochromatic source is equal to the absolute value of the normalized Fourier transform of the source. For a uniform circular source of angular subtense θ, the coherence or modulation factor is

$$K_\theta = \frac{2J_1(\pi s\,\theta/\lambda)}{\pi s\,\theta/\lambda}$$

The effect of varying the shear distance on the shear-modulation product is shown in figure 5.21 for two different types of reference object. The wavefront error is inversely proportional to this product. The first chart represents a reference object illuminated by reflected radiation, such as a solar-illuminated satellite. In this case, the received power is proportional to the angular size of the object, and the shear can be optimized to maintain a constant wavefront measurement error irrespective of object size. Note that the curves for 0.5, 1, and 2 arc second sources all have the same peak amplitude, although the peaks occur at different shear values. In effect, the additional photon flux from a large object can be made to compensate for the drop in modulation due to its larger angular size.

The second chart covers an object such as a laser beacon, in which the received power remains constant even though the apparent size of the object may vary because of changes in the scattering cross-section, which are due to turbulence in the optical path. In this case, an increase in the angular size of the reference source inevitably reduces the shear-modulation product (increases the error), although there is still an optimum value of shear in each case. These charts show the necessity for optimizing the shear in shearing interferometers.

To find the minimum wavefront error, we substitute for SNR and K in equation (5.39) to obtain

$$\sigma_\phi = \frac{\pi}{\sqrt{2}k_s\lambda J_1(\pi s\,\theta/\lambda)} \qquad (5.46)$$

The minimum wavefront measurement errors σ_ϕ occur at the maxima of $J_1(\pi s\theta/\lambda)$; that is, when $\pi s\theta/\lambda$ = 1.84, 5.33, and so on. The smallest optimal value of shear for a circular source of angular subtense θ is therefore

$$s = 0.586\,\lambda/\theta \qquad (5.47)$$

For a rectangular source, the optimal value of shear is s = 0.500 X/θ. When a Ronchi grating is used at the focal plane to generate the shear, this optimal shear corresponds to a grating period equal to exactly twice the size of the reference image. In other words, the optimal condition is for the image of the reference source to be exactly one half cycle of the grating, which is also evident from physical reasoning.

It is of great importance to avoid the minima of the periodic functions that determine the value of K_θ, as at these points the signal modulation falls to zero and the wavefront error becomes infinitely large. With a rectangular source, this condition occurs when the image of the source is exactly one cycle of the grating. When using a field stop with large extended sources, modulation due to the field stop itself is eliminated by adjusting it to cover an integral number of grating cycles.

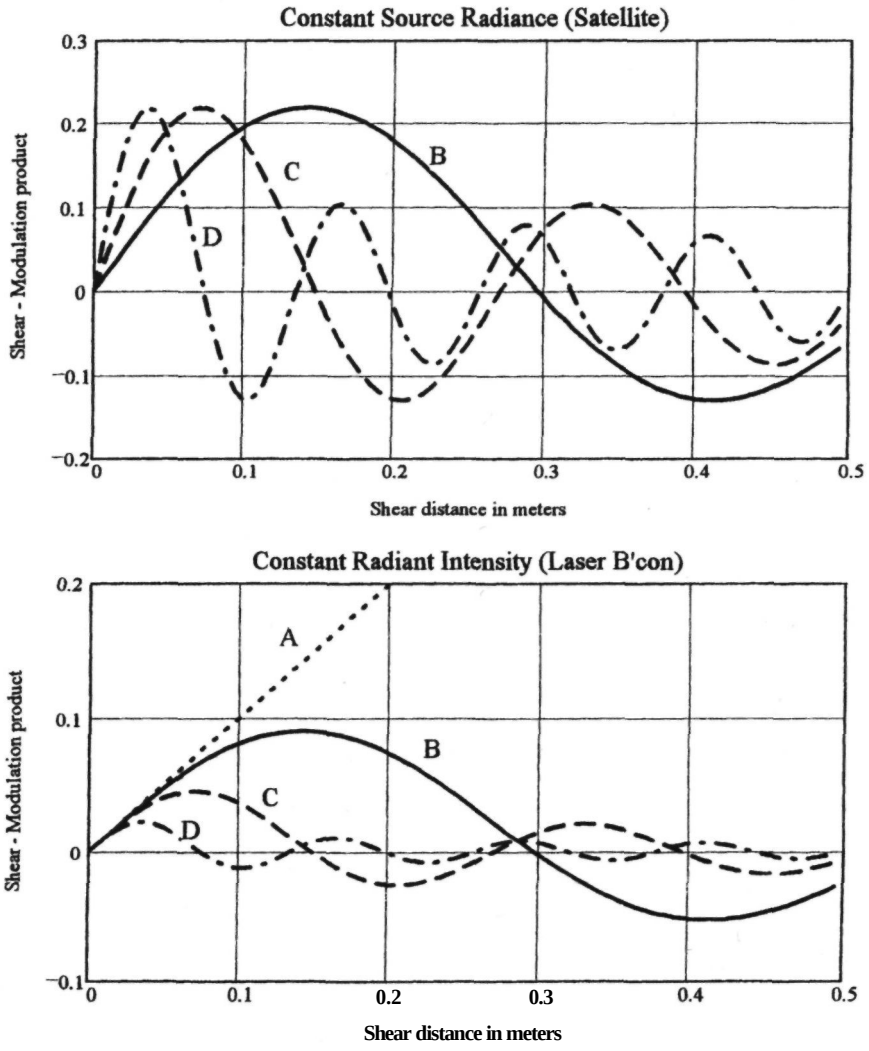
In the case of a point reference source, different considerations apply. The modulation factor K_θ is then unity, independent of the shear. The maximum value of shear that can be used depends on the residual wavefront error: as the shear is increased, the correlation between the interfered wavefronts decreases, reducing the modulation factor K_d. When the adaptive optics loop is first closed, the wavefront errors are large, and a small value of sjd is appropriate. After convergence, when the residual wavefront errors are small, a larger value of sjd may be used. For point sources, the maximum shear is normally equal to the subaperture size; that is, $(s/d)_{max} = 1$. Methods of adjusting the shear are discussed in the following sections.

5.4.5 *Variable Shear Interferometer*

The AC shearing interferometer is realized in practice using rotating radial gratings, which allow the shear distance to be varied by adjusting the radius at which the optical beams intersect the grating. Two gratings are normally used, for x and y shear, with the grating lines oriented in orthogonal directions. One axis of the shearing interferometer used in the Compensated Imaging System (CIS) is depicted in figure 5.22. The reference frequency used for measuring phase is obtained from an auxiliary light source passing directly through the grating. To utilize all of the incident light, each radial grating was made with reflecting lines, forming a second interferogram in which the modulation is 180° out of phase with that formed by the transmitted light. Outputs from the two sets of detectors on each axis were combined coherently.

The gratings used in the CIS were 10 cm in diameter with 480 radial lines, rotating at 5000 rpm to give a modulation frequency of 40 kHz. The shear was variable over a range of about 5 to 1, with a maximum shear of 1/14 of the aperture, equivalent to 11.4 cm at the telescope's 1.6-m primary mirror. The maximum shear corresponds to the size of the wavefront sensing zones.

The photon detectors used in the CIS were end-on photomultiplier arrays, one for each gradient-sensing



A point source
B source diameter 0.5 arc second
C source diameter 1.0 arc second
D source diameter 2 **arc seconds**

Figure 5.21 Variation of shear-modulation product with shear distance for two types of reference source. The wavefront error is inversely proportional to this product. The first chart represents a reflecting object such as a solar-illuminated satellite. The received power is proportional to the angular size of the object, and the shear can be optimized to maintain a constant error irrespective of object size. The second chart covers an object such as a laser beacon, from which the received power remains constant even if its angular size varies due to turbulence in the optical path. In this case, an increase in the angular size of the reference source inevitably increases the error.

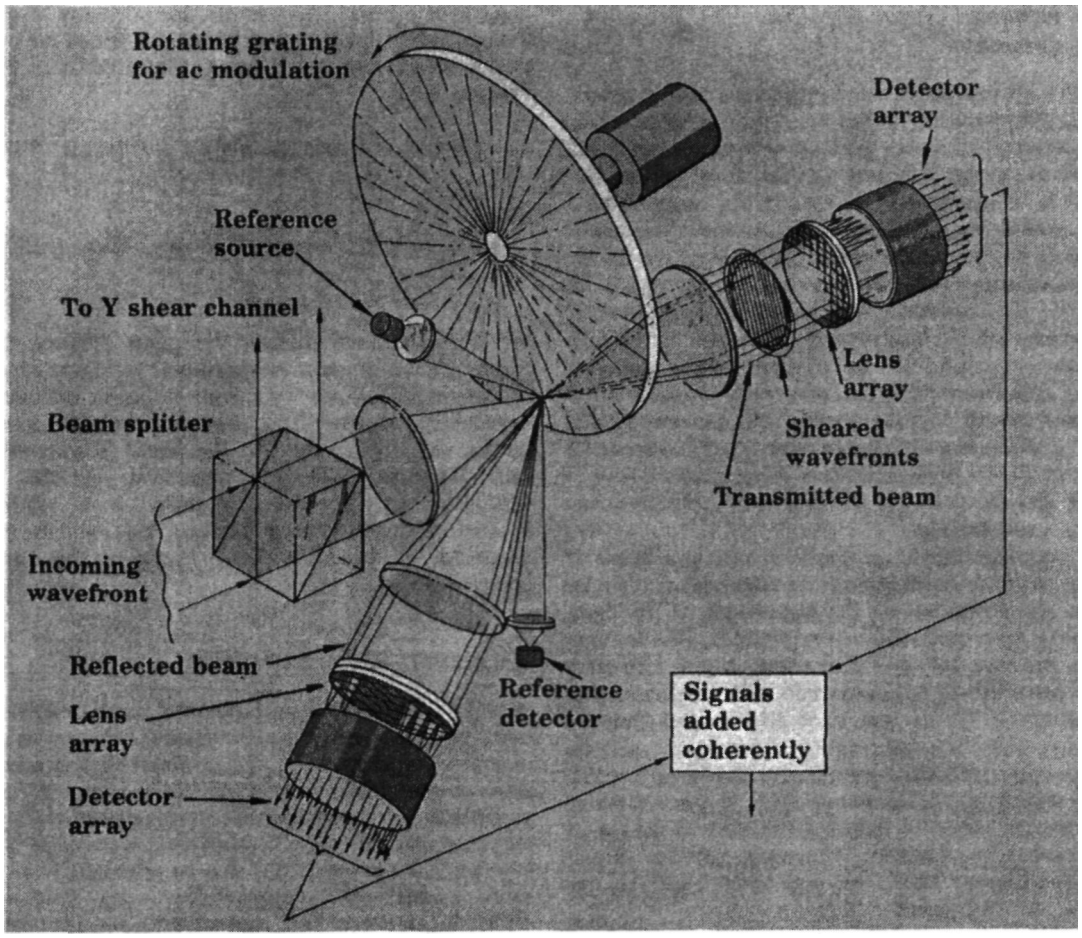


Figure 5.22 Variable-shear interferometer used in the Compensated Imaging System.

zone of the aperture. Each of the four arrays comprised 152 detectors, for a total count of 608. To provide complete coverage of the aperture, it was divided up using crossed arrays of cylinder lenses, the net result being a close-packed square array of spherical lenses with the photomultiplier tubes at their focus. The detectors operated in the pulse-counting mode at rates of up to 20 MHz. The phase measurement process consisted of reading the sine and cosine values of the reference signal (generated directly from the grating) at the instant that each signal photon was detected. These values were accumulated over the integration time of 100 μs and then passed to a look-up table that contained the phase values corresponding to the accumulated sine and cosine functions.

Good features of the radial grating AC shearing interferometer are its wide spectral band, excellent tolerance of scintillation, the availability of variable shear (which enables the dynamic range of the slope measurement to be adjusted in real time to accommodate changes in the turbulence strength), and the ability to operate with extended reference sources. Another useful feature is the capability for rejecting background light by using a field stop. By adjusting the width of the field stop to an exact multiple of the grating period, modulation from a bright background can be reduced to an arbitrarily small value. This feature has been employed in a solar telescope using detail on the solar surface as the wavefront sensor reference; it also allows the use of adaptive optics on bright stars during daytime. The AC shearing interferometers are best suited for use with continuous light sources. With pulsed reference sources, complications arise because of conflicts between the pulse rate and the temporal modulation frequency. In such cases, a version of the shearing interferometer using spatial modulation can be used.

5.4.6 Spatial Carrier Shearing Interferometry

Although the AC shearing interferometer has several desirable features, it has the drawbacks of using mechanically rotating gratings and of not being suitable for pulsed reference sources. In wavefront sensors for astronomical adaptive optics, some method of modulating the intensity is desirable because it greatly facilitates the detection process, easing the requirements on detector calibration and giving the sensor some immunity to scintillation. Spatial carrier shearing interferometry was developed to provide these capabilities in an efficient and economical way. The technique employs a mask or modulation device located near the pupil plane, the basic principle of which has already been explained in section 5.2 above. In this section, matters relating to the type of gratings and detection schemes used in practical systems are considered.

When an optical grating is placed in a beam of light, it produces multiple diffraction orders at angles that are dependent on the wavelength of the light. Special gratings having a sine-wave amplitude function produce only two diffraction orders, but these are very difficult to make. It is much easier to make gratings or masks that have hard edges, but the abrupt transitions of amplitude and phase generate high-order diffraction components that interfere with the desired low-order components. In the AC interferometer, temporal filtering was used to get rid of unwanted modulation frequencies. When spatial modulation is employed, the same goal can be achieved with spatial filtering.

The canonical arrangement of a spatial carrier shearing interferometer is shown in figure 5.23. The grating is placed in a collimated beam near a conjugate image of the pupil at which the wavefront disturbance is located. The afocal optical system of two lenses (or mirrors) reimages the pupil at the detector plane. At the intermediate focal plane, the Fourier transform of the pupil plane is formed. At this plane, the spatial frequency spectra of the wavefront disturbance and the grating are present, enabling unwanted diffraction orders to be eliminated by spatial filtering. A suitable grating for this instrument is a half-wave phase Ronchi, consisting of equispaced bars with phase shift alternating between 0 and 1/2 wave. This type of grating produces no zero order and no even orders. A simple spatial filter, consisting of a rectangular aperture at the Fourier plane, cuts out the third- and higher order sidebands, leaving just the two first-order components that interfere to produce the shearing interferogram at the detector plane. The shear is given by $s = \Delta z\,\lambda/g$, where Δz is the displacement of the grating from the pupil plane. When Az = 0, the shear is zero and the grating is imaged directly onto the detector. A set of parallel interference fringes is produced because of the tilt between the interfering wavefronts. The detector elements are arranged to sample the positions of the fringes, using three or four pixels per fringe period. For a one-dimensional grating, the intensity at the detector is

$$I(x, y) = I_0 + K_x \cos\left[2\omega_x x + k\,s\frac{\partial}{\partial x}W(x, y)\right] \quad (5.48)$$

where

ω_x - $2\pi/g$ is the spatial frequency of the grating,
k = $2\pi/\lambda$,
K_x = modulation factor

Wavefront disturbances in the pupil displace the fringes by an amount proportional to the product of the shear and the wavefront slope, as indicated by equation (5.48). These displacements are sensed by the detector pixels, and the phase is computed using the three- or four-bin algorithms described in section 5.4.3. It is convenient to locate the grating in the collimated beam because this allows the shear to be varied without changing the basic fringe spacing.

5.4.7 Multiplexing Techniques

Wavefront slope is a vector quantity, so slope sensors must make two sets of measurements to determine its magnitude and direction. To facilitate computation, slope measurements are usually made in two orthogonal axes, normally requiring two sets of gratings and detectors. By using multiplexing techniques, the two sets of measurements can be obtained from a single grating and a single detector array. This simplifies the hardware and eliminates the need for optical beam splitters with their attendant transmission loss.

Multiplexing is a simple matter with spatially modulated sensors, because the gratings and detectors themselves are two-dimensional. A spatial carrier shearing interferometer, such as that described in section 5.4.6, is easily converted to two dimensions by using a grating with two sets of crossed lines. When two orthogonal half-wave phase Ronchi gratings are added together, they create a three-level grating, but, as zero- and one-wave delays are essentially the same, the grating degenerates into a half-wave checkerboard pattern. The intensity at the detector is then

$$I(x, y) = I_0 + K_x \cos\left[2\omega_x x + k\,s\frac{\partial}{\partial x}W(x, y)\right] + K_y \cos\left[2\omega_y y + k\,s\frac{\partial}{\partial y}W(x, y\right] \quad (5.49)$$

The interference fringes are now displaced in both the x and y directions according to the wavefront slope, as shown in figure 5.24. These displacements are sensed by a two-dimensional detector array using 4×4 pixels per fringe element. The pixels are

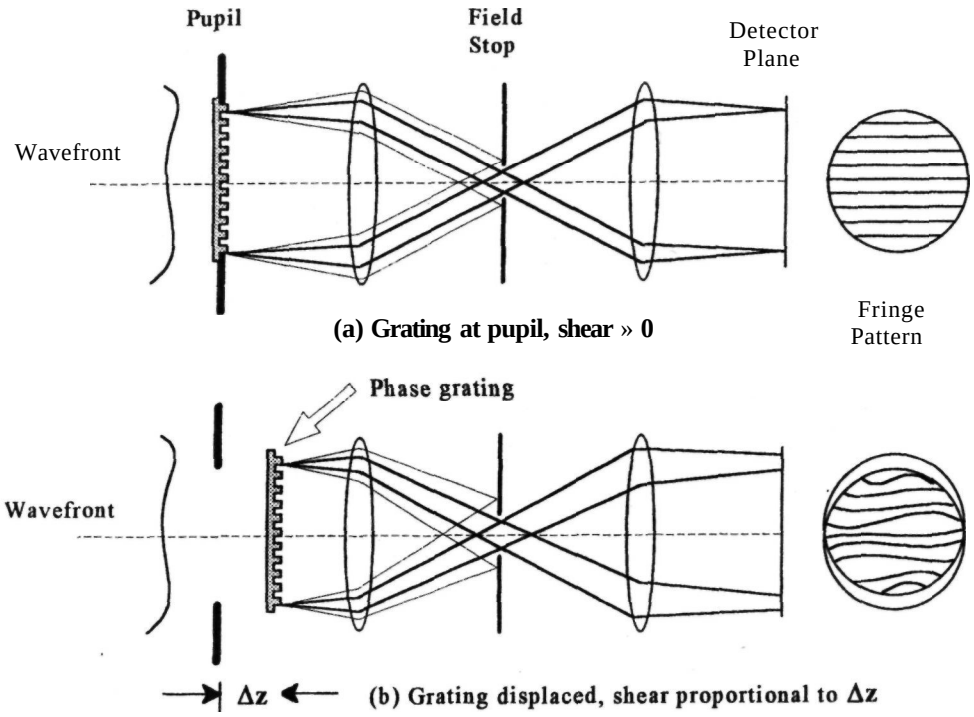


Figure 5.23 Spatial carrier shearing interferometer.

assembled into two four-element bins in x and y, with the pixel values averaged in the orthogonal directions.

Multiplexing temporally modulated shearing interferometers is more complicated, because of the one-dimensional nature of the variable, time. The two channels must be separated by using different temporal modulation frequencies. A grating pattern is needed that diffracts light simultaneously in two orthogonal directions, and, at the same time, modulates the two beams at different temporal frequencies that can be separated by electrical filters. The solution, described by Horwitz [1990], is to use two logarithmic spirals, of opposite sense, as shown in figure 5.25. A grating composed of families of these curves has properties similar to a radial grating, but in two dimensions. These properties include the following: (1) the period varies linearly with radius; (2) the line orientation remains constant with radius; and (3) each spiral has a constant number of lines. But, the number of lines is different in the two spirals, so, at a given rotation speed, two different temporal frequencies are generated. A grating of this type consists of a checkerboard pattern with each "square" being defined by one of the family of spirals. An AC shearing interferometer using a log spiral grating has been used successfully for astronomical wavefront sensing [Horwitz 1991, Wild et al. 1993, Wild, Kibblewhite, Fang Shi et al. 1994].

5.4.8 Large-Shear Interferometers

The shearing interferometers described previously employ relatively small shear values in order to measure wavefront slope, which is averaged over the shear distance. To operate in this mode, the wavefront slope must be reasonably constant over the shear distance, otherwise the fringe contrast falls and measurement accuracy is compromised. Shearing interferometers can also be used with large shear values to measure differences in phase or optical pathlength. One useful application is the phasing of individual elements of a segmented mirror. In this case, the wavefront slope is nearly zero over large areas. A pure slope sensor, such as the Shack-Hartmann sensor, is insensitive to phase differences between its subapertures and cannot be used for this purpose.

Figure 5.26 shows how a lateral shearing interferometer is used for phasing two panels of a segmented mirror. In this case, the shear distance is adjusted to produce interference between the light reflected from two panels, which need not be contiguous. A rotating radial grating, located at the focus of the mirror, may be used to produce the shear and also to generate AC modulation. Phase ambiguities are eliminated by the use of white light or multiwavelength illumination. The resulting intensity variations are measured at two locations at the detector plane D. The normalized signal is given by

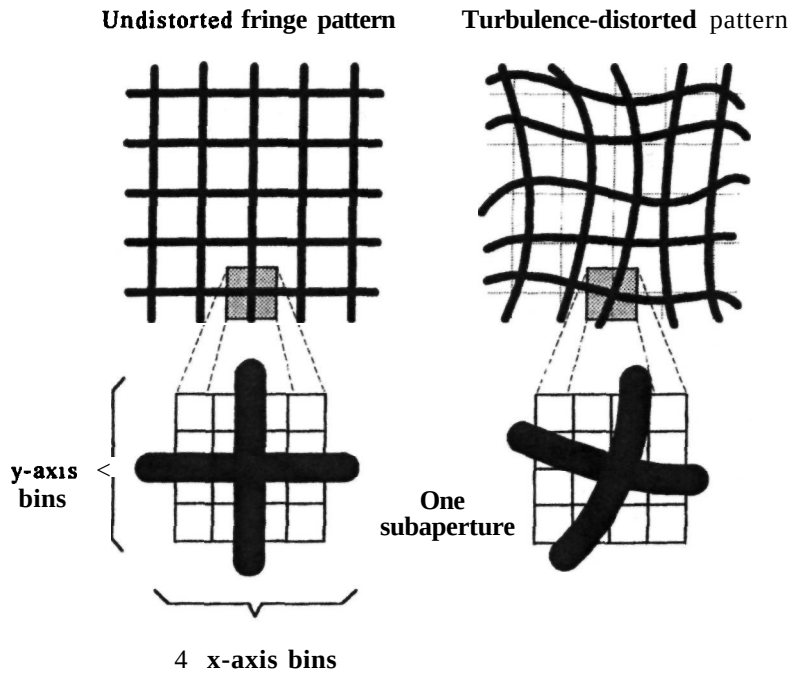


Figure 5.24 Fringe patterns produced in a spatially multiplexed shearing interferometer. Wavefront slope variations displace the interference fringes in two dimensions, enabling simultaneous, *x* and *y* slope measurements to be made in each subaperture.

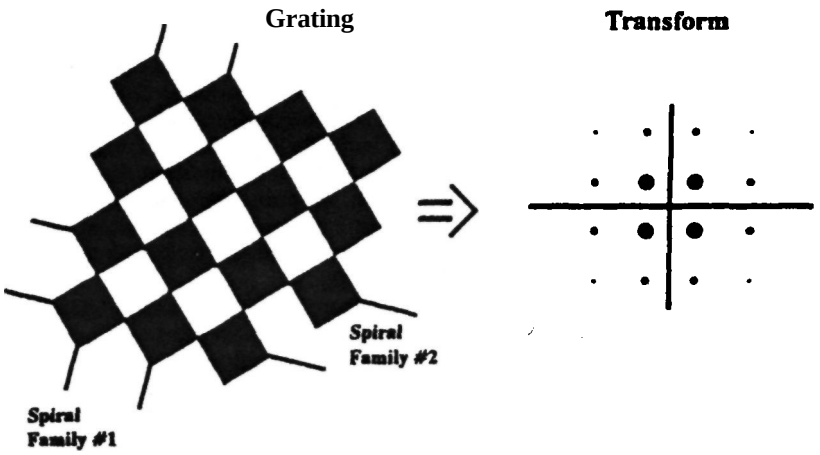


Figure 5.25 Temporally multiplexed shearing interferometer using a logarithmic spiral grating [Horwitz 1990]. The grating has a checkerboard pattern that generates overlapping (sheared) diffraction images of the optical pupil at angles of ±45°, as shown in the transform pattern. These patterns are modulated at different temporal frequencies, allowing them to be separated by filtering.

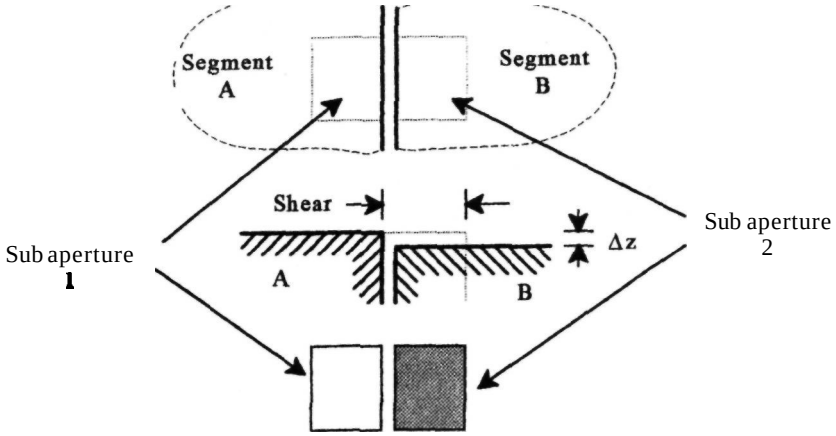


Figure 5.26 Large-shear interferometer for phasing panels of a segmented mirror.

$$I(t, \lambda) = (1/2)[1 + K \cos(\omega t + \phi(\lambda)] \qquad (5.50)$$

where

K = modulation factor
$\phi(\lambda)$ = electrical phase angle due to interference between the two beams.

In subaperture 1, both beams are reflected from segment A, so the phase angle $\phi_1(\lambda) = 0$. In subaperture 2, the beams are reflected from different segments, resulting in a phase angle $\phi_2(\lambda) = 2k\Delta z$, where $k = 2\pi/\lambda$ and Az is the displacement between the segments. The shear distance must be at least equal to the subaperture size, but its actual value is irrelevant, provided that the panels are not tilted.

The value of Az is found by measuring the difference between the phase shifts in subapertures 1 and 2. The phase shifts at different wavelengths are simultaneously nulled only when the displacement Az is zero. The modulation factor K depends on the residual wavefront errors in the interfered beams. In order to measure the displacement accurately, higher order errors, such as wavefront tilt and defocus, must first be removed.

5.5 Curvature Sensing

S.5.1 Curvature from Slope Measurements

The zonal approach to optical wavefront sensing consists of dividing the wavefront into subapertures, within each of which some property, such as displacement, slope, or curvature, is measured. Zonal slope sensing, which involves measurement of the first-order derivatives $(\partial z/\partial x + dz/dy)$, has proved to be very successful in adaptive optics. From a single snapshot of the wavefront slopes over the optical pupil, the entire wavefront can be reconstructed, up to the cutoff frequency of the sampling interval. The slope data require reconstruction to obtain the wavefront surface values, but the additional errors incurred are small.

Wavefront curvature can be determined from measurements made with a slope sensor in four adjacent subapertures, as shown in figure 5.27. Two orthogonal slope measurements are made in each subaperture, giving

$$\begin{aligned} \mathcal{C}_x &= \partial^2 z/\partial x^2 = c[(x1 - x2) + (x4 - x3)] \\ \mathcal{C}_y &= \partial^2 z/\partial y^2 = c[(y1 - y4) + (y2 - y3)] \end{aligned} \qquad (5.51)$$

where c is a constant of proportionality. The $\mathcal{C}_x$ and $\mathcal{C}_y$ terms represent cylindrical curvatures in the x and y directions, respectively. The wavefront curvature defined in this way is centered at the origin of four slope-sensing subapertures as shown. From the same set of eight slope measurements, the following additional wavefront functions can be determined:

$$\begin{aligned} \mathcal{C}_{x+y} &= c[(x1 - x3) + (y1 - y3)] \\ \mathcal{C}_{x-y} &= c[(x4 - x2) + (y2 - y4)] \\ \mathcal{J}_x &= x1 + x2 + x3 + x4 \\ \mathcal{J}_y &= y1 + y2 + y3 + y4 \end{aligned} \qquad (5.52)$$

The $\mathcal{J}_x$ and $\mathcal{J}_y$ terms are the overall wavefront slopes in the x and y axes, while $\mathcal{C}_{x+y}$ and $\mathcal{C}_{x-y}$ represent cylindrical curvature about the +45° and —45° axes, respectively. The four curvature components may be combined to form the more familiar optical aberrations as follows:

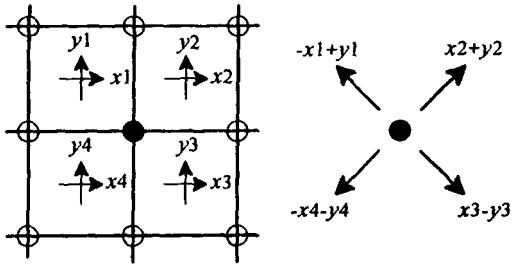


Figure 5.27 Derivation of wavefront curvature from wavefront slopes.

$$\begin{aligned}\text{Defocus} &= \mathcal{C}_x + \mathcal{C}_y = \mathcal{C}_{x+y} + \mathcal{C}_{x-y}\\ \text{Astigmatism}\,(0°) &= \mathcal{C}_x - \mathcal{C}_y\\ \text{Astigmatism}\,(45°) &= \mathcal{C}_{x+y} - \mathcal{C}_{x-y}\end{aligned} \tag{5.53}$$

The set of eight slope measurements made in four subapertures therefore yields five orthogonal wavefront aberrations. This process employs wavefront slope measurements to derive curved surfaces. The question then arises as to whether wavefront curvature may be measured directly, without first measuring slope.

Curvature sensing uses the second derivative of the wavefront as the basis for measuring randomly distorted wavefronts, instead of the first derivative as in slope sensing. Wavefront curvature exists in two dimensions, and if the wavefront surface is z(.v, *y),* then the local curvature can be expressed as

$$\mathcal{C}(x, y) = \partial^2 z/\partial x^2 + \partial^2 z/\partial y^2 \tag{5.54}$$

Although wavefront curvature is generally different in *x* and >', the term "curvature sensing" usually implies direct measurement of spherical curvature over contiguous zones of the wavefront. When restricted in this way, curvature is a scalar quantity that can be measured directly, rather than being derived from conventional slope measurements. Not all wavefront shapes can be described in terms of spherical curvature; astigmatism, for example, consists of equal and opposite cylindrical curvatures in two orthogonal axes.

5.5.2 Direct Curvature Sensing

Curvature sensing as described by Roddier [1988a] is based on direct measurement of spherical curvature, a rotationally symmetrical scalar quantity that is relatively easy to measure and correct. Curvature sensing is most useful when used in conjunction with a bimorph mirror that produces a variable curvature deformation in response to a control signal. It may seem surprising that a rotationally symmetrical scalar element can be used to measure (and reproduce) randomly distorted wavefronts. In fact, some additional data are required, specifically the radial gradients at the edges of the aperture. A knowledge of these gradients is essential to correct nonspherical distortions such as astigmatism and coma.

A simple curvature sensor employs two detector arrays located at the near and far sides of the detector plane. Local wavefront curvature causes differences in the intensity at corresponding detector locations, producing error signals proportional to the curvature. At the edges of the aperture, the intensity differences correspond to the edge gradients required to set up the boundary conditions; in the case of a circular aperture, these are radial tilts.

The method of measuring wavefront curvature is shown in figure 5.28. An incoming wavefront W, nominally a plane wave, is focused by the telescope objective L, of focal length Z, on to the focal plane *F*. The distribution of intensity is measured at two planes P_1 and P_2, symmetrically displaced from plane *F* by distance *p*. Let the incoming wavefront have local curvature $C_w = 1/r_w$ where r_w is the local radius of curvature of the wavefront over a small area a_0. The curved wavefront will come to a focus at a distance z_c, given by

$$z_c = \frac{Z\, r_w}{Z + r_w} \tag{5.55}$$

The focal shift is then

$$\Delta z = Z - z_c = \frac{Z^2}{Z + r_w} \tag{5.56}$$

Z + r_w

For a beam area of a_0 at the objective, the corresponding areas that will be illuminated at planes P_1 and P_2 are

$$\begin{aligned} a_1 &= a_0\left(\frac{p - \Delta z}{Z - \Delta z}\right)^2\\ a_2 &= a_0\left(\frac{p + \Delta z}{Z - \Delta z}\right)^2\end{aligned} \tag{5.57}$$

If *H* is the irradiance in watts per meter squared at the aperture, over area a_0, then the irradiances at planes P_1 and P_2 are

$$\begin{aligned} H_1 &= H\left(\frac{a_0}{a_1}\right)^2\\ H_2 &= H\left(\frac{a_0}{a_2}\right)^2\end{aligned} \tag{5.58}$$

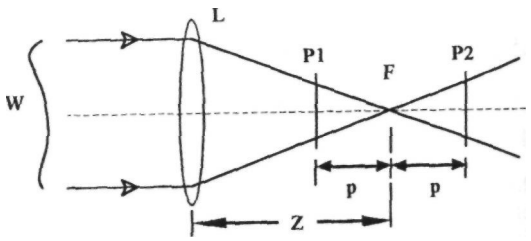


Figure 5.28 Curvature sensor using displaced focal planes.

Defining the signal A/ as the normalized difference in irradiance at planes P_1 and P_2, we obtain

$$\Delta I = \frac{H_1 - H_2}{H_1 + H_2} = \frac{a_2^2 - a_1^2}{a_2^2 + a_1^2} = \frac{2p\,\Delta z}{p^2 + \Delta z^2} \tag{5.59}$$

The value of Az is given in equation (5.56). As the wavefront curvatures being measured are extremely small compared with that produced by the telescope objective (that is, $r_w \gg Z$), then we may use the approximation $Az \approx Z^2/r_w$. Also, $p^2 \gg Az$, enabling A/ to be expressed as

$$\Delta I = \frac{2Z^2}{p\,r_w} = \frac{2Z^2 C_w}{p} \tag{5.60}$$

The signal A/ is therefore directly proportional to the local curvature C_w of the input wavefront, and is independent of the absolute value of the irradiance at the input aperture.

This analysis is based on geometrical optics, and certain conditions must be fulfilled to assure its validity. The irradiance distribution at planes $P\backslash$ and P_2 is a blurred replica of the pupil, the blur size being determined by the diffraction caused by turbulence at the input aperture. If the scale size of the turbulence is r_0, then the diffraction angle is λ/r_0 and the blur size at plane P_1 is $\lambda(Z - p)/r_0$. It is necessary for this blur to be small compared with the areas over which the curvature is to be measured; that is $r_0\,p/Z$ These requirements lead to the condition

$$p \geq \frac{\lambda Z^2}{\lambda Z + r_0^2} \tag{5.61}$$

This condition is easily satisfied in astronomical telescopes, with the minimum value of the displacement p being typically in the range of 1-20cm. At optical wavelengths, $r_0^2 \gg \lambda Z$, so the condition may be simplified, in practice, to $p > \lambda Z / r_0^2$.

Low-order distortions of the incoming wavefront, such as defocus or astigmatism, produce large differences in irradiance at the edges of the images at planes $P\backslash$ and P_2, because of their difference in size. These edge signals are proportional to the radial gradients at the boundary of the wavefront and are used to control the tilt at the edges of the wavefront corrector.

5.5.3 Focus Modulation

The implementation of a curvature sensor with two separate detector planes is clumsy and inefficient because of the need for a beam splitter. It also requires calibration of the two detector arrays. An elegant and efficient alternative [F. Roddier 1990a] is to use a single detector array with optical modulation of the focus between the two planes. This is achieved by using a variable curvature mirror located at the nominal focus, as shown previously in figure 5.8. When this mirror is flat, the detector is reimaged at the telescope pupil. When the mirror is activated, its curvature is sinusoidally modulated by a small amount in positive and negative directions at a constant frequency, causing the focus to shift between planes P, and P_2. Since the variable curvature mirror is close to the focal plane, its aberrations are unimportant with small fields of view. A simple and convenient implementation is to use a reflective membrane attached to a moving-coil loudspeaker. The use of focus modulation not only eliminates the need for precise detector calibration, but also allows the amount of modulation to be varied in real time, to accommodate changes in turbulence strength. A sensor of this type with 13 subapertures, using photon-counting avalanche diode detectors, is described by Graves et al. [1994].

5.5.4 Performance of Curvature Sensors

Curvature sensors provide a scalar output that represents the two-dimensional wavefront curvature over each subaperture. The standard deviation of a single curvature measurement due to photon noise when using a point reference source is shown by Roddier et al. [1988] to be

$$\sigma_p = \frac{p}{Z^2 N_p^{1/2}} \tag{5.62}$$

where N_p is the photon count. It is assumed that the photon detector averages the light from the reference source over r_0-sized areas. Using the minimum value of $p > \lambda Z^2/r_0^2$ from equation (5.61), the measurement error is

$$\sigma_p = \frac{\lambda}{r_0^2 N_p^{1/2}} \tag{5.63}$$

A single curvature measurement is equivalent to four adjacent wavefront tilt measurements, two in the x direction and two in the y direction. When comparing the performance of a curvature sensor with that of a conventional slope sensor, it is necessary to include the reconstruction process. The photon error in a single subaperture of a curvature sensor is similar in magnitude to that of an equivalent Hartmann-Shack tilt sensor, but the error propagation in the reconstruction process is much greater for curvature measurements than for tilt measurements. Roddier et al. [1988] found that the error coefficient for curvature sensing increases as the square of the array size, whereas for wavefront tilt sensing it increases logarithmically. The reason for this difference is that wavefront curvature due to atmospheric turbulence tends to be statistically independent in adjacent areas, so the error variances simply add up over the aperture in the reconstruction process. This is the price paid for the simplicity of the approach. Wavefront slope, on the other hand, is correlated over larger areas and the redundancy in the measurements is put to good use in the reconstruction process

by using multiple paths to connect the evaluation grid.

Because the reconstruction errors grow with aperture size, curvature sensing is competitive in performance with slope sensing only for small arrays of up to about 50 elements. This is sufficient to provide compensation at IR wavelengths for apertures up to about 6 m under good seeing conditions. An experimental wavefront curvature system built at the University of Hawaii Institute for Astronomy [Roddier et al. 1994] employs a curvature sensor with 13 photon-counting avalanche diodes to measure seven curvature terms and six edge slopes. The illumination at two planes displaced from focus is compared by modulating the focus of the sensor at a rate of 1 kHz, by means of a vibrating membrane mirror. The data are processed by a 13 x 13 matrix multiplier and then fed via a high-voltage amplifier to a 13-element bimorph mirror. A closed-loop bandwidth of 100 Hz is achieved. The performance of this system is described in chapter 10.

5.6 Wavefront Sensing with Extended Sources

5.6.1 *Reference Source Characteristics*

What characteristics are needed in a reference source for adaptive optics? One tends to think of small, bright spots of light, and, indeed, an unresolved star is the ideal reference source for a wavefront sensor. But much of the visible structure of the universe consists of extended matter with low-contrast details. Even nearby objects, such as the Sun, the Moon and the planets, have very low surface contrast. The solar granulation, for example, has a contrast (A///) of less than 5% when observed through average turbulence. The question may be asked whether such low-contrast extended objects can be used for wavefront sensing in the "self-referencing" mode, or whether it is necessary, in such cases, to use laser beacons as the reference sources.

The answer is very simple: if an r_0-sized subaperture can resolve sufficient detail in the reference object to measure the angular fluctuations produced by the atmosphere, then the wavefront errors over the whole telescope aperture can be measured and corrected. The proposition may be stated more precisely in the following terms: if the angular motion of the reference object can be tracked to a precision of a times the diffraction limit in each subaperture of a zonal compensation system, then, in principle, it should be possible to compensate the angular resolution of the whole aperture to α times its diffraction limit, excluding other sources of error.

Theoretically, it is irrelevant whether the detail being tracked is a bright spot on a dark background (reference star or laser beacon) or a dark spot on a bright background (solar granulation or pore). In practice, wavefront sensors differ greatly in their ability to use different kinds of reference objects. It is well known that Shack-Hartmann sensors using quadrant detectors are easily confused by multiple objects. In general, it is much easier to use conventional bright spots as reference sources because there is no background light to confuse the measurement. Negative reference sources are best handled with AC-modulated wavefront sensors (such as the AC shearing interferometer) because of the difficulty in calibrating detectors over a large intensity range.

It is useful to distinguish between two different kinds of extended reference objects: (1) compact sources, which are resolvable, but whose angular subtense is less than the isoplanatic angle; and (2) extended sources, which exceed the isoplanatic angle. With compact sources, which include laser beacons, conventional wavefront sensors can generally be used, making minor adjustments to accommodate the finite size of the reference object. For large extended sources, which are generally of low contrast, special wavefront sensor configurations are necessary to restrict the field of view and to handle the low modulation levels.

5.6.2 *Wavefront Sensing with Compact Reference Sources*

Shack-Hartmann sensors generally will operate with compact reference sources without modification. Sources smaller than the isoplanatic angle are, by definition, only a few arc seconds in diameter, and therefore not much larger than the diffraction limit of the Hartmann subapertures. The result of using a reference source of finite size is to increase the wavefront error in proportion to the source size, as shown in equation (5.16).

With shearing interferometers, it is necessary to optimize the shear in order to get the best performance with extended sources. The quantity to be maximized is the product of shear and modulation. As shown in section 5.4.4, the optimum shear distance for a circular reference source is $0.586\lambda/\theta$, and that for a rectangular source is $0.5\lambda/\theta$, where θ is the angular size of the source. In shearing interferometers employing a Ronchi grating to generate the shear, this occurs when the image of a rectangular source spans exactly one half-period of the grating.

The operation of curvature sensors with extended sources has been confirmed by Roddier et al. [1994], who report successful operation using a nebula of about 1 arc second diameter as the reference source.

For solar adaptive optics, there are two approaches to wavefront sensing, depending on the application. Pores and small spots of a few arc seconds diameter have relatively high contrast and have been used successfully as reference objects with a Hartmann sensor employing a quadrant detector. In the system described by Acton and Smithson [1992], the sensor had a zoom capability to vary the

effective field of view on the solar surface. They found that compensation was effective only with reference spots (lock points) smaller than 5 arc seconds. The degree of correction decreased with radial distance from the lock point, as expected. Good compensation was obtained only in the immediate vicinity of isolated features of relatively high contrast, although some improvement was usually visible across the entire recorded field of 30 arc seconds.

The second approach to solar wavefront sensing is to use correlation of low-contrast image detail as the basis for wavefront tilt determination. This has more general application, as it is independent of the image structure within the wavefront sensor field of view, provided only that some structure exists in the directions in which wavefront tilt is to be sensed. Sensors of this type are described in the following section: they may, in principle, be used with any extended source.

5.6.3 Wavefront Sensing with Large Extended Sources

The use of shearing interferometers with large extended sources goes back to the early days of adaptive optics, when the CIS was built to obtain high-resolution images of low-Earth-orbit satellites. The size of these objects often exceeded the isoplanatic angle, and field stops were incorporated into the wavefront sensor design to exclude extraneous light. Similar wavefront sensors have been used for solar compensation [Hardy 1987] and for daylight imaging of bright stars using the 60-inch telescope at Mount Wilson Observatory [Shelton et al. 1993].

In the AC lateral shear interferometer, a rotating radial Ronchi grating is used at the image plane to produce the sheared beams. When using an extended source, an arbitrary field stop inserted at this location would normally produce large AC modulation, overwhelming that due to the reference object. This modulation is eliminated by adjusting the shear distance (or the width of the field stop) so that it corresponds to a null in the sinc function, as described in section 5.4.4 and shown in figure 5.21. The null condition occurs when the field stop spans exactly one period of the grating, which is equivalent to the condition $e = \lambda / s$.

The maximum signal modulation occurs when the dominant spatial frequency in the reference object is equal to that of the grating. The process can be regarded as a cross-correlation of the grating with the reference object. As the reference object moves randomly within the field stop, because of atmospheric turbulence, the grating position or phase giving maximum correlation is determined, enabling the wavefront tilt angle of the reference to be measured. In this case, the correlation is performed using only one spatial frequency at a time, although this frequency can be varied.

A more general approach to wavefront sensing with extended sources has been described by von der Luhe et al. [1989]. The system has been used for tracking and motion stabilization of the solar granulation. The same principle could be extended to measurement of higher order wavefront aberrations. The method consists of generating a reference image $I_R(\mathbf{x})$ of the area of interest and comparing it with live images $I_L(\mathbf{x})$ obtained with a fast matrix detector. Their cross-covariance is given by

$$CC_{LR}(\Delta) = \iint d\mathbf{x} I_L(\mathbf{x}) I_R(\mathbf{x} + \Delta) = \mathcal{F}^{+}\{\mathcal{F}^{-}[I_L(\mathbf{x})] \times \mathcal{F}^{-*}[I_R(\mathbf{x} + \Delta)]\} \quad (5.64)$$

where

x = two-dimensional spatial coordinate,
A = two-dimensional image displacement or spatial lag,
$\mathcal{F}^{+}$ and $\mathcal{F}^{-}$ = forward and inverse Fourier transforms
* = complex conjugation

The cross-covariance is maximized when the images match, giving a lag of Δ_{max}, from which the displacement of the live image is determined. The value of Δ_{max} is independent of the structure of the image. Computation of the cross-covariance function is implemented by a special-purpose digital computer using fast Fourier transforms. The displacement error is then conditioned and applied to a steering mirror that recenters the live image on the matrix detector. The reference image is updated periodically to accommodate any structural changes.

The performance of this correlation tracker on solar telescopes has been described by Rimmele et al. [1991]. It is used to stabilize the image motion of small areas of about 5 x 5 arc seconds anywhere on the Sun, using a small two-axis agile mirror. Residual image motion is reduced from about 0.5 arc second rms to 0.05 arc second rms. The imaging matrix is a 32 x 32 Reticon diode array operating at 976 frames per second, of which a 16 x 16 subfield is actually processed. The processing delay is 3.4 ms, which limits the closed-loop bandwidth to 40 Hz. The performance of the system is adequate for a meter-class solar telescope.

This technique could, in principle, be extended to the sensing of higher order wavefront aberrations by using a correlation tracker of this type in each subaperture of a Shack-Hartmann sensor. For example, a 128 x 128 detector array could be partitioned into 8 x 8 subapertures, each containing a 16 x 16 imaging field. The major practical problem to be overcome in implementing such a system is the huge data-processing load. Using a 16-port CCD operating at 1 MHz per port, the data could be read out in 1 ms, but a massive parallel processor would be required to perform the cross-correlation in comparable time.

One method of reducing the computational load is to use optical correlation. A single transmitting mask

in the focal plane, as shown in figure 5.29, performs a two-dimensional cross-correlation in all subapertures simultaneously, greatly simplifying the entire process.

A wavefront measurement technique for extended incoherent light sources using this approach has been proposed by von der Liihe [1988]. The transmitting mask contains a reference pattern that converts displacements of the live image into intensity changes that are sensed by a detector array. If the intensity distribution at the image plane is /(x), a suitable mask transmittance is

$$M(\mathbf{x}) = A + B[I(\mathbf{x} - \Delta) - I(\mathbf{x} + \Delta)] \qquad (5.65)$$

where

A and B = constants limiting the transmission of the mask between 0 and 1,
A = image displacement comparable to the correlation scale of the intensity distribution

The intensity distribution produced by this function is effectively the derivative of the reference image in the direction of the displacement A. Displacement of the live image in the positive direction caused by a wavefront tilt in **a** subaperture increases the transmitted intensity measured at the detector plane in the corresponding subaperture, and vice versa.

A possible method of implementing this wavefront sensor is shown in figure 5.30. A liquid crystal is used as the transmission mask, enabling the reference image to be updated frequently. The image data required for the mask are obtained from a video camera. To eliminate effects of intensity variations, the mask is switched between two states, positive and negative, by modulating the plane of polarization at a high rate by means of the electro-optic modulator. This results in two alternating intensity values at the detector; these values are synchronously demodulated. Displacements of the live image in the positive or negative directions with respect to the reference image then result in positive or negative outputs from the demodulator. When the images match, the output is nulled, independent of the absolute intensity.

5.7 Photon Detectors for Wavefront Sensing

5.7.7 *Requirements*

One of the most pressing needs in astronomical adaptive optics is to improve the sensitivity of wavefront-measuring systems, so that objects of lower brightness can be used as reference sources. The limiting magnitude of an adaptive optics system depends on many factors, including the turbulence conditions, subaperture size, and integration time, as well as the transmission of the optical components and the efficiency of the wavefront sensor. Maximization of wavefront sensor efficiency is a worthwhile goal for both natural-star and laser-beacon adaptive optics. For systems using natural reference stars, improving the detector sensitivity, especially at IR wavelengths, increases the potential sky coverage. Laser beacons operate at visible and ultraviolet (UV) wavelengths, so, in this case, improving the performance of short-wavelength detectors will reduce the beacon laser power requirements.

The efficiency of all wavefront sensors is governed by the signal-to-noise ratio produced by the photon detector. In section 5.3.2, it was shown that the signal-to-noise ratio is determined by the photon (shot) noise of the detected signal, the photon gain of the intensifier (if any), and the amplifier noise. An ideal detector would have 100% quantum efficiency, with either infinite photon gain or zero amplifier noise. For adaptive optics, it should also be capable of handling relatively high photon-counting rates. An aperture of 20 x 20 cm, for example, collects a flux of about 100,000 photons per second from a star of m_v as 10, which can cause problems with some photon-counting detectors. Because of the short integration times used for wavefront sensing (typically several milliseconds), the actual number of photons counted per measurement is on the order of 100 per subaperture.

Needless to say, there are no perfect detectors. Practical detectors have either low quantum efficiency with virtually no noise (such as photomulti-

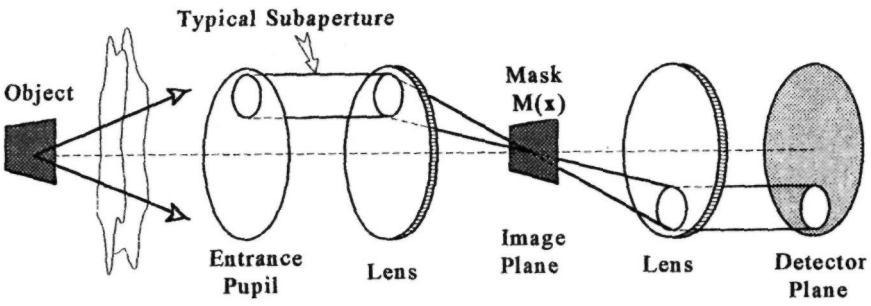


Figure 5.29 Wavefront slope sensing system using optical correlation.

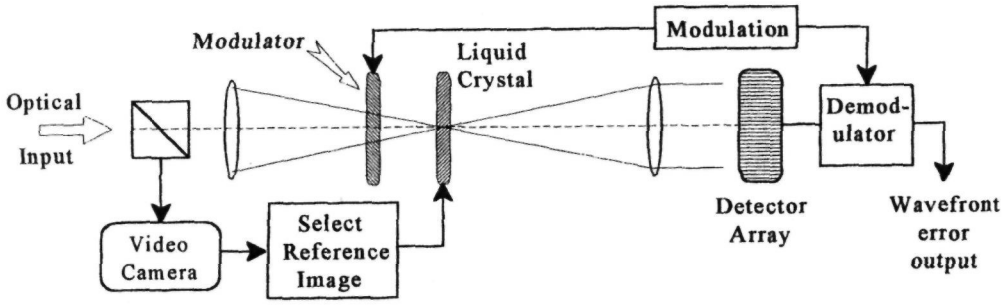


Figure 5.30 Proposed wavefront sensor using optical correlation with a liquid crystal. The liquid crystal is located in the Fourier plane and contains the reference image, which is updated in real time [von der Liihe, 1988].

pliers, intensified devices, and avalanche diodes) or high quantum efficiency with amplifier noise (such as CCDs). A comparison of the signal-to-noise ratios of intensified detectors, avalanche photodiodes, and silicon CCDs versus input flux is shown in figure 5.31. At very low input levels, intensified detectors and avalanche photodiodes have better performance because of the absence of noise, but, for higher, inputs, straight silicon devices are superior because of their better quantum efficiency. The crossover point depends mainly on the read noise of the CCD; as this noise is reduced to a few electrons per pixel, directly illuminated CCDs become competitive, even at the lowest input levels. The application of these photon detectors to wavefront sensing is briefly reviewed in the following sections.

5.7.2 CCD *Detectors*

The high quantum efficiency and stable geometry of CCDs makes them particularly well suited for use in wavefront sensors. At low photon counts, the signal-to-noise ratio is dominated by amplifier noise, which in the case of CCD detectors is the "read" noise produced in converting the charge packets into voltages at each output port. The CCD read noise is proportional to the clock rate of the output ports and so can be minimized by using multiple ports, with parallel readout.

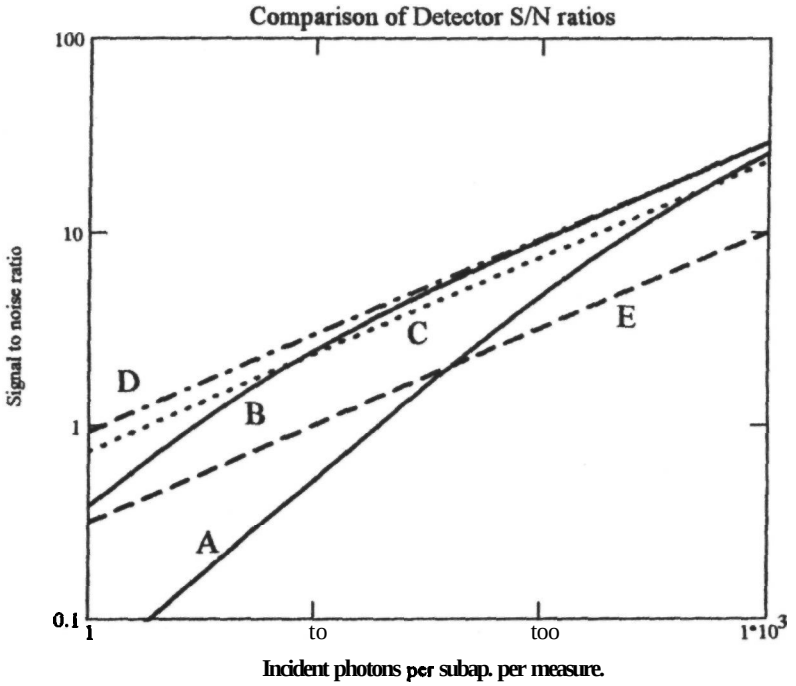
Low-noise CCDs having two or four output ports, with readout noise of 5–10 electrons per pixel, frame rate of 2 kHz, and quantum efficiency of over 80%, have already been developed, as mentioned in section 5.3.4. The next generation of CCDs for wavefront sensing has the goal of reducing the readout noise to about 1 electron rms at frame rates of 1500–2000 Hz. Devices are being developed with special pixel configurations to suit specific sensors: for example, a radial pattern for compatibility with curvature sensors.

There are two distinct areas of technology development, the CCD chip itself and the camera, the latter containing the supporting electronics, cooling, data transfer, and control functions. Several organizations are developing high-frame-rate, low-noise CCD chips suitable for wavefront sensors. The CCID-11 device, developed at MIT Lincoln Laboratory and described by Beletic et al. [1993], is a 64 x 64 pixel, backside-illuminated CCD with 80-90% quantum efficiency at 450–750 nm. It operates at a maximum frame rate of 3 kHz with four output ports. The read noise is 6 electrons rms at 1 MHz, 12 electrons at 2 MHz, and 25 electrons at 5 MHz. The pixel size is 24 x 24 μm and the response is uniform over the array to within a few percent.

The approach at Jet Propulsion Laboratory (JPL)/Caltech [Levine et al. 1994] is to read out each column of a 32 x 64 pixel array separately, using a skipper amplifier, which is capable of multiple nondestructive readout of the charge packets. The CCD is a three-phase frame-transfer device with 32 output ports and a frame rate of 1500 Hz. Experimental devices achieved read noise of 8 electrons; it is hoped to reduce this to about 2.5 electrons after 10 reads of the skipper amplifier on subsequent designs.

Geary [1994] describes two CCD designs using 16 ports, with 64 x 64 and 128 x 128 arrays of 25 μm pixels. The 64 x 64 device has a 5-kHz frame rate with a 2.5-MHz maximum pixel rate, while the 128 x 128 device is designed to run at 1-kHz frame rate with 1-MHz output rate per port.

A radical departure in CCD design has been described by Geary and Luppino [1994], using a circular rather than a rectilinear pixel array. The application is curvature sensing, for which circular symmetry is appropriate. Low readout noise is achieved by using the minimum number of pixels, together with multiple ports to keep the pixel rate down to about 100 kHz, at which a readout noise of 2 electrons rms should be possible. The CCD has 24 angular sectors with 12 radial zones, divided into four quadrants. The radial pitch is 30 μm, and the azimuthal pitch varies from 15 μm at the inner edge to 100 μm at the outer edge of the active area. An



A CCD, QE = 0.85, read noise = 8 e per pixel (4 pixels)
B CCD, QE = 0.85, read noise = 1 e per pixel (4 pixels)
C avalanche photodiode, QE = 0.9, detection threshold 0.6
D avalanche photodiode, QE = 0.9, detection threshold 0.95
E photomultiplier, QE = 0.1

Figure 5.31 Signal-to-noise ratio comparison of detectors used in wavefront sensors. The comparison is based on four detector readouts per measurement. The given number of incident photons per subaperture per measurement is divided into four bins, to each of which the read noise (if any) is added. The signal and noise of the four bins are then combined to get the overall signal-to-noise ratios.

additional ring of 12 zones serves as the frame-storage area. The serial register is wrapped around the storage area.

A camera design to support high-frame-rate CCD chips has been described by Beletic et al. [1993]. The camera head contains a thermoelectric cooler and the electronics necessary to run the CCD and extract the data. A high-speed fiber optic data link is employed to transfer the data to the sequence controller and user's station, which can be several kilometers distant. The user has complete control of the CCD through the fiber optic link, allowing the frame rate and operating mode (binning) to be changed as required.

5.7.3 Avalanche Photodiodes

Avalanche photodiodes employ the large and noiseless gain in photocurrent obtained when hole-electron pairs created by absorbed photons collide with ions to create additional hole-electron pairs. The avalanche effect occurs when the reverse bias voltage on the device nears breakdown level. In the Geiger mode [Brown et al. 1986], the bias is adjusted so that individual photons can be counted. In principle, the bias can be adjusted to give a detection probability of near unity, but thermal carriers cause spontaneous breakdown. The high intrinsic quantum efficiency of silicon is therefore tempered by the need to reduce the detection probability to about 0.6 in order to avoid spontaneous breakdown; this results in an effective quantum efficiency of about 50%. The maximum counting rate is determined by the time required to quench the avalanche effect. Using passive quenching, the minimum quenching time is about 500 ns, limiting the count rate to about 1.5 MHz.

The breakdown voltage of avalanche diodes also varies with counting rate, because of heating of the

photodiode junction. Active cooling is necessary to keep the junction temperature below −20°C in order to minimize dark current. The use of arrays at low light levels is complicated by the fact that the avalanche process generates a small amount of light that can trigger breakdown in adjacent elements, so these devices are packaged as separate modules, coupled to the optical system through fibers.

In spite of these problems, avalanche diodes have been used successfully for signal detection and telescope guidance [Nightingale 1991]. Avalanche diode modules produced by EG&G Optoelectronics Canada are used in the curvature sensing system developed at the University of Hawaii [Graves et al. 1994], and have been used for fast guiding at the Phillips Laboratory Starfire Optical Range [Fugate et al. 1993].

Avalanche diodes with characteristics specially tailored to adaptive optics are now being developed at EG&G [Dautet et al. 1993]. The effective quantum efficiency has been increased to over 85% at 0.5 μm and the use of active quenching allows photon-counting rates of 20 MHz. These devices have several advantages over CCDs, including zero read noise and the fact that they can be electronically gated, which is significant when using laser beacons. The data can also be read out in parallel, eliminating the time required for frame transfer and readout in CCD arrays.

S.7.4 *Photomultiplier Arrays*

Photomultiplier arrays were used in the first large-scale adaptive optics system, the Compensated Imaging System (1982), which employed four photon-counting arrays, each consisting of 152 end-on photomultipliers [Hardy 1993]. This detector is described in section 5.4.5. A similar detector using arrays of 60 photomultipliers was used in the Atmospheric Compensation Experiment (ACE); the detector was built in the early 1980s and later tested on the 60-inch telescope at Mount Wilson Observatory [Shelton et al. 1993]. At the time these systems were designed, the photomultiplier was the only viable method of photon counting at the relatively high rates required for adaptive optics. The resulting detector was bulky and expensive, and the photocathodes limited its quantum efficiency to about 10% in the visible band.

Compact end-on photomultipliers and multianode devices are currently available, but their quantum efficiency is still much lower than that of silicon detectors, and they require relatively high voltages for the dynode structure. With the availability of silicon devices that have quantum efficiencies of 80–90% at visible wavelengths, there appears to be little incentive to consider the use of photomultipliers for future adaptive optics systems.

5.7.5 *Intensified Devices*

One of the first steps in the development of high-sensitivity detectors was to combine an image intensifier with a silicon photodetector array, using a fiber optic faceplate to couple the light from the target phosphor into the array. The gain in these devices was sufficient to eliminate amplifier or readout noise in the detector arrays, but the quantum efficiency was limited by the photocathode. These intensified arrays had many practical problems, not least of which was obtaining good optical transmission between the two separate devices involved. The presence of even the smallest air gap between the fiber optic faceplate and the detector window caused a large loss of light.

A more successful development has been the electron-bombarded CCD, which consists of a photocathode and electron accelerator front-end integrated with a silicon detector array [Richard et al. 1990]. The electrons emitted by the photocathode are accelerated and focused directly on the CCD itself. The energy gain of the electrons, G, can be as high as 2000, which reduces the effect of readout noise on the CCD to negligible proportions. Unfortunately, the price paid for this reduction is a heavy one, because the quantum efficiency of photocathodes at visible wavelengths is generally much lower than that of silicon, the former being typically about 10% compared with over 80% for CCDs with direct illumination. The high-energy electrons are liable to damage the CCD structure, although special types have been developed to work in this mode. A detector of this type is used in the Come-On Plus adaptive optics system [Rousset et al. 1994].

5.7.6 *Infrared Detectors*

While most adaptive optics systems employ wavefront sensors operating at visible wavelengths, there is an incentive for wavefront sensing in the IR bands, because there are many bright sources in the 1.25, 1.62, and 2.2 μm bands that have no visible counterparts. Several IR cameras have been built for astronomical imaging at wavelengths between 1 and 5 μm, but these are mostly unsuitable for use in wavefront sensors. Imaging cameras generally have a large number of pixels and are designed for long exposures with relatively slow readout rates. Frame rates of over 100 Hz are required for adaptive optics, resulting in high readout noise.

Solid-state arrays using indium antimonide (InSb) and mercury-cadmium-telluride (HgCdTe) have been developed for astronomical imaging. A 64 x 58 pixel, InSb focal plane array built by Santa Barbara Research Center (SBRC) has been described by Orias et al. [1986]. This device operates in the 1-5 μm spectral band, with a maximum quantum efficiency of 80% at 2 μm. The readout noise is about 300 electrons per pixel. The design and performance

of cameras using this focal plane array have been described by Fowler et al. [1987] and McCarthy et al. [1990].

A hybrid IR camera has been used in a Shack-Hartmann wavefront sensor in the Come-On Plus program [Gendron et al. 1991]. This camera employs a 64 x 64 HgCdTe diode array mounted on a silicon CCD. The quantum efficiency at 2.2 μm is reported to be greater than 50%, with readout noise of 30 electrons at 100-Hz frame rate. Rockwell International of Anaheim, CA have developed 256 x 256 pixel HgCdTe arrays operating in the 1-2.5 μm spectral band. They are used in the NICMOS (Near Infrared Camera/Multi-Object Spectrometer) installed on the Hubble Space Telescope in 1997.

5.8 Neural Networks

5.8.1 Introduction

A neural network is a general form of information processor, used in cases where a logical measurement process or control scheme is difficult to define. Neural networks may be "trained" to perform specific tasks by presenting them with a large number of practical cases and adjusting their parameters until a useful relation is obtained between the input and output. This approach is radically different from the conventional method of designing control schemes, in which specific characteristics of a process are selected a priori and exploited to obtain the required result. In adaptive optics, for example, wavefront errors are usually measured by spatially subdividing the beam and determining local properties of the wavefront, such as gradient or curvature. This method requires the zones to be contiguous, or at least separated by less than the coherence distance of the wavefront.

Arrays of large telescopes offer the potential for a huge leap in angular resolution if the problem of combining optical data from isolated apertures can be overcome. Conventional adaptive optics cannot be used because the wavefronts from each aperture may be uncorrelated, resulting in significant absolute phase (piston) and overall tilt errors. Recovery of piston errors is possible using pairs of in-focus and out-of-focus images with phase-retrieval techniques [Gonsalves 1982, Paxman and Fienup 1988], but higher order aberrations are difficult to measure because of the nonlinear nature of the optical imaging process. Angel, Wizinowich et al. [1990] have shown that a neural network can be trained to recognize wavefront aberrations from the intensity distributions of the image pairs. In principle, a network of this type can provide outputs corresponding to either a zonal or modal representation of the wavefront, thus functioning as a combined wavefront sensor and reconstructor.

5.8.2 Principle of The Perceptron

The perceptron is a type of neural network that consists of a parallel processor with three layers, as shown in figure 5.32(a). The input data, which in this case are the intensities from each pixel of the in-focus and out-of-focus detector arrays, are applied to the input nodes /,. The input nodes are connected through a weighting network to an array of "hidden" nodes H_j. The weights $W_{i,j}$ are linear transmission factors determined during the training process; they comprise the first layer of the processor. The inputs to the second layer are

$$H_j = \sum_i I_i W_{ij} \tag{5.66}$$

The second or "hidden" layer implements the nonlinear sigmoid function that acts as a threshold or switch, controlling the transmission between the input side H_j and the output side H_i'. This function is

$$f(u_j) = \frac{1}{1 + e^{-u}} = \frac{1}{1 + e^{-(H_j - \theta_j)}} \tag{5.67}$$

where θ_j is the threshold for node/ The value of this threshold is determined during the training process. The outputs of the second layer are then

$$H_j' = H_j f(u_j) = \frac{H_j}{1 + e^{-(H_j - \theta_j)}} \tag{5.68}$$

These signals pass through another set of weighting functions W_{jk}, determined in the training process, to the output nodes O_k. The output at node k is

$$O_k = \sum_j W_{jk} H_j' \tag{5.69}$$

The externally supervised training process employs a back-propagation algorithm that repeatedly presents input training patterns and adjusts the parameters of the network, W_{ij}, θ_j, and W_{jk}, to minimize the error signal (the difference between the network output and the known wavefront). This process involves a large number of trials, typically 10^5 to 10^6.

5.8.3 Application to Adaptive Optics

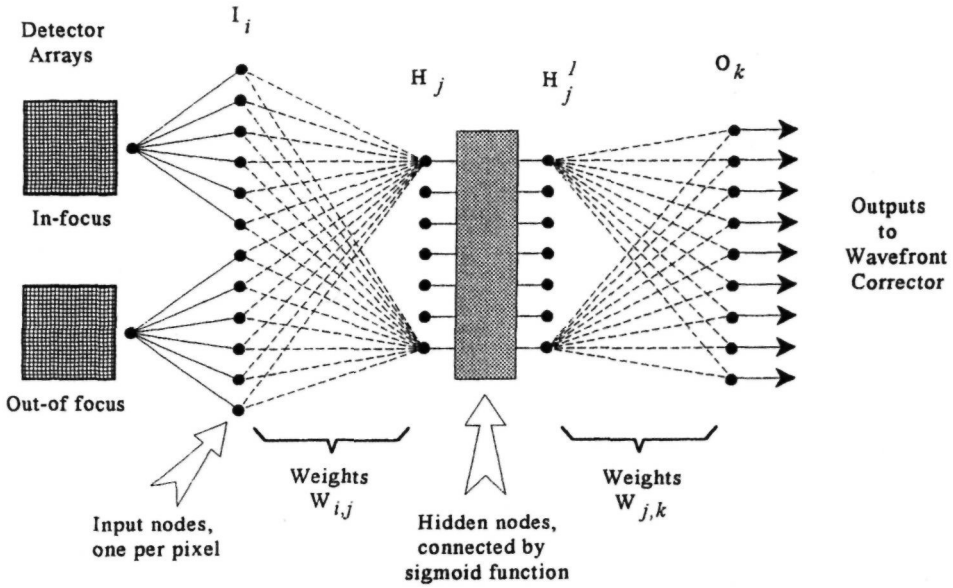
The use of a neural network to control the wavefront corrector in an adaptive optics system is shown in figure 5.32(b). In practice, two separate detector arrays may not be necessary. Barrett et al. [1992] have proposed a system for time-sharing a single detector array using a variable focus device with a birefringent lens. Using simulations and non-real-time experiments, they found that using two 4×4 detector arrays with 32 input nodes and 32 hidden nodes, the system was capable of estimating the first 12 Zernike modes.

The following are some of the advantages of neural networks:

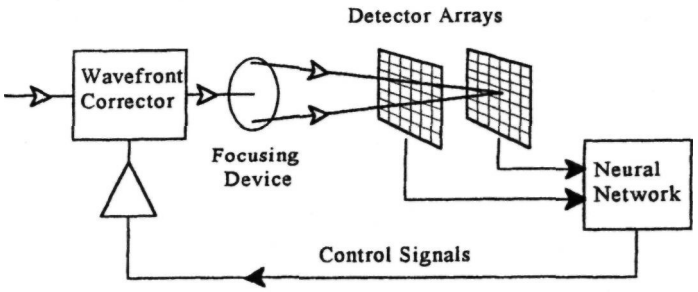
- The architecture is parallel, so the processor operates at high speed.

- The optical system is simple.
- The neural network operates on the point spread function of the imaging system, so it includes all the optics.

The algorithm and hardware are easily adaptable to changing conditions.



(a) Principle of Neural Network (Perceptron)



(b) Implementation for Adaptive Optics

Figure 5.32 Principle of the perceptron type of neural network and its application to controlling a wavefront corrector in adaptive optics.

6 Wavefront Correctors

6.1 Introduction

Since the invention of the telescope in the early seventeenth century, the need for lenses and mirrors of high accuracy and stability has been a major challenge to scientists and engineers. While the size of telescope optics has steadily increased, the surface accuracy that must be achieved remains constant at a small fraction of the wavelength of light. The classic solution to maintaining the surface figure of large telescope mirrors against the onslaught of gravitational and temperature changes has been to use the most rigid and stable materials available: glass, quartz, and ceramics.

A revolution has recently occurred in the development of telescope optics. Large primary mirrors now employ flexible faceplates that are actively controlled to maintain their figures. In addition, small high-speed deformable mirrors are included in the optical train to compensate atmospheric turbulence in real time. To perform these tasks, wavefront correctors must combine the high degree of mechanical stability essential for optical components with the additional capability of making precisely controlled adjustments to an optical wavefront. These requirements have led to the development of a totally new class of optical device, the evolution of which started in the late 1960s and is still in progress. The most successful wavefront correctors combine specially developed materials and processes into a single integrated structure. The absence of suitable wavefront correctors was one of the main reasons for the long hiatus between Babcock's initial concept of adaptive optics in 1953 and its first flowering in the early 1970s.

Wavefront correctors are of two basic types: (1) inertial components in which an optical surface is mechanically moved to change the optical pathlength of a beam of light; and (2) refractive components in which the index of an optically transmitting material is changed to modify the velocity of propagation — again, changing the optical pathlength. The most common inertial components are deformable mirrors, which have high optical efficiency, wide spectral bandwidth, and a large dynamic range. Because of these desirable features, they are used almost universally in adaptive optics systems today. There are three main families: discrete actuator mirrors using a continuous faceplate, bimorph mirrors in which the actuator is combined with the faceplate, and segmented mirrors. The disadvantages of inertial components result from their mechanical nature, such as limited spatial resolution and temporal frequency response, although these are sufficient for most adaptive optics applications.

The main advantages of refractive devices stem from their solid-state nature; there are no inertial restrictions and their intrinsic resolution is determined by molecular structure. Such devices are usually electrically controlled; using microelectronics technology, spatial resolutions in the micrometer range are achievable. Liquid crystals appear to be the most promising refractive devices for astronomical adaptive optics. They are best suited for operation in the infrared (IR) band, because of the

relatively large dispersion at visible wavelengths. Their slow response time, which was initially a major impediment, can be improved by using dual-frequency operation. Because of the potential advantages of refractive devices, their development is continuing, and they may, in future, challenge deformable mirrors as the preferred wavefront corrector for astronomical adaptive optics at longer wavelengths.

The spatial characteristics of various types of wavefront corrector are compared in figure 6.1, which shows the range of zone spacings, degrees of freedom, and optical apertures for the main technologies employed. The large active mirrors used as telescope primaries may be several meters in diameter, with actuator spacings of 100 mm or more. Conventional deformable mirrors for real-time turbulence compensation are typically 50-200 mm in diameter and use discrete actuators with spacings of 7-10 mm. The largest practical number of discrete actuators is in the region of 10^4. Deformable mirrors can be made as monolithic devices, using a pattern of integrated addressing electrodes. By this means, the actuator spacing can be reduced to about 1 mm; however, the stroke of such devices is limited by the stress produced in bending the faceplate. To achieve even higher spatial resolution and more correction zones, it is necessary to use refractive devices, such as liquid crystals, in which the spatial resolution is determined by the addressing electrodes, for which scale sizes in the micrometer range are possible using microcircuit technology.

6.2 Wavefront Correctors for Astronomy

6.2.1 Requirements

The requirements for wavefront correctors used in astronomical adaptive optics are generally specified in the following terms: (1) the number of actuators, or degrees of freedom, and their spatial arrangement; (2) the dynamic range of the wavefront correction; (3) the spectral range; (4) the temporal frequency response of the corrector and its driving electronics; and (5) the surface flatness or lowest residual error obtainable. In addition to these first-order requirements, several additional items are usually specified, including: (6) the influence function of the actuators; (7) the hysteresis; and (8) the power dissipation in the corrector and drivers. There may also be requirements on the reliability and repairability of the wavefront corrector, which is often the most expensive component in an adaptive optics system. The significance of these wavefront corrector specifications will now be discussed briefly.

Number and Spatial Arrangement of Actuators

The number of actuators or degrees of freedom within the active area is the basic descriptor of all wavefront correctors, as it determines the order of compensation that the corrector can produce. In zonal correction devices, a square array of actuators is usually employed, as this is compatible with the

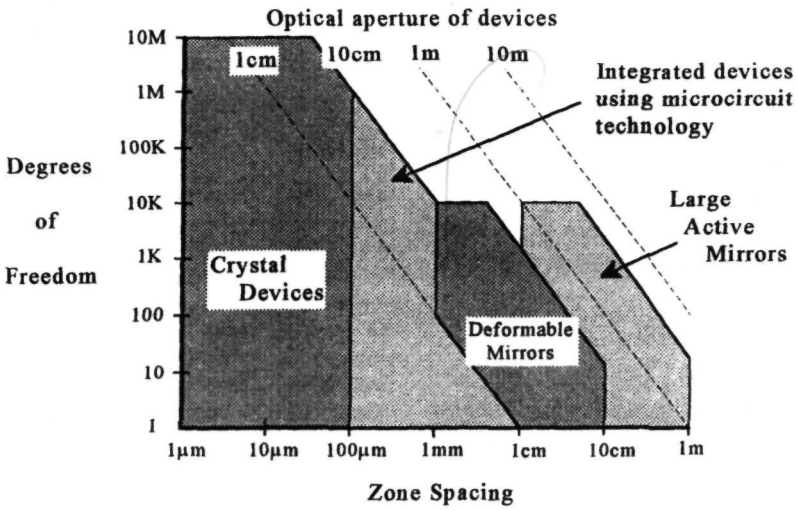


Figure 6.1 Wavefront corrector characteristics. This chart depicts the spatial characteristics of generic types of wavefront corrector, ranging from devices such as liquid crystals that function on a molecular scale, to large active mirrors having dimensions of several meters.

pattern of sensing zones in photon detectors, such as CCDs. Square arrays do not fit neatly into circular apertures; hexagonal arrays produce more uniform actuator spacing and are preferable if a suitable wavefront sensor is available. When related to the size of the telescope pupil, the actuator spacing d in relation to the value of r_0 (the coherence length of the wavefront disturbance), determines the "fitting error" of the corrector and is consequently an important parameter of the adaptive system. Full wavefront compensation with zonal compensation requires an effective actuator spacing of $1.0–1.5\ r_0$.

Dynamic Range of the Wavefront Correction

The dynamic range or *stroke* must be sufficient to compensate the peak to valley excursions of the wavefront, which is usually specified as ±2.5 times the standard deviation, giving a .994 probability that wavefront excursions will be corrected. The wavefront excursions to be corrected are proportional to $(D/r_0)^{5/6}$, where D is the optical aperture and the turbulence coherence length r_0 is specified at 0.5 μm. When specified in terms of optical pathlength (meters), the wavefront excursion due to turbulence is independent of wavelength. The stroke required is greatly reduced by compensating overall tilt with a separate fast-steering mirror. The stroke of zonal deformable mirrors for astronomical adaptive optics is typically ±2 μm, giving a peak-to-valley wavefront correction of 8 μm. This is sufficient to compensate an 8-m aperture with moderate turbulence (r_0 = 0.07 m), assuming a separate tip-tilt corrector.

Spectral Range of the Corrector

Wavefront sensing is usually performed in the visible band at 0.4–0.7 μm, while the science observations may use IR bands between 1 and 4 μm. The optical pathlength corrections must be constant over the entire spectral range. Efficient operation over wide spectral bands is readily accomplished with reflective devices such as deformable mirrors, but is more difficult with refractive correctors because of dispersion.

Temporal Frequency Response of the Corrector

The temporal spectrum of turbulence is determined by the wind velocity profile in the atmosphere and, in common with other atmospheric parameters, it can vary widely within short periods of time. The temporal response of a wavefront corrector may be limited by the device itself, such as the response time of a liquid crystal, or it may be caused by limitations in the driver electronics, such as the current sourcing capacity of drivers for piezoelectric actuators. It is important therefore to specify the frequency response of the whole correction subsystem, which is designed as part of the adaptive optics servo loop. Small deformable mirrors have a bandwidth in the region of 1 kHz.

Minimum Residual Error of the Wavefront Corrector

The minimum error may be specified under passive or active conditions. If the telescope is to be used while the adaptive optics system is inoperative, and if the corrector cannot be optically bypassed, it is necessary for its passive wavefront quality (no drive signals) to be very good, or at least compatible with that of the main optical system. On the other hand, if the reliability of the correction device is not at issue, and there is sufficient dynamic range to accommodate the expected optical errors, then the optical quality can be specified with active flattening. In this case, wavefront errors in the correction device are removed dynamically by a calibration process.

Influence Function of the Actuators

In zonal deformable mirrors, the influence function defines the response of each actuator. The shape of the influence function determines how well the deformable mirror is able to compensate wavefront errors. It defines the minimum error after compensation, usually a ripple with a period equal to the actuator spacing. Influence functions should mesh together with the minimum of ripple to produce a smooth combined response. The shape of the influence function is controlled mainly by the relative stiffness of the faceplate and actuators, and by the design of the attachment points. There is typically a 10% response at the adjoining actuator.

Hysteresis

Most of the actuators used for deformable mirrors are made from ferroelectric and ferromagnetic materials, which all show some degree of hysteresis. Piezoelectric ceramics such as PZT (lead zirconate titanate) typically have 10-20% hysteresis, while for the electrostrictive material PMN (lead magnesium niobate) it is much smaller, usually 1-3%. In closed-loop adaptive optics systems, a moderate amount of hysteresis is relatively unimportant, as it is overcome by the large gain of the feedback loop. But in open-loop systems, which rely on accurate calibration of the wavefront sensor, hysteresis is detrimental to system performance. Hysteresis can be reduced considerably by driving the actuators with charge amplifiers, in which the input voltage controls the amount of electric charge transferred to or from the actuator.

Power Dissipation in the Corrector and Drivers

Power dissipation in the actuators can produce unwanted heating in a wavefront corrector. The ferroic actuators used in most deformable mirrors have relatively low power dissipation, because they act electrically as capacitors. Power losses are due mainly to dissipation in the dielectric. At the other extreme, electromagnetic force actuators, such as the solenoid "voice coils" used on large active mirrors, have relatively high power dissipation due to ohmic heating of the coil. A constant current is required to maintain the force. Linear driver amplifiers may dissipate more heat than the actuators themselves, but as the drivers can be located remotely, the telescope environment is not compromised.

Typical requirements for three different types of wavefront corrector used in astronomical adaptive optics are summarized in table 6.1. The three applications are:

1. turbulence compensation in the B, V, and R visible bands (0.4–0.7 μm);
2. turbulence compensation in the J, H, K, and L infrared bands (1.25–4 μm); and
3. wideband adaptive secondary mirror.

The optical pathlength variations produced by turbulence are independent of wavelength, so the requirement for surface deformation is the same in the visible and IR bands.

6.2.2 Types of Wavefront Corrector

The main types of wavefront corrector used in adaptive optics are listed in table 6.2. Most wavefront correctors used today are zonal deformable mirrors employing an array of actuators to deform a continuous faceplate. Deformable mirrors have been made in sizes up to 2000 actuators. Segmented (zonal) mirrors have been made with up to 500 segments, comprising 1500 degrees of freedom. Modal mirrors have been developed for specific purposes, such as compensating systematic distortion in laser wavefronts, but they are generally restricted to low-order correction up to Zernike mode 6 (astigmatism), for which the optimal D/r_0 ratio is 6. For astronomical adaptive optics, they would therefore be usable with apertures only up to about 1 m at visible wavelengths, but with much larger apertures at IR wavelengths. Zonal deformable mirrors are normally used with conventional wavefront tilt sensors. Bimorph mirrors employing local curvature control are compatible with adaptive optics systems employing curvature sensing.

6.3 Actuators

6.3.1 Actuator Types

Actuators are conveniently divided into two categories: *force* actuators exert a controllable force that is ideally independent of their displacement, while *displacement* actuators produce a controllable displacement that is ideally independent of the reaction. These categories represent the two extremes of mechanical stiffness, with force actuators having very low (ideally zero) stiffness, while displacement actuators possess very high (ideally infinite) stiffness. Practical actuators fall somewhere in between. The important factor is the actuator stiffness in relation

Table 6.1 Wavefront Corrector Requirements for Astronomical Adaptive Optics

Requirement	Visible	IR	Adaptive Secondary Mirror
Spectral band, μm	0.4–0.8	1–4	0.4-35
Number of actuators	50–1000	6-200	100 + (zonal)
Diameter, cm	10-30	5-20	20-100
Control mode	Zonal	Modal or zonal	
Interactuator spacing, cm	0.5–1.0		
Actuator array pattern			
Surface deformation:			
overall, μm peak	±2	±2	
interactuator, μm peak	±1.5	±1.5	
Interactuator coupling, %	5-10	5-10	
Bandwidth, Hz (incl. drivers)	1000	300	
Actuator hysteresis, %	1-3	1-3	
Surface figure accuracy:			
passive, μm rms	0.05-0.02		
active, μm rms	0.03-0.01		
Optical efficiency, %	> 90	>95	>90
Surface smoothness, angstroms	15–40		

Table 6.2 Wavefront Correctors Used in Adaptive Optics

Segmented mirrors	
Piston only	Square or hexagonal array; one actuator per segment
Piston plus tilt	Square or hexagonal array using three dedicated actuators per segment Triangular or hexagonal array using three shared actuators per segment
Monolithic piezoelectric mirrors	Thin glass faceplate bonded to PZT block with integral electrodes
Thin-plate deformable mirrors	Glass or quartz faceplate Discrete axial actuators, using bonded multilayer stacks of piezoelectric or electrostrictive material
Bimorph mirrors	Local curvature produced by piezoelectric actuators bonded directly to the back of a thin faceplate
Modal mirrors	Bending moments applied overall to a continuous faceplate to produce low-order (Zernike) corrections
Adaptive secondary mirrors	Employed in Cassegrain and Gregorian telescopes Combines IR chopping capability with tip-tilt and higher order corrections. Relatively large size
Membrane mirrors	Metal or plastic reflecting membrane Deflected by electrostatic force or air pressure
Integrated wavefront correctors	Single-crystal silicon facesheet with monolithic PLZT[a] actuator structure, 1-mm spacing, (Itek).
Dirigible optics	Actively controlled rigid lenses and mirrors Wavefront corrections are produced by *directed* displacements and tilts
Electro-optical correctors	Nematic liquid crystals

[a] PLZT, lanthanum-modified lead zirconate titanate.

to that of the object being controlled, such as the faceplate of a deformable mirror.

Hydraulic and electromagnetic actuators are good examples of force actuators. Hydraulic actuators employ a fluid, the pressure of which can be controlled, and use a piston to transfer the force to the object, while electromagnetic actuators employ an electric current to generate a magnetic field that reacts against a permanent magnet to produce the force. Electromagnetic actuators are built in both linear (voice coil) and rotational (stepper motor) configurations. When a motor is used in conjunction with a screw mechanism, its effective stiffness is greatly increased by the mechanical advantage of the screw, converting it into a displacement actuator in which the position of the nut is precisely controlled by rotation of the screw. The stiffness of a screw is determined by its cross-sectional area, its length, and the elastic modulus of the materials. The stiffness can be reduced by placing a spring in series with the actuator. Electromechanical actuators of this kind are used for active control of the surfaces of large primary mirrors. Ferroic devices, in which strain is generated in the bulk material by an applied electric or magnetic field, are the most commonly used type of displacement actuator for deformable mirrors. Piezoelectric, electrostrictive, and magnetostrictive actuators are examples of this category.

Thermal actuators have been developed for special purposes where the power dissipation and relatively slow response are unimportant. Available devices include:

1. bimetal alloy thermal actuators, which are made in both strips and coils, that produce linear or rotational motion;
2. shape-memory alloy thermal actuators, which employ nickel-titanium alloys that change their shape with temperature due to a crystalline phase change;
3. high-output paraffin (HOP) thermal actuators, which employ a phase change in paraffin wax to produce a linear motion.

These devices are all simple and reliable, with long cycle lives, making them potentially useful for control of large active mirrors, providing that the thermal

energy is contained. In their present form, they are unsuitable for use as actuators in the small, high-bandwidth deformable mirrors used for turbulence compensation. This situation could change with further development to reduce their size and thermal inertia.

Either force or displacement actuators can be used to control a deformable mirror. Force actuators have the advantage that small axial displacements of the actuator itself do not significantly change the applied force. This is important in large active mirrors as it relaxes the stiffness requirements for the supporting structure. The main disadvantage of force actuators is that the surface displacement depends on the stiffness of the faceplate, which may not be known precisely. For this reason, force actuators require feedback or calibration devices to ensure accurate operation.

Displacement actuators are used primarily in segmented and deformable mirrors less than about 50 cm in diameter. A rigid and dimensionally stable reference surface or baseplate is essential with these actuators, because any deflections due to thermal, structural, or gravitational changes are transmitted through the actuators to the faceplate. Ferroic devices are well suited for use as displacement actuators in deformable mirrors because of their small size, excellent stability, and high stiffness. Characteristics of these devices are listed in table 6.3 and are discussed in the following sections.

6.3.2 Piezoelectric Actuators

The direct piezoelectric effect is the creation of an electric charge in a material under an applied stress. The inverse piezoelectric effect is the production of strain-inducing stress as the result of an applied electric field. These effects were originally observed in quartz crystals. The material used for deformable mirror actuators is lead zirconate titanate, $Pb(Zr,Ti)O_3$, commonly known as PZT. It has a strong piezoelectric effect, with a maximum strain, A///, more than 0.001. The PZT is made in the form of a ceramic that is initially isotropic with randomly oriented dipoles. To produce piezoelectric effects, it must be poled, which is accomplished at a temperature of about 150°C by applying a direct current (DC) electric field to align the dipoles parallel to the field. After poling, the piezoelectric deformation is linear and bipolar, with hysteresis typically between 10% and 20%.

The key parameter in determining performance is the piezoelectric constant d_{33}, which relates the mechanical strain to the applied electric field. The strain for an unloaded piezoelectric element is given by

Table 6.3 Characteristics of Ferroic Actuators

Material Property	Piezoelectric	Electrostrictive	Magnetostrictive
	PZT	PMN	Terfenol-D
Composition	$Pb(Zr_{0.52},Ti_{0.48})O_3$	$Pb(Mg_{0.33}Nb_{0.67})O_3$	$Tb_{0.3}\,Dy_{0.7}\,Fe_{1.9}$
Maximum strain x 10^{-6}	1000	1000	2000
Strain coefficient	d_{33}= 5x 10^{-10} m V^{-1}	m = 3 x 10^{-15} m V^{-2}	λ = 5 x 10^{-9}oersted^{-2}
Hysteresis, %	10-20	< 1	10-20
Thermal expansion coefficient x 10^{-6} per °C	2-5	2	12
Elastic modulus, x 10^{10} N m^{-3}	5	12	3
Density, 10^3 kg m^{-3}	7.5	7.3	9.2
Tensile strength, 10^6 N m^{-3}	76		
Curie temperature, °C	200-350	0	370
Temperature range, °C	0–100	$T\pm10$ r = 0–40	-10 to > 100
Dissipation factor, %	3	8	
Typical actuator			
Active element length, mm	20	20	50
diameter, mm	6	6	6
Number of layers	130	130	N/A
Layer thickness, μm	150	150	N/A
Sensitivity	100 nm V^{-1}	100 nm V^{-1}	25 μm A^{-1}
Max. drive	150 V	150 V	2.0 A
Typical capacitance, μF	1	4	N/A
Force for 10% strain reduction, N	450	700	

N/A, not applicable

$$\begin{aligned} \Delta l &= \text{length of element} \times d_{33} \times \text{electric field} \\ &= l\, d_{33} V/l \\ &= d_{33} V \end{aligned} \tag{6.1}$$

where

l = length of the element
V - applied voltage

To increase the stroke for the same applied voltage, the PZT is made in the form of disks of thickness t, stacked with alternating polarity as shown in figure 6.2. The disks are in series mechanically, but are connected in parallel electrically so that the same voltage is applied across each disk. In this case, the stroke of the complete stack of N disks of total length Nt is

$$\begin{aligned} \Delta z &= N t d_{33} V/t \\ &= N d_{33} V \end{aligned} \tag{6.2}$$

The value of rf_{33} is typically in the region of 5×10^{-1} m V^{-1}. The stroke of a piezoelectric actuator can be made arbitrarily large by increasing N. However, the minimum disk thickness is limited by the maximum electric field that can be applied to PZT before breakdown occurs. If this field is $E_b = V/t_{min}$, then the maximum strain achievable is

$$\Delta z/z = E_b d_{33} \tag{6.3}$$

A typical value of E_b is 2×10^6 V m^{-1}, which gives $\Delta z/z = 0.001$; that is, the maximum strain for PZT is about 1000 parts per million.

A ceramic material, PZT is made in many shapes and sizes, such as slabs, disks, rods, and tubes. A disadvantage of PZT is the large voltage required to obtain a useful stroke, which for deformable mirror actuators is about 4 μm. By using multilayer stacks of thin disks with intervening electrodes, as mentioned above, the required field can be achieved with relatively low voltages. The first actuators made in this way employed springs to preload the stack of disks, so they were always under compression. This arrangement caused gradual dimensional changes or "creep." The next step in the development of PZT actuators was to bond together a number of thin disks to form a solid block several centimeters in diameter, which was then diced vertically to produce an array of separate multilayer actuators. This procedure was a major advance, because the actuators could both push and pull the mirror faceplate and were under no stress in the null position. In addition, the actuators had identical ancestry, because they were all formed from the same original set of disks, producing good uniformity in terms of sensitivity, thermal expansion, and aging. The final step in PZT actuator development has been to produce completely monolithic stacks, formed of 100 or more layers with integral electrodes, each layer about 150 μm thick, which are fired as a single unit. Typical devices are 20 mm long and 6 mm in diameter, producing a stroke of 15 μm (unloaded) for an applied voltage of 150.

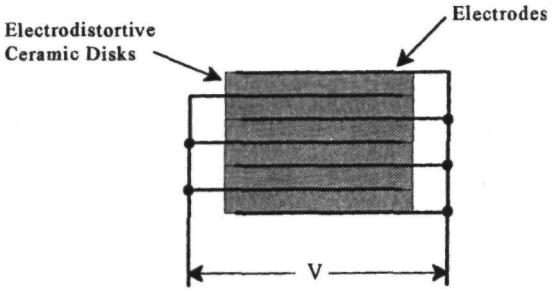


Figure 6.2 Linear actuator using stacked electrodistortive ceramic disks. The disks are in series mechanically, but are connected in parallel electrically so that the full voltage is applied to each disk.

The response of a PZT actuator is shown in figure 6.3. A good feature of PZT is its ability to operate over a wide temperature range of —25 to +100°C. Disadvantages include its relatively high hysteresis and the fact that, because it is a poled material, it tends to lose sensitivity over time.

6.3.3 Electrostrictive Actuators

Electrostriction is a property of all dielectrics, but it can only be observed in non-piezoelectric materials. It is a second-order effect in which the strain is proportional to the square of the applied electric field. The electrostriction constant m relates the mechanical strain to the applied electric field. The strain for an unloaded electrostrictive element of length / is given by

$$\begin{aligned} \Delta l &= \text{length of element} \times m \times (\text{electric field})^2 \\ &= l m (V/l)^2 \\ &= m V^2/l \end{aligned} \tag{6.4}$$

For a multilayer device consisting of N disks of thickness t, the stroke is

$$\Delta z = N m t (V/t)^2 \tag{6.5}$$

The electrostrictive material lead magnesium niobate, $Pb(Mg_{0.33},Nb_{0.67})O_3$, commonly known as PMN, is a ferroelectric material in which the large electrostrictive strain is due to a huge dielectric constant rather than to the electrostrictive coupling constant itself. The strain response closely tracks the dielectric constant as a function of temperature. The Curie temperature of PMN is near 0°C. At the optimal working temperature of around 25°C, the material has a sensitivity of 375 ppm strain at $600 V mm^{-1}$, maximum strain of about 0.1%, hysteresis of less than 1%, and thermal expansion of less than 1 ppm per °C, which closely matches that of low expansion glasses. No electric poling is necessary, as is the case with PZT, making the material inherently stable. These charac-

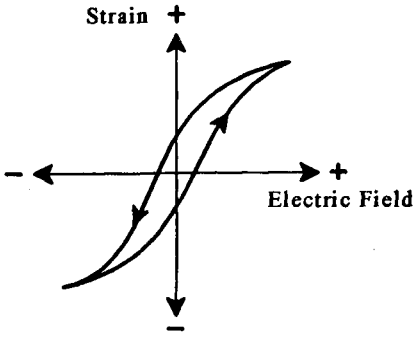


Figure 6.3 Response of typical PZT actuator.

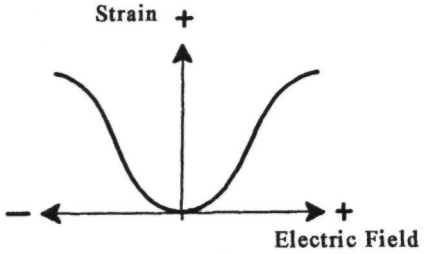


Figure 6.4 Typical response of PMN actuator.

teristics make PMN almost ideal for deformable mirror actuators.

Multilayer PMN actuators using thin layers that are diffusion bonded during the final sintering process were developed during the 1980s by Litton Itek Optical Systems [Ealey and Davis 1990]. These actuators were used successfully in a series of high-performance deformable mirrors ranging in size up to 2000 actuators [Ealey and Washeba 1990]. The actuators were built in cylindrical sections about 10 mm in length and 4 mm in diameter, each containing about 40 layers. A typical longitudinal response function is shown in figure 6.4. The transverse strain is about half of the longitudinal value.

The PMN actuators have two main drawbacks: nonlinear response and temperature sensitivity. The quadratic relation between strain and applied voltage is not a serious problem, especially in closed-loop feedback systems, where it affects the loop gain but not the null. It is usually handled by biasing the operating point to half of the peak input voltage, allowing bipolar control signals. The temperature sensitivity is a far more serious limitation, as the characteristics of PMN change markedly as the operating temperature drops toward the Curie temperature [Blackwood et al. 1991]. The sensitivity increases to a maximum around —10°C, with a value of about twice that at 25°C, and the hysteresis rises to about 15%. With high-capacitance actuators such as PMN, the hysteresis can be minimized by using a charge amplifier as a driver. It has been found that placing a capacitor in series with the actuator also minimizes hysteresis.

6.3.4 *Magnetostrictive Actuators*

Magnetostriction is the generation of strain in ferromagnetic materials in response to an applied magnetic field; it is the magnetic analog of electrostriction in dielectric materials. Magnetostriction was discovered as a property of nickel in the 1840s and magnetostrictive nickel alloys were subsequently used as sonar transducers. However, the strain produced in these materials is much smaller than that available in piezoelectric devices. The discovery of rare earth alloys with giant magnetostrictive properties [Clark and Belson 1972] has resulted in the development of linear actuators with properties suitable for deformable mirrors. The largest magnetostrictive strains, on the order of 1 x 10^{-2} are produced at cryogenic temperatures. Present interest centers on alloys of terbium (Tb), dysprosium (Dy), and iron (Fe), known as Terfenol-D, which produce magnetostrains exceeding 1 x 10^{-3} at room temperature, comparable with current piezoelectric and electrostrictive devices. The unloaded strain, proportional to the square of the magnetic field, is given by

$$\Delta l = l\lambda H^2 = l\lambda\left(\frac{4\pi NI}{1000 l_S}\right)^2 \qquad (6.6)$$

where

l = active length, m
λ = magnetostrictive coefficient, oersted^{-2}
H = magnetic field intensity, oersteds
N = number of turns on field coil
I = current, A
l_S = linkage distance of coil winding, m

Prestressing the Terfenol-D with an axial load increases the strain.

The composition of Terfenol-D may be varied to obtain different properties. The generic composition is $Tb_xDy_{1-x}Fe_y$, where x is approximately 0.3 and $y < 2$. The behavior of Terfenol-D at room temperature changes when a compressive stress greater than 5 MPa is applied [Clark et al. 1988]. Magnetostriction characteristics of Terfenol-D at three compressive stresses are shown in figure 6.5 [Clark 1991]. Under compression, there are three distinct regions:

1. At low fields, the levels of magnetization and magnetostriction are very small.
2. As the the field increases, the magnetic moment jumps to a region having much higher magnetostriction, resulting in an abrupt increase in the strain.

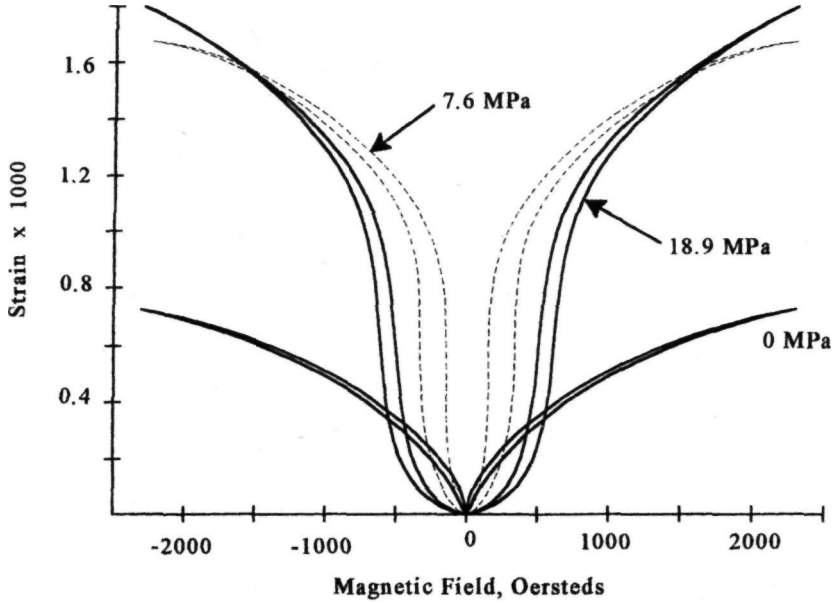


Figure 6.5 Magnetostriction of Terfenol-D at 20° C at three compressive loads [Clark 1991].

3. At high fields, the strain becomes quasi-linear.

As the compressive stress increases, a higher field is required to produce a given strain. The highly non-linear characteristics of Terfenol-D in region 2 may be used to trigger a large change in strain with a small increment of magnetic field.

Commercial Terfenol-D drive elements are made under license by Edge Technologies, Inc., using two patented processes. The free-stand zone melt (FSZM) technique produces twinned single crystal rods of 3-7 mm diameter. Grain-oriented rods of 8-50 mm diameter are produced by the modified Bridgman method. In both processes, the magnetostrictive growth axis is aligned with the transducer rod axis.

The construction of a magnetostrictive actuator is shown in figure 6.6; typical response curves are shown in figure 6.7 [Goodfriend and Shoop 1991]. To achieve quasi-linear operation with AC inputs, the operating point can be magnetically biased as shown, by including permanent magnets in the flux return path.

Terfenol-D actuators have a high energy density and produce large forces. The clamped force may be estimated from

$$F = EA\lambda H^2 \tag{6.7}$$

where

E = elastic modulus (2.5-3.5 x 10^{10} N m^{-2} for Terfenol-D)
A = the area of the transducer rod

Prestressing the actuator with a compression spring increases the strain as mentioned above.

Terfenol-D will operate over a large temperature range, from about 0 to 100°C. The magnetostrictive constant peaks at about 0°C, dropping off rapidly at lower temperatures. It remains fairly constant at temperatures up to about 50°C, then gradually falls to zero at about 400°C. The hysteresis of magnetostrictive devices using Terfenol-D is relatively high, around 20%. The temporal frequency response of magnetostrictive actuators is limited mainly by the resonant frequency of the structure containing the magnetic circuit. Typical actuators have a frequency response of 0 to 3000 Hz.

Several types of magneostrictive actuators using Terfenol-D are available commercially from ETREMA Products, Inc., a subsidiary of Edge Technologies, Inc., Ames, Iowa.

6.3.5 PLZT Actuators

To achieve the stroke of several micrometers needed for compensation of atmospheric turbulence, the piezoelectric and electrostrictive actuators described above require multiple layers. Multilayer actuators are not only difficult to manufacture, but also present problems in making reliable electrical connections to the electrodes in each stack. The space required for these electrical connections in deformable mirrors using discrete actuators puts a lower limit of about 5 mm on the center-to-center distance between multilayer stacks.

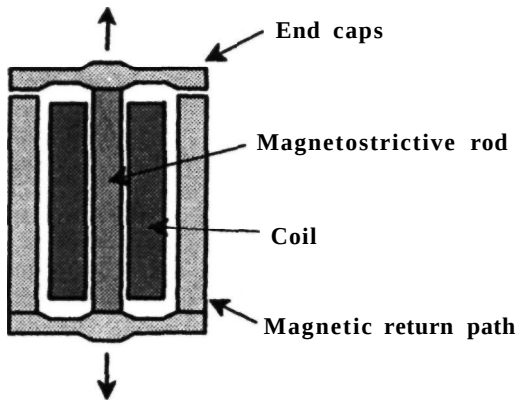


Figure **6.6** Construction of a magnetostrictive actuator. There are two modes of operation, depending on the type of magnetic return path. Magnetostriction is a quadratic function, and, with a nonpolarized magnetic path, the strain is unipolar. The addition of a magnetic bias in the return path produces a bipolar response.

It is possible that deformable mirrors with as many as 2500 actuators may be required, in future, for adaptive compensation of 8- or 10-m telescopes at visible wavelengths. With 5-mm actuator spacing, discrete-actuator mirrors of this size would be at least 250 mm in diameter and would be extremely expensive to assemble and to finish optically. Production costs of deformable mirrors could be reduced dramatically if microcircuit fabrication techniques employing a thin sheet of electrodistortive material were used to form the actuators and electrodes. High-density adaptive mirrors with interactuator spacings of less than 1 mm should be possible.

Some preliminary work on an integrated wavefront corrector device of this type was performed at Itek Optical Systems in the late 1980s, with United States Air Force (USAF) funding. It was reported by Ealey and Wheeler [1990]. The key to high-density mirrors is the development of a single-layer actuator that can operate at low voltage, preferably less than 100 V. The material investigated was PLZT, lanthanum-modified lead zirconate titanate. This can be tailored to produce many different properties, including electro-optic, photo-refractive, photostrictive, electrostrictive, and piezoelectric, depending on its composition and method of processing. Of prime interest is the electrostrictive mode in which huge strains, on the order of 1%, are achievable. This large strain capability is accompanied by large hysteresis, which approximates a square loop. One way of looking at a square hysteresis loop is to consider it as an analog memory: the strain stays where it is set after the drive voltage is removed until the device is charged or discharged by the application of a different voltage. Thus, in principle, even very large

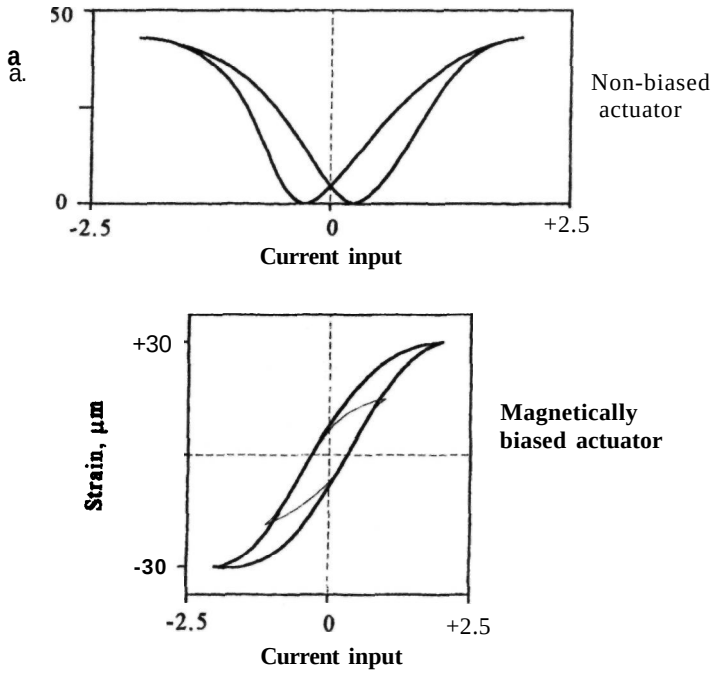


Figure *6.7* Response curves of magnetostrictive actuators.

amounts of hysteresis can be handled by an appropriate electronic driving system.

The material properties and transfer function for one type of PLZT are shown in figure 6.8. The experimental integrated wavefront corrector device, employing a 10 x 10 array of actuators with a nominal layer thickness of 200 μm, achieved a stroke of 3 μm with 300 V applied. Special techniques were developed to fabricate this integrated wavefront corrector, using a single layer of PLZT material. The individual actuators, spaced at 1 mm, were formed by cutting a square pattern of microgrooves in the layer. The PLZT layer was mounted on a glass interconnect module containing a two-dimensional array of connecting electrodes, which served also as a rigid baseplate. A silicon mirror faceplate bonded to the top of the PLZT provided the common electrode. This fabrication technique is scalable to produce large arrays.

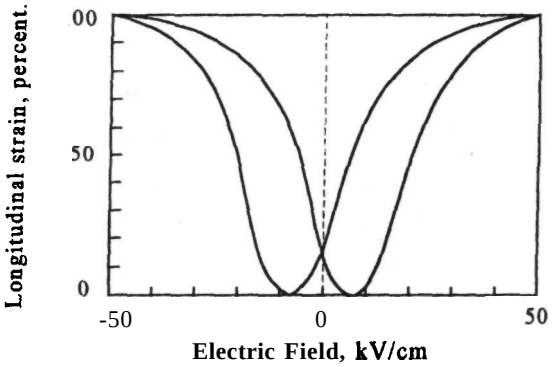


Figure **6.8** Transfer function of PLZT.

6.4 Discrete-Actuator Deformable Mirrors

6.4.1 Mechanical Design

The structure of a typical zonal deformable mirror using a continuous faceplate is shown in figure 6.9. The actuators are mounted on a massive baseplate, which is much stiffer than the mirror faceplate. The integrity of the baseplate must be maintained over the full range of environmental conditions to which the deformable mirror is exposed. The actuators are usually multilayer stacks of piezoelectric or electrostrictive ceramic material; these are diffusion-bonded with interlayer electrodes to which the control voltage is applied. Actuators of this type have been described in section 6.3. At the top of each actuator is a coupling that forms the interface between the actuator and the faceplate. The couplings act as flexures to accommodate local tilts in the faceplate; the contact areas with the faceplate are carefully sized to control the influence function of the actuators. In some cases, these couplings may be integral with the faceplate.

The faceplate itself is normally made from a stable, low-expansion material, such as quartz or Corning ULE ™. The same material is often used for both the faceplate and the baseplate, in order to match their thermal properties. The faceplate should have a high yield strength to minimize creep, and also a high tolerance to fatigue to maximize its operating

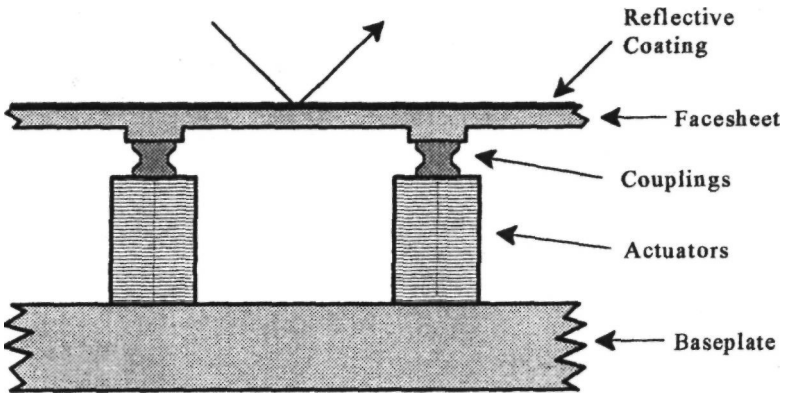


Figure 6.9 Construction of a zonal deformable mirror with discrete actuators and continuous faceplate. The baseplate is often made of the same material as the faceplate to eliminate differential thermal expansion. Thrust pads may be machined into the faceplate to control the shape of the influence function produced by each actuator.

life. Properties of typical faceplate materials are shown in table 6.4.

The mechanical design of a deformable mirror involves compromises among many conflicting factors. For example, the facesheet stiffness should be high to maintain its shape during polishing, but sufficiently low to deflect when pushed or pulled by the actuators. The bending stiffness of the actuators should be high to keep mechanical resonances well above the working bandwidth, but at the same time the faceplate must be allowed to tilt smoothly to interpolate between adjacent actuators. To resolve these conflicting requirements, it is appropriate to perform the design process in two steps. To start with, analytical models of each component are used to establish a first-order design of the deformable mirror, and to make the tradeoffs necessary to achieve the specified performance. The relation between basic physical parameters is immediately obvious in analytic expressions, especially when presented graphically, making the process of optimization direct and intuitive. When the first-order design is satisfactory, then a more rigorous analysis can be made, using a finite element model, to verify the design and perform the fine tuning. The first-order design process gives considerable insight into the factors that influence the performance of deformable mirrors and will now be outlined.

The axial stiffness of an actuator is given by

$$K_A^a = AE_A/l \tag{6.8}$$

where

A = cross-sectional area of actuator, m^2
E_A = elastic (Young's) modulus, N m^{-2}
l = length of actuator, m

Maximizing the axial stiffness of the actuators minimizes the interactuator coupling and maximizes the lateral resonant frequency of the structure.

The bending stiffness of an actuator clamped at one end is given by

$$K_A^b = 3E_A I_A/l^3 \tag{6.9}$$

where I_A is the moment of inertia of the cross-section, which is $d^4/12$ for a square actuator and $\pi d^4/64$ for a round actuator, d being the width or diameter. The bending stiffness constrains tilting of the faceplate and therefore affects the shape of the influence function of adjoining actuators. It also modifies the lateral resonant frequency of the faceplate.

The deflection of thin plates has been treated in the literature and formulas have been derived for many cases [Timoshenko and Woinowsky-Krieger 1959]. Unfortunately, these formulas involve many parameters and are cumbersome to use. For first-order design of deformable mirrors, it has been found more convenient to estimate the faceplate stiffness and deflection from beam theory, in which the effects of the main parameters are more clearly expressed. While this is obviously only an approximation, it has been found to yield good correlation with results from finite element analysis and actual measurements [Ealey and Wellman, 1989].

The actuator structure can be modeled using two cases, as shown in figures 6.10 and 6.11, which represent the two extreme cases of actuator bending stiffness. In figure 6.10, the actuator bending stiffness is very large compared with that of the faceplate, which is effectively clamped at both of the adjoining actuators. The faceplate stiffness in this case is

$$K_F^c = \frac{24EI}{a^3} \tag{6.10}$$

Table 6.4 Properties of Faceplate Materials

Material	Elastic Modulus (N m^{-2} x 10^{10})	Density (kg m^{-3})	Thermal Expansion (x 10^{-6} per°C)	Thermal Conductivity (W m^{-1} per °C)	Poisson's Ratio
Aluminum	7.1	2,700	23	180	0.35
Beryllium	30	1,870	12	150	0.05
Cer-Vit"	9.2	2,490	0.11	1.7	0.25
Copper	12	8,940	17	391	0.37
Glass (Pyrex) 7060	6	2,320	5		0.24
Molybdenum	32	10,100	5	145	0.32
Nickel	21	8,900	14	5.7	0.34
Quartz		2,200	0.6		
Silicon	12.4	2,330	2.5	150	0.26
Silicon carbide	33	2,960	2.1	105	0.13
ULE™*	6.8	2,210	0.031	1.3	0.17
Zerodur'	9	2,520	0.11	1.6	0.24

"Cer-Vit, low expansion ceramic/glass.
ULE™ ultra low expansion fused silica, Corning Inc, New York, USA.
'Zerodur, low expansion glass/ceramic, Schott Glaswerke, Mainz, Germany.

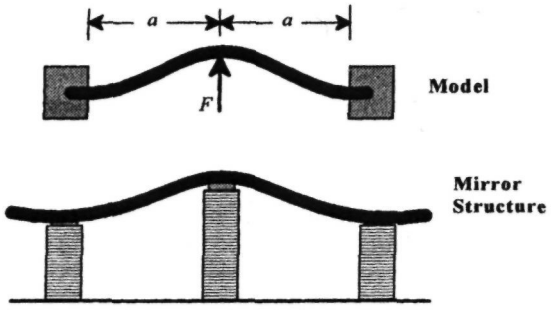


Figure **6.10** Mirror structure with high transmission of actuator stiffness to the faceplate.

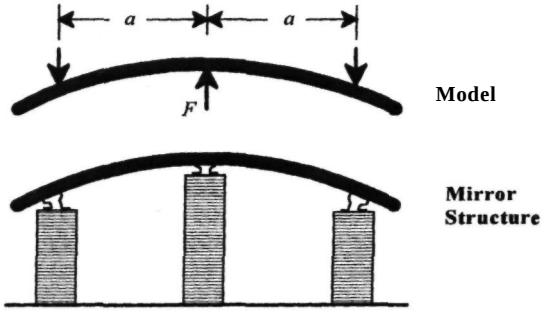


Figure **6.11** Mirror structure with low transmission of actuator stiffness to the faceplate.

where

a = actuator spacing
E = elastic modulus of faceplate
I = moment of inertia of faceplate beam section

If a force F is applied by the center actuator as shown, then the resulting deflection for the clamped case is

$$y_c = \frac{F}{K_F^c}\left[2\left(\frac{x}{a}\right)^3 - 3\left(\frac{x}{a}\right)^2 + 1\right] \qquad \textbf{(6.11)}$$

where x is the distance measured from the center actuator.

Figure 6.11 represents the condition in which the actuator bending stiffness is very small compared with that of the faceplate; this occurs when flexures are used to connect the actuators to the faceplate. The condition is modeled by a simply supported (unclamped) beam, for which the stiffness is

$$K_F^u = \frac{6EI}{a^3} \qquad (6.12)$$

The resulting deflection for the unclamped case is

$$y_u = \frac{F}{K_F^u}\left[\tfrac{1}{2}\left(\frac{x}{a}\right)^3 - \tfrac{3}{2}\left(\frac{x}{a}\right)^2 + 1\right] \qquad (6.13)$$

The influence functions produced in these two cases are shown in figure 6.12. Note that in the freely supported case, the deflection goes negative beyond the adjoining actuator. It should also be noted that the modulus of elasticity and the thickness of the faceplate (represented by the parameters E and I in the above equations) affect the magnitude of the deflection, but not its shape. The main factors controlling the shape of the influence function are the actuator bending stiffness and, to a lesser extent, the size and shape of the pusher pad that transfers the actuator force to the faceplate.

In practical deformable mirrors, the actual shape of the faceplate deflection falls in between these two models. It is convenient to define a parameter fi to specify the ratio of the actuator bending stiffness to that of the faceplate. When $fi = 1$ the actuator bending stiffness is much greater than that of the faceplate, corresponding to the clamped case of figure 6.10. At the other extreme, the condition $fi \to 0$ corresponds to zero actuator stiffness, as in figure 6.11. The actual faceplate deflection is then

$$y_a = \beta y_c + (1-\beta)y_u = \frac{Fa^3}{24EI}\left[2\left(\frac{x}{a}\right)^3 - (6-3\beta)\left(\frac{x}{a}\right)^2 - 3\beta + 4\right] \qquad (6.14)$$

Note that the cubic term is independent of fi. Influence functions for several values of fi are shown in figure 6.12.

To determine the stroke capability and the interactuator coupling of the deformable mirror, it is necessary to model a section of the faceplate containing the adjacent actuators. A typical case for a square array is shown in figure 6.13, which takes into account the four nearest neighbors. The stiffness of the faceplate K_F as a function of fi can be determined from equation (6.14) by evaluating it for $x = 0$:

$$K_F = \frac{K_F^c K_F^u}{(1-\beta)K_F^c + \beta K_F^u} = \frac{F}{y_{pk}} = \frac{24EI}{a^3(4-3\beta)} \qquad (6.15)$$

The actuator coupling C_A is defined as the ratio of the faceplate deflection at an adjacent actuator, y_d, to that of the peak deflection due to the energized actuator, y_{pk}

$$C_A = \frac{y_d}{y_{pk}} = \frac{1}{4(K_A^a/K_F)+1} \qquad (6.16)$$

The presence of coupling between actuators often improves the accuracy with which a random wavefront can be compensated by a deformable mirror. If significant coupling (more than a few percent) exists, it must be taken into account in the design of the wavefront reconstructor, because it is a potential source of instability in the adaptive optics feedback loop.

The peak deflection of the faceplate will always be less than the free stroke of the actuators, because of the stiffness of the faceplate. The faceplate itself is

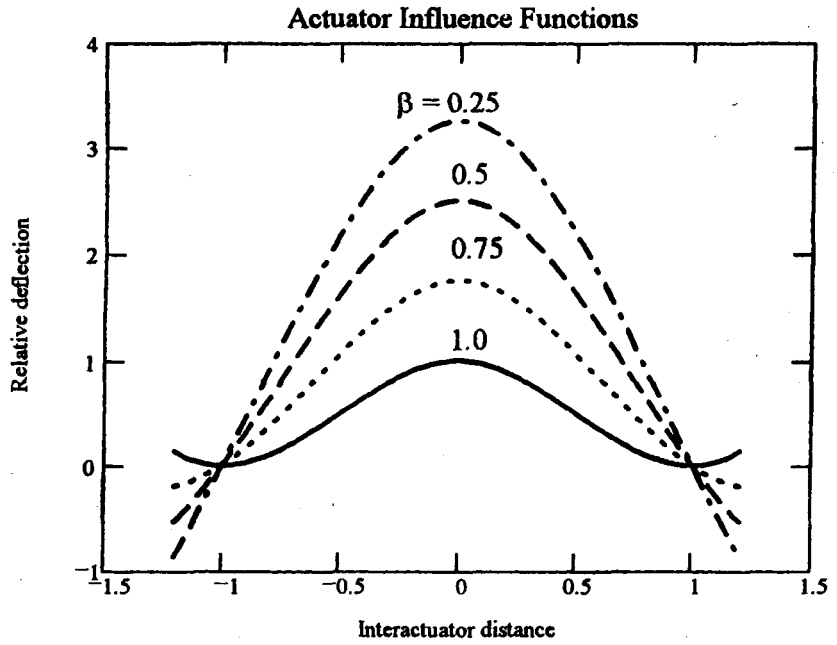


Figure 6.12 Influence functions of a zonal deformable mirror with a continuous faceplate and discrete actuators. The shape of the influence function depends on the parameter *p*, the ratio of the actuator bending stiffness (as modified by the flexure coupling the actuator to the mirror) to the stiffness of the faceplate. The faceplate thickness determines the stroke of the mirror, but not the shape of the influence function.

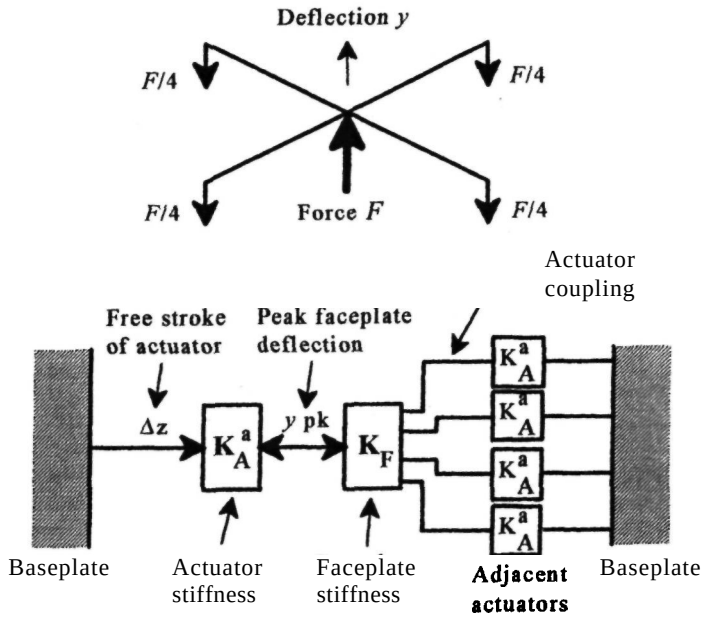


Figure 6.13 Faceplate deflection model: actuator with four nearest neighbors.

supported by the other actuators, the stiffness of which must also be taken into account. If the peak deflection of the faceplate is y_{pk} and the equivalent stiffness of the faceplate is K_F^{eq}, then it is necessary for the free stroke of the actuators to be

$$\Delta z = y_{pk}\left[1 + (K_F^{eq}/K_A^a)\right] \tag{6.17}$$

where

$$K_F^{eq} = \frac{4K_A^a K_F}{4K_A^a + K_F} \tag{6.18}$$

The procedure outlined above enables the preliminary design of a deformable mirror to be carried out and the major parameters established. Before proceeding to the detailed design, it is advisable to verify the structural integrity of the design, including the stability of the actuators under compressive loading, the stresses in the actuators and faceplate, and the effect of temperature variations. It is also necessary to verify the dynamic properties of the deformable mirror, especially the resonant frequency produced by lateral bending of the actuators.

The maximum stress in a beam occurs at the sections having the maximum bending moment and is given by

$$\sigma = \frac{Mt}{2I} \tag{6.19}$$

where

M = bending moment
t = thickness
I = moment of inertia of the section

In a deformable mirror faceplate, the maximum bending moment occurs at the actuator locations. Modeling the deformable mirror faceplate in one dimension as a beam, the maximum bending moment for case 1 (clamped) is *Fa/4*, while for case 2 it is *Fa/2*. Using the factor β, the bending moment may be expressed in the form

$$M_F^m = \frac{Fa}{2(1+\beta)} \tag{6.20}$$

$$M_F^m = \frac{y_{pk}K_F a}{2(1+\beta)} \tag{6.21}$$

Substituting for M in equation (6.19), and using equation (6.15), an expression for the maximum stress in the faceplate is obtained:

$$\begin{aligned}\sigma_{max} &= y_{pk}\left[\frac{K_F a t}{4(1+\beta)I}\right] \\ &= y_{pk}\left[\frac{6tE}{a^2(1+\beta)(4-3\beta)}\right]\end{aligned} \tag{6.22}$$

The faceplate stress is therefore directly proportional to the faceplate thickness t, for given values of peak deflection y_{pk} and actuator spacing a. The faceplate stress is lower for case 2 ($\beta = 0$, actuators with low bending stiffness), because of the smaller force required to achieve a given deflection.

The bending moment transferred from the faceplate to the actuators depends directly on the value of β, as can be seen from figures 6.10 through 6.12. When $\beta = 1$, all of the faceplate bending moment is transferred. The actuator bending stress is therefore

$$\sigma_A^b = \frac{\beta M_F^m t_A}{2I_A} \tag{6.23}$$

where t_A is the thickness of the actuator in the plane of bending, and the maximum faceplate bending moment M_F^m is given in equation (6.21).

The axial stress in the actuators of cross-sectional area A is

$$\sigma_A^a = \frac{F}{A} = \frac{y_{pk}K_F}{A} \tag{6.24}$$

As a final check on the mechanical integrity of the actuators, their margin of safety against buckling under compressive force should be evaluated. Using the Euler formulas, the safe load depends on the end conditions. In a deformable mirror, it is reasonable to assume that the lower end of the actuator is fixed and the top end is constrained laterally but free to tilt. For these conditions, the maximum compressive force that the activator will withstand is

$$F_A^m = \frac{2\pi^2 E_A I_A}{l^2} \tag{6.25}$$

The factor of safety is then

$$n = \frac{F_A^m}{F} = \frac{F_A^m}{y_{pk}K_F} \tag{6.26}$$

6.4.2 Thermal Considerations

Temperature variations can produce two kinds of problems in deformable mirrors: surface irregularities in the faceplate and bending stresses in the actuators. The faceplate is usually supported entirely by the actuators, so gross thermal expansion of the actuators displaces the faceplate as a whole without causing stresses. However, mismatches in the coefficients of thermal expansion of individual actuators may cause irregularities in the faceplate as a function of temperature. These are of concern only if the operating temperature of the deformable mirror varies significantly from the temperature at which it was polished and tested. If this temperature difference is ΔT, and the standard deviation of the actuator temperature coefficients is σ_m, then the faceplate surface irregularity is

$$\Delta z = \sigma_m \alpha_A l \Delta T \quad \text{meter rms} \tag{6.27}$$

where α_A is the mean coefficient of thermal expansion of the actuators. It is of interest to find the value of σ_m necessary to keep the wavefront errors due to mirror temperature variations below a specified value. In the case of piezoelectric actuators with $\alpha_A = 2 \times 10^{-6}$ per °C, $l = 20$ mm, and $\Delta T = 10$°C, the value of $\sigma_m = 0.025$ is required to achieve a wave-

front error of 1/25 wave rms at). = 0.5 μm In other words, the temperature coefficients of the actuators must be matched to within 2.5% rms.

A more serious thermal problem in deformable mirrors is caused by differences in the coefficients of thermal expansion of the baseplate and faceplate, which produce bending stresses in the actuators. In this case, the relevant temperature range covers not only the operating temperature of the mirror, but also the temperature cycles used in fabrication and coating operations. The relative displacement produced by mismatch of the temperature coefficients is

$$AL = L(\alpha_F \sim \alpha_B)\Delta T \tag{6.28}$$

where L is the longest span between actuators. The displacement of each of the outermost actuators will be $AL/2$. The maximum stress in these actuators caused by temperature will be

$$\sigma_A^T = \frac{Mt_A}{2I_A} \tag{6.29}$$

where the bending moment M is given by

$$M = \frac{\Delta L l K_A^b}{2} \tag{6.30}$$

Substituting for AL and K_A^b, the maximum stress is

$$\sigma_A^T = \frac{3L(\alpha_F - \alpha_B)\Delta T E t_A}{4l^2} \tag{6.31}$$

6.4.3 Dynamical Considerations

The operating bandwidth for deformable mirrors used in adaptive optics systems for turbulence compensation at visible wavelengths extends up to about 500 Hz; at IR wavelengths, it may be only 100 Hz. Servo-system design and implementation are simplified by keeping the natural resonances of mechanical devices, such as deformable mirrors and two-axis tracking mirrors, much higher than the operating bandwidth, a factor of 10 being a desirable goal. To keep natural resonances above 1 kHz, the mechanical structure of these mirrors must be small and stiff.

Three vibration modes are of primary interest:

1. The first bending mode of the substrate. The stiffness and resonant frequency of the substrate are both proportional to its thickness. Thick substrates are necessary to provide the required stiffness, so there is usually no problem in keeping the natural resonance well above the operating frequency.
2. Interactuator bending modes of the faceplate. The natural frequency of the interactuator bending mode depends on the interactuator spacing, typically about 1 cm, and on the faceplate thickness, typically about 3 mm. These values produce resonant frequencies well over 1 kHz.
3. The first bending mode of the actuators. This is generally the lowest natural resonance in a deformable mirror and consequently drives the design. It depends on the actuator stiffness and spacing, as well as on the faceplate stiffness and geometry.

The natural resonant frequencies of an actuator in the axial and bending modes are given by the following expressions:

$$f_A^a = \frac{1}{21}\left(\frac{E}{\rho}\right)^{1/2} \tag{6.32}$$

$$f_A^b = \frac{3.52}{l^2}\left(\frac{EI}{\rho A}\right)^{1/2} \tag{6.33}$$

where p is the mass density of the material. The fundamental resonant frequency (Hz) of a thin plate is given by

$$f_P = \frac{Bt}{2\pi a^2}\left[\frac{E}{\rho(1-\nu^2)}\right]^{1/2} \tag{6.34}$$

where

- t - thickness, m
- a = diameter or side of the plate, m
- p = mass density, kg m^{-3}
- E = elastic modulus, N m^{-2}
- v = Poisson's ratio
- B = a constant depending on the plate shape and method of support

Approximate values for the resonant frequencies of the baseplate and the faceplate, which are useful in the first-order design of a deformable mirror, can be determined using equations (6.32)–(6.34). In the case of the faceplate, the interactuator area can be defined as a square of side a (equal to the actuator spacing), or as a circle of diameter $\sqrt{2}a$. The method of support will vary with the location of the area on the mirror; near the center it will tend to be clamped at all edges, while at the periphery it will be free at one or two edges. The lowest resonant frequencies are likely to occur in interactuator areas that are partially free at the edge. As with all other parameters of the deformable mirror, the resonant frequencies predicted by this method should be verified by finite element analysis in the final design.

6.4.4 Optimization of Continuous-Plate Deformable Mirrors

With the large number of parameters and interrelationships described above, even the first-order design of a continuous-plate deformable mirror appears to be a formidable task. The design process is facilitated by using tradeoff charts that allow the range of possible values of critical parameters to be established quickly. The process of optimization has been outlined by Ealey and Wellman [1989].

The facesheet thickness is one of the most important design parameters to be established. Starting

with a set of actuator characteristics, a fixed actuator spacing, and the material properties of the facesheet, a tradeoff chart relating facesheet thickness to other deformable mirror properties may be constructed, as shown in figure 6.14. The constrained parameters, in this case, are facesheet ripple, facesheet stress, and actuator coupling. The facesheet ripple R is the surface deformation produced during the final polishing operation, by force on the polishing lap. The surface deflection produced by a given force may be determined from equation (6.14) and plotted against facesheet thickness. Similarly, the facesheet stress 5 for a given actuator stroke may be calculated using equation (6.22) and also plotted against facesheet thickness. The coupling between adjacent actuators, C, depends on the bending stiffness ratio β of the actuators to the facesheet. For a given actuator design and spacing, it may be determined from equation (6.16) and also plotted against facesheet thickness.

Setting a maximum value R_{max} for the facesheet ripple establishes a lower limit to the facesheet thickness, as shown in figure 6.14. For a given deflection, the facesheet stress varies directly with thickness, so the allowable stress S_{max} establishes an upper limit to faceplate thickness. If this value is not at least equal to the minimum required to satisfy the ripple specification, then the actuator design or spacing will require modification. The final choice of faceplate thickness, between these limits, may be based on the required actuator coupling C. In practice, several iterations of the tradeoff may be required to obtain a set of parameters that satisfy all requirements.

6.4.5 Deformable Mirror Performance

The design procedure outlined in the preceding text has been applied to many continuous-facesheet deformable mirrors using discrete actuators. An example is the 97-actuator deformable mirror described by Thorburn and Kaplan [1991]. The construction of this mirror is shown in figure 6.15 and a photograph of the finished device is shown in figure 6.16.

The mirror facesheet is supported entirely by 97 actuators, spaced at 7 mm, that form an 11×11 square array with the corners removed. The outer ring of actuators is used to control the edges of the faceplate and is outside the useful aperture of 63 mm diameter. The facesheet is of ultralow-expansion quartz, with a silver coating. Integral pusher pads are machined into the back of the faceplate to provide coupling to the actuators. These pads are carefully dimensioned to provide the required stiffness for control of the influence function.

The electrostrictive actuators are cylindrical PMN ceramic multilayer stacks that are fired to form monolithic structures. The actuators are not compressively stressed and are bonded to both the baseplate and facesheet, on which they exert a bipolar force. The useful actuator stroke is $\pm 2\ \mu m$. The baseplate is made of the same material as the faceplate to eliminate differential thermal expansion; it is extremely stiff to provide the required stability. Actuator connections are made with printed circuit cards that slide into slots in the baseplate.

The electrical drivers are an integral part of the deformable mirror system. The PMN actuators have a quadratic response and are operated with a bias of 85 V to bring the operating point into the linear region of the response. To obtain the required stroke of $\pm 2\ \mu m$ the actuators are driven with ± 35 V, which is produced by a driver input of ± 10 V. Characteristics of the complete wavefront corrector are summarized in table 6.5

6.5 Segmented Mirrors

6.5.1 Types of Segmented Mirror

Segmented mirrors were the first kind of wavefront correctors to be developed. The simplest type are

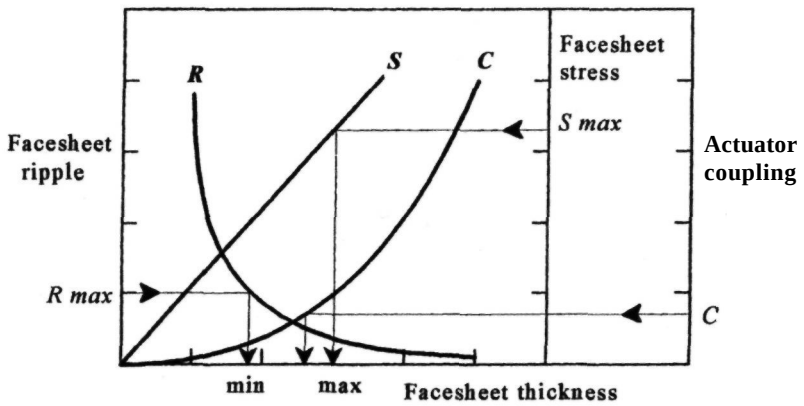


Figure 6.14 Optimization of facesheet thickness for a given actuator spacing and actuator stroke.

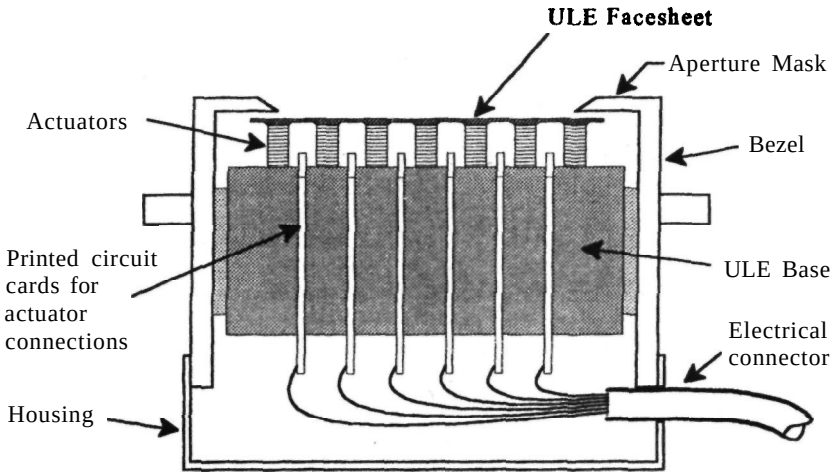


Figure **6.15** Construction of a discrete actuator deformable mirror [Thorburn and Kaplan 1991]. This cross-section shows the relative thicknesses of the face-sheet and baseplate, and the method of making a large number of connections to the actuators, using printed circuit boards (PCBs), which are partially slotted into the baseplate between the rows of actuators.

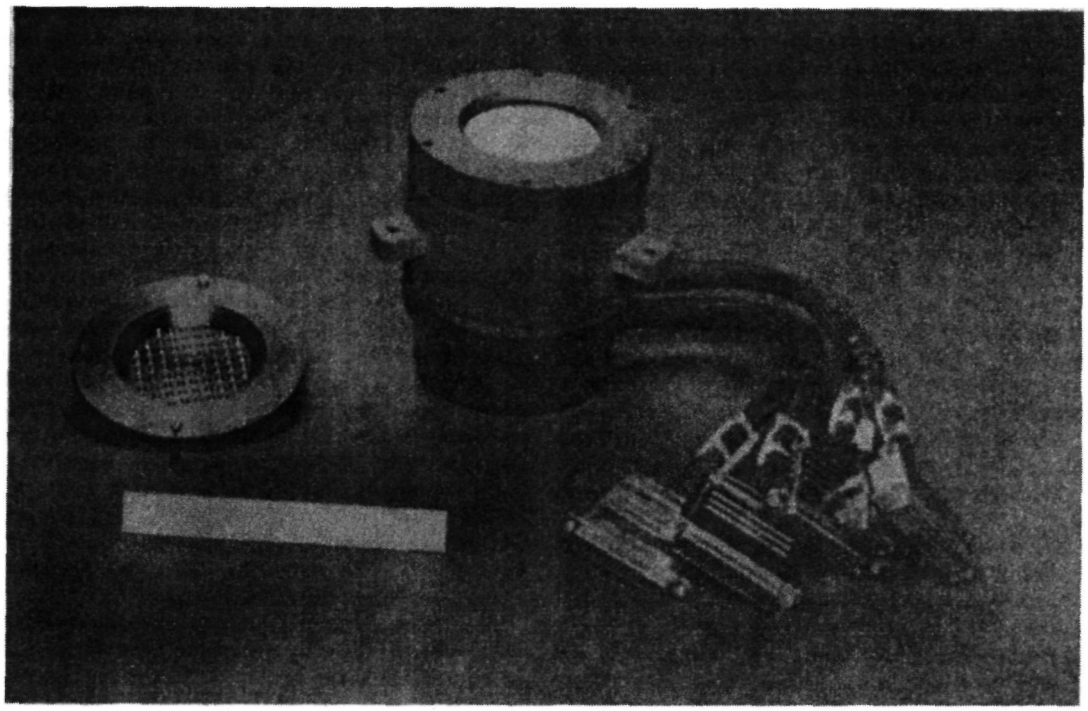


Figure **6.16** Photograph of a 97-actuator deformable mirror [Thorburn and Kaplan 1991]. At the left is the optical alignment mask, pierced with 97 holes exactly centered on the actuators.

Table 6.5 Characteristics of 97-Actuator Deformable Mirror

Working aperture	63 mm
Actuator spacing	7 mm
Stroke (total)	4.0 μm
(adjacent actuators)	2.8 μm
Response uniformity	±7% of full stroke
Hysteresis	1.5% of full stroke
Interactuator coupling	10% to nearest neighbor
Surface figure accuracy	0.019 μm rms
Surface smoothness	16 angstrom, rms
Reflectivity, 0.5-1.5 μm	> 0.95
Overall size of mirror package	12.5 cm diameter x 15 cm long
Actuator capacitance	1 μF
Maximum slew rate	1 V μs^{-1}
Small signal bandwidth	> 1 kHz (complete system)
Input voltage range	±10 V

known as piston-only correctors, employing one actuator per segment to correct the average optical phase over each zone. As the wavefront distortions due to atmospheric turbulence are smoothly varying functions, piston-only correctors inevitably produce wavefront discontinuities at the edges of each segment. Because of their crude correction capabilities, piston-only segmented mirrors are rarely used in adaptive optics today.

The next step in the development of segmented mirrors was to give each segment 3 degrees of freedom by supporting it on three independent actuators. In this way, the segments could be adjusted in tip and tilt, as well as absolute phase, resulting in much better fitting of the wavefront. Step discontinuities at the edges are greatly reduced by this means.

Segmented mirrors were used originally for single-wavelength applications, primarily in laser systems. For this type of operation, in which the coherence length of the radiation is many wavelengths, it is possible to make phase corrections modulo one wavelength, so that a phase error of say, 1.5 waves, is compensated with a phase shift of 0.5 wave. This strategy reduces the dynamic range required by the segmented mirror to ±0.5 wave, irrespective of the actual pathlength error, provided that the coherence length of the radiation, $\lambda^2/\Delta\lambda$, is not exceeded. Segmented mirrors of this type, for use in monochromatic light at visible and near-IR wavelengths, employ actuators that have a stroke of only 1 or 2 μm.

If a deformable mirror of this type is used with broadband light, the effect of truncating the stroke is to produce a wavefront error proportional to the product of the spectral bandwidth and the dynamic range of the wavefront. For a wavefront having an excursion of n waves rms, the mean-square error due to truncation is

$$\sigma^2 = (2\pi n)^2(\Delta\lambda/\lambda)^2 \quad \text{rad}^2 \tag{6.35}$$

where

$\Delta\lambda$ = spectral bandwidth
λ = mean wavelength

For astronomical imaging, it is essential to compensate the full optical pathlength error so that light of all wavelengths can be brought to a common focus. The peak-to-valley wavefront excursion due to atmospheric turbulence is between 4 and 8 μm, depending on the ratio of D/r_0, and requires a mirror stroke of up to 4 μm peak to peak. Segmented mirrors with long-stroke actuators suitable for white-light compensation have been built, using three degrees of freedom per segment. Hulburd et al. [1991] describe a mirror with 512 square segments that uses tubular piezoelectric actuators, each of which is split into three independently controlled sections. The mirror, shown in figure 6.17, is assembled from modules, each containing 4 x 4 or more actuators. These modules are removable for repair or replacement. An auxiliary figure sensor is required to monitor the mirror segments for offsets and drifts, which must be removed to achieve a uniphase surface. For white-light operation, the figure sensor must be capable of measuring the piston errors unambiguously, which necessitates a large dynamic range. This is achieved by using a two-wavelength interferometer, which, in effect, synthesizes a much longer wavelength. Because the mirror segments are freely supported, the full force of each actuator is available to move each segment. The temporal response of segmented mirrors is consequently very good, the response time of the 512-actuator mirror to a step function being about 100 μs.

A different type of segmented mirror, developed for use in solar adaptive optics, has been described by Acton and Smithson [1991]. This mirror, shown in figure 6.18, employs 19 hexagonal segments, each 2.8 cm across. Each segment is controlled by three piezoelectric actuators, referred to as a triad, mounted in a common Invar case. These actuators,

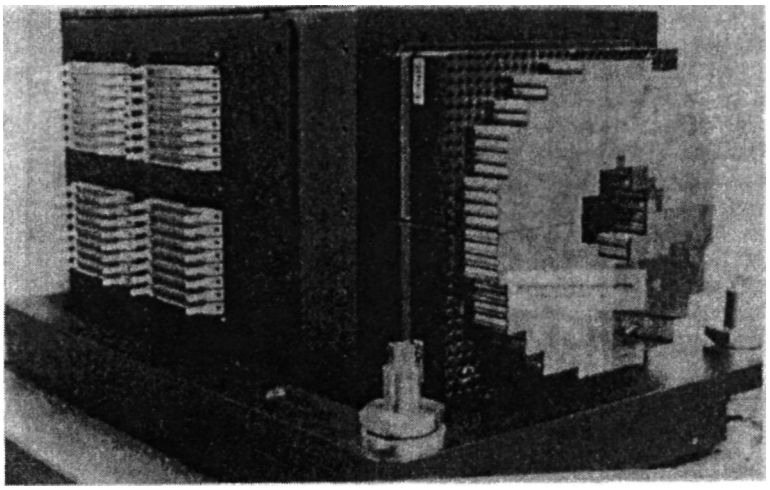


Figure 6.17 Photograph of a 512-segment deformable mirror [Hulburd et al. 1991]. Each segment is supported on a piezoelectric actuator with 3 degrees of freedom, giving tip, tilt, and piston adjustments. The overall diameter of the mirror is 22 cm.

made by Physik Instrumente, Germany, are stabilized by strain gauges, giving them a linear response with no measurable hysteresis. The temporal bandpass of each segment is 300 Hz at the half-power point. The surface of the adaptive mirror is monitored with a Michelson white-light interferometer. Color fringes occur when 2π phase errors are present on individual segments, enabling phase ambiguities to be resolved.

In the mirrors described in the preceding text, each segment is driven by dedicated actuators, using three actuators or degrees of freedom per segment. In principle, it is possible to share the actuators between several segments, using the triangular or

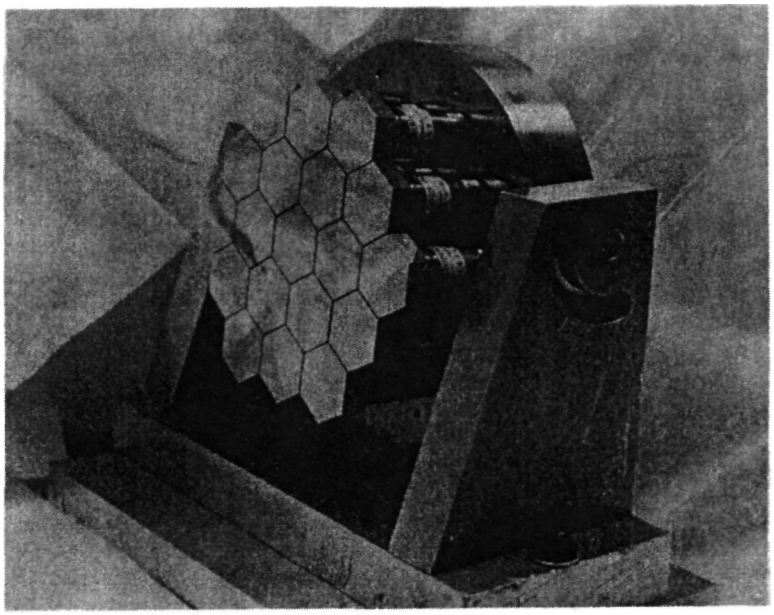


Figure 6.18 Adaptive mirror with 19 hexagonal segments. This device was used in a solar image compensation system [Acton and Smithson 1992].

hexagonal geometry shown in figure 6.19. Each segment is still controlled by three actuators. Using the triangular geometry, the number of segments is twice the number of interior actuators, while in the hexagonal case the number of segments equals the number of interior actuators. In each case, additional actuators are required at the periphery of the mirror. The wavefront fitting errors are different for each configuration. These shared-actuator configurations are unsuitable for small wavefront correctors because, once the mirror has been assembled, there is no way to accommodate variations in segment thickness at the actuator locations.

6.5.2 Segmented Mirror Characteristics

Segmented mirrors have two characteristics that are different from other wavefront correction devices:

- The number of actuators (degrees of freedom) is not necessarily the same as the number of subapertures or zones; consequently, wavefront fitting errors depend on the specific segment geometries and methods of control.
- The segments are separated by small gaps.

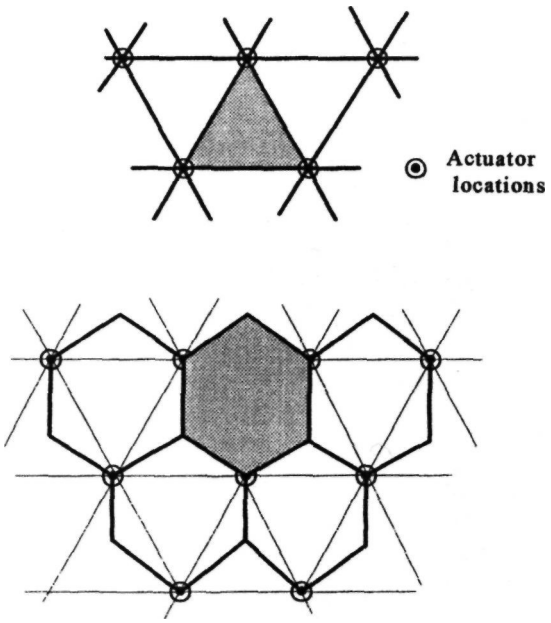


Figure 6.19 Segment patterns with shared actuators. The triangular geometry has twice as many segments as interior actuators, while for hexagonal segments the number of interior actuators and segments is equal. A major problem with shared actuators is the difficulty of accommodating variations in segment thickness at the actuator locations.

For Kolmogorov turbulence, the wavefront fitting error variance is given by

$$\sigma_{\text{fit}}^2 = a_{\text{F}}\left(\frac{d}{r_0}\right)^{5/3} \tag{6.36}$$

where

a_{F} = constant depending on the segment geometry and control mode
d = characteristic size of the segment

Values of a_{F} for some common segmented mirror configurations are given in table 6.6.

Care must be taken when comparing the fitting errors of segmented and continuous-plate mirrors. The complexity of an adaptive optics system is measured by the number of degrees of freedom (or actuators) that have to be controlled. The number of segments required to achieve a given fitting error may be smaller than the number of actuators required to achieve the same performance with a continuous-plate mirror. But if the total number of degrees of freedom is compared, continuous-plate mirrors are generally more economical. This subject is discussed further in section 9.4.5.

Gaps in a segmented mirror are inevitable. They have two effects on the imaging capabilities of the system: (1) a direct loss of energy through the gap; and (2) a reduction in Strehl ratio due to scattering of light by the edges of the segments. When the gap area is small compared with the full aperture, there is no significant broadening of the central core. The diffracted light appears in the form of streaks radiating from the core of the image in a geometrical pattern corresponding to the segment geometry. The energy distribution of the diffraction pattern in the direction normal to the length of the crack has the form

$$I_{\text{gap}} \propto \frac{a^2b^2}{A^2}\left[\frac{\sin(kax)}{kax}\right]^2 \tag{6.37}$$

where

a and b = width and length of the crack, respectively
A = full aperture area
k = $2\pi/\lambda$

6.5.3 Large Segmented Mirrors

The comments above relate to small segmented mirrors used for adaptive wavefront control. Much larger segmented mirrors have been developed for use as primary mirrors in astronomical telescopes and laser beam directors. Because of the large difference in size and mass, the construction and control of these large segmented mirrors is markedly different from those used in high-speed adaptive mirrors. The 10-m Keck telescopes are currently the largest in existence, employing an electromechanical position-sensing and feedback system to maintain the relative

Table 6.6 Segmented Mirror Fitting Error

Configuration	Coefficient a_F rad^2	Actuators per Segment
Piston only (square segments)	1.26	1
Piston only (circular segments)	1.07	1
Piston plus tilt (square segments)	0.18	3
Piston plus tilt (circular segments)	0.14	3
Shared actuators (triangular segments)		0.5
Shared actuators (hexagonal segments)		1
(Continuous deformable mirror)	0.28	1

alignment of the 36 segments. The 4-m Large Active Mirror Program (LAMP) mirror [Plante, 1991] employs seven segments, each of which is mounted on a hexapod support structure with fully controllable actuators. In addition, the faceplate of each segment is actively controlled using 50 electromagnetic actuators. The purpose of active control on mirrors of this size is to maintain the surface figure and not to compensate external wavefront errors.

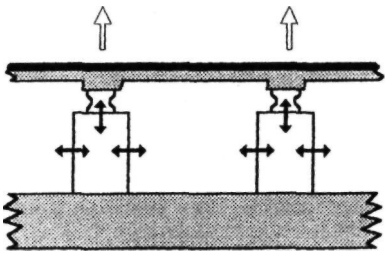
6.6 Bimorph Mirrors

6.6.1 *Principle of Operation*

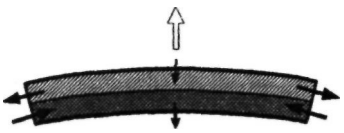
Bimorphs are combinations of two materials, the relative sizes of which can be varied with a stimulus such as temperature or applied voltage. A distinction is made between bimorph actuators and bimorph mirrors. Bimorph actuators employ two piezoelectric layers, or one piezoelectric and one inert layer, to produce a controllable force or large linear displacement. They are discrete devices, capable of being used in the continuous-plate and segmented mirrors described previously.

Bimorph mirrors are devices in which the mirror faceplate itself is a continuous bimorph structure consisting of two dissimilar layers, the shapes of which can be locally controlled. When piezoelectric material is used for one or both of the layers, the application of a voltage causes the areas under the electrodes to change in size, resulting in spherical bending of the faceplate. The deformation is produced by forces within the plate itself, which needs only to be simply supported. There is no reaction structure and the actuators do not need individual adjustment. The construction of bimorph mirrors is therefore much simpler than that of discrete-actuator mirrors. Individual actuators are formed by a pattern of electrodes on the surface of the bimorph. It must be kept in mind that bimorph structures are sensitive to dimensional changes from all causes, including temperature variations and aging. The bimorph faceplate itself must be fabricated with great care to avoid such problems. Also, it should be noted that bimorph structures cannot assume the shapes of all Zernike polynomials without the application of edge gradients.

The analysis and design of bimorph mirrors is complicated by the fact that bimorph structures produce thickness changes in addition to bending, these effects being of opposite polarity. The operation of a discrete-actuator mirror is compared with that of a bimorph structure in figure 6.20. When using electrodistortive materials such as PZT or PMN, volume is conserved so that dimensional changes in the axial (poling) direction are accompanied by compensating changes in the lateral directions. With discrete actuators, only the axial motion is coupled to the mirror



Discrete **actuators: Only the axial deflection is coupled to the facesheet.**



Bimorph structure: Bending **and** thickness **changes both deflect the facesheet,** but in **opposite directions.**

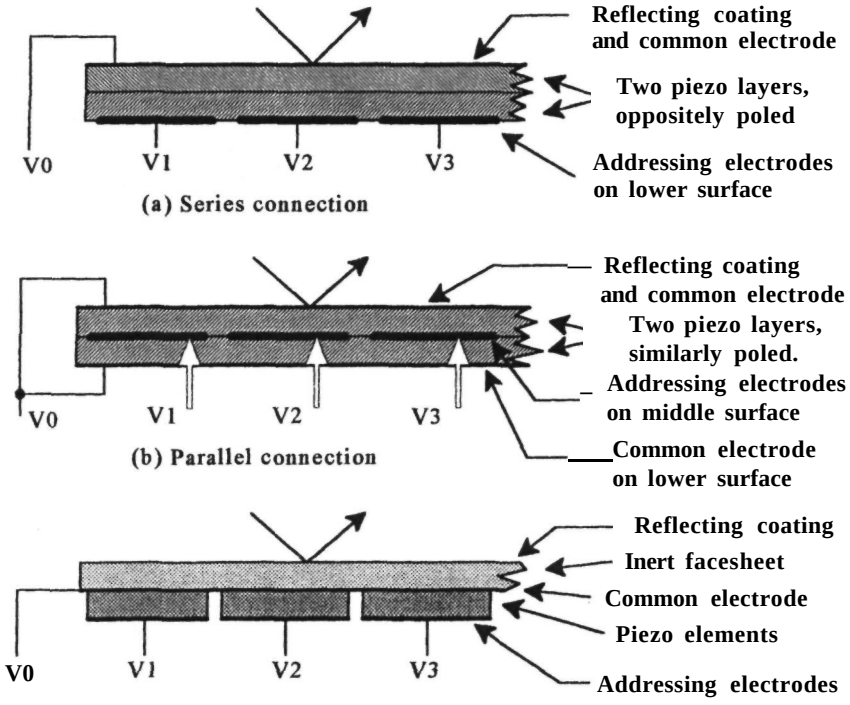
Figure 6.20 Comparison of discrete-actuator and bimorph structures.

faceplate. The lateral force is transmitted to the baseplate, but, as this is deliberately made very thick, the bending is negligible.

In bimorph mirrors, both the axial and lateral strains **are** coupled to the faceplate. The primary deflection mechanism is bending due to differences in the lateral strains. In the case shown in figure 6.20, the top layer has expanded laterally and the bottom layer has contracted laterally, resulting in the top mirror surface taking convex curvature. But, the thickness of the top layer has decreased, so that the net change in the position of the mirror surface depends on the relative magnitude of the two effects. A dominant bending effect can be achieved by making the size of the controlling zones (electrodes) much larger than the thickness of the bimorph. The spatial resolution is consequently limited, with the result that practical bimorph mirrors are restricted to fewer than 100 actuators. The spatial frequency limitation of bimorph mirrors is discussed further in section 6.6.3.

There are several possible bimorph configurations, three of which are shown in figure 6.21. In each case, an array of controlling voltages is applied to a pattern of electrodes deposited on one of the bimorph elements, using a continuous electrode as the common element. The use of two piezoelectric layers has the advantage of doubling the sensitivity when they are oppositely poled; it also minimizes thermal distortions due to a mismatch in the coefficients of thermal expansion. Piezoelectric ceramic materials are hard and reasonably stable, so they can be optically figured and polished in the same way as conventional optical materials. In configuration (a) of figure 6.21, the piezoelectric elements are oppositely poled and the control voltage is applied across both elements in series. The addressing electrodes are easily accessible at the back of the bimorph structure. When a voltage is applied, the transverse piezoelectric effect causes one layer to expand and the other to contract, producing spherical deformation of the plate.

The applied voltage necessary to achieve a specified deformation can be reduced by using the parallel connection shown in configuration (b), in which the poling directions are the same for both elements. In this case, the addressing electrodes are buried between the layers, which may cause problems with addressing leads, especially with a large number of electrodes. Configuration (c) uses a single piezoelectric layer in combination with a passive substrate, which can be of the same material in order to minimize differences due to temperature and aging. This arrangement is sometimes referred to as a unimorph, although the second inert layer is essential **to** its operation. A variation of this scheme employs an array of separate piezoelectric disks bonded to the back of the faceplate in the form of a mosaic. A thin continuous conducting film between the sub-



(c) **Unimorph**

Figure 6.21 Bimorph mirror configurations

strate and the piezoelectric disks acts as a common electrode.

6.6.2 Bimorph Mirror Characteristics

The equations describing the electromechanical behavior of piezoelectric material [Mason 1950] may be stated as

$$T_i = C^{d}_{i,j} S_j - d_{n,i} D_n \tag{6.38}$$

$$E_m = -d_{m,j} S_j + \beta_{m,n} D_n \tag{6.39}$$

where T_i and S_j are the stress and strain tensors, and the variables E_m and D_n are the components of the electric field and the electric displacement field, respectively. The constants in these equations are the mechanical coefficients of elasticity $C^{d}_{i,j}$ the piezoelectric constants $d_{n,i}$, and the inverse dielectric permittivities, $\beta_{m,n}$, which are all known properties of the material. The indices ;' and j run from 1 to 6, while the indices m and n represent the three spatial directions: 1 being the x direction, 2 the y direction, and 3 the z direction. The bimorph is oriented with its surfaces in the x,y plane, with z representing the direction parallel to the optical axis in which the displacements are required. The origin of the z axis (z = 0) is defined by the boundary surface between the two elements of the bimorph.

The matrix equations (6.38) and (6.39) represent nine algebraic equations with 18 unknowns, so additional relationships are needed to obtain a determinate set. The analysis is considerably simplified by assuming that the material is mechanically homogeneous and isotropic in the x,y plane, that the inverse dielectric permittivity matrix is diagonal, and that the piezoelectric constants are all zero except rfjj, d_{32} and d_{33}, with $d_{31} = d_{32}$. These properties are typical of most piezoelectric materials. The first assumption implies that most of the elastic constants $C^{a}_{i,j}$ are zero and that the remaining terms are not independent of each other. A further simplification is made by assuming that the tensile stress in the z direction (T_3) is zero, because the top and bottom surfaces of the bimorph are unconstrained.

When a voltage is applied to the bimorph in the direction of poling, the first element expands in the x,y plane and contracts in the z direction, while the second element contracts in the x,y plane and expands in the z direction. The changes in the x,y plane produce bending about the center line of the bimorph, while the z-axis changes cause thickness variations in the elements. The surface deflection of a bimorph mirror is the sum of these two effects, which are of opposite polarities. For voltages applied over areas with dimensions much greater than the thickness of the bimorph, the deflection due to bending is the dominant effect; this is the result intuitively expected from a bimorph. But, in the case of voltages that are applied over small areas, the change in thickness can be the larger effect. Because these effects are of opposite polarity, there is obviously an electrode size at which the two effects cancel out, setting an upper limit to the spatial frequencies that can be compensated.

If the displacement of a point on the middle surface of the bimorph caused by bending is $\psi_0(x,y)$, and the displacement of a point due to thickness change is $\psi_1(x,y,z)$, then the total displacement can be expressed as

$$u_z(x, y, z) = \psi_0(x, y) + \psi_1(x, y, z)$$

$$u_x(x, y, z) = -z \frac{\partial \psi_0}{\partial x} \tag{6.41}$$

$$u_y(x, y, z) = -z \frac{\partial \psi_0}{\partial y} \tag{6.42}$$

Using the piezoelectric equations (6.38) and (6.39), Kokorowski [1979] showed that the displacements of a bimorph mirror for an applied voltage $V(x,y)$ are

Bending effect

$$\nabla^2 \psi_0(x, y) = -\frac{4V(x, y)}{a_0 t^2} \tag{6.43}$$

Thickness change

$$\psi_1(x, y, z) = \frac{2V(x, y) z^2 a_1}{a_0 t^2} \tag{6.44}$$

where t is the total thickness of bimorph.

At the reflecting surface, when z = //2, the thickness change becomes

$$\psi_1(x, y, z) = \frac{V(x, y) a_1}{2a_0} \tag{6.45}$$

In these equations, a_0 is a piezoelectric constant given by the identity

$$a_0 \equiv (d_{31} - b d_{33}) + \frac{\left[C^{d}_{11} - b\left(C^{d}_{13} - C^{d}_{44}\right)\right]\left[\beta_{33} - d^2_{33}/C^{d}_{33}\right]}{d_{33}\left(b + C^{d}_{55}/C^{d}_{33}\right) - d_{31}} \tag{6.46}$$

and tf| is a dimensionless constant given by

$$a_1 \equiv b + \frac{d_{33}\left[C^{d}_{11} - b\left(C^{d}_{13} - C^{d}_{55}\right)\right]}{C^{d}_{33}\left[d_{33}\left(b + C^{d}_{55}/C^{d}_{33}\right) - d_{31}\right]} \tag{6.47}$$

where

$$b = C^{d}_{13}/C^{d}_{33}$$

The solution to equation (6.43) may be obtained analytically, or numerically using finite element analysis. The analytic expression for surface displacement is

$$u_z(x, y, -t/2) = -\frac{V(x, y) a_1}{2a_0} + \frac{2}{\pi a_0 t^2} \iint dx'\, dy' \ln |(x, y) - (x', y')|\, V(x', y') \tag{6.48}$$

For a rotationally symmetrical bimorph system using a circular actuator (electrode) of radius r_a, in which r_s is the radius of the supporting ring and r_m is the radius of the mirror, the mirror deflection at any radius r, caused by bending, is given by

$$\psi_0(r) = \frac{V}{a_0 t^2}\left[r_a^2 - r^2 - 2r_a^2 \ln\left(\frac{r_a}{r_s}\right)\right] \qquad \text{for } 0 < r < r_a \tag{6.49}$$

$$\psi_0(r) = \frac{V}{a_0 t^2}\left[-2r_a^2 \ln\left(\frac{r_a}{r_s}\right)\right] \qquad \text{for } r_a < r < r_m \tag{6.50}$$

The mirror deflection due to thickness change is

$$\psi_1(r) = \frac{-Va_1}{2a_0} \tag{6.51}$$

The deflection of a piezoelectric bimorph mirror for different values of actuator radius is shown in figure 6.22.

Bimorph mirrors are able to achieve relatively large strokes with low applied voltages, the stroke being inversely proportional to the square of the thickness. They are also compatible with wavefront curvature sensors, because error voltages representing wavefront curvature can be applied to a bimorph mirror with a minimum of processing; the basic wavefront reconstruction required is performed by the bimorph elements themselves.

However, there are boundary conditions that must be satisfied in order to compensate some wavefront shapes. Low-order aberrations, such as astigmatism (cylindrical curvature) and spherical aberration, require the application of edge gradients to the bimorph. This can be achieved by making the active diameter of the bimorph larger than the optical aperture, with additional electrodes to produce the edge gradients. Schwartz et al. [1994] have pointed out that additional computation is required to determine the edge corrections when discrete actuators are employed.

6.6.3 Spatial Frequency Response

It is of interest to determine the maximum number of zones or degrees of freedom that a bimorph mirror will support. Because bimorphs are continuous structures, it might be thought that the spatial resolution is limited only by the number of addressing electrodes

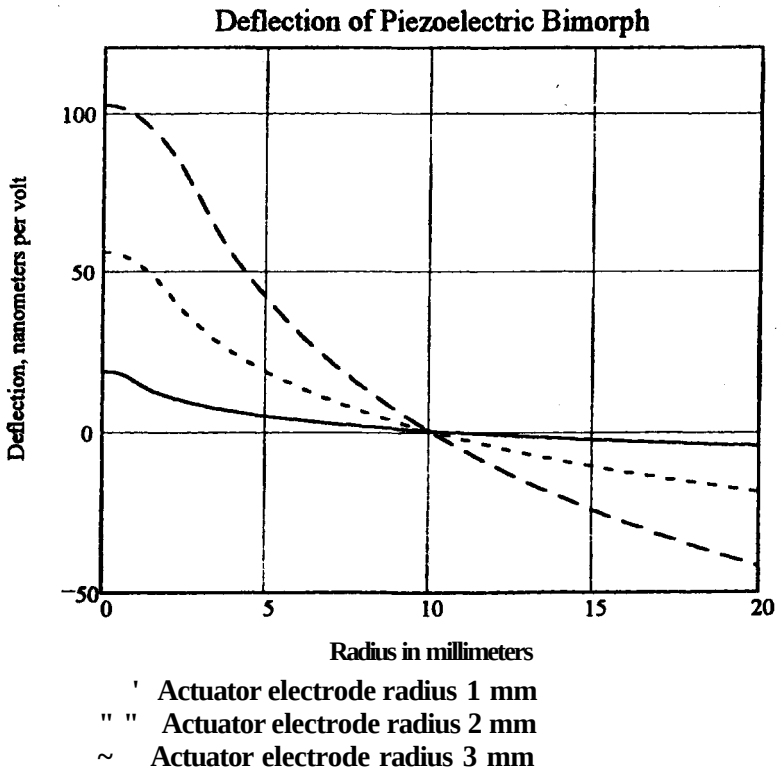


Figure 6.22 Deflection of a piezoelectric bimorph using circular electrodes. The material is PZT type 1, the total thickness of the two layers is 1 mm, and the support ring radius is 10 mm.

that can be applied to the surface. Unfortunately, this is not the case with piezoelectric bimorphs. As the electrode spacing is reduced, the surface deflection due to bending decreases until it becomes comparable to that due to thickness changes in the layers, which is independent of electrode spacing. As these deflections are of opposite polarity with respect to the applied voltage, a point is reached at which the net response is zero.

To determine the spatial frequency properties of a bimorph mirror, the Fourier transform of equation (6.48) can be found. Kokorowski [1979] shows that this can be expressed as

$$U_z(\mathbf{k}) = \Phi(\mathbf{k})\left[\frac{4}{a_0 t^2 k^2} - \frac{a_1}{2a_0}\right] \qquad (6.52)$$

where $\Phi(\mathbf{k})$ is the Fourier transform of the voltage distribution on the mirror electrodes and $k = 2\pi/\Lambda$, where A is the spatial wavelength. Equation (6.52) gives the amplitude of the deformation of the mirror surface due to a sinusoidal distribution of applied voltage. The first term in the expression is associated with bimorph bending and is proportional to the square of the spatial wavelength. The second term is due to thickness changes and is independent of spatial wavelength.

The spatial frequency response of a piezoelectric bimorph using type 1 ceramic material is shown in figure 6.23. In the first chart, the two terms are plotted separately. Note that the bending sensitivity depends on both the bimorph thickness and the spatial frequency, whereas the thickness sensitivity is constant for a given material. As these effects are of opposite polarity, the frequency response drops to zero when they are equal, as shown in the second chart in figure 6.23. For a 1-mm-thick bimorph using the specified material, the two effects are equal at a spatial frequency of about 0.5 cycles per millimeter, equivalent to a periodic actuator spacing of 1.1 mm. To make efficient use of the bending effect, it should be at least 10 times greater than the thickness variations, suggesting a minimum actuator spacing of $1.1 \times \sqrt{10} = 3.5$ mm.

The second factor limiting the total number of actuators is the aspect ratio of the faceplate. Faceplates having a large aspect ratio are difficult to polish flat and also have gravity sag problems. An aspect ratio of 40 is about the practical limit. For a 1-mm-thick faceplate, the number of actuators across the diameter is then $40/3.5 = 11$, giving a total for a circular aperture of about 100.

Bimorph devices, by their very nature, are sensitive to differences in the thermal expansion of the constituent layers. Ideally, the coefficients of thermal expansion of the two layers should either be zero, or precisely matched, to eliminate temperature effects. The simplest way to meet this goal is to use identical materials from the same production batch for the two layers.

6.6.4 Bimorph Mirrors in Practice.

Bimorph adaptive mirrors with 19 and 37 elements, for use in astronomical adaptive optics systems, have been built and tested at the National Optical Astronomy Observatories, Tucson, Arizona [Forbes 1989]. The 37-element mirror employs a PZT bimorph 23 mm in diameter and 1.0 mm thick, with a hexagonal electrode array. The electrodes are photodeposited on the PZT using microcircuit technology. The first resonance of the device is at 11.6 kHz, enabling a working bandwidth of 2 kHz.

Bimorph mirrors for adaptive optics, using 13 and 19 actuators, have been developed at Laserdot [Gaffard et al. 1994].

6.7 Adaptive Secondary Mirrors

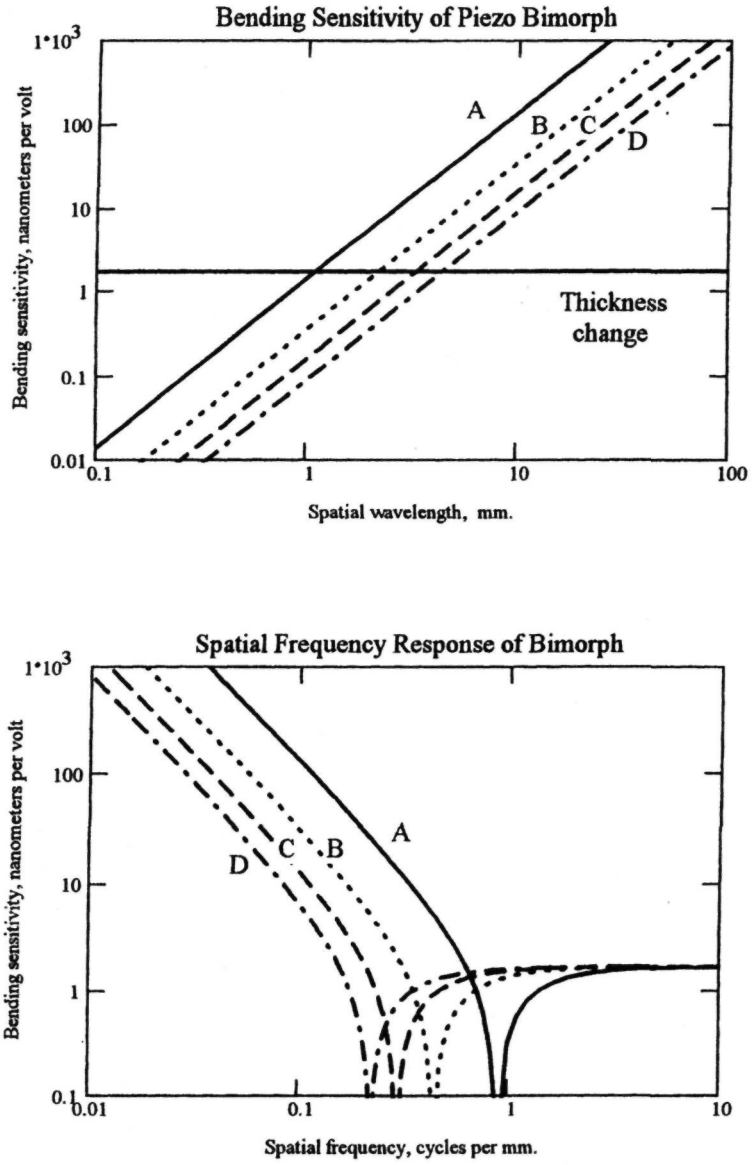
6.7.1 General Considerations

Wavefront compensation may be implemented in an astronomical telescope by means of an adaptive secondary mirror. An adaptive secondary mirror replaces the original fixed secondary mirror, with the following benefits:

1. An existing mirror surface is used for the adaptive corrector, eliminating some of the losses associated with the extra components required for adaptive optics.
2. Adaptive secondaries can include a chopping capability for background suppression in IR observations.
3. The secondary mirror is common to all the telescope imaging paths, such as Nasmyth, Cassegrain, and coude, with the exception of prime focus, and therefore makes adaptive optics available for most observations.
4. Adaptive secondary mirrors are an effective method of upgrading older instruments without major changes to the telescope structure and optics.

There are some disadvantages, including the greater weight of an adaptive mirror and the heat dissipation of the actuators, which are located, unfortunately, in the center of the entrance pupil. This area is very sensitive to turbulence, making an effective thermal control system essential. A more fundamental disadvantage is that the location of the secondary mirror in Cassegrain telescopes is far from a conjugate image of atmospheric turbulence, resulting in a smaller isoplanatic angle than if the corrector were placed at an image of the entrance pupil. While this reduction is small in most cases, its significance is that it denies the possibility of increasing the isoplanatic angle by the use of multiconjugate compensation.

Adaptive secondary mirrors are usually larger than conventional deformable mirrors and present



A thickness 0.5 mm
B thickness 1.0 mm
C thickness 1.5 mm
D thickness 2.0 mm

Figure 6.23 Spatial response of bimorph mirrors. The first chart compares the bending sensitivity, which increases with spatial period (twice actuator spacing), with the thickness change, which remains constant. When these are equal, the net sensitivity drops to zero, as shown in the second chart. Thin bimorphs have the best spatial frequency response, but become increasingly fragile as their aspect ratio increases.

some unique design problems. Secondary mirrors generally have a diameter of 300-1000 mm, compared with the 100–200 mm typical of conventional deformable mirrors. The interactuator spacing is therefore large, leading to problems with ripple in faceplate fabrication, as well as a low frequency of natural resonance. Also, secondary mirrors have a curved faceplate, which can complicate fabrication and adjustment.

Another major difference between adaptive secondaries and conventional deformable mirrors is the large overall tilt requirement of the secondary mirror for IR chopping, which is typically ±10 or more arc seconds on the sky, about four times greater than that required for turbulence compensation. The chopping motion can be implemented by using all of the actuators to tilt the faceplate, in which case the outer actuators require a stroke of about ±100 μm, which is very large for piezoelectric and electrostrictive actuators.

Adaptive secondary mirrors must be packaged very compactly, with the entire support structure and actuator array inside the mirror "faceprint." With on-axis telescope designs, the space surrounding the active area of the secondary mirror must be kept clear so as not to obstruct the primary mirror, leaving no space for the peripheral actuators normally used for edge control. A possible solution is to increase the actuator density toward the edge of the mirror. Because of the large amplitude of the chopping motion, secondary mirrors require dynamic balancing, which necessitates a second set of actuators with dummy loads, moving antiphase to the primary actuators.

6.7.2 Optical Considerations

There are two conditions for the optical imaging of the wavefront corrector in an adaptive optics system. The first condition is that the wavefront corrector must be imaged onto the wavefront sensor to establish the spatial relationship between the actuators and the measurement zones. This is easy to implement, because the optical components involved are all fixed within the system. The second condition is that the wavefront corrector should be located at a plane conjugate to the center of turbulence, in order to maximize the isoplanatic angle. If all turbulence were concentrated in a single layer conjugate to the deformable mirror, then the isoplanatic angle would be unlimited. In such a case, it would be an enormous advantage to have the wavefront compensator at the correct location. In practice, turbulence is distributed throughout the propagation path and a compromise is necessary when using a single corrector.

Conventionally, the wavefront corrector is located near or at a conjugate image of the entrance pupil of the telescope, which is usually the primary mirror. This is the optimum location if most of the turbulence occurs at low levels. The definition of isoplanatic angle θ_0 (section 3.7) implicitly assumes that the wavefront corrector is located at the telescope aperture, or at a plane conjugate to it.

The effect of varying the position of the wavefront corrector on the isoplanatic angle was addressed in section 2.5.2, in which the effective isoplanatic angle θ_E was determined as a function of the displacement of the corrector from the pupil. This relation is shown in figure 2.11. For a Cassegrain telescope, the effective isoplanatic angle θ_E is slightly less than θ_0, while for a Gregorian telescope it is slightly more. The use of an adaptive secondary mirror does not significantly change the isoplanatic angle from its basic value θ_0, but neither does it allow the large increase that is possible by optimal conjugate imaging.

6.7.3 Current Designs

Adaptive secondary mirrors are being studied for possible use in new 8-m telescopes, such as the Large Binocular Telescope (LBT) and the Gemini Telescopes. In addition, active secondary mirrors with fast-steering (tip-tilt) capability are planned for upgrading the performance of the United Kingdom Infrared Telescope (UKIRT) 3.8-m and the University of Hawaii (UH) 2.2-m telescopes.

The adaptive secondary mirror for the LBT [Salinari et al. 1993] will have active, adaptive, and chopping capabilities. A Gregorian configuration has been studied, with a secondary-mirror diameter of about 800 mm. The adaptive mirror structure is attached to the secondary mounting flange by a hexapod support. A substrate of low-thermal-expansion material provides the basic shape reference, to which the thin meniscus faceplate is attached, using about 400 actuators in a triangular pattern. The faceplate thickness is only 2 mm and it is laterally stabilized to the substrate by means of a flexure in the center. The actuators are electromagnetic voice coils, each having a capacitive position sensor. The mirror can be used in the active mode (without the real-time wavefront sensor) to compensate static aberrations in the optical path of the telescope. The fitting error of this mirror in the adaptive mode has been modeled with simulated wavefronts corresponding to r_0 = 13 cm at A = 0.5 μm The average residual error of 74 nm for the mirror alone corresponds to Strehl ratios of 0.64 at λ = 0.7 μm and 0.95 at λ = 2.2 μm Dynamical analysis showed that faceplate resonances were above 1 kHz, with the lowest mode occurring between adjacent actuators (the "quilting" mode), which have the relatively large spacing of 38 mm.

A study performed for the Gemini Telescopes [Bigelow et al. 1993] adopted a different approach for the 1-m-diameter secondary mirror, combining adaptive wavefront compensation at IR wavelengths with a vibration-balanced chopping capability. The proposed mirror employs a thicker faceplate (10 mm)

with magnetostrictive (Terfenol-D) actuators, which combine a long stroke capability with high mechanical stiffness. In both the LBT and Gemini designs, chopping motion is implemented using the adaptive actuators to rotate a thin faceplate as a rigid body about its center of gravity.

Real-time compensation of wavefront tilt (angle-of-arrival variations) and defocus can be accomplished with only three actuators, which can also serve as the chopping drivers in a secondary mirror. This approach has been taken in new secondary mirrors designed for the UKIRT 3.8-m and UH 2.2-m telescopes. For this application, the faceplate is inherently stiff and is driven by three piezoelectric actuators with an operating bandwidth of about 250 Hz. Compensation for the angular momentum of the mirror, which is particularly important in the chopping mode because of its large amplitude, is performed by a second set of three actuators driving a balance mass in the opposite direction. The larger and slower motions in 5 degrees of freedom required for collimation and focus are implemented by a hexapod support.

6.8 Membrane Mirrors

6.8.1 Basic Principles

Membrane mirrors represent an extreme case of deformable thin plate mirrors; the membrane itself has no inherent stiffness, so that tension must be applied to maintain a flat surface. Very small forces are required to displace a membrane, and deflection is usually achieved without physical contact, using electrostatic actuators. All membrane devices require rigid support at their periphery. Membrane mirrors developed for adaptive optics employ one large membrane spanning the full aperture, supported only at the edge in the same way as a pellicle. Their purpose and characteristics are different from those of membrane light modulators, which are used to control the spatial distribution of light on a pixel-by-pixel basis. Modulation devices have a rigid substrate that is physically divided into discrete cells or pixels, which are individually deflected. There is purposely no overlap between individual cells, so that smooth overall deflections cannot be produced. A large-amplitude ripple at the pixel spacing is always present because the mirror surface is pinned between actuators. The maximum deflection achievable is limited to that of a single cell.

These limitations are not encountered in all modulation devices. A solid-state deformable mirror device (DMD) has been developed by Texas Instruments. Tiny silicon mirror elements less than 100 μm^2 are deflected electrostatically by integrated electronic circuits. In some respects, the DMD is a miniature segmented mirror, and could possibly be used for correction of high-order wavefront errors, in conjunction with a separate, low-order corrector.

In contrast to optical modulation devices, the full-aperture membrane mirrors used in adaptive optics wavefront correctors can be deformed into very smooth shapes with large amplitudes, and are therefore well suited to compensate the low-order components of wavefronts distorted by atmospheric turbulence.

The equation of motion of an ideal membrane with no viscous damping, when an external stress $F(r, t)$ is applied, is [Yellin 1976, Grosso and Yellin 1977]

$$\frac{\partial^2 z}{\partial t^2} = \frac{T_m}{B}\nabla^2 z + \frac{P(r,t)}{B} \tag{6.53}$$

where

- z = displacement normal to membrane, m
- t = time, s
- T_m = membrane linear tension (stress x thickness), N m^{-1}
- B = membrane mass per unit area (density x thickness), kg m^{-2}
- ∇ = two-dimensional Laplacian operator
- P = pressure, N m^{-2} (Pa)

At a temporal frequency ω, the equation of motion may be written in the form

$$\nabla^2 z + \frac{B}{T_m}\omega^2 z = \frac{-P(r)}{T_m} \tag{6.54}$$

The steady state deflection ($\omega = 0$) is

$$\nabla^2 z = \frac{-P(r)}{T_m} \tag{6.55}$$

This relation defines the basic property of a membrane: to produce a given deflection, z, the pressure applied must be proportional to the Laplacian of the deflection.

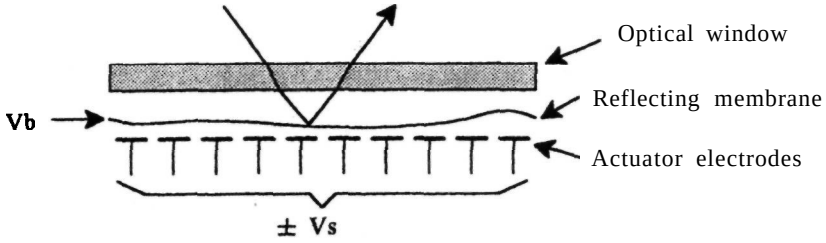
For a membrane of radius r_m, the deflection at radius r produced by the pressure F applied by a central actuator of radius r_a is given by

$$z = \frac{P}{4T_m}\left[r^2 - r_a^2 + 2r_a^2 \ln\left(\frac{r_a}{r_m}\right)\right] \qquad \text{for } r < r_a \tag{6.56}$$

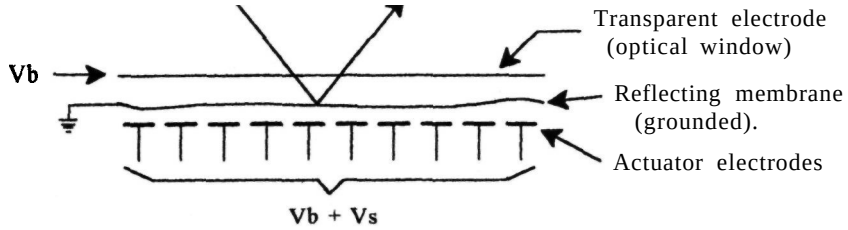
$$z = \frac{P}{4T_m}\left[2r_a^2 \ln\left(\frac{r}{r_m}\right)\right] \qquad \text{for } r_a < r < r_m \tag{6.57}$$

These equations enable the shape of the membrane to be determined for circularly symmetrical cases. Note that these equations are similar to those for a bimorph mirror.

Two basic electrode configurations used in electrostatically driven membrane mirrors are depicted in figure 6.24. The simplest arrangement, shown in figure 6.24(a), employs an array of fixed electrodes in close proximity to the membrane. A fixed bias, usually several hundred volts, is maintained between the membrane and the mean voltage on the electro-



(a) Biased field system



(b) Balanced field system

Figure 6.24 Membrane mirror electrode configurations.

des. The most convenient way to achieve this is to make the mean electrode voltage zero and to apply the bias to the membrane. Bipolar control voltages applied to the electrodes produce local changes in the electric field that change the pressure exerted on the membrane.

The pressure produced by an electric field between two flat electrodes is

$$P = \frac{\varepsilon_0 V^2}{d^2} \tag{6.58}$$

where

ε_0 = permittivity, 8.85 x $10{-}12$ F m ,
V = applied voltage
d = separation of the electrodes

In the arrangement of figure 6.24(a), there is a constant force on the membrane because of the bias voltage, resulting in a parabolic mirror surface. When a signal voltage V_S is applied to an electrode, the local pressure on the membrane changes by an amount AP given by

$$\Delta P = 2\varepsilon_0 \frac{V_B}{d^2} V_S \tag{6.59}$$

The pressure on the membrane, and therefore its deflection, is directly proportional to the product of the bias voltage V_B and the signal voltage V_S, so the sensitivity of the device can be increased by using a large bias voltage. The limit is set by the incidence of ionization in the gap, which causes corona, and ultimately electrical breakdown. Bias voltages of 250-500 have been used in prototype units. The deflection of the membrane for a given signal voltage V_S can then be determined by substituting ΔP for P in equations (6.56) and (6.57).

The second electrode configuration, shown in figure 6.24(b), is the balanced field system in which a transparent layer that functions as a bias electrode is located above the reflecting membrane. The membrane is normally grounded and a bias potential V_B (usually several hundred volts) is applied to both the window and the actuator electrodes. When the control signals are zero, the forces on each side of the membrane are equal, resulting in a flat mirror surface. To deflect the mirror, control voltages are superimposed on the bias on each actuator, resulting in net actuator voltages of $V_B \pm V_S$.The electrostatic forces on the membrane are therefore locally unbalanced and it moves to restore equilibrium.

For this configuration, the net pressure exerted by an actuator on the membrane is

$$P = \varepsilon_0 \left[\frac{(V_B + V_S)^2}{d_1^2} - \frac{V_B^2}{d_2^2} \right] \tag{6.60}$$

where d_1 and d_2 are the electrode spacings. When $V_S << V_B$ and $d_1 = d_2$, this expression reduces to equation (6.59).

6.8.2 Membrane Characteristics

Membranes for deformable mirrors must be optically flat, be capable of taking a high reflectivity coating, and have a high resonant frequency. The lowest resonant frequency (the drumhead mode) for a circular membrane of diameter D is given by

$$f_0 = \frac{2.4}{\pi D}\sqrt{\frac{T_m}{B}} \qquad (6.61)$$

It is necessary to keep membrane resonant frequencies well above the operating bandwidth, which, for some adaptive optics systems, can extend up to 1 kHz. Apart from using as small a diameter as possible, it is necessary for membrane materials to have high tensile strength and low density, in order to maximize the value of T_m/B. Although the membrane thickness does not explicitly appear in equation (6.61), it is desirable to keep it as small as possible to minimize the tension that must be applied to the membrane.

Membranes suitable for use in deformable mirrors have been fabricated from many different materials, including titanium, titanium alloys, nickel, beryllium, molybdenum, and silicon. Non-metallic membranes have been made using nitrocellulose and Mylar™. Metal membranes are produced by evaporating the material onto a polished flat substrate in a vacuum chamber and subsequently floating it off, as described by Yellin [1976]. The thickness of metal membranes is typically about 1 μm, and for nitrocellulose it is about 5 μm. In most cases, a high-efficiency reflective coating is added.

6.8.3 Practical Considerations

Many experimental membrane mirrors have been built in the last 25 years, but they have yet to be accepted for use in astronomical adaptive optics. There are several reasons for this, the most obvious being the fragile nature of the membranes and the difficulty of damping out unwanted vibrations. Damping is accomplished by adjusting the air pressure (density) surrounding the membrane. This requires a sealed mirror cavity with a window, adding two more optical surfaces. At atmospheric pressure, membranes are sluggish. Critical damping requires pressures of only about 20 mm Hg.

While metal membranes can achieve high resonance frequencies of several kilohertz because of their high strength, they are expensive to make. Nitrocellulose and plastic membranes are much cheaper, but have much lower resonant frequencies.

A prototype membrane mirror designed for the Subaru 8-m Telescope has been described by Takami and Iye [1994]. It uses the biased field configuration shown in figure 6.24(a). The membrane is 2-μm-thick nitrocellulose, 50 mm in diameter. The working area is 25 mm^2 with a 5 x 5 array of actuators. A bias voltage of 500 is used, which curves the membrane. The optical curvature is compensated by using a plano-convex lens as the window. The control voltages are ±200. With the membrane cavity evacuated, strong resonances occur at 1.6 and 3.8 kHz. These resonances are suppressed by introducing 5–10 torr of air pressure, resulting in a flat frequency response up to 3.3 kHz.

A membrane deformable mirror has been developed at Johns Hopkins University as part of their adaptive optics program [Clampin et al. 1991]. This device has an active diameter of 37 mm, with 61 central and 30 boundary actuators. The operating voltage is 200 ± 100V, giving a stroke of ±3 μm with 1 KHz bandwidth.

6.9 Refractive Wavefront Correctors

6.9.1 Types of Refractive Corrector

Refractive wavefront correctors are devices in which the refractive index or optical retardation can be varied in response to a stimulus, such as applied voltage. Refractive devices in which the optical pathlength is controlled by rigid body motions are covered in section 6.10.

Several materials, both crystalline and liquid, exhibit birefringence as a result of their molecular structure. When light passes through such materials, it is split into two orthogonally polarized components, known as the ordinary and extraordinary rays, which have different indices of refraction. Electro-optic materials have the capability of changing their index of refraction by the application of an electric field. Such devices, including bismuth silicon oxide crystals and liquid crystals, are used as light valves or image display devices by sandwiching them between crossed polarizers so that, normally, the light is blocked. Application of a voltage across the crystal changes the extraordinary refractive index, which rotates the plane of polarization in the crystal, allowing light to pass through. The change in the refractive index generates optical pathlength changes that can be used to compensate for atmospheric turbulence.

6.9.2 Liquid Crystal Correctors

Liquid crystals are materials that have an intermediate phase between crystalline and liquid, retaining some of the ordered characteristics of the crystalline state, while having the flow properties of a liquid. In liquid crystals, the ordering occurs in one or two dimensions, rather than the three-dimensional ordering that is characteristic of solid crystals. This intermediate state is known as the mesophase and occurs because the liquid crystal molecules, which are rod-like with a length-to-width ratio of more than 3:1, are polarized in such a way that the intermolecular forces

keep them aligned in one direction, called the director. At high temperatures, liquid crystals lose all order and become ordinary liquids.

Liquid crystals are classified as nematic, smectic, or cholesteric. Nematic crystals have molecules that line up in one dimension but are randomly ordered in the other two. Smectic crystals have molecules that line up in a single direction, in well-defined rows. Cholesteric crystals are made up of layers, each consisting of parallel molecules, with successive layers rotated so that the molecules have a spiral configuration. A key property of liquid crystals is that the molecular structure in the mesophase is changed when an electric field is applied, producing a change in the index of refraction. In nematic crystals, the index in the direction of the extraordinary ray can be varied between the ordinary value, n_0, and the extraordinary value, n_e. The difference between these values is the birefringence, An.

Certain types of nematic liquid crystals exhibit large changes in refractive index for quite low applied voltages, with typical values of An being between 0.1 and 0.2 [Wu 1987]. With a cell thickness of 100 μm, this produces an optical path change of 10–20 μm peak to peak, which is sufficient to compensate atmospheric turbulence in telescopes of the 8-m class, assuming that the overall tilt component is removed separately. Liquid crystals have many desirable features, such as small size (1-5 cm), high spatial resolution (electrode size can be as small as 10 μm), low-voltage operation (25 V), no moving parts, and low cost.

The construction of a typical liquid crystal cell is shown in figure 6.25(a). When used for wavefront compensation, the correction zones are defined by an array of transparent electrodes deposited on the windows of the cell. Each correction zone produces a local change in optical pathlength, acting as a piston corrector. Liquid crystals are not well suited to produce overall wavefront tilts, but may be combined with a tip-tilt mirror to form a complete compensation system. This can be accomplished by mounting the liquid crystal directly on the mirror surface so that it operates in the double-pass mode.

The refractive index change is obtained for only one axis of polarization, that of the extraordinary ray. The initial orientation of nematic liquid crystals (which determines the polarization axis) is controlled by surface treatment of the sides of the cell. Microscopic grooves in the surface cause the molecules to line up in the direction of the grooves. There are two methods of operating with unpolarized light: either using two liquid crystal cells in series, oriented with their extraordinary axes at 90°, or using a single cell with a flat mirror in double-pass configuration, as shown in figure 6.25(b). For double-pass operation, a 1/4 wave plate is inserted between the cell and the mirror to rotate the plane of polarization of the reflected beam by 90° [Love 1993]. Each polarization then experiences one passage through the cell.

While liquid crystals appear to have great potential for wavefront correction, two technical issues must be addressed before they can be seriously considered for use in astronomical adaptive optics. The first issue concerns the spectral transmission and dispersion of liquid crystals, which determine the spectral bands over which they may be used. The second issue concerns the response time, specifically whether liquid crystals can be made to operate fast enough for real-time wavefront compensation. These aspects have been studied by a group at the Osservatorio Astrofisico di Arcentri, Firenze, Italy [Bonaccini et al. 1990, 1991, Esposito et al. 1993], with encouraging results that will be briefly summarized.

6.9.3 Properties of Liquid Crystals

For any wavefront correction device in adaptive optics, the essential properties to be established are the dynamic range, the spectral range, and the temporal response. The basic property of liquid crystals is the birefringence, $\Delta n = n_e - n_o$, the difference between the refractive index of the extraordinary and ordinary rays. In the visible and near-IR regions, Wu [1987] has determined that the birefringence at wavelength X may be expressed as

$$\Delta n = G(T)\left[\frac{\lambda^2\lambda_R^2}{(\lambda^2 - \lambda_R^2)}\right] \qquad (6.62)$$

where

- G = parameter dependent on the properties of the liquid crystal
- T = temperature
- λ_R = mean resonance wavelength (which is in the ultraviolet region)

At visible wavelengths, when $X > \lambda_R$, the above expression reduces to Cauchy's formula:

$$\Delta n = G\lambda_R^2\left[1 + \left(\frac{\lambda_R}{\lambda}\right)^2 + \left(\frac{\lambda_R}{\lambda}\right)^4 + \ldots\right] \qquad (6.63)$$

The parameter G, which is temperature-dependent, is given by

$$G = gNZS(f^*_{par} - f^*_{perp}) \qquad (6.64)$$

where

- g = a constant independent of the liquid crystal material
- N = number of molecules per volume
- Z = average number of active electrons per liquid crystal molecule
- S = order parameter of the liquid crystal
- $(f^*par - f^*perp)$ = differential oscillator strength at the mean resonance wavelength λ_R (which is usually around 300 nm)

The value of Z is linearly proportional to the electric field. The values of G and λ_R can be determined experimentally by measuring the birefringence at two or more specific visible wavelengths. With this

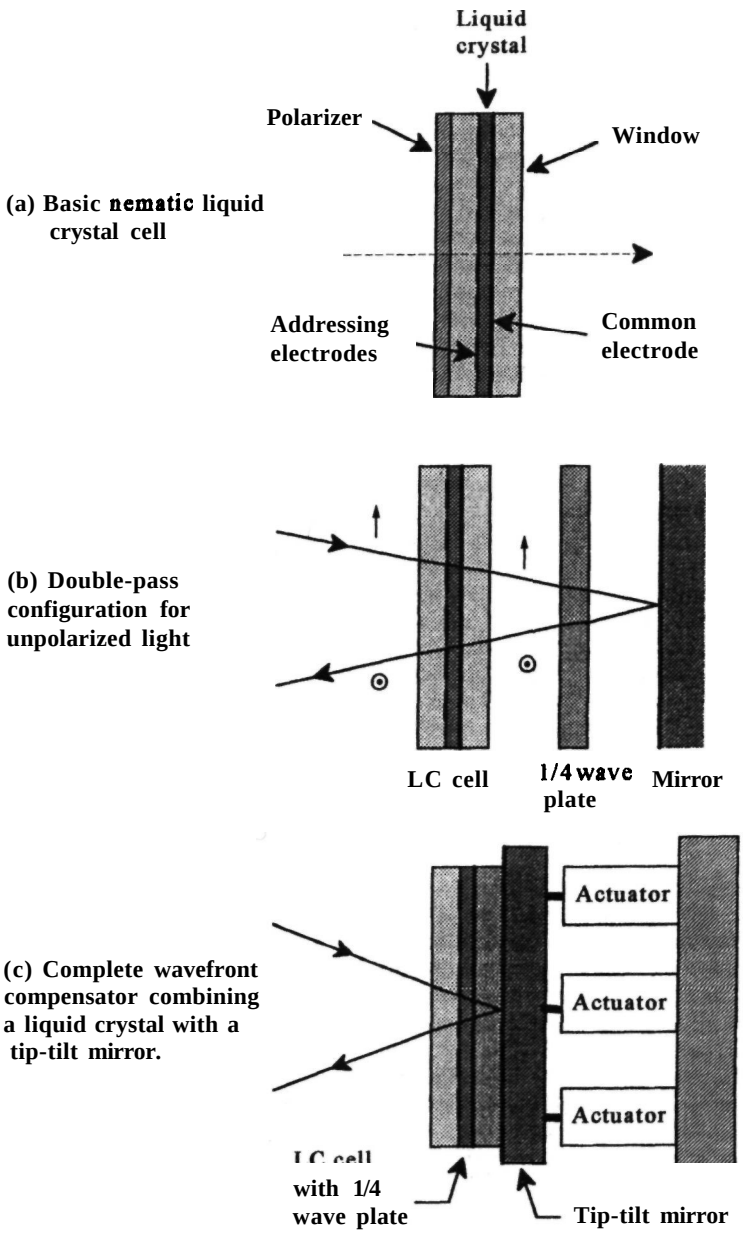


Figure 6.25 Liquid crystal (LC) wavefront compensators.

information, the value of An can then be calculated at other wavelengths using equation (6.62). At 1R wavelengths, when $X >> \lambda_R$, the birefringence An is given by $G\lambda_R^2$, and is essentially independent of/.. The birefringence of several liquid crystals is shown as a function of wavelength in figure 6.26. It is seen that the birefringence is almost constant above 2 μm, but that considerable dispersion is present at wavelengths below about 1 μm.

The value of An is not necessarily constant along the ray path through a liquid crystal cell, as the molecules become aligned parallel to the windows at the two boundaries. The molecule tilt distribution across the cell can be calculated as a function of position and applied voltage, and, from this, the function An(z) is calculated. The dynamic range or effective stroke of the cell is then given by

$$\delta = \int_0^L \Delta n(z)\, dz \tag{6.65}$$

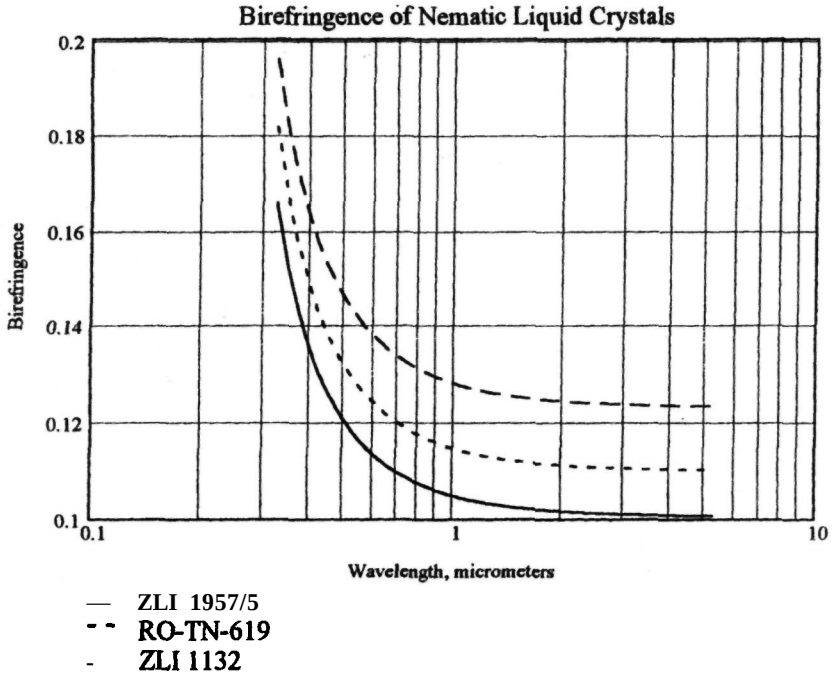


Figure 6.26 Birefringence of three nematic liquid crystals, as a function of wavelength.

where L is the thickness of the cell. An approximate value for S is given by L An. For the liquid crystal ZLI 1132 at a wavelength of 1.25 μm, the value of An is 0.127. Using a 50-μm-thick cell in double pass, the effective stroke is approximately 13 μm peak to peak, which is more than adequate to compensate an 8-m telescope. Equation (6.62) indicates that An increases at shorter wavelengths, leading to the hope that liquid crystals could be used for adaptive optics at visible wavelengths. Unfortunately, this is only possible over narrow spectral bands, because of dispersion.

The spectral bandwidth over which liquid crystals are useful for wavefront compensation is determined by the variations of the refractive indices n_o and n_e with wavelength. First, it is necessary to specify the center wavelength λ_c and the maximum wavefront correction error that can be tolerated at the edges of the spectral band. If this error is, say, ±0.1 wave, then the maximum pathlength error is $\pm 0.1\ \lambda_c$ and the corresponding change in refractive index is $\pm 0.1\ \lambda_c/L$. The spectral range enclosing this allowable dispersion can then be read off the curve. For the liquid crystal material Merck ZLI 2772, Bonaccini et al. [1990] report that a center wavelength of 2.3 μm provides a correction capability over the range 1.7–4.7 μm. The slope of the refractive index curves becomes steeper at shorter wavelengths, showing that the spectral ranges for a given dispersion error get narrower as the visible band is approached.

The spectral transmission of a liquid crystal cell depends on the window and electrode materials, as well as on the liquid crystal itself. In general, cell thicknesses are less than 0.1 mm and the nematic liquid crystal compounds are highly transparent in the visible and IR as far as 25 μm. However, each material has specific absorption bands because of molecular resonances. For the electrodes, indium tin oxide conductive coatings of about 150-ångstrom thickness have good transparency over the visible and near-IR bands. The resistance of 2000 ohms per square is satisfactory, given the very low current required.

The response time of liquid crystals is the parameter that needs the most attention, especially the decay time, which is normally three to four times longer than the rise time. Bonaccini et al. [1990] derive the following expressions for the rise time t_r and fall time t_f:

$$t_r = \frac{\gamma_1 d^2}{(\Delta\varepsilon V^2/4\pi) - K_{11}\pi^2} \tag{6.66}$$

$$t_f = \frac{\gamma_1 d^2}{K_{11}\pi^2} \tag{6.67}$$

where

γ_1 = viscosity
d = cell thickness
$\Delta\varepsilon$ = dielectric anisotropy at frequency ν of the applied electric field
V = applied voltage
K_{11} = splay elastic coefficient

The rise and fall (decay) times are both dependent on the cell thickness d and the elastic coefficient K_{11}, and are damped by the viscosity γ_1. The rise time has an additional term in the denominator proportional to the applied voltage, which must overcome the elastic force to turn the cell on. This driving term is several times larger than the elastic term, resulting in a rise time of a few milliseconds. But the fall time has only the elastic coefficient in the denominator, resulting in a much slower decay.

Equations (6.66) and (6.67) appear to be conveying bad news. The rise and fall times t_r and t_f both depend directly on cell thickness d; the stroke also depends on d, so it does not seem possible to get both a fast response and a large stroke from a liquid crystal. Fortunately, there is a method of driving liquid crystals that will circumvent this problem.

The decay time problem can be solved using the dual-frequency effect, which exploits the fact that in some nematic liquid crystal mixtures, the dielectric anisotropy Ae reverses its sign above a threshold frequency, ν_0, which is usually around 10 kHz. This sign reversal causes the electric field to align the molecules orthogonally to its direction, so that it acts as a restoring force. This technique is implemented, in practice, by driving the cell with two sinusoids, at frequencies above and below the threshold frequency ν_0. The amplitude ratio of the sinusoids then determines the actual tilt angle ϕ of the molecules, which controls the phase shift of the device. Equal rise and fall times have been obtained using this approach, with a slew rate of about 20 μm s^{-1} [Bonaccini et al. 1991].

The temporal response of liquid crystal wavefront correctors may be further improved by using a closed-loop servo with capacitive feedback to drive the individual zones. The capacitance of each zone is directly proportional to the optical path difference produced in the cell. A high-gain local feedback loop using this capacitance as a sensing element can therefore be used to improve the response time of the device.

Although much work remains to be done on developing liquid crystals for use as wavefront correctors, the results so far are positive. In view of the potential advantages of such devices in terms of high spatial resolution and low cost, it is hoped that practical devices will become available in the near future.

6.10 Tracking Mirrors

6.10.1 Requirements

Compensation of overall wavefront tilt, or the angle of arrival, is the simplest type of adaptive optics, and it has been used in astronomical telescopes since the 1950s. One of the first tilt correctors employed a gimbal-mounted quartz plate, tilted on two axes by small servomotors, to maintain the image of a star on the jaws of a spectrograph slit [Babcock et al. 1956]. Most tilt correctors in use today use flat mirrors that are mechanically tilted by electromagnetic or piezoelectric actuators. The temporal bandwidth required to compensate wavefront tilt is on the order of 100 Hz, but the mechanical resonances must be much higher, so the physical size of these devices is relatively small, the aperture usually being less than 10 cm.

The main performance requirements to be specified for astronomical tracking mirrors are:

1. maximum tilt or travel,
2. servo bandwidth or slew rate,
3. optical quality,
4. positioning accuracy (open loop),
5. angular resolution (precision),
6. tracking accuracy (closed loop),
7. jitter or noise level,
8. mechanical envelope,
9. external reaction forces.

The angular tilt error due to atmospheric turbulence is determined by the D/r_0 ratio of the telescope. The single-axis mean-square angular tilt in radians squared, over an aperture of diameter D, is given by

$$\sigma_\alpha^2 = 0.182\left(\frac{D}{r_0}\right)^{5/3}\left(\frac{\lambda}{D}\right)^2 \tag{6.68}$$

where r_0 is the turbulence coherence length at wavelength λ. The tilt angle is actually independent of wavelength, because r_0 varies as $\lambda^{6/}$. When an optical beam is reduced in diameter, the pathlength errors are unchanged, with the result that tilt angles are magnified by the ratio D/d, where d is the diameter of the tilt correction mirror.

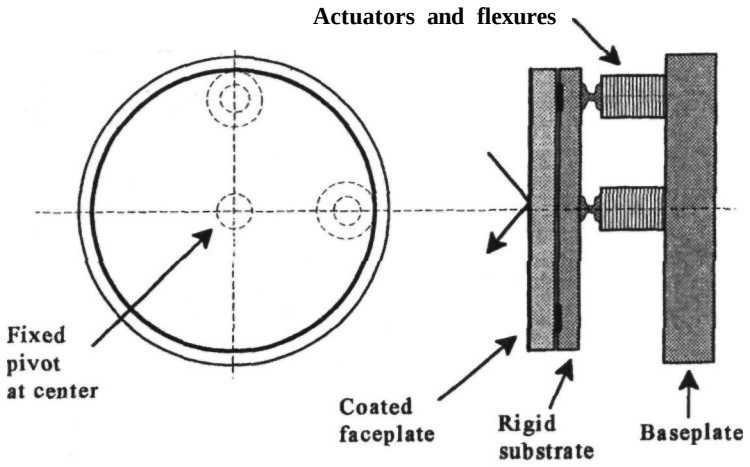
The peak angular excursion of the wavefront may be estimated using 2.5 standard deviations, which is exceeded only 0.5% of the time with Gaussian statistics. The peak-to-peak mechanical tilt requirement for the tracking mirror is half of this value, being equal to

$$\alpha_{PP} = \pm 1.25\frac{D}{d}\sigma_\alpha \tag{6.69}$$

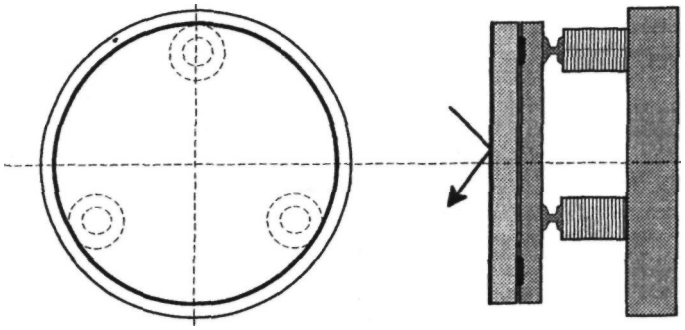
In the case of a 4-m telescope with r_0 equal to 0.1 m at X — 0.5 μm the overall tilt angle at the primary mirror has a standard deviation of 0.24 arc seconds rms. The peak-to-peak correction capability required for a tracking mirror of 0.1 m diameter is then ±12 arc seconds.

6.10.2 Types of Tracking Mirror

Typical drive systems for two-axis mirrors are shown schematically in figure 6.27. In each case, the substrate is supported by three flexures that constrain it mechanically, while allowing small angles of tilt. The stiffness of the flexures is much less than that of the substrate, to avoid distorting the structure. The faceplate support must be designed carefully to avoid transmitting any strain from the substrate to the mirror surface.



(a) Fixed pivot with **two peripheral actuators.**



(b) Three equispaced **actuators**

Figure 6.27 Drive systems for two-axis mirrors.

It is usually necessary to eliminate axial displacement of the mirror surface caused by tilt. In the first scheme, this requirement is achieved by placing a fixed pivot flexure at the center of the aperture, with two actuators located 90° apart. This arrangement is easy to implement, but it is mechanically unsymmetrical and it produces undesirable coupling between the actuators. The second scheme employs three identical actuators, symmetrically located at 120° spacing. It is more stable mechanically than the first scheme. All three actuators are driven, requiring a coordinate transformation of orthogonal tilt-correction signals. In some cases, it is necessary to eliminate the angular momentum produced by a tracking mirror. This is done by adding a counterbalance mass which is driven antiphase to the mirror, as shown in figure 6.28.

Tracking mirrors usually employ linear actuators, which may be electromagnetic, piezoelectric, or magnetostrictive. These devices have been described in section 6.3. The largest strokes are obtained with electromagnetic and magnetostrictive actuators, but current must be supplied to maintain the deflection, leading to heat dissipation in the coil windings. Piezoelectric and electrostrictive actuators are capacitive devices, with small internal losses. Although a large electric current may be required to change the voltage across the actuators, the power dissipation within electrodistortive devices is usually small at the frequencies used for wavefront tilt correction. An alternative drive system employs a **gimbal,** which, for the small angles required, may also be implemented with flexures.

Magnetic suspension has some potential advantages in mirrors for astronomical applications. By eliminating mechanical flexures, such mirrors are given the freedom to move in three dimensions, a useful capability for telescope secondaries. The

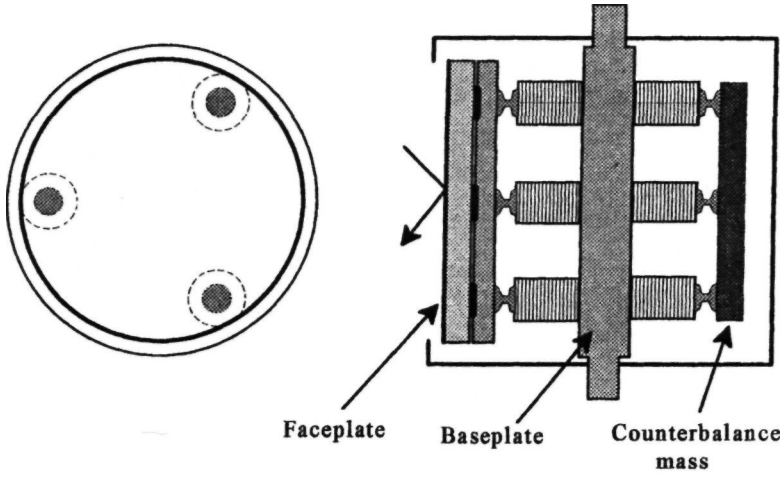


Figure 6.28 Tip-tilt mirror with compensation for angular momentum.

servo systems required for magnetic suspension are considerably more complex than those employed in mechanically constrained mirrors [Medbury and Gupta, 1991].

One of the main problems in designing fast tracking mirrors is to get the natural resonance of the structure well above the operating bandwidth, preferably greater than 1 kHz. This requires minimizing the moving mass and using a stiff support system. In general, higher bandwidth devices consume more power. Position sensors are often required to provide real-time monitoring of the mirror angle and also to provide dynamic feedback to the servo. Eddy current sensors are suitable for this purpose.

Several types of piezoelectric tracking mirrors suitable for astronomical adaptive optics have been described by Marth et al. [1991].

6.11 Dirigible Optics

6.11.1 General Description

Dirigible optics are rigid optical components that are steered or moved in real time to compensate wavefront distortions. Their wavefront correction capability is generally limited to low-order modes, such as tilt, defocus, astigmatism, and coma. A tip-tilt mirror may be considered an elementary device of this type, but the term is intended to describe powered components, both lenses and mirrors, that may be physically moved in five of their six possible degrees of freedom. Components with optical power, such as curved mirrors and lenses, are moved laterally to compensate wavefront tilt, and displaced axially to correct defocus. Higher order corrections can be obtained by tilting these components with respect to the optical axis, as described by Hardy and Wallner [1994].

There are several advantages in using rigid components to compensate low-order aberrations. The optical elements employed are small (typically less than 100 mm in diameter), intrinsically stable, and they are routinely manufactured to a high degree of perfection. In many cases, inexpensive off-the-shelf components can be used. Each component spans the entire beam collected by the telescope (at a reduced pupil diameter), so that wavefront phase corrections are implemented in a smooth and predictable way over the whole aperture. This approach completely eliminates the local irregularities that can occur in zonal wavefront compensation devices, such as the actuator footprints that are found in some deformable mirrors and the gaps that scatter light in segmented mirrors. Another practical advantage of dirigible optics is that the wavefront changes produced by displacing the elements are differential rather than absolute, so the stability requirements are less stringent than with deformable mirrors. The physical motions are larger, being measured in millimeters and degrees rather than micrometers and arc seconds, thereby making the elements easy to control.

6.11.2 Single Active Lens

Overall wavefront tilt produces image motion; it is normally compensated with a two-axis tracking mirror. An alternative method of compensating overall tilt is to decenter a lens in the imaging path. The motions required to compensate the tilt component of atmospheric turbulence are small, and no unwanted aberrations are introduced because the lens remains on-axis to the displaced image. In fact, it can be argued that using an existing lens in the optical path to compensate for overall tilt produces

less overall image degradation than adding a separate tip-tilt mirror.

A suitable location for the tilt correction function is the relay lens, following the prime focus of the telescope primary, as shown in figure 6.29. The field lens located at the prime image plane reimages the telescope pupil at the active lens. This lens collimates the beam at a small diameter, typically a few centimeters, and relays it to the camera objective. Tilt correction involves moving the active lens in its x, y plane to restore the direction of the collimated beam parallel to the optical axis. For input wavefront tilts of θ_x, θ_y in star space, the image displacements at the prime focus are $Ax = \theta_x F$ and $Ay = \theta_y F$, where F is the telescope focal length. To restore the collimated beam parallel to the optical axis, it is evident that the lens must also be moved by the amount Ax, *Ay.* Note that the lens itself is operating on-axis as far as the image is concerned; the displacement produces a small increase in its cone angle.

Overall defocus of the wavefront is easily compensated in the conventional way by moving the active lens in its z axis. For a peak defocus of p_d waves, the axial displacement required (for paraxial conditions)

$$\Delta z = 8 p_d \lambda N^2 \tag{6.70}$$

where

λ = wavelength
N = f-number of the active element

Tilting a rigid lens produces the classic off-axis aberrations of astigmatism and coma, which can be used to compensate for distortions of the incoming wavefront. The field angles used in compensated imaging for astronomical telescopes are extremely small, so that lens aberrations dependent on field angle (such as astigmatism and coma) are effectively produced only by the deliberate tilt angle of the lens. When a lens is tilted, its focal length is shortened in the plane of tilt, perpendicular to the tilt axis. For a lens tilt angle of α, the change in focal length is proportional to αF while the coma is proportional to $\alpha F/N^2$.It is evident that independent values of astigmatism and coma cannot be realized with a single tilted lens. Furthermore, a single lens has only two independent tilt axes, while astigmatism and third-order coma together have four independent parameters. It is therefore necessary to use two active lenses in order to provide independent compensation of astigmatism and coma.

6.11.3 Two-Lens Active Corrector

The use of multiple active lenses for wavefront compensation has been investigated by Hardy and Wallner [1994]. A two-lens active corrector and its slope vector diagram are shown in figure 6.30. Each lens has five degrees of freedom: $x, y,$ and z displacements, plus tilts about two orthogonal axes, α and β. The lenses are tilted at orientations ϕ_1 and ϕ_2 by amounts equal to M_1 and M_2, where

$$\begin{aligned} \phi_1 &= \tan^{-1}(\beta_1/\alpha_1) \qquad M_1 = (\alpha_1^2 + \beta_1^2)^{1/2} \\ \phi_2 &= \tan^{-1}(\beta_2/\alpha_2) \qquad M_2 = (\alpha_2^2 + \beta_2^2)^{1/2} \end{aligned} \tag{6.71}$$

The resulting wavefront distortions are:

x-tilt

$$T1 = k_T[M_1 \cos\phi_1 + M_2 \cos\phi_2]$$

y-tilt

$$T2 = k_T[M_1 \sin\phi_1 + M_2 \sin\phi_2]$$

Defocus

$$D = k_D[M_1^2 + M_2^2]$$

x-astigmatism

$$A1 = k_A[M_1^2 \cos 2\phi_1 + M_2^2 \cos 2\phi_2]$$

y-astigmatism

$$A2 = k_A[M_1^2 \sin 2\phi_1 + M_2^2 \sin 2\phi_2]$$

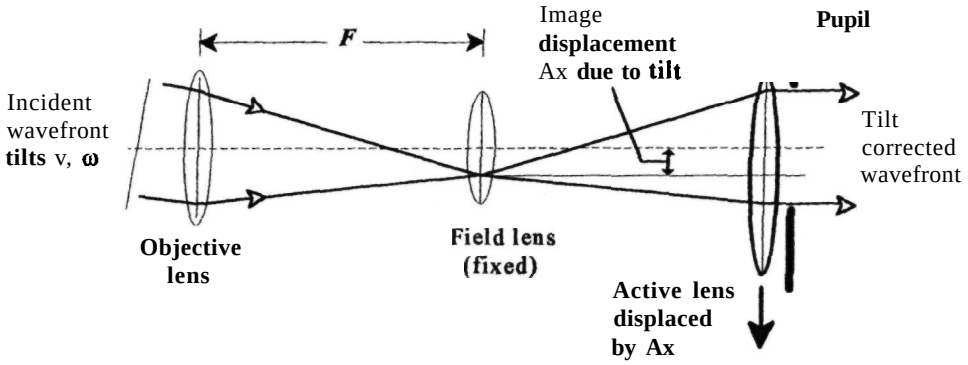


Figure 6.29 Wavefront tilt correction using an active lens.

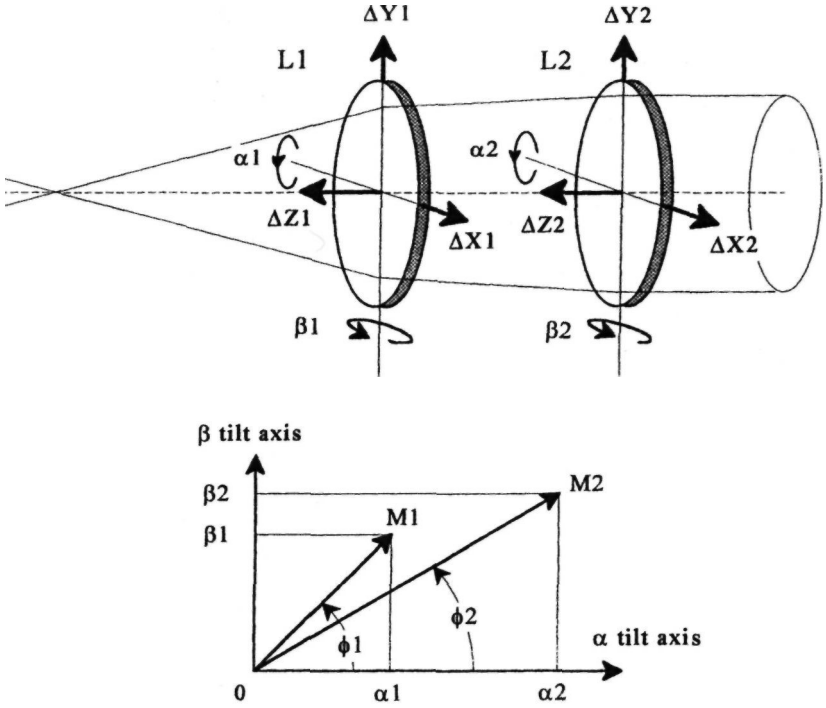


Figure 6.30 Two-lens active corrector and slope vector diagram. Independent control of astigmatism and coma is obtained by tilting the two lenses independently with tilt magnitudes M_1 and M_2 at orientation angles ϕ_1 and ϕ_2. These tilts are produced by orthogonal rotations a and β on each lens as shown. The lenses may also be displaced by A.x, *Ay*, and Az to compensate image motion and defocus.

x-coma

$$C1 = k_C[M_1 \cos\phi_1 + M_2 \cos\phi_2]$$

y-coma

$$C2 = k_C[M_1 \sin\phi_1 + M_2 \sin\phi_2]$$

Spherical

$$S = k_S[M_1^2 + M_2^2] \tag{6.72}$$

The coefficients k_T, k_D, k_A, k_C, and k_S depend on the focal length and focal ratio of the lenses.

It is necessary to control these functions independently in order to compensate random wavefronts, such as those produced by atmospheric turbulence. Overall wavefront tilt (image motion) is compensated by moving one or both active elements laterally in their *x* and *y* axes, as described above. Defocus is compensated by moving one or both elements in the z axis. Spherical aberration is corrected by equal and opposite displacements of the active lenses, maintaining the overall focus constant. The lens displacements required to make these wavefront corrections are orthogonal and the wavefront corrections produced are essentially independent.

To produce astigmatism at any angle ϕ without coma, the two lenses are tilted equally in opposite directions, so that

$$\begin{aligned} \phi_1 &= \phi \\ \phi_2 &= \phi + \pi \\ M_1 &= M_2 = M \end{aligned} \tag{6.73}$$

The resulting wavefront distortions are

$$\begin{aligned} A1 &= k_A[M^2 \cos 2(\phi) + M^2 \cos 2(\phi + \pi)] \\ &= 2k_A[M^2 \cos 2\phi] \\ A2 &= k_A[M^2 \sin 2(\phi) + M^2 \sin 2(\phi + \pi)] \\ &= 2k_A[M^2 \sin 2\phi] \\ C1 &= k_C[M \cos(\phi) + M \cos(\phi + \pi)] = 0 \\ C2 &= k_C[M \sin(\phi) + M \sin(\phi + \pi)] = 0 \end{aligned} \tag{6.74}$$

To obtain coma at any angle ϕ without astigmatism, the lenses are tilted at ±45° to the direction of the desired coma, which causes the astigmatism to cancel out. Then

Table 6.7 Two-Lens Active Wavefront Compensator

Aberration	Lens displacements		Peak values for 4-m aperture
	L1	L2	r_0 = 10 cm at X - 0.55 μm
x tilt	+ x	$+x$	0.20 mm
y tilt	$+y$	$+y$	0.20 mm
Defocus	+ z	+z	5.10 mm
Astigmatism, 0°	+ α	$-a$	14.5°
Astigmatism, 45°	+ $(\alpha + \beta)$	$-(\alpha + \beta)$	14.5°
x coma	+ $(\alpha + \beta)$	+ $(\alpha - \beta)$	6.9°
y coma	+ $(\alpha + \beta)$	$-(\alpha - \beta)$	6.9°
Spherical	+ Z	$-z$	

$$\begin{aligned} \phi_1 &= \phi + \pi/4 \\ \phi_2 &= \phi - \pi/4 \\ M_1 &= M_2 = M \end{aligned} \tag{6.75}$$

so that

$$\begin{aligned} A1 &= k_A[M^2 \cos 2(\phi + \pi/4) + M^2 \cos 2(\phi - \pi/4)] = 0 \\ A2 &= k_A[M^2 \sin 2(\phi + \pi/4) + M^2 \sin 2(\phi - \pi/4)] = 0 \\ C1 &= k_C[M \cos(\phi + \pi/4) + M \cos(\phi - \pi/4)] \\ &= \sqrt{2}M \cos\phi \\ C2 &= k_C[M \sin(\phi + \pi/4) + M \sin(\phi - \pi/4)] \\ &= \sqrt{2}M \sin\phi \end{aligned} \tag{6.76}$$

In this case, although the astigmatic terms are zero, an overall defocus component of magnitude $D_A = k_D[2M^2]$ is produced. This is compensated by shifting both lenses in their z axes.

The magnitude of the higher order residual aberrations produced by lens tilt has been investigated by optical ray tracing. Residual aberrations are defined as those remaining after subtraction of astigmatism and coma (the desired corrections), and after nullifying the low-order aberrations produced by lens tilt (wavefront tilt, defocus, and spherical aberration), which are removed by x, y, and z displacements. The required corrections are proportional to the tilt angles and are therefore predictable and easily implemented. The residual wavefront error produced by a lens tilt of 5° was 0.005 μm, or about 1/100 wave at visible wavelengths. The residual aberrations produced by lens tilts of this magnitude are therefore negligible.

The active lens motions required to compensate the basic wavefront aberrations produced by atmospheric turbulence with a Kolmogorov spectrum are summarized in table 6.7. The peak values (2.5σ) are those encountered in a 4-m telescope with $f/16$ focal ratio, for a turbulence strength of r_0 = 10 cm, at a wavelength of 0.55 μm. The peak excursions are seen to be small and easily implemented, even for a relatively large telescope.

Wang and Markey [1978] analyzed the effectiveness of modal compensation of atmospheric turbulence. Their results show that correction of low-order wavefront modes through Zernike term Z_8 (third-order coma) produce the best resolution when the aperture size was $6.76r_0$, giving a resolution improvement of about 13 times compared with a large uncompensated aperture. For good seeing conditions, (r_0 = 15–20 cm at λ = 0.5 μm) the two-lens active corrector should therefore compensate an aperture of 1-1.4 m at visible wavelengths, while in the IR at 2.2 μm, it should compensate apertures of 6–8 m.

7 Laser Beacons

7.1 Introduction

7.1.1 The Need for Reference Beacons

All adaptive optics systems require a means for measuring the propagation path to provide the data to drive the wavefront compensator. Conventionally, this is achieved with a distant source of radiation that traverses the turbulent medium, bringing to the wavefront sensor a record of the wavefront distortion encountered during its journey. In the case of orbiting satellites, for which the first image compensation systems were designed, the solar-illuminated spacecraft themselves functioned as the reference source. By chance, these objects fulfill the two main requirements for adaptive optics reference sources: they are bright, and their size and orbit are such that they are usually contained within a single isoplanatic patch.

In astronomy, there are relatively few objects of scientific interest that combine these special characteristics of size and brightness: the Galilean satellites of Jupiter are rare examples. Most objects of current interest in astronomy are dim extended sources, such as gaseous clouds, protoplanetary disks, star clusters, and nebulae, with no nearby reference source. To make adaptive optics available for the majority of astronomical observing tasks it is therefore necessary for the user to provide a suitable reference source. Many different terms have been coined for such sources, including artificial stars, synthetic beacons, and laser guide stars. As they are not stars, and do not, in fact, perform the function of guiding (for which natural stars are required), the man-made reference sources will be described in this book as laser beacons.

Laser beacons were originally developed for a purpose quite remote from astronomy, specifically for the direction of high-energy laser beams to targets in or above Earth's atmosphere. To obtain a useful concentration of energy, the wavefront degradation suffered by the beam in propagating through the atmosphere must be very small, much less than can be tolerated for imaging. A wavefront error of 1/10 wave rms, for example, reduces the peak intensity of a focused beam to 67% of its unaberrated value, although the size of the spot is almost unchanged. It is therefore essential, for efficient power transfer through the atmosphere, to have a means of compensating laser beams for turbulence. Using the principle of optical reciprocity, if the propagation path is measured and corrected for a beam traveling from a distant target to the telescope, then the same correction applied to an outgoing beam will cause the beam to be focused at the target. In the case of orbiting objects traveling at high velocities, the measurement of the appropriate propagation path presents some difficulties. If the turbulence is measured using a beacon on or near the target, then by the time the light has made a round trip from the beacon to the ground and back again, the target will have moved to a new position, probably outside the isoplanatic patch of the measured wavefront. To take account of the finite velocity of light, the upgoing laser beam must be pointed ahead of the apparent target position and the wavefront must be measured in this direction.

The geometry is shown in figure 7.1. The point-ahead angle is $\alpha = 2V/c$ where V is the velocity of the target and c is the speed of light. For low-Earth-orbit satellites, α is typically about 10 arc seconds, which considerably exceeds the isoplanatic angle at visible wavelengths. A reference beacon is required at the location that the object will occupy when the laser beam arrives. With a cooperative target, it is possible to attach the beacon to the target itself; this was done for some of the ground-to-space laser propagation experiments using rockets. For the more general case of uncooperative targets, a beacon must be created in the desired location ahead of the target.

The use of a beacon approximating a point source is appropriate for transmitting laser beams to a target, but it produces inadequate sampling of the turbulence for the very distant objects observed in astronomy. The basic function to be performed is measurement of the turbulent propagation path and there are other ways to achieve this. For example, a probe beam transmitted through the telescope aperture may be configured so that wavefront measurements can be made with scattered light during the upward path of the beam toward the object. This approach offers two important advantages over the use of focused beacons:

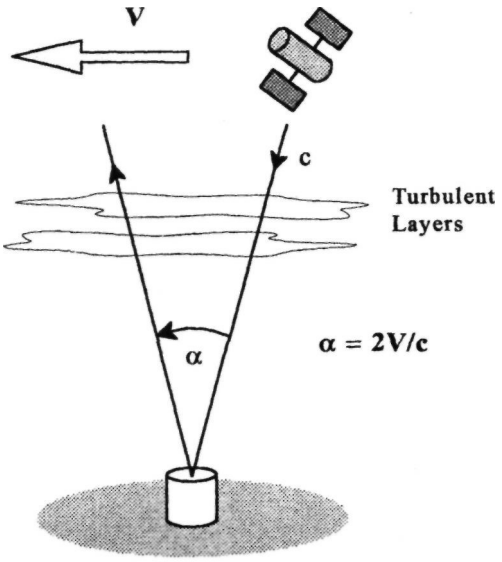


Figure 7.1 Point-ahead geometry for moving targets.

1. The measurement is made over the whole cylindrical aperture, so that all turbulence in the path is measured and not just that in a conical area below the beacon.
2. The three-dimensional structure of the turbulence may be resolved by range gating, leading to more efficient wavefront correction and the possible use of distributed (multiconjugate) compensation to increase the field angle.

The probe beam approach was one of the first methods to be proposed for laser wavefront measurement [Hutchin 1991] and it is described in the following section.

7.1.2 A Brief History of Laser Beacons

In common with conventional adaptive optics, the development of laser beacons for wavefront measurement went through a period of experimentation in which numerous approaches were tried before a practical direction emerged.

In the late 1970s, it became evident that the potential applications of adaptive optics, both for imaging and for laser energy transfer, were severely limited by the lack of suitable reference sources. The basic problem was to measure the wavefront disturbance due to turbulence in any desired direction through the atmosphere. For compensated imaging, the measured path must be in the same isoplanatic patch as the observed object. For directed laser systems with rapidly moving objects, the measured path must be pointed ahead of the target, as explained above.

Studies of methods that did not rely on natural sources for measuring atmospheric distortion were encouraged by the Strategic Technology Office of the Advanced Research Projects Agency (ARPA). The first idea, devised at Itek in 1978 by R. A. Hutchin, was called the "ray method" and is illustrated in figure 7.2. Laser pulses are fired from multiple subapertures of the telescope, tracing out the beam path through the atmosphere. The rays are deviated by the turbulence, and their positions, made visible by molecular scattering, are tracked from two sides. The effects of atmospheric turbulence in the viewing path are minimized by measuring the differential displacements of the rays. This approach is more basic than using a distant reference source because it allows a three-dimensional model of the turbulence to be reconstructed. Potentially, this would enable multiconjugate compensation of turbulence over a significant field angle, eliminating the anisoplanatic problem.

Experiments based on this concept, sponsored by ARPA, were performed by Itek at the Rome Air Development Center's (RADC) test range at Verona, New York, in 1980. These tests showed that implementation of the ray method would be extremely difficult. While it might be useful at infrared (IR) wavelengths, using a small number of probes and a slow frame rate, it did not appear to be feasible at visible and shorter wavelengths where a much larger number of probes would be required. The search for other approaches was therefore continued.

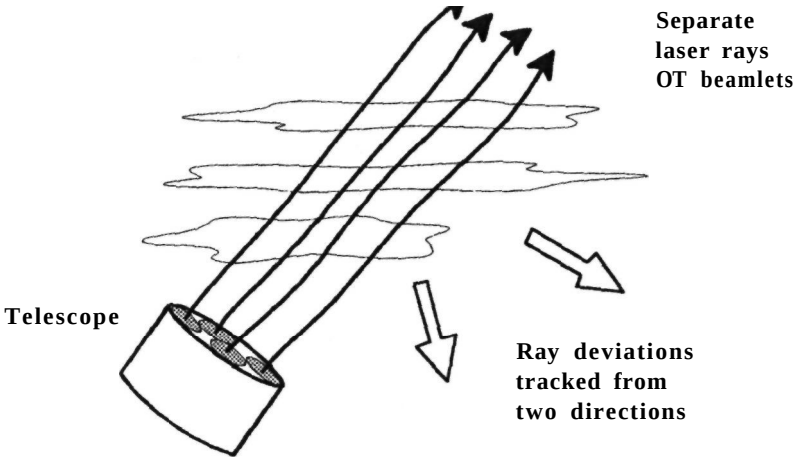


Figure 7.2 Ray method of probing atmospheric turbulence.

In 1981, the idea of creating an artificial star in the atmosphere using the Rayleigh (molecular) scattering from a focused laser beam was independently proposed by J. Feinlieb of Adaptive Optics Associates (AOA) and Hutchin of Itek. A similar idea was also proposed by R. Hunter et al. of Western Research Corporation [Benedict et al. 1994]. This method, which originally was called the "Astral" or A-method, is depicted in figure 7.3. To achieve a reasonable scattering cross-section, the beacon must be located at an altitude where air density is still appreciable, typically between 10 and 20 km. It was realized from the start that using a single low-altitude beacon within the atmosphere was inefficient, because turbulence below the beacon is sampled only in a cone-shaped volume, and turbulence above the beacon is not sampled at all. Sampling efficiency can be improved somewhat by using multiple beacons within the beam, but the best solution is to use a near-parallel beam.

Because of the sampling problem with Rayleigh beacons, Hutchin continued the search for improved methods of measuring atmospheric turbulence with a laser. In the summer of 1981, he conceived the "shear method" or S-method, shown in figure 7.4, in which full-aperture pulsed laser beams are transmitted from the telescope along the optical path to be measured. Each laser pulse consists of two coherent components mutually tilted at a small angle, giving rise to a set of (initially) equally spaced interference fringes normal to the direction of tilt. As the beams propagate through the atmosphere, the wavefront distortion they encounter is encoded in the form of displacements in the fringe positions. The fringes are made visible by Rayleigh scattering. Snapshots of the "interferograms" produced can be obtained at any altitude by a range-gated camera looking through the telescope aperture. The pulses would normally be fired in pairs, producing two sets of orthogonal fringes. The fringe displacements measured from each pair, at a given altitude, represent the accumulated wavefront slopes in two orthogonal directions and may be reconstructed to give the complete wavefront phase at any altitude, using algorithms similar to those used in shearing interferometers.

The shear method produces a three-dimensional representation of the turbulence in the whole optical path, allowing wide-field compensation with multiple correctors. Analysis of this approach showed that best results would be obtained using "bootstrapping"; that is, compensating the lower layers first and gradually working up through the beam. The principle of the shear method was verified by simulations and laboratory experiments. It was found that the system was not as robust as simpler methods that use focused beacons, the wavefront measurements being sensitive to small imperfections in operation.

In spite of the drawbacks of the A-method using Rayleigh scattering, it has many advantages, including the fact that it is compatible with conventional wavefront sensors and may be used over a wide range of wavelengths for which lasers are readily available. In 1982, D. L. Fried calculated the effects of the undersampling due to a conical beam, for which the term "focal anisoplanatism" had been coined by T. R. O'Meara. Based on the recommendations of the scientists working on the project, R. Benedict at ARPA decided to proceed with the development of a Rayleigh backscatter system using a pulsed laser.

To reduce the measurement error due to focal anisoplanatism, which becomes a major problem with large apertures, the use of multiple beacons was investigated. The main challenges here were, first, to find a method of measuring the wavefront from each beacon separately within the allowable time, and then to combine the data into an overall wavefront error signal that would be sent to the deformable mirror or other correction device. Two methods were investigated: "stitching" and "mer-

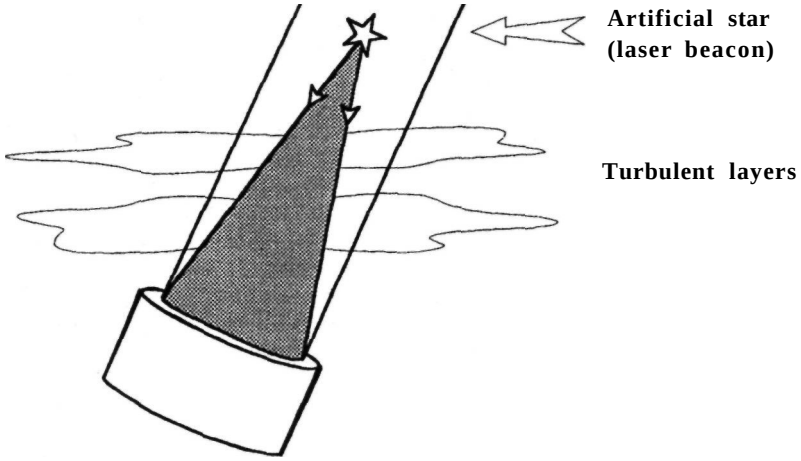


Figure 7.3 Turbulence measurement with a single beacon (A-method).

ging." The stitching method measures the wavefront over sections of the main aperture immediately below each beacon and then determines the best way to combine the measurements. The merging technique measures the wavefront over the entire telescope aperture for every beacon and computes the best overall estimate. Theoretically, merging gives a slightly better correction than stitching, but the computational load is heavier.

In June 1982, a major advance in the development of laser beacons ocurred when W. Happer, a member of the Jasons, a group of university research scientists that works on problems of interest to the Department of Defense, suggested that laser beacons created **in** the mesospheric sodium layer at 90 km altitude might be used for atmospheric turbulence compensation. A revised version of the Jason reports on this subject written in 1982 and 1984 has recently been published [Happer et **al.** 1994]. The main advantage of using such high-altitude laser beacons is the reduction in focal anisoplanatism compared to that encountered with low-altitude Rayleigh beacons.

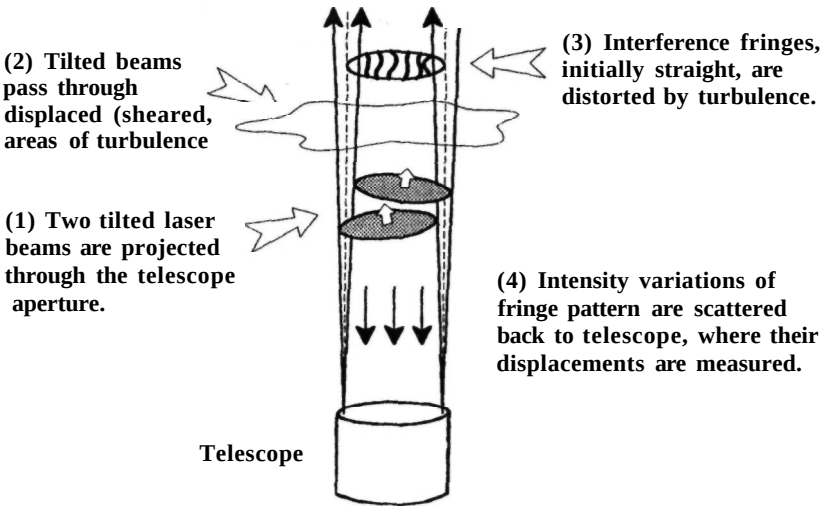


Figure **7.4** Shear method of turbulence measurement (S-method).

With the establishment of two possible methods of generating laser beacons, the next step was to confirm that reliable wavefront measurements could be made using this new idea. The first confirmation came in August of 1983 when R. Fugate's team at the Air Force Weapons Laboratory (now Phillips Laboratory) compared the wavefront measurements made with a Rayleigh beacon at 10 km altitude with those from a natural star. The laser used was a doubled neodymium-yttrium aluminum garnet (Nd:YAG) operating at 532 nm. The wavefront slope measurements were made with a Hartmann sensor equipped with 18 subapertures. The measurements were consistent with theoretical predictions made by Fried. This milestone was followed in September 1984 with the first measurements made with a sodium layer beacon at an altitude of 90 km. This experiment was conducted by R. Humphreys and C. Primmerman of the Massachusetts Institute of Technology (MIT) Lincoln Laboratory at the White Sands Missile Range. A pulsed dye laser tuned to the sodium D_2 line was used to create the beacon. Wavefront tilt measurements made in two adjacent subapertures were compared with similar measurements made with a natural star. Again, good agreement was obtained with theory.

The final step in this investigation was to incorporate the laser beacons into closed-loop adaptive optics systems. In the mid-1980s, a flashlamp-pumped pulsed dye laser operating at 0.512 μm was installed by MIT Lincoln Laboratory at the Air Force Maui Optical Site (AMOS) for Rayleigh beacon experiments [Zollars 1992]. The usable energy was about 3 joules per pulse, at a repetition rate of 5 Hz. This laser was used with the Short Wavelength Adaptive Techniques (SWAT) adaptive optics system [Barclay et al. 1992] to perform the first turbulence compensation experiments with a laser beacon. The SWAT system employed a telescope of 0.6-m aperture with a 241-actuator deformable mirror and a 218-subaperture Hartmann wavefront sensor. Because the laser pulse rate was much lower than the rate at which atmospheric distortion changes, the compensation was performed on a pulse-by-pulse basis; that is, without using information from previous pulses, as is usually the case in closed-loop control systems.

In August 1988, the first compensated stellar images were obtained with the SWAT system, using a Rayleigh beacon. Later tests with a single laser beacon at an altitude of 6 km produced images having Strehl ratios of approximately 0.4, close to the predicted performance. Subsequent experiments were made with two laser beacons, and it was shown that the images obtained were better, by a small margin, than those obtained with a single beacon, thus establishing the feasibility of multiple-beacon operation.

Closed-loop compensation using a laser beacon with a high pulse rate was demonstrated in 1989 by Fugate's team at Phillips Laboratory, using the 1.5-m telescope at Starfire Optical Range. A copper vapor laser was employed, operating at 5000 pulses per second with an average power of 75 W. Laser wavelengths were 0.5106 and 0.5782 μm and the beacon altitude was 10 km. The closed-loop control bandwidth of this "Generation 1" compensation system was 65 Hz. In these experiments, the Strehl ratios achieved with the laser beacon were compared with those using a natural guide star. The average Strehl ratios obtained were 0.127 using the laser beacon and 0.200 using the natural guide star. These compare with theoretical predictions of 0.122 and 0.33, respectively. In later experiments (in 1992) with the "Generation 2" system using increased laser power and improved adaptive optics, Strehl ratios of about 0.5 were obtained with a single laser beacon, producing star images with a full-width at half-maximum (FWHM) of 0.13 arc seconds.

Experiments using single and multiple Rayleigh beacons have also been conducted by Thermo Electron Corporation employing short-wavelength excimer lasers. To eliminate focal anisoplanatism, Sandler [1992a] has proposed using an array of 20 beacons, using the concept of periodic interferometric weighting, also known as the "D-method." This system has so far been tested only in the laboratory.

The existence of the U.S. Government-sponsored work on laser beacons described above was not revealed publicly until 1991. In the meantime, similar ideas had been published in the open literature, notably by Foy and Labeyrie [1985]. Their paper presented the concept of using the backscatter from pulsed laser beams, focused high in the atmosphere, to measure wavefront distortion in adaptive optics systems. Both Rayleigh scatter and resonant scatter from the sodium layer were considered.

Based on this suggestion, Thompson and Gardner [1987] conducted an experiment at the Mauna Kea Observatory, in which a laser beacon was produced in the mesospheric sodium layer by using a pulsed dye laser tuned to 589 nm. Following this achievement, Gardner, Thompson, and Welsh published several papers covering the theory and practice of laser beacons using both the Rayleigh and sodium resonance scattering modes. They investigated the characteristics of the sodium layer, analyzed methods of implementing adaptive optics using laser beacons for astronomical telescopes, and calculated the expected performance.

The experiments and analyses conducted since 1980 have established, beyond any doubt, the potential value of laser beacons for implementing adaptive optics in astronomy. Because of the cost and complexity of such systems, the initial research on laser beacon systems was conducted at government-funded laboratories.

Several research groups are now pursuing the development of laser beacons, including the

University of Chicago, the University of Illinois, Lawrence Livermore National Laboratory, and the Steward Observatory of the University of Arizona. The general objective is to develop reliable low-cost laser beacon systems for implementing adaptive optics at good observing sites. For astronomy, the major problem is to obtain sufficient sky coverage, because natural guide stars are still needed for absolute pointing, even when laser beacons are used to compensate the wavefront within the telescope aperture. The use of laser beacons has, in effect, enlarged the isoplanatic patch for adaptive optics from a few arc seconds to several arc minutes, but this restriction has not yet been entirely removed.

7.1.3 Outline of Approaches and Problems

The adaptive optics components and system configurations used with laser beacons are basically similar to those employed with natural reference sources; such systems are usually designed to operate in either mode. There are additional sources of error that must be taken into account with laser beacons, however, and these produce differences in the system design and optimization procedures.

When using laser beacons with adaptive optics, the following factors must be considered:

- The beacons are formed within Earth's atmosphere, where they are optically near to the telescope aperture. Consequently, they do not sample the full telescopic beam when it is focused on a distant object, nor any turbulence above the beacon. This condition is known as the cone effect, and the resulting error in wavefront measurement is called focal anisoplanatism.
- Even when using laser beacons, a fixed (natural) reference source is required to compensate image motion. The reason is that the position of a laser beacon projected from a ground-based telescope is randomly displaced by atmospheric turbulence, and cannot be pointed accurately at an object outside the atmosphere without the aid of a reference, also outside the atmosphere. If it is bright enough, the science object itself may be used, but this is rarely the case. For long exposures, a significant part of the total wavefront error is contained in overall tilt, which must be compensated with great accuracy. For tilt compensation, it is therefore necessary to use a fixed reference source within the same (tilt) isoplanatic patch as the science object.
- The lasers used for generating adaptive optics beacons have special requirements. The pulse repetition rate of the laser is the controlling factor in the design of the adaptive optics feedback loop. To provide continuous wavefront correction, the laser must have a pulse rate approximately 10 times higher than the compensation bandwidth, requiring a rate of several kilohertz. In addition, lasers generating sodium beacons must be accurately tuned to the sodium D_2 line (589.1 nm), with specific pulse shapes and spectral content.
- The use of laser beacons complicates the telescope optical system more than does conventional adaptive optics, because additional optical components are required for beam sharing with the laser. Unwanted scattered light, as well as light from the beacon itself, must be rejected by the science imaging path.

7.2 Physical Principles of Laser Scattering

7.2.1 Types of Scattering

Scattering at the atomic level occurs when a photon of the incident radiation field is annihilated while a photon of scattered radiation is created. If the scattered frequency is the same as the incident frequency, then the scattering is termed elastic. If the quantum state of the atom is changed in the process, then the radiation is changed in frequency and the scattering is said to be inelastic. Scattering from larger particles in the atmosphere is basically a reflection of the incident radiation, with no change in frequency.

Four different modes of scattering of radiation by gaseous media are recognized:

1. Rayleigh scattering, in which radiation is scattered from atoms or molecules, with no change in frequency;
2. Raman scattering, in which radiation is scattered by atoms or molecules, with a change in frequency;
3. Mie scattering, in which radiation is scattered from small particles or aerosols of size comparable to the wavelength of the incident radiation, with no change in frequency;
4. Resonance scattering, in which radiation matched in frequency to a specific atomic transition is scattered, with no change in frequency.

A related process is fluorescence, in which radiation matched to a specific atomic transition is absorbed and emitted at a lower frequency.

The processes of interest for adaptive optics beacons are molecular Rayleigh scattering within the atmosphere and resonance scattering in the mesospheric sodium layer at 90 km. The presence of dust, fog, haze, and clouds causes Mie scattering, which occurs mostly in the lower atmosphere

(below 35 km) and varies unpredictably. The Mie scattering cross-section can have a large value, producing backscatter that competes with the desired returns.

7.2.2 The Lidar Equation

The process of using a light beam to probe a medium using the backscattered energy as a function of range is known as lidar (light detection and ranging). Lasers have been used for remote sensing of the atmosphere since the early 1960s. The use of lasers to create beacons for adaptive optics is directly related to the technology of remote sensing.

The lidar equation defines the energy detected at the receiver because of the scattering process as a light beam propagates through a medium. The basic form of the equation is:

Number of photons detected = number of transmitted photons

x probability that a photon is scattered

x probability that a scattered photon is collected

x probability that a collected photon is detected

+ background photons

With certain simplifying assumptions, described below, the lidar equation may be expressed symbolically in the form

$$N(z) = \left(\frac{E\lambda}{hc}\right)(\sigma_B n(z)\Delta z)\left(\frac{A_R}{4\pi z^2}\right)(T_o T_A^2 \eta) + N_B \quad (7.1)$$

where

$N(z)$ = expected number of photons detected in range interval Az
E = laser pulse energy, J
λ = optical wavelength, m
h = Planck's constant, 6.626 x 10^{-34} Js
c = velocity of light, 3 x 10^8 m s^{-1}
σ_B = effective backscatter cross-section, m^2
$n(z)$ = number density of scatterers at range z, m^{-3}
Δz = receiver range gate length, m
A_R = area of receiving aperture, m^2
z = range at center of range gate, m
T_o = transmission of optical components in transmit and receive paths
T_A = one-way transmission of atmosphere between telescope and beacon
η = quantum efficiency of photon detector at wavelength X
N_B = number of background and noise photoelectrons

The assumptions made in equation (7.1) are that the range gate Az is small compared with z, so that $n(z)$ and T_A may be assumed to be constant over the distance Az, and that the scattering process is linear so that σ_B is constant. The first assumption is justified in the case of laser beacons, which are generated by a focused beam, using a range gate in the detector optical system to select a small region around the focus. In the case of sodium beacons, there may be significant layering within the focal volume, which is accommodated by using an average value of $n(z)$. The resonance fluorescence process in sodium is easily saturated, so care must be taken to use an appropriate value of σ_B. Note that all factors relating to optical efficiency have been grouped into a single multiplication term.

The two factors «(z) and σ_B in equation (7.1) are specific to the type of scattering employed and are discussed in the following sections.

7.2.3 Rayleigh Scattering

The efficiency of a scattering process is characterized by its scattering cross-section. Rayleigh scattering is not isotropic and the cross-section depends on the polarization of the incident radiation, as shown in figure 7.5. In the forward and backward directions (scattering angles 0 and *n),* the parallel and perpendicular components add, while in the plane perpendicular to the direction of propagation (scattering angle 7r/2), only the perpendicular component is scattered. This is the reason that sunlight scattered from the sky at right angles to the Sun is polarized.

When using laser beacons, the backscattering cross-section is of primary interest. The backscattering cross-section for Rayleigh scatter is given by

$$\sigma_B^R = \frac{\pi^2(n^2-1)^2}{N^2\lambda^4} \quad (7.2)$$

where

n = refractive index of the medium
N = atom density of the medium
X = wavelength

For the mixture of gases that characterize the atmosphere below 100 km, Measures [1984, p. 42] gives the Rayleigh backscattering cross-section as

$$\sigma_B^R(\lambda) = 5.45\left[\frac{550}{\lambda\,(\text{nm})}\right]^4 \times 10^{-32}\ \text{m}^2\ \text{sr}^{-1} \quad (7.3)$$

The λ^{-4} dependence indicates that Rayleigh scattering is much more effective at shorter wavelengths, which is the reason that the sky is blue.

The product of the Rayleigh backscatter cross-section and atmospheric density is given by Gardner et al. [1986] as

$$\sigma_B^R n_R(z) = 3.6 \times 10^{-31}\,\frac{P(z)}{T(z)}\,\lambda_L^{-4.0117} \quad (7.4)$$

where

$P(z)$ = atmospheric pressure at range z, millibars
$T(z)$ = atmospheric temperature at range z, K

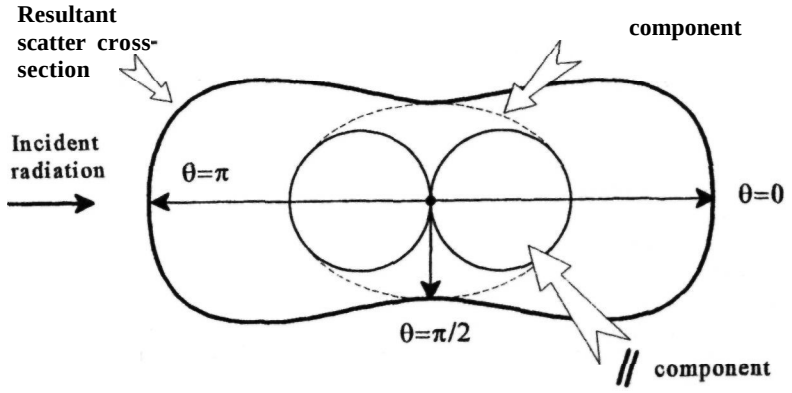


Figure 7.5 Rayleigh scattering cross-section.

The brightness of a Rayleigh beacon, for a given pulse energy, is proportional to the air density in the illuminated volume, which decreases exponentially with altitude. For this reason, Rayleigh beacons are limited to moderate altitudes in the atmosphere, generally in the range of 8-12 km.

The geometry of a Rayleigh beacon is shown in figure 7.6. The angular size of a beacon formed at the focus of a converging beam is determined by the depth of the beam path covered by the range gate in the receiver. To minimize the energy requirements, it is necessary to maximize the probability that incident photons will be scattered, which means that the greatest possible scattering depth should be used. However, the size of the beacon should not exceed the diffraction limit of the wavefront sensor subapertures, otherwise there is a penalty in the signal-to-noise ratio of the wavefront sensor.

From the geometry of figure 7.6, it can be shown that the maximum depth is

$$\Delta z = \frac{2\Delta\alpha\, z^2\, D_P}{\left[D_P^2 - (z\,\Delta\alpha)^2\right]} \tag{7.5}$$

where $\Delta\alpha$ is the angular size of the beacon and D_P is the diameter of the laser projection aperture, which is not necessarily the same as the telescope aperture.

The optimum angular size of the beacon is determined by the subaperture size of the wavefront sensor, or by the value of r_0, whichever is smaller. This condition is expressed as

$$\begin{aligned} \Delta\alpha &= 2.44\frac{\lambda}{d} \qquad \text{when } d \le r_0 \\ &= 2.44\frac{\lambda}{r_0} \qquad \text{when } d > r_0 \end{aligned} \tag{7.6}$$

When $\Delta\alpha$ is limited by turbulence, the maximum value of Az may be approximated as

$$\Delta z \approx \frac{4.88\lambda z^2}{D_P\, r_0} \tag{7.7}$$

7.2.4 Sodium Resonance Fluorescence

The existence of a layer in the mesosphere that contains the alkali metals sodium, lithium, potassium, and calcium has been known since the 1920s. The layer is located at altitudes between 80 and 110 km, with a peak near the mesopause at around 92 km. This region of Earth's atmosphere is generally inaccessible to direct measurements, being too high for balloons and too low for spacecraft. The structure of the layer has been revealed by observations made with rocket-borne instrumentation and also with laser remote sensing from the ground. The first laser measurements of sodium were made by

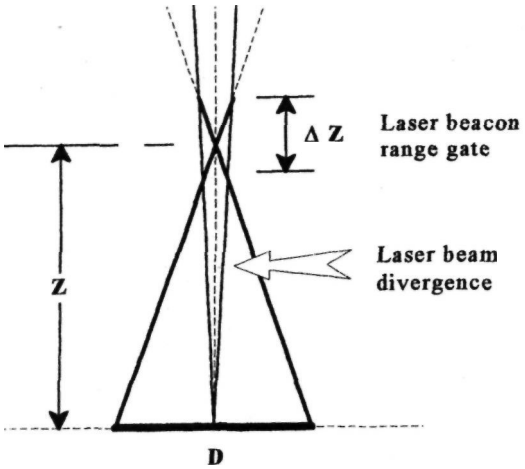


Figure 7.6 Geometry of Rayleigh beacon.

Bowman et al. [1969] using a laser tuned to 0.589 /um. The metallic sodium is believed to be of meteoric origin. The peak density varies between 10^3 and 10^4 cm^{-3}. Because of the relatively high density and large backscatter cross-section of sodium atoms, laser probing has proved to be a very effective method of measurement. These characteristics, together with the high altitude of the layer, also make sodium useful for generating laser beacons for adaptive optics.

The scattering profile measured by a laser tuned to the sodium line is modeled as a function of height in figure 7.7. The return for altitudes up to 75 km is dominated by Rayleigh scattering, which is proportional to atmospheric density. Below about 35 km, there is also Mie scattering from dust and aerosols, which enhances the return and causes local irregularities in the profile. Because of the exponential decrease in air density, it becomes increasingly difficult to make Rayleigh measurements at higher altitudes. The strong peak at around 92 km caused by resonance fluorescence in the sodium layer, indicates that this is a much more efficient process than Rayleigh scattering.

Lidar measurements of the seasonal variations in the sodium column abundance and centroid height have been made over a period of several years by Gardner et al. [1986]. The column abundance, measured at mid-latitudes in the Northern Hemisphere, shows a winter maximum of about 10^{14} m^{-2} and a summer minimum of about 3 x 10^{13} m^{-2}. The centroid height varies between 90 and 94 km, with a mean of 92 km; there is no consistent annual variation. The sodium density profile has been measured by Beatty et al. [1988] and has the distribution depicted in figure 7.8. The average thickness (FWHM) of the layer is about 10 km. Properties of atomic sodium and the atmospheric sodium layer are summarized in table 7.1.

Although other metals present in the mesosphere show resonant backscatter at visible and near-ultraviolet (near-UV) wavelengths, their densities and/or backscatter cross-sections are much lower than those of sodium, so they are of less interest for generating laser beacons.

Resonance scattering differs from Rayleigh scattering in two ways: (1) the incident laser radiation is tuned in frequency to match the absorption spectrum of the atomic transitions, resulting in a greatly enhanced scattering cross-section (higher efficiency); and (2) saturation can occur when the energy density within the sodium layer becomes large enough to change the population densities of the atomic states. The transitions that are induced in sodium atoms by incident radiation are shown in figure 7.9, using a

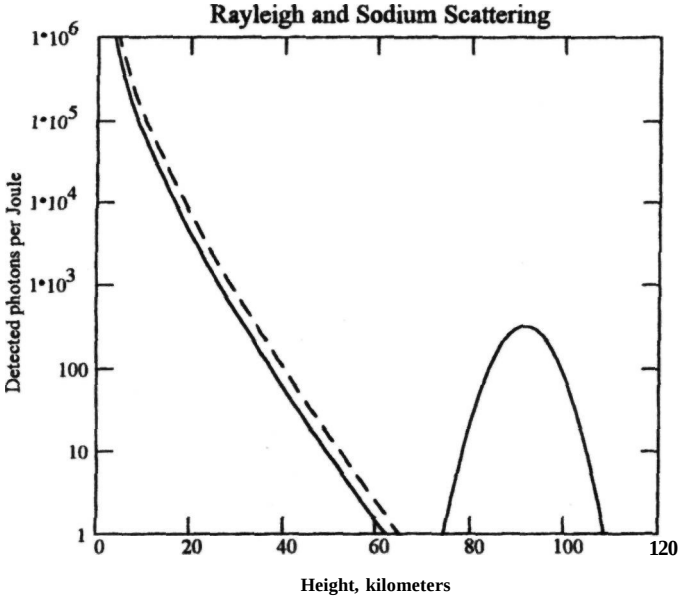


Figure 7.7 Rayleigh and sodium scattering model. The chart shows the expected signal from a laser pulse of length 1 km (3.3 μs), as a function of the mean height selected by the range gate. At a wavelength of 589 nm, returns are received from both Rayleigh scattering and from sodium resonance fluorescence, and the sodium layer produces a sharp peak at about 90 km. At other wavelengths, only the return from Rayleigh scattering is received.

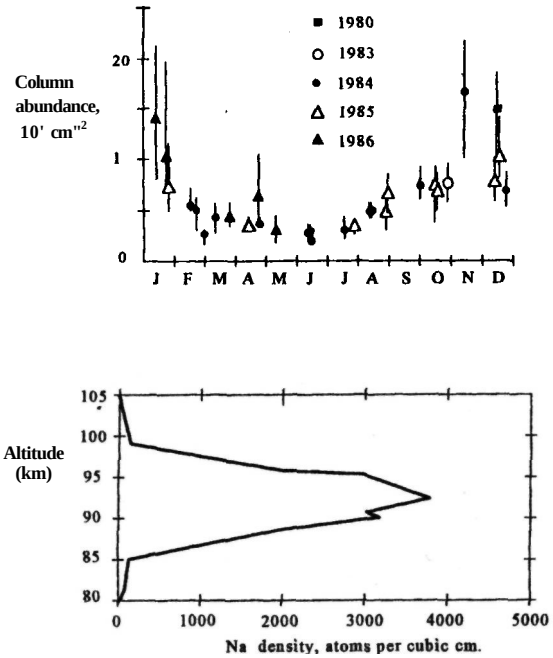


Figure 7.8 Characteristics of the mesospheric sodium layer [Gardner et at. 1986, Beatty et al. 1988].

simplified two-state model. The real sodium atom is far more complex and will be described later in connection with the design of sodium beacon lasers. However, its non-linear behavior with regard to absorption may be explained using the simplified model.

The metal atoms are normally in their lower, unexcited state at a low temperature, around 200 K. When a pulse of laser energy passes through the medium, some atoms absorb a photon and are excited to the upper state. Atoms in the upper state may de-excite back to the lower state by either spontaneous decay or stimulated emission. With spontaneous decay, a photon of the same energy (wavelength) is emitted incoherently and isotropically. The component of this radiation in the direction of the laser source is the desired resonant backscatter. The natural lifetime of excited sodium atoms is 16 ns. At low levels of incident radiation, the mean time for an atom in the excited state to absorb a photon is much longer than 16 ns, so most of the population remains in the lower state. In this regime, the backscatter is proportional to the laser energy.

Higher levels of radiation create a larger population in the excited state, allowing stimulated emission to become a competing process. In this process, atoms in the excited state absorb photons, which are re-emitted coherently; that is, with the same frequency, phase, and direction as the original photon. Stimulated emission does not produce backscattered photons. As the incident energy increases, a greater proportion of excited atoms absorb a photon and produce stimulated emission before they decay naturally. The net result is that as the population in the excited state increases, the proportion of atoms available to produce backscatter decreases; in other words, the medium saturates.

The real sodium atom has two hyperfine levels in the ground state and four hyperfine levels in the upper state, as depicted in figure 7.10, splitting the D_2 line into six components. The only significant energy difference exists between the two levels of the ground state. As a consequence, the six lines appear in two groups, producing a double peak, with the main components separated by 1.772 GHz. The spectral lines are broadened by two mechanisms: (1) the lifetime of the atoms in the excited state, whether because of natural decay (homogeneous broadening) or collisions; and (2) Doppler shifts due to a velocity component along the line of sight (inhomogeneous broadening).

The finite lifetime of the excited atoms produces homogeneous line broadening with the characteristic Lorentzian profile as a function of angular frequency ω, given by

$$F(\omega) = \frac{1}{\pi} \frac{(1/\tau_n)}{(\omega_0 - \omega)^2 + (1/\tau_n)^2} \qquad (7.8)$$

Table 7.1 Properties of the Atmospheric Sodium Layer

Parameter	Symbol	Value
Atomic sodium		
D_2 line wavelength	λ_{Na}	589 nm
Natural lifetime	τ_n	16.2 ns
Degeneracy correction factor	$g_u/(g_l + g_u)$	0.67
Atmospheric sodium layer		
Mean altitude	Z_{Na}	92 km
Depth (FWHM)	AZ	10 km
Column density	N	5×10^{13} m^{-2}
Emission cross-section	σ	4×10^{-16} m^2

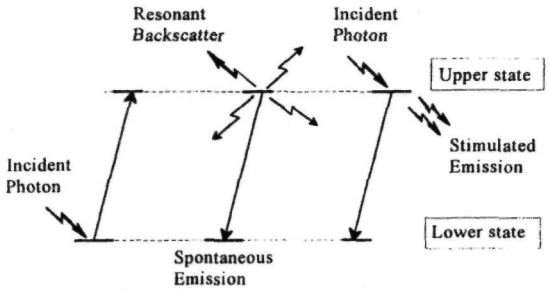


Figure 7.9 Absorption and emission in sodium (simplified two-level model). Atoms in the lower state are raised to a higher state when a photon is absorbed. Spontaneous decay produces resonant backscatter, mostly in the desired direction. If an excited atom absorbs an incident photon, then stimulated emission is produced in the forward direction, which is lost to the receiver.

where ω_0 is the atomic resonance angular frequency and τ_n is the natural lifetime of the atom in the excited state. The half-width at half-maximum (HWHM) of this profile is

$$HWHM = \Delta\omega_n = \frac{1}{\tau_n} \tag{7.9}$$

The natural line width is given by

$$\Delta\lambda_n = \frac{\lambda_{Na}^2}{c}\frac{1}{2\pi\tau_n} \tag{7.10}$$

For the sodium D_2 line, $\Delta\lambda_n$ is 1.15×10^{-14} m, equivalent to a temporal bandwidth of 10 MHz. The natural lifetime of excited sodium atoms is much shorter than the mean time between collisions in the sodium layer (about 10 μs), so that the line broadening due to collision-induced transitions is negligible.

The spectral line broadening due to Doppler shifts depends on the velocity of the atoms, which is a function of the temperature of the medium. The observed frequency due to a velocity component v along the line of sight is

$$\omega = \omega_0\left(1 + \frac{v}{c}\right) \tag{7.11}$$

where ω_0 is the frequency when at rest. Measures [1984, p. 98] shows that for thermal motion with Maxwellian distribution, the probability that a photon will be observed with an angular frequency in the interval $(\omega, \omega + d\omega)$ is given by the Gaussian distribution

$$G(\omega) = \frac{1}{\beta\sqrt{\pi}} e^{-(\omega-\omega_0)^2/\beta^2} \tag{7.12}$$

where

$$\beta = \left[\frac{2kT\omega_0^2}{mc^2}\right]^{1/2} \tag{7.13}$$

k = Boltzmann's constant, J/K^{-1}
T = temperature, K
ω_0 = rest frequency of atom, rad s^{-1}
m = mass of electron
c = velocity of light, ms~'

The Doppler-broadened spectral line has a HWHM given by

$$\Delta\omega_D = \beta(\ln 2)^{1/2} = 0.833\beta \tag{7.14}$$

For a sodium atom at 200 K, the Doppler-broadened spectral linewidth is about 1 x 10 m, about 100 times greater than the homogeneous broadening due to natural decay time. Doppler broadening is therefore the dominant factor in determining the linewidth.

The structure of the Doppler-broadened sodium D_2 absorption line is shown in figure 7.11 for temperatures of 150, 200, and 250 K. The areas of the two components of the split line have a 3:5 ratio, due to the fact that the $F' = 1$ and $F' = 2$ ground states have, respectively, three and five momentum levels, as shown in figure 7.10. Atoms spontaneously decaying from the excited states therefore have 3/8 and 5/8 probabilities of decaying to the $F' = 1$ and $F' = 2$ levels. The Doppler broadening of the lines increases with temperature, which can vary from 130–150K in the summer to 210–230 K in the winter [Gardner et al. 1986]. This sensitivity to temperature is, in fact, the method used to determine the temperature of the sodium layer, by comparing the absorption cross-sections at two frequencies corresponding to the peak and valley of the response.

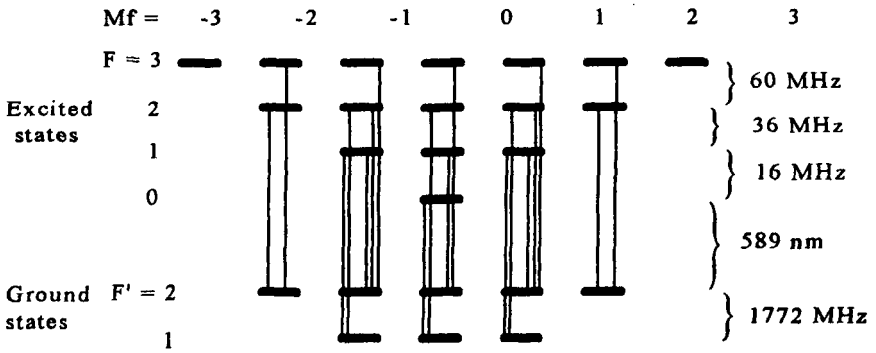
7.2.5 Saturation Effects in the Sodium Layer

Nonlinear effects in the Doppler-broadened absorption line of sodium have been analyzed by Welsh and Gardner [1989b]. The analysis is greatly simplified if the laser pulse can be modeled as separable in time, space, and frequency. This is generally the case for most lasers in which the bandwidth is greater than the reciprocal of the pulse width. However, this simplification may not apply to narrow-band lasers that have macropulse outputs consisting of trains of very short pulses. The special considerations for such devices are treated in section 7.2.6.

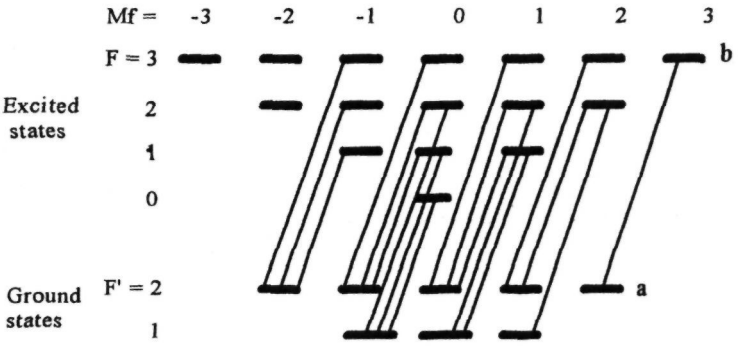
Assuming separability, the total photon flux density produced by the laser pulse may be expressed as a function of the temporal pulse shape, beam cross-section, and frequency profile:

$$P(t, \rho, \omega) = N_L(t)I(\rho)g(\omega) \tag{7.15}$$

where



A. Linear Polarization



B. Circular Polarization

Figure 7.10 Energy levels of the D_2 line of sodium showing the optical transitions induced by linearly and circularly polarized light. The frequencies between adjacent hyperfine sublevels are shown at the right.

$N_L(t)$ = total number of incident photons per unit time, s

$I(p)$ = distribution of photon flux per unit area, m^{-2}

$g(\omega)$ = distribution of photon flux per unit angular frequency, rad~' s

An effective absorption cross-section $\sigma_{eff}(\omega_d)$, which depends on the Doppler-broadened spectrum of sodium atoms, weighted by the spectrum of the laser pulse may then be defined as:

$$\sigma_{eff}(\omega_d) = \int_{-\infty}^{\infty} \sigma(\omega - \omega_d) g(\omega)\, d\omega \qquad (7.16)$$

where

ω_d = Doppler angular frequency shift, rad s^{-1}

$\sigma(\omega - \omega_d)$ = Doppler broadened absorption cross-section of Na, m^2 (shown in figure 7.11)

The changes in the population of upper-state atoms caused by the laser pulse may then be expressed as:

Gain from absorption =
$$[n(\omega_d) - n_u(t, \rho, \omega)]\sigma_{eff}(\omega_d) I(\rho) N_L(t) \qquad (7.17)$$

Loss from spontaneous emission =
$$\frac{n_u(t, \rho, \omega)}{\tau_n} \qquad (7.18)$$

Loss from stimulated emission =
$$n_u(t, \rho, \omega)\alpha\, \sigma_{eff}(\omega_d) I(\rho) N_L(t) \qquad (7.19)$$

where

$n(\omega_d)$ = total density of Na atoms per unit Doppler angular frequency shift, m^{-3} rad^{-1} s

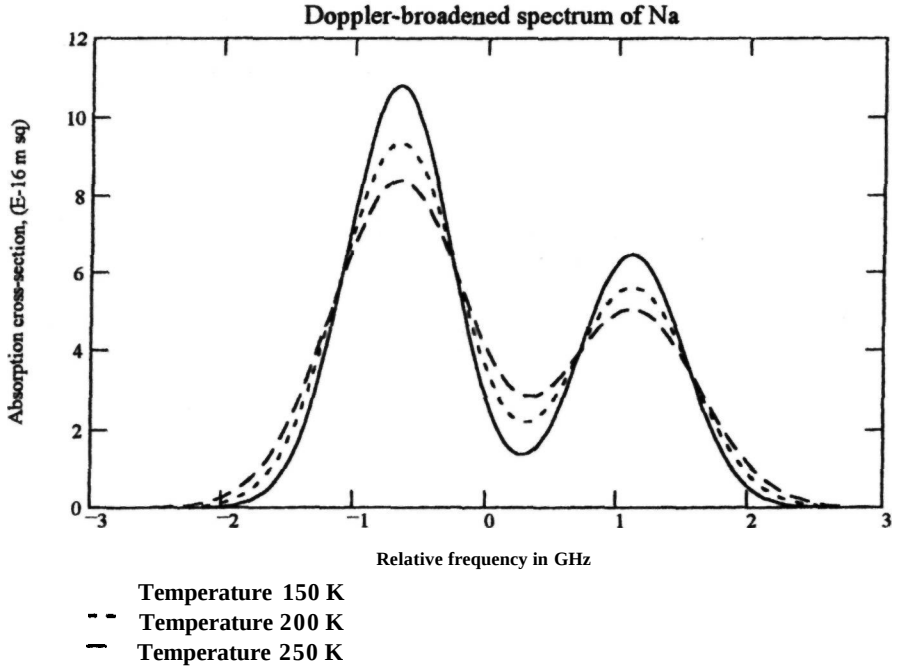


Figure 7.11 Absorption cross-section of sodium D_2 line for temperatures of 150, 200, and 250 K. The center wavelength is 589.158 nm and the separation between the peaks is 1772 MHz.

$n_u(t, p, \omega)$ = density of upper-state atoms per unit Doppler angular frequency shift, m rad^{-1} s

α = ratio of the degeneracies for the lower and upper states

The backscattered energy is proportional to the number of lower-state atoms that absorb photons. Equation (7.17) shows that as the population of the upper state increases, the density of atoms in the lower state is depleted, reducing the backscattering.

To optimize the efficiency of the scattering process, it is necessary to control three critical factors: the laser energy, the laser linewidth, and the pulse length. Referring to the lidar equation (7.1), in order to minimize the laser energy required to achieve a given photon count, the backscatter cross-section σ_B must be maximized. For a given laser energy, the parameters to be optimized are the laser spectral linewidth and the pulse length.

The total scattering cross-section for the case of negligible saturation is

$$\sigma_T = \int_{-\infty}^{\infty} \frac{n(\omega_d)}{n_0} \sigma_{eff}(\omega_d)\, d\omega_d \qquad (7.20)$$

where

n_0 = is the total integrated density of Na atoms, m^{-3}, and $\sigma_{eff}(\omega_d)$, the effective scattering cross-section, is defined in equation (7.16).

Saturation effects can be quantized by the saturation time τ_S, defined by Megie et al. [1977], which is a measure of the characteristic time of stimulated emission. The level of saturation within the sodium layer is indicated by the ratio τ_S/τ_n. When the saturation time is larger than the natural decay time τ_n, the rate of stimulated emission is small compared with the rate of natural decay, so that saturation effects are negligible, as shown by the lower curve in figure 7.12. When the saturation time is much shorter than the natural decay time, the rate of stimulated emission overwhelms the natural decay rate, resulting in significant saturation effects, as shown by the upper curve in figure 7.12.

The saturation time is defined as

$$\tau_S = \frac{2\sqrt{12}}{\sigma_T}\frac{\Delta t}{W} = \frac{2\sqrt{12}}{\sigma_T}\frac{\Delta t \pi a^2 hc}{E\lambda T_A} \qquad (7.21)$$

where

Δt = root-mean-square duration of laser pulse, s

σ_T = total effective absorption cross-section of atoms, m^2

W = average photon flux density from laser, photons rcT^2 s

a = root-mean-square radius of laser beam, m

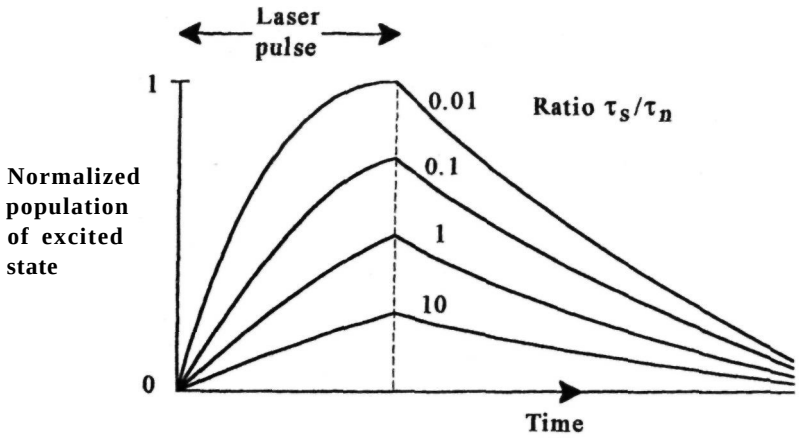


Figure 7.12 Saturation of sodium layer.

Using the saturation time defined above, Welsh and Gardner [1989b] derived expressions for the reduction in backscattered energy due to saturation for three different combinations of laser beam cross-section and pulse shape:

1. Uniform beam cross-section and rectangular temporal pulse shape;
2. Gaussian beam cross-section and rectangular pulse shape;
3. Gaussian beam cross-section and exponential pulse shape.

These results can be used to define a new backscattering cross-section, σ_S, which includes the effects of saturation and which can be used in the lidar equation.

For the usual case of a Gaussian beam and rectangular laser pulse, the backscattering cross-section is

$$\sigma_S = \int_{-\infty}^{\infty} \frac{\tau_S \sigma_T n(\omega_d)}{2\tau_n n_0} \ln\left[\frac{2\tau_n \sigma_{eff}(\omega_d)}{\tau_S \sigma_T} + 1\right] d\omega_d \quad (7.22)$$

The design process can be simplified by choosing a value for the ratio τ_S/τ_n. Equation (7.21) shows that for a given pulse energy, the value of τ_S can be increased (that is, saturation can be reduced) by increasing the pulse duration. Welsh and Gardner found that the laser pulse energy required to achieve a given number of photon counts at the wavefront sensor decreases monotonically as the value of τ_S/τ_n is increased from 1 to 10. However, long pulses are difficult to generate with most lasers, and a good compromise appears to be a τ_S/τ_n ratio of 2 corresponding to a pulse length of 32 ns. The optimum laser linewidth is somewhat less than the Doppler-broadened sodium spectral line, with a value of 600 MHz FWHM being near optimum. Specific values will depend on the parameters of the adaptive optics system.

7.2.6 Sodium Beacons Using Short Pulses

As mentioned above, special considerations apply to the generation of sodium beacons with narrow-band, short-pulse lasers. To obtain the highest scattering cross-section, the maximum number of sodium atoms must be excited by each laser pulse. To achieve this goal, the laser energy should be effectively distributed over the Doppler-broadened sodium absorption line, which has an FWHM bandwidth of about 2GHz.

With many types of lasers, the spectral bandwidth is determined independently of the pulse length. For a single rectangular pulse, the frequency spectrum has the form

$$G(f) = A\frac{\sin(\pi t_p f)}{\pi t_p f} \quad (7.23)$$

where

A = pulse area
t_p = pulse length
f = temporal frequency

The first null of this function occurs at a frequency of $f_0 = 1/t_0$. For lasers with pulse lengths exceeding about 10 ns, the corresponding pulse bandwidth of up to 100 MHz is much smaller than the spectral linewidth, so the pulse width has little effect on the spectral distribution of the laser output.

Certain types of laser, such as free-electron lasers (FELs) and mode-locking lasers, produce their output in the form of macropulses, or trains of very short micropulses. Mode locking is a process in which multiple modes, existing in the laser cavity and normally having random phases, are forced to have the same phase. The output then consists of a train of narrow micropulses, less than 1 ns in duration. The mode frequency spacing produced by a cavity of length l is $c/2l$, where c is the velocity of light. If there are N modes, each producing intensity

/, then the output for incoherent addition will be NI. When the modes are locked, the coherent addition produces a peak intensity of N^2I; that is, N times higher than for uncorrelated modes.

The micropulse duration in this case is $t_0 = 2l/Nc$ and the micropulse spacing is $T = 2l/c$. The overall spectral width is $1/t_0 = N/2l$, which is the same as that of the conventional system with random modes. The difference is that instead of a long pulse of constant intensity, the energy is now concentrated in a train of micropulses that have $\sqrt{N}$ of the duration and A' times the intensity. The higher peak intensity increases the efficiency of nonlinear operations, such as mixing.

Because the frequency spectrum of the laser pulse is not continuous, but is concentrated in N narrow lines, only a fraction of the pulse energy is available to excite the sodium atoms, whose absorption spectrum is distributed continuously over the Doppler-broadened range. To increase the absorption efficiency, Humphreys et al. [1992] have suggested phase modulating the positions of the micropulses to produce a more continuous distribution.

7.3 Wavefront Measurement with Atmospheric Beacons

7.3.1 Beacon Geometry

The process of measuring a turbulence-degraded wavefront by means of laser beacons generated within Earth's atmosphere will now be considered. Because beacons are relatively close to the telescope, the light rays used to measure the turbulence travel along different paths from those traveled by the object being observed, producing a measurement error called focal anisoplanatism. Several additional errors are also encountered when using laser beacons, including the overall tilt (image motion) caused by the use of a separate guide star, as well as the wavefront measurement errors caused by the limited brightness and finite size of the laser beacons themselves. These errors are discussed in detail in section 7.4. They must be combined with the basic errors common to all adaptive optics systems in order to optimize the overall design and performance of compensated telescopes using laser beacons.

The basic geometry of a laser beacon is depicted in figure 7.13. Light rays from the distant object are normal to the telescope aperture D. The beacon B is at height H and is assumed to be on the optical axis in the same direction as the distant object. The angle between the beacon and the object, as seen from a point on the telescope aperture at radius p, is $\theta = \tan^{-1}(\rho/H)$. At the center of the aperture, the angle is zero, and it reaches a maximum at the outer edge of the aperture, where $\rho = D/2$.

There are two significant differences between the focal anisoplanatism produced by this geometry and the angular anisoplanatism between two distant objects outside the atmosphere: (1) With a beacon, the offset angle θ is proportional to the aperture radius, whereas with two distant objects it is constant over the telescope aperture; and (2) turbulence above the beacon is not measured. The error due to focal anisoplanatism can be assessed in the following way. From section 3.7, it is known that the mean-square wavefront error due to anisoplanatism is proportional to the five-thirds power of the angular separation, multiplied by the five-thirds moment of the turbulence distribution:

$$\sigma_\theta^2 = \text{const} \times \theta^{5/3} \times k^2 \int_0^H C_N^2(h)\, h^{5/3}\, dh$$

Substituting for θ and using $\bar{\mu}_{5/3}(H)$ for the five-thirds turbulence moment up to the beacon height H, the following relation is obtained

$$\sigma_{FA}^2 = \text{const} \times \left(\frac{D}{2H}\right)^{5/3} k^2 \bar{\mu}_{5/3}(H)$$

This expression shows the basic dependencies of focal anisoplanatism. The full expression, derived later, contains additional terms accounting for turbulence above the beacon and removal of overall piston and tilt components. The mean-square wavefront error due to focal anisoplanatism depends on the ratio D/H, whereas with angular anisoplanatism it depends on the offset angle θ. Clearly, focal anisoplanatism penalizes large telescopes more than small ones, but the error for any size of telescope is minimized by placing the beacon at the maximum possible height.

7.3.2 Average Phase and Tilt Removal

In evaluating the wavefront errors produced by focal anisoplanatism with laser beacons, it is necessary to remove the average phase (piston) and the overall tilt components. The average value of the optical phase has no significance for incoherent imaging systems

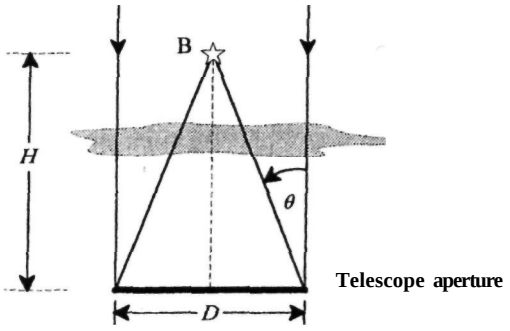


Figure 7.13 Focal anisoplanatism.

using a single aperture, and may therefore be ignored. The reason for excluding overall tilt is that a laser beacon projected from the ground cannot function as a reference source for measurement of overall tilt or image position, because its position is randomly displaced by atmospheric turbulence. This displacement cannot be sensed by measuring the position of a beacon projected through the same aperture. The principle of optical reciprocity states that light will trace the same path in either direction in a static medium. Consequently, if a laser beam is projected on-axis through a telescope aperture, the backscattered light will always appear to be on-axis when viewed through the same aperture, irrespective of any wavefront distortion, assuming that the wavefront does not change within the propagation time, which is generally true for atmospheric beacons. The result is that no overall tilt error would be recorded by a sensor using the same aperture. If a separate aperture were employed to project the beacon, then, again, its random motion due to turbulence would add to that of the sensing telescope, preventing any meaningful measurement of the position of the science object.

In order to project a beam from the ground in the same direction as an exo-atmospheric object, it is necessary to have a reference source in the same frame of reference as the object; that is, beyond the atmosphere. A laser beacon can only serve as a reference source for wavefront aberrations ocurring within the telescope aperture itself, starting with the defocus error. Such higher order wavefront errors are all measured with reference to the mean tilt across the aperture and are therefore independent of the absolute position of the reference source. This subject is of great importance in adaptive optics and is discussed further in section 7.4.

An additional complication arises when multiple beacons are employed. In this case, unmeasured tilt differences between the beacons are reconstructed into higher order wavefront errors across the aperture. These errors are described in section 7.3.5 and must be taken into account when determining the performance of multiple-beacon systems.

7.3.3 Computation of Error Due to Focal Anisoplanatism

Exact calculation of the focal anisoplanatism error produced by laser beacons has proved to be a difficult task. The classic analytical approach using wavefront structure functions that depend only on differences in position within the telescope aperture becomes unwieldy because the beacon beam converges to (nearly) a point, while the telescope beam is collimated. One of the first investigations was that of Belsher and Fried [1984], in which it was shown that the mean-square wavefront error due to focal anisoplanatism, in radians squared of phase, can be expressed as

$$\sigma_{\mathrm{FA}}^2 = \left(\frac{D}{d_0}\right)^{5/3} \tag{7.24}$$

where D is the telescope aperture and the value of d_0 depends on the operating wavelength, zenith angle, beacon altitude, and vertical distribution of turbulence. The quantity rf_0 may be considered as the diameter of the aperture over which the wavefront error due to focal anisoplanatism is 1 rad^2. The expression for d_0, derived by Belsher and Fried, scales as $\lambda^{6/5}\ \cos^{3/5}(\zeta)$, where ζ is the zenith angle.

Computed values of d_0 as a function of beacon altitude for three different atmospheric turbulence profiles at a wavelength of $0.5\,\mu\mathrm{m}$ are given in a reissue of the original paper [Fried and Belsher 1994]. For each turbulence profile, the value of d_0 increases almost linearly with beacon height. The slope of the curves appears to be very sensitive to high-altitude turbulence; that is, it depends on higher order turbulence moments. Wavefront errors due to focal anisoplanatism have also been estimated by Gardner et al. [1990] and Welsh and Gardner [1991] for specific adaptive optics system configurations, using results based on conventional anisoplanatism.

When designing adaptive optics systems, it is more useful to employ an analytic approach, in which the results are expressed directly in terms of the aperture diameter, turbulence moments, beacon height, zenith angle, and operating wavelength. With such a tool, the influence of each parameter on the system performance is directly apparent and optimization can be carried out more readily.

A powerful method of analyzing the optical effects of turbulence using transverse filtering has been described by Sasiela [1988]. An outline of this method has been given in chapter 3. It is particularly useful in cases involving two beams propagating in different directions, which is the case with anisoplanatic errors. In this approach, wavefront phase errors are expressed in the form

$$\sigma^2 = 0.207k^2 \int_0^\infty dz\, C_{\mathrm{N}}^2(z) \int d\kappa f(\kappa)\, g(\gamma, \kappa, z) \tag{7.25}$$

where

- $k = 2\pi / \lambda$ is the temporal wave number of the propagating wave
- $C_{\mathrm{N}}^2(z)$ = turbulence strength as a function of range, z
- K = is the spatial wave number transverse to the z direction,
- $f(\kappa)$ - two-dimensional turbulence spectrum, which is normally assumed to be the Kolmogorov spectrum $f(\kappa) = \kappa^{-8/3}$

The essence of this approach is the use of filter functions in the spatial domain that depend on the specific geometry of the beam paths. The filter function is generalized in equation (7.25) as $g(y, \kappa, z)$, where y is a propagation parameter and κ is the transverse

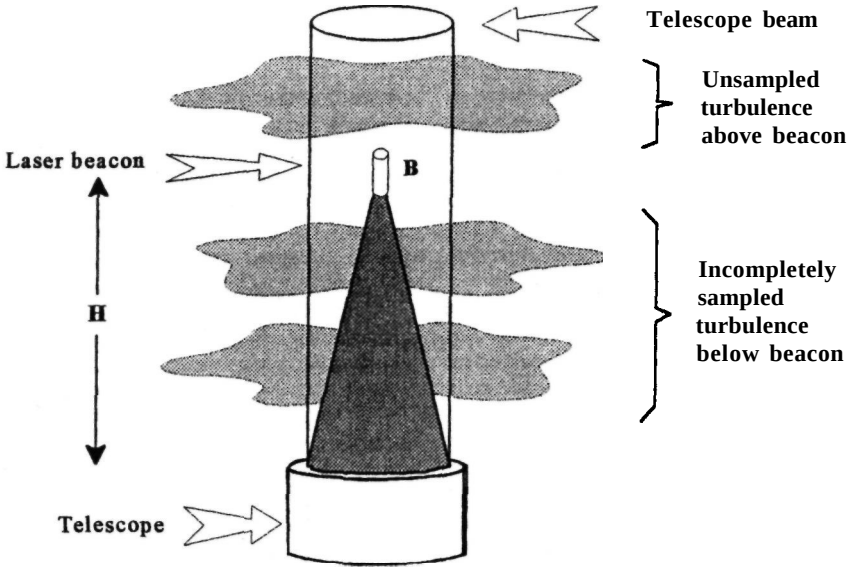


Figure 7.14 Laser beacon measurement errors.

wave number. Three integrations must be performed: the z integration over the axial range, and a double integral in κ over the transverse spatial variables.

The beam paths associated with a laser beacon within the atmosphere are depicted in figure 7.14. The path traversed by the light from the science object to the telescope is only partially sampled by the beacon. Turbulence above the beacon is totally unmeasured. Below the beacon, the rays reaching any point on the aperture, from the object and the beacon, pass through the turbulence at different angles. The measurement error of a single beacon therefore consists of two components: the unmeasured wavefront error above the beacon and the anisoplanatic wavefront error below the beacon. Because the laser beacon is not used to measure overall piston or tilt, these components (the first three terms in the Zernike expansion) are subtracted from the calculated error.

The specific wavefront errors involved are defined as follows:

σ_u^2 = Total unmeasured wavefront error variance above the beacon,
σ_{up} = Piston component of wavefront error above the beacon,
σ_{ut}^2 = Overall tilt component of wavefront error above the beacon,
CTL = Total anisoplanatic wavefront error variance below the beacon,
σ_{Lp} = Piston component of anisoplanatic error below the beacon,
CTLI = Overall tilt component of anisoplanatic error below the beacon.

The net mean-square wavefront measurement error using a single beacon is then

$$\sigma_B^2 = \sigma_u^2 - \sigma_{up}^2 - \sigma_{ut}^2 + \sigma_L^2 - \sigma_{Lp}^2 - \sigma_{Lt}^2 \qquad (7.26)$$

The overall tilt error encountered when using laser beacons will be called *conic tilt error* to distinguish it from the overall tilt component of angular anisoplanatism. This error is significant when multiple beacons are used. The conic tilt error includes contributions from above and below the beacon and is defined as

$$\sigma_{ct}^2 = \sigma_{ut}^2 + \sigma_{Lt}^2 \qquad (7.26a)$$

Computation of the wavefront error above the beacon is straightforward. The piston- and tilt-removed wavefront error corresponding to the first three terms in equation (7.26) may be determined directly from the relation

$$\sigma_{uptr}^2 = 0.057\, k^2\, D^{5/3} \int_H^\infty dz\, C_N^2(z) \qquad (7.27)$$

The mean-square value of the anisoplanatic error below the beacon can be computed using equation (7.25). Sasiela [1994] has derived the filter functions required to determine the total error, as well as the unwanted piston and tilt components that must be subtracted. Assuming operation in the near field, the mean-square value of the anisoplanatic error below the beacon is given by

$$\sigma_L^2 = 1.303\, k^2 \int_0^L dz\, C_N^2(z) \int_0^\infty d\kappa\, \kappa^{-8/3} F(\kappa) \qquad (7.28)$$

where

$$F(\kappa) = \left\{1 - 2\frac{2J_1(aw)}{aw}\frac{2J_1(Dw)}{Dw}J_0\left(\frac{\kappa bz}{L}\right) + \left[2\frac{J_1(aw)}{aw}\right]^2\right\} \tag{7.29}$$

and where

a = beacon diameter
b = beacon offset from telescope axis
D = diameter of telescope aperture
L = beacon range
w = beam weighting factor, $\kappa(L - z)/2L$, m^{-1}

If the beacon is assumed to be a point source in the same direction as the science object, then $a = 0$ and $b = 0$, and the filter function simplifies to

$$F(\kappa) = 2\left\{1 - \frac{2J_1(Dw)}{Dw}\right\} \tag{7.30}$$

This function includes the uniform phase (piston) and tilt components.

The filter functions for extracting the overall piston and tilt components from the focal anisoplanatic error below the beacon are given by

$$F_P(\kappa) = \left\{\frac{2J_1(\kappa D/2)}{\kappa D/2} - \frac{2J_1(Dw)}{Dw}\right\}^2 \quad \text{(Piston)}$$

$$F_T(\kappa) = \left\{\frac{4J_2(\kappa D/2)}{\kappa D/2} - \frac{4J_2(Dw)}{Dw}\right\}^2 \quad \text{(Tilt)} \tag{7.31}$$

The corresponding wavefront errors may then be evaluated by substituting expressions (7.30) and (7.31) in equation (7.28). The integration is performed using Mellin transform techniques [Sasiela and Shelton 1993a,b]. The solution is expanded in a power series in the parameter z/L, allowing the answer to be expressed in terms of turbulence moments. Definitions of the full and partial turbulence moments were given in chapter 3 and are repeated here for convenience:

The full turbulence moment is

$$\mu_m = \int_0^\infty dz\, C_N^2(z) z^m$$

The partial turbulence moments at height H are

Upper moment

$$\mu_m^+(H) = \int_H^\infty dz\, C_N^2(z) z^m$$

Lower moment

$$\mu_m^-(H) = \int_0^H dz\, C_N^2(z) z^m$$

These moments are defined at the zenith; that is, at $\zeta = 0$. At other angles, they should be multiplied by $\sec^{m+1}(\zeta)$.

Sasiela [1994] finds that a good approximation to the focal anisoplanatism below the beacon for typical turbulence distributions is obtained using only the first terms of each series. On this basis, the error components are

Total

$$\sigma_L^2 \approx k^2 D^{5/3}\left[0.500\frac{\mu_{5/3}^-(H)}{H^{5/3}}\right] \tag{7.32}$$

Piston

$$\sigma_{Lp}^2 \approx k^2 D^{5/3}\left[0.0833\frac{\mu_2^-(H)}{H^2}\right] \tag{7.33}$$

Tilt (two-axis)

$$\sigma_{Lt}^2 \approx k^2 D^{5/3}\left[0.368\frac{\mu_2^-(H)}{H^2}\right] \tag{7.34}$$

The turbulence above the beacon [equation (7.27)] may be expressed in terms of the zero-order moment of the turbulence above the beacon:

$$\sigma_{uptr}^2 = \sigma_u^2 - \sigma_{up}^2 - \sigma_{ut}^2 = k^2 D^{5/3}[0.057\mu_0^+(H)]$$

Combining these equations, the final expression is obtained for the focal anisoplanatic error, which is the total wavefront error due to a beacon at height H, with piston and tilt removed:

$$\sigma_{FA}^2 = k^2 D^{5/3}\left[0.057\mu_0^+(H) + 0.500\frac{\mu_{5/3}^-(H)}{H^{5/3}} - 0.452\frac{\mu_2^-(H)}{H^2}\right] \tag{7.35}$$

This wavefront error is scaled by the factor $k^2 D^{5/3}$, independent of the beacon altitude and turbulence profile. It may therefore be expressed as a function of $(D/d_0)^{5/3}$, as mentioned previously [equation (7.24)]. The value of d_0 implied by equation (7.35) is then

$$d_0 = \left\{k^2\left[0.057\mu_0^+(H) + 0.500\frac{\mu_{5/3}^-(H)}{H^{5/3}} - 0.452\frac{\mu_2^-(H)}{H^2}\right]\right\}^{-3/5} \tag{7.36}$$

This equation gives results that agree with the numerically evaluated d_0 values of Belsher and Fried to within 1%, which would appear to confirm the validity of both methods. The advantage of the analytical method is that it enables d_0 to be determined directly from the turbulence moments defined in section 3.3.6. Values of the parameter d_0 for several turbulence models and observation wavelengths are given in table 7.2.

7.3.4 Evaluation of Beacon Sampling Errors

The wavefront error incurred by the use of a single laser beacon as the reference source, combining the effects of undersampling the turbulence both above

Table 7.2 Values of Focal Anisoplanatism Parameter d_0

Turbulence Model	Zenith Angle (degrees)	Beacon height (km)	d_0 @ 0.50 μm (m)	d_0 @ 0.70 μm (m)	d_0 @ 1.65 μm (m)	d_0 @ 2.20 μm (m)
Hufnagel	0	10	0.56	0.84	2.35	**3.32**
		92	2.00	3.00	**8.40**	**11.86**
	45	10	0.46	**0.68**	1.91	**2.70**
		92	1.63	2.44	6.82	9.64
Mauna Kea background	0	10	1.33	2.00	5.58	7.89
		92	**4.43**	6.63	18.55	26.20
	45	10	1.08	1.62	4.54	6.41
		92	3.60	5.39	15.07	21.28
Mauna Kea average	0	10	0.87	1.30	3.64	5.14
		92	**3.34**	5.00	**14.00**	19.78
	45	10	0.71	1.06	2.96	4.17
		92	2.72	4.07	11.38	**16.07**

and below the beacon, is conveniently described as beacon sampling error and sets certain inescapable limits to the design and performance of adaptive optics systems. The beacon sampling error depends on beacon altitude, telescope aperture, zenith angle, wavelength, and turbulence distribution, as shown in equation (7.35). To gain some perspective on the utility of laser beacons in astronomy, the effects of these parameters on the beacon sampling error is reviewed in this section. This is only one component of the total wavefront error of an astronomical adaptive optics system; its actual value in any specific system will be determined in the overall design optimization process, which is covered in chapter 9.

Even if the sampling error were the largest component in the error budget, it is unlikely that it could be allowed to exceed about 1 rad^2, (corresponding to a Strehl ratio of 0.37). This value can be used as a guide in evaluating the performance and configuration of laser beacon systems. Beacon altitudes of 8-12 km are appropriate for Rayleigh scattering, while altitudes of 90–100 km are used for sodium resonance scattering.

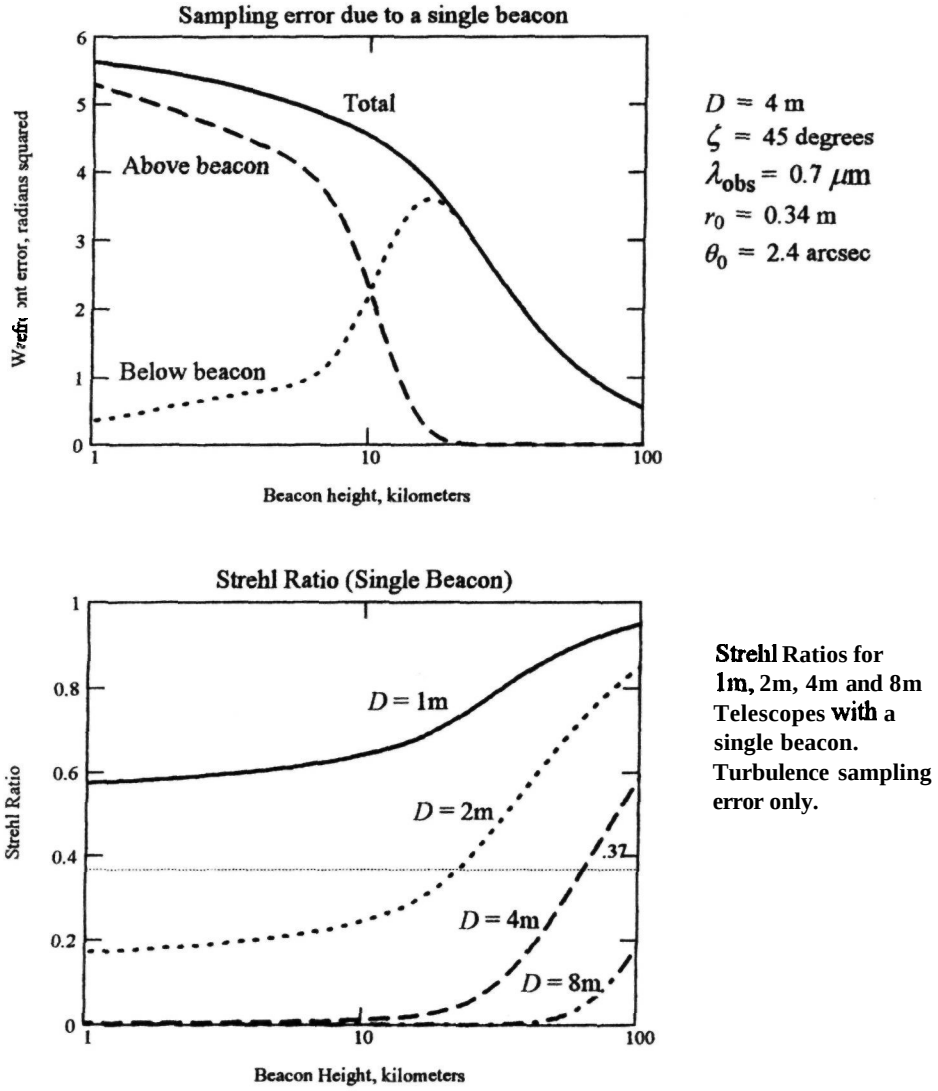
Figures 7.15 through 7.18 show beacon sampling errors and the corresponding Strehl ratios plotted against beacon altitude at wavelengths of 0.7 and 1.65 μm, for turbulence models representing good and average seeing at a high-altitude site. Each figure shows wavefront sampling errors above and below the beacon, as well as the corresponding Strehl ratios for telescope apertures of 1, 2, 4, and 8 m. The shape of the sampling error curves is interesting. Between about 8 and 20 km, the error below the beacon actually gets worse as its altitude increases, although the total error continues to decrease monotonically. This is explained by the presence of high-level turbulence that is not well sampled when the beacon is located close to it. Although the error below the beacon increases in this region, the error above it drops more rapidly, resulting in a lower overall error. It is also seen that the upper component of turbulence (above the beacon) essentially disappears at an altitude of about 20 km, leading to a rapid decrease in sampling error at higher altitudes. These effects favor the use of high-altitude sodium beacons.

Figures 7.15 and 7.16 show results for wavelengths of 0.7 and 1.65 μm with very good seeing, using the Mauna Kea background turbulence model described in section 3.3.4. This model represents the best conditions at a high-altitude site, giving $r_0 = 0.34$ m and $\theta_0 = 2.4$ arc seconds. The horizontal line is drawn at a Strehl ratio of 0.37, corresponding to 1 rad^2 of phase error. The intersection of this line with the curve for each aperture size gives the minimum height at which a single beacon can provide a sampling error of 1 rad^2. This aperture size is also the value of d_0 for the corresponding beacon height.

Figure 7.15 shows that to achieve a Strehl ratio of 0.4 at an observation wavelength of 0.7 μm, a single Rayleigh beacon covers an aperture of about 1.5 m, while a sodium beacon covers about 5 m. For these conditions, a 4-m telescope requires at least seven Rayleigh beacons, but only a single sodium beacon. An 8-m telescope may need as many as 30 Rayleigh beacons, but only four sodium beacons. The feasibility of multiple beacons is discussed in section 7.3.5.

At longer wavelengths, the situation improves considerably. In the H-band at 1.65 μm, figure 7.16 shows that a 4-m telescope is on the borderline with one Rayleigh beacon, while one sodium beacon should suffice, even for an 8-m telescope.

High-level turbulence is usually present, even at the best sites; beacon errors for the Mauna Kea average turbulence conditions are shown in figures 7.17 and 7.18. At an observation wavelength of 0.70 μm, the value of d_0 for a beacon height of 10 km is close to 1 m, necessitating large arrays of Rayleigh beacons. A 4-m aperture would be consid-



Strehl Ratios for 1m, 2m, 4m and 8m Telescopes with a single beacon. Turbulence sampling error only.

Figure 7.15 Beacon sampling errors and Strehl ratios for the Mauna Kea background (low turbulence) profile at $X = 0.70\ \mu m$, as a function of beacon height, for apertures of 1, 2, 4 and 8 m. The horizontal line represents 1 rad of phase error.

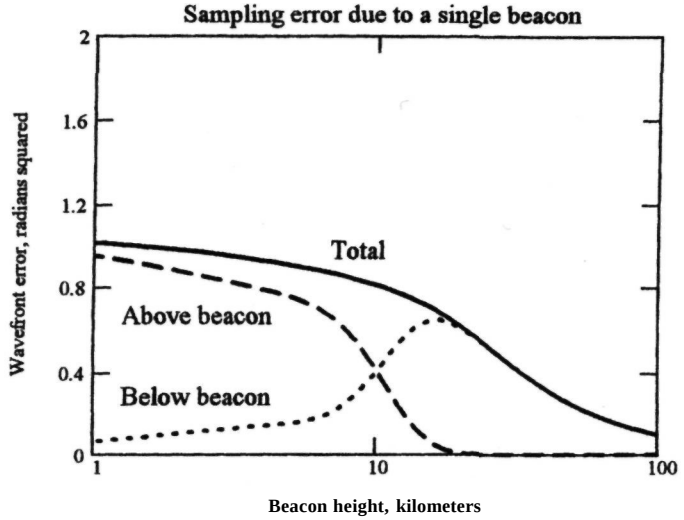
erably improved with one sodium beacon, while an 8-m telescope would require four. Again, there is considerable improvement at longer wavelengths, as shown in figure 7.18. For observations at $1.65\ \mu m$, the value of d_0 for a beacon altitude of 10 km is about 3 m, so a Rayleigh array of two or three beacons may suffice for a 4 m telescope, whereas for sodium a single beacon should satisfy even an 8-m telescope.

The impact of focal anisoplanatism on adaptive optics systems is shown effectively by plotting d_0 against beacon altitude. Figure 7.19 shows such a plot with observing wavelength as the parameter, using the Mauna Kea average turbulence model. The relationship is approximately linear, with a ratio of about 4:1 between the d_0 values for beacons at 90 km and 10 km.

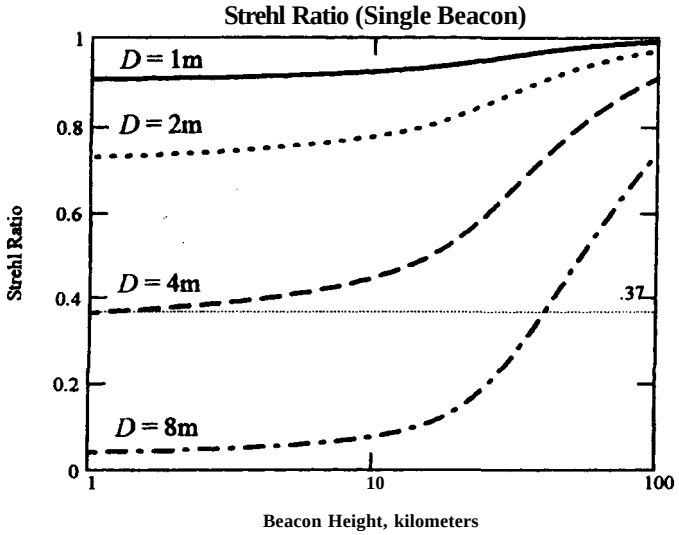
A direct comparison between single Rayleigh and sodium beacons for average turbulence is shown in figure 7.20, in which the Strehl ratio due to beacon sampling is plotted against aperture size. The advantage of using sodium beacons is clear.

7.3.5 Multiple Beacons

The results obtained above show that while a single sodium beacon may suffice for a 10-m aperture at IR



D = 4m
ζ = 45 degrees
λ_{obs} = 1.65 μm
r_0 = 0.34 m
θ_0 = 2.4 arcsec



Strehl Ratios for 1m, 2m, 4m and 8m Telescopes with a single beacon. Turbulence sampling error only.

Figure 7.16 Beacon sampling errors and Strehl ratios for the Mauna Kea background (low turbulence) profile at A = 1.65 μm, as a function of beacon height, for apertures of 1, 2, 4 and 8 m. The horizontal line represents 1 rad of phase error.

wavelengths in the best seeing conditions, multiple beacons will be necessary to implement adaptive optics compensation at visible wavelengths, especially if Rayleigh scattering is used. This is not a welcome prospect, because multiple beacons introduce a host of technical problems, as well as being additional sources of wavefront error. The simplest concept for implementing multiple beacons is to use an array of point sources within the telescope beam, as shown in figure 7.21. Each section of the telescope aperture, of size D_B, has a corresponding beacon. The beacons are generated by projecting laser beams from the ground, either through the telescope itself or through a closely related aperture. By using multiple beacons, the aperture area serviced by each beacon is reduced to a dimension comparable to d_0, minimizing the wavefront measurement error due to focal anisoplanatism. The wavefront measurements made with each beacon must then be combined or "stitched" together to obtain the best estimate of the overall wavefront. This is not a trivial process, as will become apparent.

Although the model depicted in figure 7.21 is convenient to analyze, it is not necessarily the best way to implement multiple beacons. To optimize the stitching process, the wavefront from each beacon should

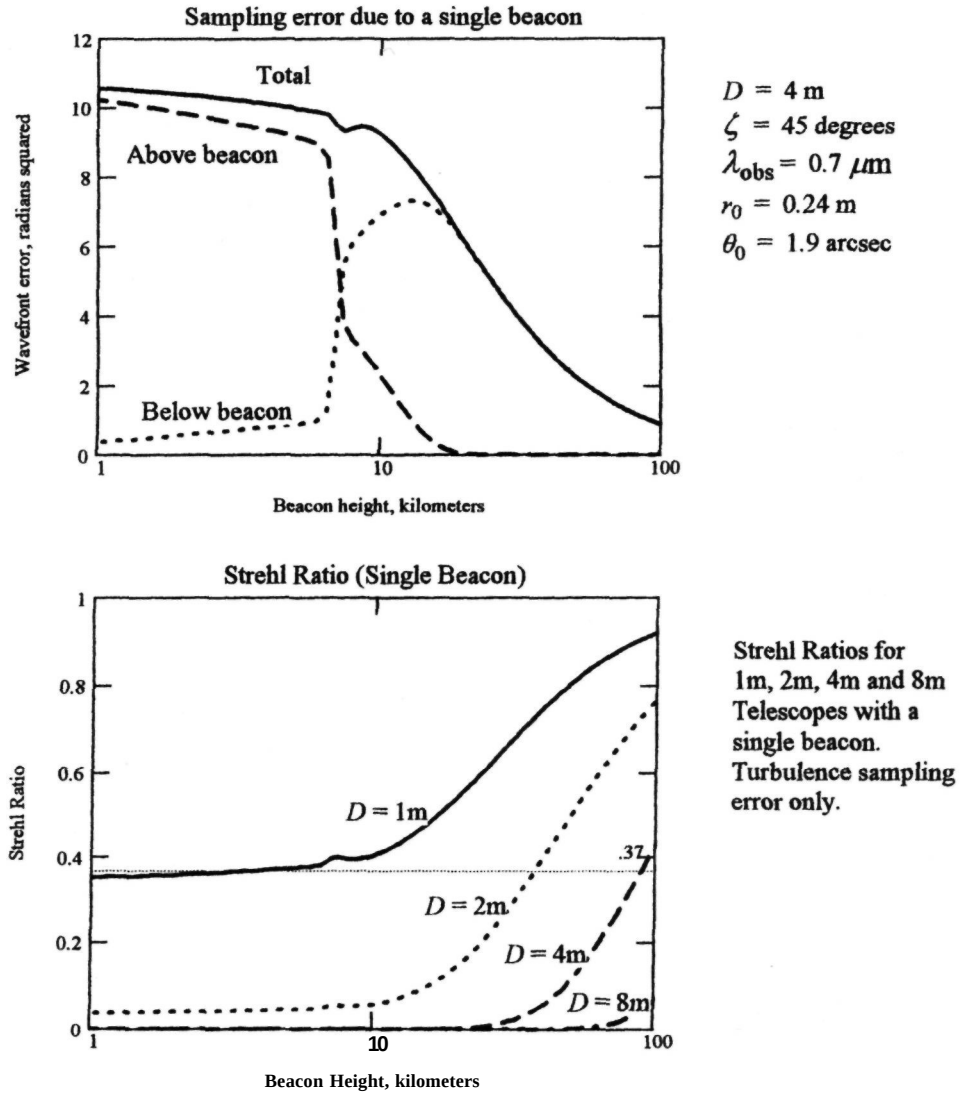


Figure 7.17 Beacon sampling errors and Strehl ratios for the Mauna Kea average turbulence profile at A = 0.70μm, as a function of beacon height, for apertures of 1, 2, 4 and 8 m. The horizontal line represents 1 rad of phase error.

ideally be measured over the whole telescope aperture, maximizing the overlap. The wavefront estimate at each location in the aperture is then the weighted sum of the measurements from all beacons. In the following treatment, the model of figure 7.21 is assumed.

Wavefront Measurement with Multiple Beacons

The major difference between wavefront measurements with one beacon and measurements with multiple beacons is that with a single beacon covering the whole aperture, the overall tilt component is discarded, because the image position must be stabilized by an external fiducial source. Overall tilt measurement errors associated with a single beacon are therefore irrelevant. This is not the case when multiple beacons are used, because the tilts belonging to each subsection of the aperture are an essential part of the wavefront measurement. These subsection tilts represent higher order wavefront components across the whole aperture, and therefore must be determined and corrected accurately. To minimize random position errors, multiple beacons should be projected by,

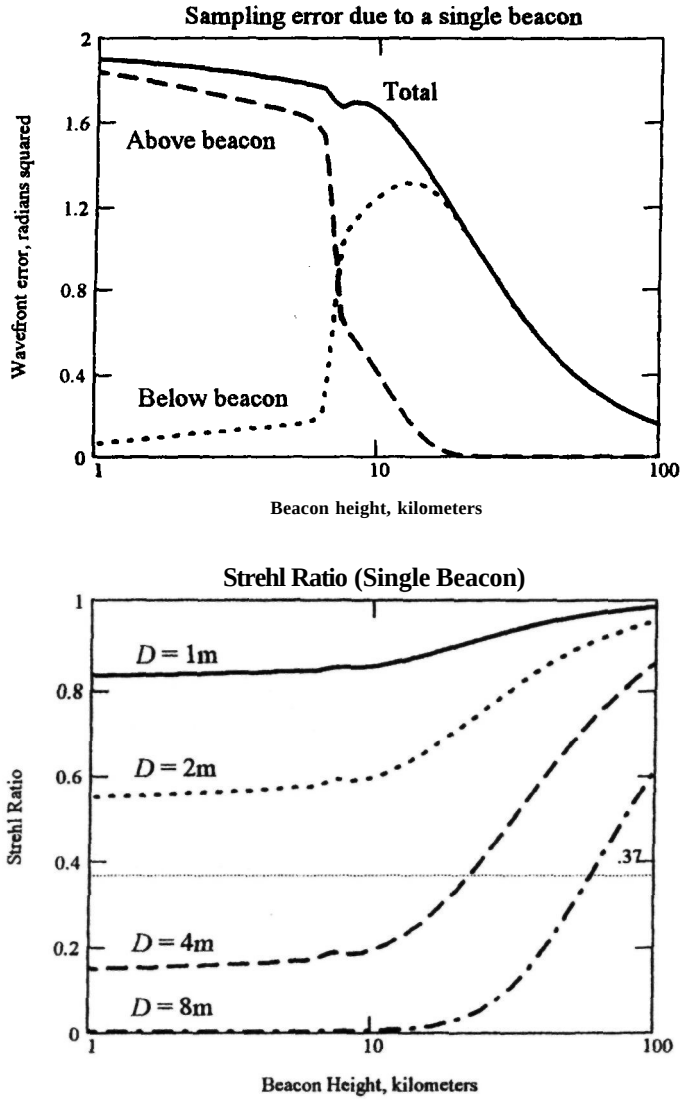


Figure 7.18 Beacon sampling erorrs and Strehl ratios for the Mauna Kea average turbulence profile at $\lambda = 1.65\,\mu m$, as a function of beacon height, for apertures of 1, 2, 4, and 8 m. The horizontal line represents 1 rad of phase error.

or at least measured by, a common aperture, which should be as large as possible. The worst possible scheme would be to project and measure the position of each beacon through the same elements of an array of small subapertures, because each uplink and downlink beam would then traverse the same path through the atmosphere. Subaperture tilts would be unmeasured and no useful information on the overall wavefront would be obtained.

There are two sources of error in measuring the subsection tilts associated with each beacon. The first of these is conic tilt error, defined in equation (7.26a), which is the overall tilt component of focal anisoplanatism, including unsampled tilt errors above the beacon. The second error is due to uncertainty in the positions of the beacons. Even though the entire beacon array may be projected and measured by the same optical aperture, the beams must be separated near their foci in order to form the individual spots. Consequently, they suffer independent, random displacements due to atmospheric turbulence.

When the wavefronts measured by multiple beacons are combined or "stitched" together, these individual tilt and position errors are added to the other measurement errors. The overall result of using multiple beacons is to reduce the error due to focal

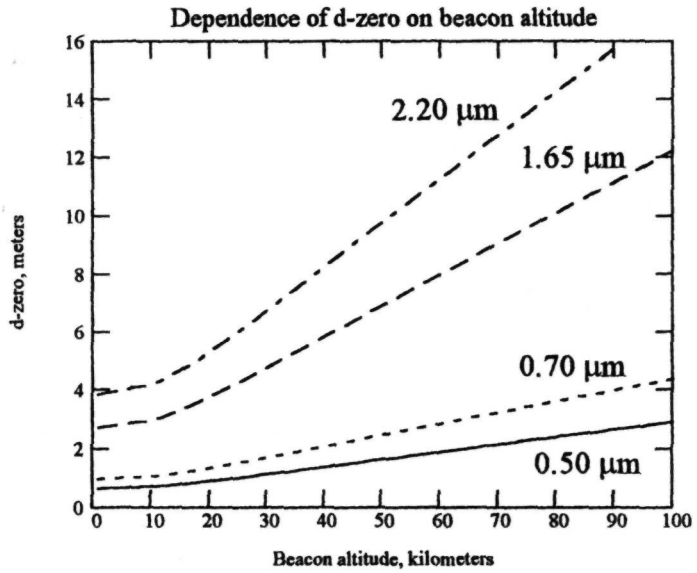


Figure 7.19 Dependence of d_0 on beacon altitude for the Mauna Kea average turbulence model at 45° zenith angle, for observation wavelengths of 0.50, 0.70, 1.65, and 2.20 *urn.*

anisoplanatism, but, at the same time, to add errors due to conic tilt and uncertainty in the beacon position. To estimate the overall improvement in performance obtained with multiple beacons, it is necessary to determine these errors.

If the wavefront error variance due to focal anisoplanatism for a single beacon over an aperture D_B, with piston and tilt removed, is defined as $\sigma_B^2(D_B)$; the conic tilt error over the same aperture, expressed as a phase variance, is $\sigma_{ct}^2(D_B)$; and the beacon position uncertainty error, expressed as a phase variance, is $\text{ffp}(Z)_B, D)$; then the measurement error due to the use of multiple beacons can be expressed as

$$\sigma_{MB}^2 = \sigma_B^2(D_B) + E[\sigma_{ct}^2(D_B)] + \sigma_P^2(D_B, D) \qquad (7.37)$$

where

E = error propagation factor that depends on the correlation between the anisoplanatic tilts
D = full aperture

The corresponding measurement error with a single beacon is $\sigma_B^2(D)$ The total number of beacons is N x N.

The wavefront phase error due to focal anisoplanatism with piston and tilt removed, σ_B^2 has been given in equation (7.35). The mean-square error is proportional to the 5/3 power of the aperture associated with each beacon, so that with multiple beacons, the error due to focal anisoplanatism is proportional to $N^{-5/3}$. Figures 7.15 to 7.18 illustrate the dependence of σ_B^2 on aperture size, beacon height, turbulence profile, and wavelength.

Conic Tilt Error

The conic tilt a^2, [equation (7.26a)], contains contributions from below and above the beacon. The lower component is computed from equation (7.34) and the upper component is given by [Sasiela 1988]

$$\sigma_{ut}^2 = 0.38k^2 D^{5/3} \mu_0^+(H) \qquad (7.38)$$

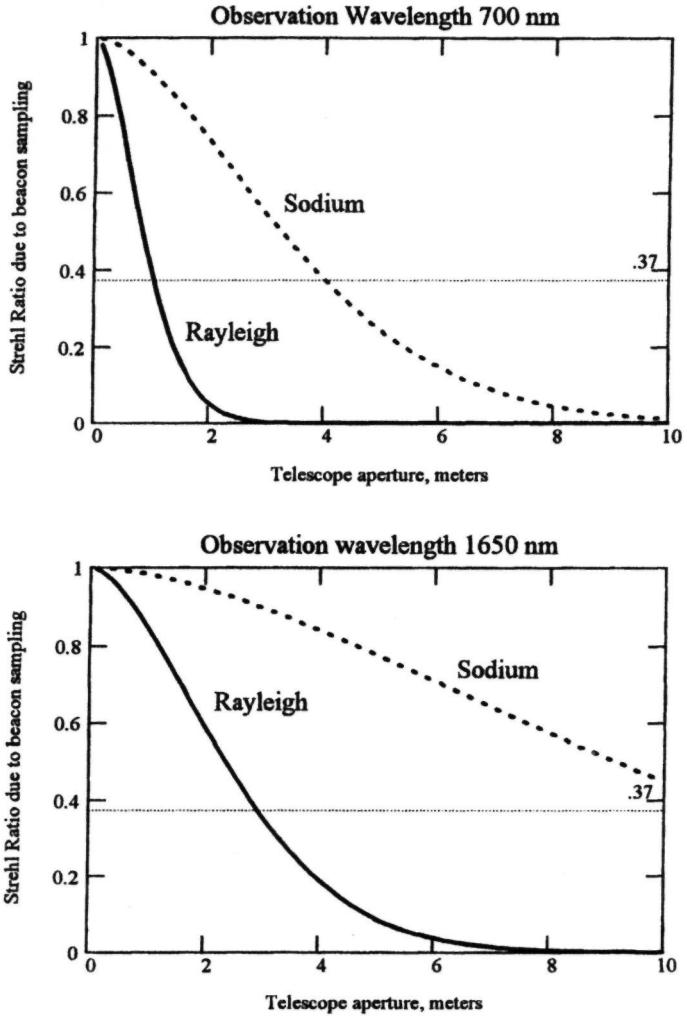
Values of σ_{Lt}^2 and σ_{ut}^2 are shown as a function of beacon height in figure 7.22, for separations of 1, 1.33, and 2 m. These separations represent the sections of the telescope aperture that would be serviced by each beacon in a multiple-beacon system. At beacon heights below about 8 km (the realm of Rayleigh beacons), most of the error comes from unmeasured tilt above the beacon. For the conditions shown, a Rayleigh beacon separation of less than 1 m is necessary to reduce the error to 1 rad^2 at a wavelength of 1.25 μm. Sodium beacons fare much better, as the error above the beacon becomes negligible and the conic tilt error drops to less than one-tenth of a radian squared.

To facilitate the calculation of conic tilt errors, it is convenient to define the parameter d_1 as the aperture diameter over which the tilt error produced by a single beacon is 1 rad^2. The conic tilt error over an aperture of diameter D_B, expressed as mean-square phase error, is then

$$\sigma_{ct}^2 = \sigma_{ut}^2 + \sigma_{Lt}^2 = \left(\frac{D_B}{d_1}\right)^{5/3} \qquad (7.39)$$

where

$$d_1 = \left\{k^2\left[0.38\mu_0^+(H) + 0.368\frac{\mu_2^-(H)}{H^2}\right]\right\}^{-3/5} \qquad (7.40)$$



Turbulence model:
Mauna Kea average,
ζ = 45 degrees,
r_0 = 0.24 m,
θ_0 = 1.9 arc sec.

Single beacons,
Rayleigh at 10 km,
sodium **at** 92 **km.**

Figure 7.20 Comparison of Strehl ratios for Rayleigh and sodium beacons for Mauna Kea (MK) average turbulence conditions, for observation wavelengths of 0.70 and 1.65 μm. The horizontal line represents 1 rad error.

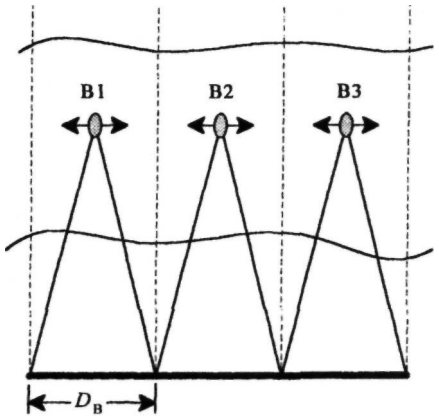
These expressions are analogous to those for focal anisoplanatism, equations (7.24) and (7.36). Values of $d\backslash$ for various operating conditions are given in table 7.3.

Beacon Position Error

The error due to beacon position uncertainty σ_P^2 will now be evaluated. The geometry is shown in figure 7.23. The beacons are at range L and height H_B, where $L = H_B \sec(\zeta)$. The beacon separation perpendicular to the telescope line of sight is b, and the full telescope aperture is D. The beacons are generated by laser beams launched from the telescope itself or a related aperture. The precision with which the position of each beacon is known depends on the size of the aperture used to launch or measure it, whichever is larger. The problem is to find the tilt component of the anisoplanatic error between two focused beams with separation b. The propagation paths between the telescope and each beacon overlap, producing correlation between the position variations due to turbulence. This is beneficial, as it is only the differences in position that are of interest; common mode position variations are equivalent to overall tilt, which is measured separately.

Figure 7.23 shows that the propagation path is divided into two sections that are separated by the beam clearance height h_c, given by

$$h_c = \frac{H_B}{(b/D)+1} \tag{7.41}$$

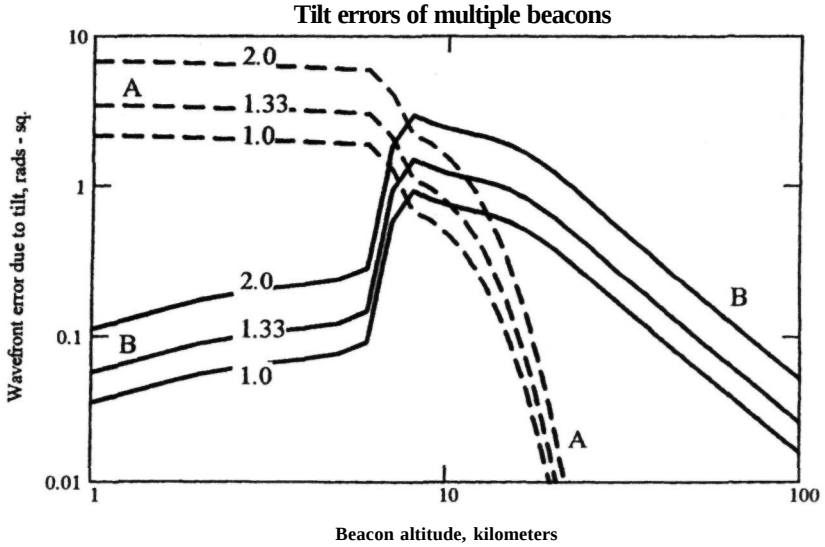


Wavefront error above beacons is not measured

Beacon position uncertainty

Tilt component of focal anisoplanatism is different for each beacon.

Figure 7.21 Error sources with multiple beacons. Even if the position of each beacon was known exactly, the turbulence above each section of the aperture is not measured exactly. The tilt components of the beacon measurement errors produce random higher order errors.



A Error above beacons, for separations of 1.0, 1.33, and 2.0 m.
B Error below beacons, for separations of 1.0, 1.33, and 2.0 m.

Figure 7.22 Wavefront errors produced by the tilt components of focal anisoplanatism. Conditions: Turbulence model Mauna Kea Average; observation wavelength 1250 nm; zenith angle 45 degrees. When multiple beacons are employed, tilt errors of this type are reconstructed into higher-order errors over the telescope aperture. They are of most concern with Rayleigh beacons, necessitating the beacons to be spaced less than 1 m apart. With sodium beacons the errors are negligible, even with a beacon spacing of 2 m.

Table 7.3 Values of Conic Tilt Error Parameter d_1

Turbulence Model	Zenith Angle (degrees)	Beacon Height (km)	rf, @ 0.50 μm (m)	d_1 @ 0.70 μm (m)	d_1 @ 1.65 μm (m)	d_1 @ 2.20 μm (m)
Hufnagel	0	10	0.23	0.34	0.68	**1.33**
		92	2.57	3.84	**7.71**	15.2
	45	10	0.18	0.27	0.55	**1.08**
		92	2.08	3.12	6.26	**12.34**
Mauna Kea background	0	10	0.49	0.73	1.47	**2.90**
		92	**5.56**	8.33	16.7	32.9
	45	10	0.40	0.60	1.20	**2.36**
		92	4.52	6.76	13.56	**26.73**
Mauna Kea average	0	**10**	0.36	0.54	1.09	2.15
		92	4.46	6.68	**13.4**	264
	45	10	0.29	0.44	0.88	1.74
		92	3.63	5.43	10.88	21.45

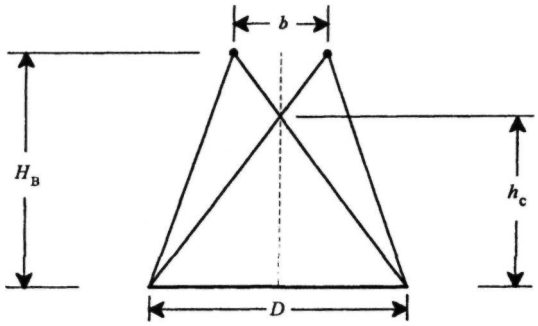


Figure 7.23 Geometry and notation for beacon position errors.

Below h_c, the beams overlap by an amount that increases to 100% at the telescope aperture. The weighting factor for differential tilt error is therefore zero at the telescope and rises to a maximum at h_c. Above h_c, the beams are separated and the weighting factor decays to zero at the beacon height.

Analytical expressions for the anisoplanatic tilt errors between two beacons with separation b at height H_B, when viewed by a common aperture of diameter D, have been developed by Sasiela [1994] using transverse filtering techniques. The angular errors depend on $D^{-1/}$ and are converted into equivalent wavefront phase errors over the sub-apertures associated with each beacon, by multiplying by $(D_B k/4)^2$ where $D_B = b = D/N$. The error component below the beam separation height, in radians squared of phase, is then

$$\sigma_{PL}^2 = k^2 D_B^2 D^{-1/3} \int_0^{h_c} dz\, C_N^2(z) F_{PL}(b, z, D, L) \quad (7.42)$$

where

$$F_{PL}(b, z, D, L) = 0.6675 \times \left[\left(\frac{bz}{DL}\right)^2 \left(1 - \frac{z}{L}\right)^{-1/3} - 2.067\left(\frac{bz}{DL}\right)^4 \left(1 - \frac{z}{L}\right)^{-7/3} - 1.472\left(\frac{bz}{DL}\right)^{14/3} \left(1 - \frac{z}{L}\right)^{-3} + \cdots\right] \quad (7.43)$$

The error component above the beam separation height, in radians squared of phase, is

$$\sigma_{PH}^2 = k^2 D_B^2 D^{-1/3} \int_{h_c}^{H_B} dz\, C_N^2(z) F_{PH}(b, z, D, L) \quad (7.44)$$

where

$$F_{PH}(b, z, D, L) = 0.7600 \times \left[\left(1 - \frac{z}{L}\right)^{5/3} - 0.6637\left(\frac{bz}{DL}\right)^{-1/3} \left(1 - \frac{z}{L}\right)^2 - 0.0031\left(\frac{bz}{DL}\right)^{-7/3} \left(1 - \frac{z}{L}\right)^4 - \cdots\right] \quad (7.45)$$

The total beacon position error variance, in radians squared of phase, is then

$$\sigma_P^2 = \sigma_{PL}^2 + \sigma_{PH}^2 \quad (7.46)$$

This function is plotted against the measurement aperture diameter D in figure 7.24, for various beacon heights and separations. The chart shows clearly the advantage of using a large aperture for measuring the beacon position. For Rayleigh beacons at heights around 10 km, the beacon position error σ_P^2 is much smaller than σ_T^2 and may usually be neglected, even when the measurement aperture is as small as 1 m. For sodium beacons at 92 km, the position error σ_P is between one-tenth and one-third the size of of.

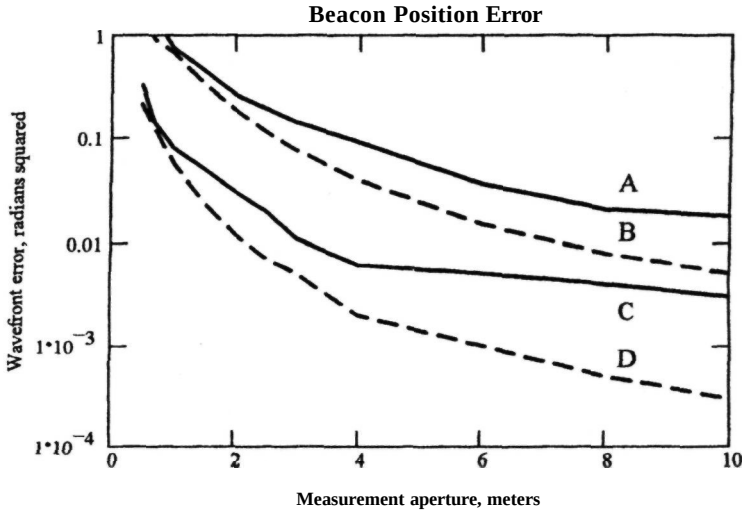
The dependence of the position error on the number of beacons is shown in figure 7.25.

Combining Multiple-Beacon Data

The wavefront measurements made with each beacon must be combined or stitched together to reconstruct the phase over the whole aperture. To illustrate this process, a system with four beacons is shown in figure 7.26. In the simplest case, each beacon is centered over the corresponding section of the full aperture. Each section has a dimension D_B, comparable in size to the characteristic dimension d_0. Within each D_B-sized section is an array of $n \times n$ subapertures, the size of each subaperture being commensurate with the turbulence coherence length r_0. If the wavefront within each D_B-sized area is measured exactly with respect to the full aperture, then the individual wavefront slopes of the subapertures will match those of the adjoining sections so that they may all be fitted or "stitched" together without any error, thereby defining the wavefront across the whole aperture.

In practice, the wavefronts are not measured exactly by the beacons because of the cone effect; in addition, their positions are not precisely known because of random displacements due to turbulence. The mean-square value of each of these errors has been calculated earlier, in terms of basic system parameters such as beacon height, aperture size, wavelength, and turbulence profile. The effect of these statistical errors on the overall reconstructed wavefront must now be determined.

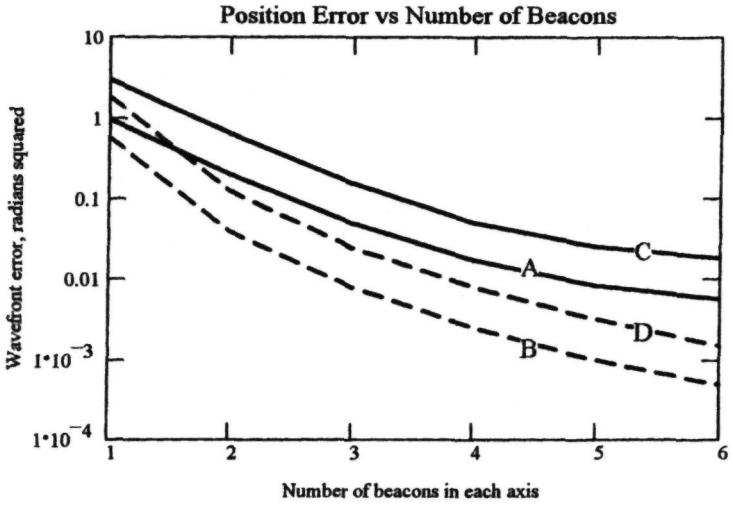
The conic tilt error σ_{ct} is calculated over wavefront areas of diameter D_B as shown in equation (7.39). The contribution of these tilts to the total wavefront error is determined with respect to the average tilt across the telescope aperture D. The overall wavefront error produced by the conic tilts therefore depends on the correlation between them. For example, if the conic tilts were perfectly correlated, they would combine to produce pure tilt across the full aperture, equivalent to zero wavefront error. On the other hand, if the conic tilts were completely uncorrelated, their mean-square values would add to produce a significant error in the reconstructed (stitched) phase values. The correlation is conveniently specified as an error propagation constant E, multiplying the beacon tilt errors as shown in equation (7.37). For perfect correlation between the beacon tilts (or when there is only a single beacon), E is zero.



A beacon height 10 km, separation 2 m
B beacon height 92 km, separation 2 m
C beacon height 10 km, separation 1 m
D beacon height 92 km, separation 1 m

Turbulence profile HV 20/20, X = 0.55 urn, ζ = 45 degrees

Figure 7.24 Wavefront errors produced by uncertainty in the position of multiple beacons, as a function of the size of the aperture used to measure the beacon positions. The relative tilt error between two beacons separated by the specified distance is converted to the equivalent wavefront error variance over aperture sizes equal to the beacon spacing. HV, Hufnagel-Valley.



A beacon height 10 km, aperture 4 m
B beacon height 92 km, aperture 4 m
C beacon height 10 km, aperture 8 m
D beacon height 92 km, aperture 8 m
Turbulence profile HV 20/20, X **= 0.55 urn,** ζ **= 45 degrees**

Figure 7.25 Wavefront errors produced by uncertainty in the position of multiple beacons, as a function of the number of beacons for Rayleigh (10 km) and sodium (92 km) beacons. Telescope apertures of 4 and 8 m are used to measure the beacon positions. The relative tilt error between two adjacent beacons is converted to the equivalent wavefront error variance over aperture sizes equal to the beacon spacing, HV, Hufnagel-Valley.

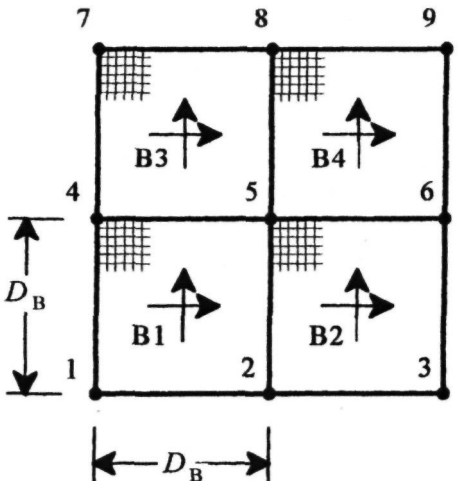


Figure 7.26 Four-beacon stitching model.

The correlation between turbulence-induced tilts in two apertures depends on the spatial power spectrum of the turbulence, the ratio of displacement to aperture size d/D, and the direction of displacement relative to the tilt axis. The tilt correlation functions do not depend on the vertical distribution or strength of the turbulence, nor on the wavelength. Correlation drops to zero only for orthogonal tilts when the beams are displaced in the direction of one of the tilt axes. If the beam displacement has a component in both orthogonal tilt axes, then the tilts become correlated. Correlation values for parallel, diagonal and orthogonal displacements in a square array are shown in figure 7.27. Sasiela [1994] has found that to a close approximation, the correlation functions for focal anisoplanatic tilt in a conical beam are the same as those for pure tilt in a collimated beam. Using these tilt correlation values, the correlation matrix $\mathbf{C_n}$ is easily constructed; the matrix for a 2 x 2 beacon array is shown in figure 7.28.

With a knowledge of the tilt correlation functions and the geometrical layout of the beacons with respect to the telescope aperture, the errors due to beacon position uncertainties can now be determined. For the simple 2×2 beacon system, if the phase values at the corners of each section of the wavefront are represented by the vector $\boldsymbol{\phi}$ and the random tilts in each section are represented by the vector n, then the measured wavefront gradients are given by the vector g, where

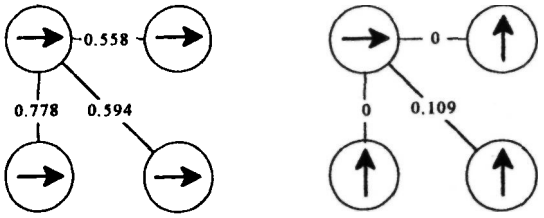


Figure 7.27 Tilt correlation factors for a square array.

$$\mathbf{g} = \mathbf{A}\boldsymbol{\phi} + \mathbf{n} \tag{7.47}$$

The reconstruction matrix A, relating the phase values to the gradients, is easily determined by inspection of figure 7.26 and is shown in figure 7.29. Each gradient is defined by the phases at the four points surrounding it. For example, in this array

$$g_1 = \tfrac{1}{2}[\phi_2 + \phi_5 - \phi_1 - \phi_4]$$
$$g_2 = \tfrac{1}{2}[\phi_3 + \phi_6 - \phi_2 - \phi_5], \quad \text{etc.}$$

The phase values giving the best (least-squares) fit to the measured gradients are given by the vector $\hat{\boldsymbol{\phi}}$, defined as

$$\hat{\boldsymbol{\phi}} = (\mathbf{A}^{\mathrm{T}}\mathbf{C}_{\mathrm{n}}^{-1}\mathbf{A})^{\oplus}\mathbf{A}^{\mathrm{T}}\mathbf{C}_{\mathrm{n}}^{-1}\mathbf{g} = \mathbf{B}\mathbf{g} \tag{7.48}$$

where T indicates the transpose of the matrix, and © indicates the generalized inverse, which is necessary when the matrix has singular values. The error in the estimated phase values is

$$\varepsilon = \mathbf{B}\mathbf{C}_{\mathrm{n}}\mathbf{B}^{\mathrm{T}} \tag{7.49}$$

A more useful parameter is the error propagator E that expresses the fraction of the beacon tilt error that shows up in the estimated wavefront when multiple beacons are used. The error propagator is given by

$$E = \frac{1}{N} Tr\left[(\mathbf{A}^{\mathrm{T}}\mathbf{C}_{\mathrm{n}}^{-1}\mathbf{A})^{\oplus}\right] \tag{7.50}$$

where *Tr* indicates the trace (the sum of diagonal elements). This value of E is then inserted in equation (7.37) to find the wavefront error due to multiple beacons. Values of E for beacon arrays up to 4 x 4 have been calculated by Sasiela [1994] and are given in table 7.4 .

7.3.6 Laser Projection Geometry

Laser beacons may be launched using many different optical configurations. The methods may be differentiated by how closely they are integrated with the optical and mechanical functions of the telescope. The most complete integration is achieved by combining the outgoing laser beam with the optical system of the telescope and projecting it through the main aperture, using a part or the whole of the primary mirror. The next step is to separate the laser beam optically from the telescope, while continuing to share the mount. In this case, the laser beam may

$$\mathbf{C} = \begin{bmatrix} 1 & a & b & c & 0 & 0 & 0 & d \\ a & 1 & c & b & 0 & 0 & d & 0 \\ b & c & 1 & a & 0 & d & 0 & 0 \\ c & b & a & 1 & d & 0 & 0 & 0 \\ 0 & 0 & 0 & d & 1 & b & a & c \\ 0 & 0 & d & 0 & b & 1 & c & a \\ 0 & d & 0 & 0 & a & c & 1 & b \\ d & 0 & 0 & 0 & c & a & b & 1 \end{bmatrix}$$

a =0.558 Parallel tilts, displaced 1 unit in direction of tilt.

b =0.778 Parallel tilts, displaced 1 unit perpendicular to tilt.

c =0.594 Parallel tilts, displaced 1.414 units at 45 degrees to tilt.

d =0.109 Orthogonal tilts, displaced 1.414 units at 45 degrees.

Figure 7.28 Tilt correlation matrix for 2 x 2 beacon array. The tilt correlation values depend on the geometrical relationship of the subapertures, as shown in figure 7.27.

$$A = \begin{vmatrix} 1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 1 \end{vmatrix} * 0.5$$

Reconstruction matrix for 2 x 2 beacon array
4 x-tilts + 4 y-tilts = 8 rows,
3x3 phase points = 9 columns.

Figure 7.29 Tilt reconstruction matrix for 2 x 2 beacon array. This matrix defines the weighting or correlation between the *x* and *y* tilts and the phase points.

Table 7.4 Beacon Tilt Error Propagator

Beacon Array	Error Propagator, *E*
2 x 2	0.426
3 x 3	0.465
4 x 4	0.957

be projected coaxially using separate optics mounted within the central area normally obscured in a Cassegrain telescope, or the laser optics may be attached to the side of the telescope tube, clear of the main aperture. Lastly, a bistatic arrangement may be used in which the laser projector is separated optically and structurally from the telescope. Whichever system is employed, it is evident that the laser beacon must be boresighted accurately to the science object being observed. Each method of projection has its advantages and disadvantages; these will be briefly reviewed in this section.

The advantage of using the full telescope aperture for projecting multiple beacons was pointed out by Foy and Labeyrie [1985]: by employing a common aperture, differences in tilt that could displace individual beacons are minimized. Indeed, as seen through the telescope (that is, with respect to the image plane), an array of spots projected and viewed through the same aperture will appear to be stationary, by virtue of the principle of optical reciprocity. In fact, the array of beacons will be displaced by turbulence, and a fixed reference star is required to point the telescope along a fixed line of sight, as discussed in section 7.4. Some differential motion is bound to occur because the individual beams must separate to form the individual images, but this is minimized by the use of a common aperture.

The use of a separate aperture to project the laser beam has both good and bad effects. The main advantage of employing a separate projection aperture is that it eliminates high-intensity laser light from the astronomical telescope. The laser projection optics and the telescope optics are both simplified and each can be optimized for its specific task. When the laser beam is physically separated, much of the Rayleigh scattering is outside the field of view of the telescope, so that a continuous-wave laser may be employed. Bistatic operation also has the potential of allowing a single laser installation to service several telescopes at one site.

There are two disadvantages in using a separate aperture to launch the laser beacon:

1. Tilt anisoplanatism produced by the separation of the telescope and laser projector causes a random displacement between the beacon and the telescope axis, even when a fixed guide star is used to point the telescope. This effect could be eliminated by using a fixed star to point the laser beam as well as the telescope.
2. The angle between the axes of the telescope and the laser projection optics causes the beacon to be elongated. The elongation may be considerable with bistatic operation, but it also occurs even when the laser is mounted coaxially with the telescope, in front of the secondary mirror. In this case, wavefront sensor subapertures within the telescope aperture are off-axis by an amount equal to their radius from the center of the primary mirror.

The basic geometry for bistatic operation is shown in figure 7.30. For simplicity, it is assumed that the

telescope T is pointed at the zenith. The laser beacon projector P is located a distance y from the telescope. The beacon is formed at mean height z, within a depth Az centered on the optical axis of the telescope. This depth corresponds either to the thickness of the scattering layer or to the receiver range gate, whichever is smaller. The beacon-pointing offset angle is $\alpha = \tan^{-1}(y/z)$

The angular separation of the laser beam from the telescope axis at any height h is

$$\theta(h) = \tan^{-1}\left[y\left(\frac{1}{h} - \frac{1}{z}\right)\right] \qquad (7.51)$$

In the case of a sodium beacon at 92 km, Rayleigh scattering produced by the laser beam at altitudes up to about 25 km will be off-axis by $29\,\mu$rad (6 arc seconds) for each meter that the laser projector is displaced. A displacement of just a few meters is therefore sufficient to keep low-altitude Rayleigh scatter out of the compensated field of view, eliminating the need for range gating and allowing the use of lasers with a very high pulse rate or even continuous outputs.

Even when a laser beacon is projected coaxially with the telescope, the laser beam is still effectively off-axis to the annular primary mirror. The geometry for this configuration is shown in figure 7.31. The radius of the projection aperture is a_b and the laser beam is assumed to have negligible divergence. The off-axis angle of the edge of the laser beam at any height h, as seen from a point on the primary mirror at radius a, is then

$$\theta_1(h) = \tan^{-1}\left(\frac{a - a_b}{h}\right) \qquad (7.52)$$

If a_b =0.25 m and h = 25 km, then the off-axis angle as seen from a mirror location at a radius of 2 m is $70\,\mu$rad or 14.4 arc seconds. If the science object and fixed guide star are within this radius, then a field stop will exclude the low-altitude Rayleigh scattering.

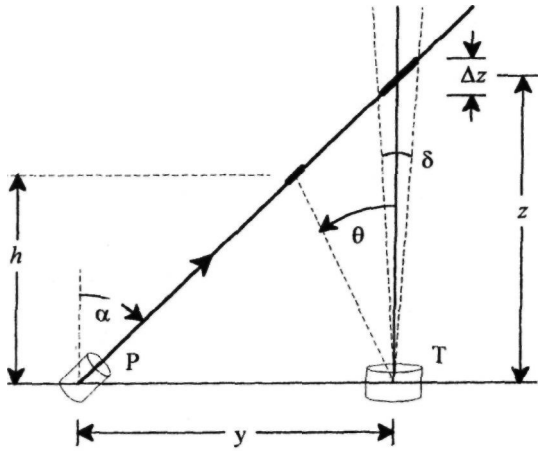


Figure 7.30 Geometry for biastic operation.

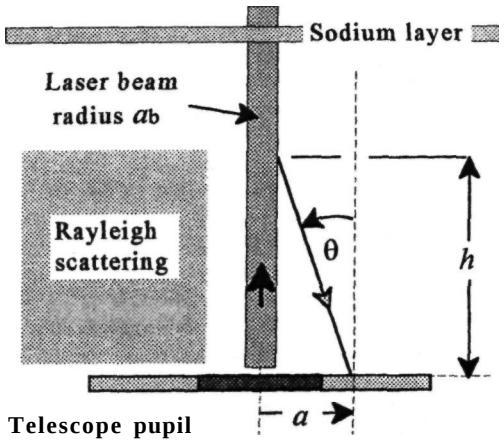


Figure 7.31 Geometry for coaxial projection of a sodium beacon.

Angular elongation of a laser beacon may be a problem with either Rayleigh or sodium beacons. Because of the thickness of the sodium layer and the need for relatively long pulses to minimize saturation, a sodium beacon actually consists of a bundle of glowing filaments, about 1 m in diameter and 10 km in length. Even when viewed from a few meters off-axis, beacons of this length have significant radial elongation. Beacon elongation has several undesirable effects: it increases the wavefront sensor errors for both Shack-Hartmann and shearing interferometers and may also produce errors because of angular anisoplanatism.

The angular elongation of the beacon in the direction of laser offset is

$$\delta = \Delta z \frac{y}{z^2} \qquad (7.53)$$

For a sodium beacon, the mean altitude z is about 92 km and the layer thickness Az approximately 10 km. The beacon elongation is therefore $1.2y\,\mu$rad, where y is the offset distance of the laser projector. To limit the elongation to $11\,\mu$rad, corresponding to the angular size of a 1-m sodium spot, the maximum laser displacement on the ground is 9 m. For larger displacements, it may be necessary to compensate for the elongation. The length of the return pulse received at the telescope is equal to twice the transit time of the light through the layer, plus the laser pulse length:

$$t_R = 2\left(\frac{\Delta z \sec\alpha}{c}\right) + t_L \qquad (7.54)$$

where

t_L = laser pulse length
c = speed of light

For a sodium beacon, the transit time is about 33 μs and the pulse length observed from the ground is

$66\,\mu s + t_L$. Because of the short pulse length, it is difficult to compensate optomechanically for pulse elongation. Beckers [1992] has suggested compensating the angular displacement of a laser beacon by adjusting the clock rate in the charge-coupled device (CCD) detectors used in the wavefront sensor to match the motion of the laser beacon. The charges would be moved along the columns of the CCD in the direction that the beacon pulse is traveling. Compensation of beacon elongation is only effective if the pulse length is short compared with the transit time through the layer; that is, when $t_L \ll t_R$ With long pulses, the spatial extent of the pulse becomes comparable to the layer thickness, and it can only be shortened by gating the receiver.

Elongation of the laser beacon may also degrade the wavefront measurement due to anisoplanatism. The isoplanatic angle is approximated as

$$\theta_0 = 0.31 \frac{r_0}{h} \left(\frac{\lambda_1}{\lambda_0}\right)^{6/5}$$

where

r_0 = turbulence coherence length at wavelength

h = mean turbulence height
λ_1 = observation wavelength

Using the criterion that the beacon elongation half-angle should not exceed θ_0, we obtain the relation $0.6y < \theta_0$, expressed in μrad. The limitation on the laser offset distance is therefore

$$y < 0.52 \times 10^6 \frac{r_0}{h} \left(\frac{\lambda_1}{\lambda_0}\right)^{6/5} \qquad (7.55)$$

Taking average values of $r_0 = 0.1$ m, and $h = 5$ km, the maximum value of y for visible observations is about 10m, while for IR observations at $1.6\,\mu$m it is 37 m.

These figures are rough estimates of the laser separations that are possible in bistatic systems. When designing a specific adaptive optics system, the wavefront errors due to laser projector offset must be calculated and included in the error budget. Even with bistatic operation, it is still highly advantageous to project all beacons in an array through the same aperture. The worst possible strategy would be to use separate projection apertures for each beacon in an array, because, in that case, the angular motions would be uncorrelated, and if the subapertures were small, the angular errors would be large.

7.4 Image Stabilization

7.4.1 Position Reference Requirements

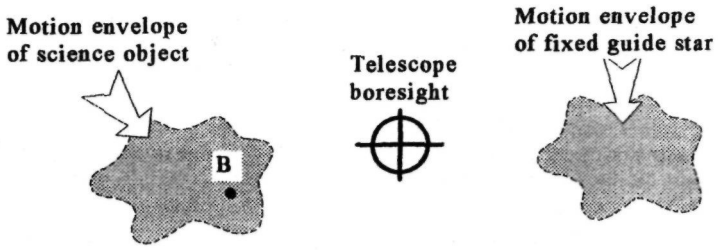
A laser beacon projected from the ground is randomly displaced by atmospheric turbulence and cannot be used as an absolute position reference to stabilize an image. This may not be important for very short exposures of less than 1/100 second or so, but for the much longer exposures typical of deep-space astronomy, the presence of even a small amount of image motion can offset any gain in resolution produced by adaptive optics. As the allowable image motion is usually specified as a fraction of the diffraction-limited resolution of the telescope aperture, *X/D*, the requirement for image stabilization becomes more critical as the aperture size increases.

To stabilize the position of an image, a fiducial source in the same frame of reference as the object is required. For astronomical objects, an exo-atmospheric source is necessary. Ideally, a star included in the science object itself should be used; the source brightness required for precise tilt tracking (typically around $m_V = 16$) is much less than that for high-order compensation. However, many objects of interest are of greater magnitude, or are diffused and otherwise unsuitable for use as reference sources. It is clear that for general observational astronomy, an independent and reliable method of image stabilization is essential.

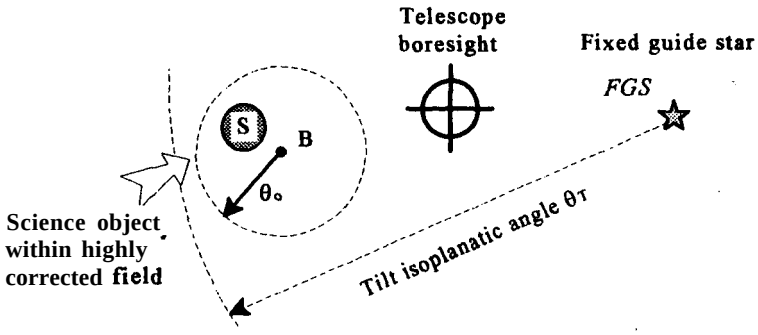
It is instructive to review the image stabilization problem using the telescope's image plane as the reference frame, because that is where the image is recorded and where its quality is evaluated. The conditions relating to image stabilization are depicted in relation to the telescope field of view in figure 7.32.

The situation before the adaptive optics is energized is shown in figure 7.32(a). It is assumed that the beacon is launched coaxially with the main telescope aperture, either sharing the optical beam path, or using a projection aperture in close proximity to it. With a common atmospheric path, the beacon will always appear to be near the center of the field of view, because backscattered light follows exactly the same path as the outgoing light through the atmosphere. On the other hand, light from the science object and a nearby reference star passes through the atmosphere in only one direction, with the result that these objects move randomly within an envelope determined by the amplitude of the tilt component of the turbulence. If they are within the same isoplanatic patch for overall tilt, their motion will be highly correlated.

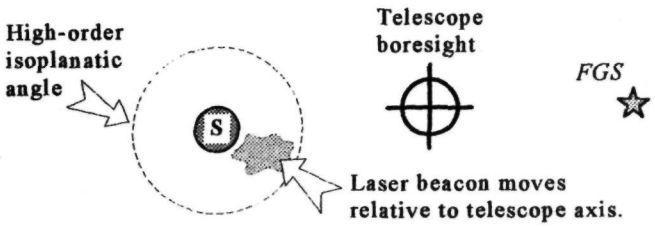
The situation when a fixed guide star(FGS) is used to correct overall image motion is shown in figure 7.32(b). The FGS is the reference source for the tilt sensor that controls the fast-tracking (tip-tilt) mirror in the adaptive optics system. The position of the FGS with respect to the optical axis can be offset as indicated. Surrounding the FGS is an area of radius θ_T; this is the isoplanatic area within which image motion is highly correlated with that of the FGS. The science object S must be within this area. Surrounding the beacon (B) is the high-order isoplanatic region of radius θ_0 which must also include the science object. This circle defines the highly compen-



(A) Laser beacon projected through telescope, with no overall tilt correction. The beacon location is fixed, but both the science object and the fixed guide star move randomly due to the overall tilt component of the turbulence.



(B) Overall image motion corrected using the fixed guide star *FGS.* **The science object must be within the high-order isoplanatic angle of the laser beacon** *B,* **as well as the tilt isoplanatic angle of** *FGS*



(C) If the laser beacon is projected through a different aperture from the imaging **system, it moves in relation to the telescope axis. This is acceptable so long** as the **laser beacon remains within the isoplanatic angle around the** science object.

Figure 7.32 Image stabilization conditions.

sated field of the telescope. The long-exposure image quality is maximized by making B and FGS as close as possible to S.

In the case of bistatic operation, when the laser beacon is projected through an aperture that is not congruent with that of the telescope, the situation is modified as shown in figure 7.32(c). Because the outgoing and return paths of the beacon are no longer identical, the position of B will move randomly in relation to the telescope axis, even when the science object is stabilized by the adaptive optics. The isoplanatic area surrounding the beacon will consequently be randomly displaced, with the possibility that it may not include the science object. With a bistatic laser projection system, the wavefront tilts are uncorrelated and may lead to considerable offsets, especially with a small laser aperture. In this case, a separate tilt-correction loop may be employed to stabilize the beacon position in the telescope field of view, driving a fast-tracking mirror in the laser projection path. A further problem with bistatic operation is that it causes angular elongation of the beacon, which may produce wavefront measurement errors. This subject has been discussed in section 7.3.6.

For long exposures, the precision of the tilt compensation provided by the fixed reference star is a critical issue. The wavefront error due to anisoplanatism increases monotonically with angular distance from the reference star and is inversely proportional to its brightness. When several fixed guide stars are available, a tradeoff may be necessary to select the best one. It is also possible to use multiple FGSs to provide some averaging of random errors, although this will increase the complexity of the tilt wavefront sensor.

When operating at visible wavelengths, as indicated in figure 7.32, the high-order isoplanatic patch of the laser beacon is usually much smaller than the tilt isoplanatic patch surrounding the FGS. It is unlikely therefore that the FGS will receive any high-order compensation at visible wavelengths—in fact, it is better not to compensate than to apply the wrong correction as this can double the mean-square error. The optical path to the FGS sensor should therefore be separated before the deformable mirror, as indicated in figure 7.33. However, the situation is quite different when an IR star is used as the FGS; in this case, the compensated isoplanatic patch surrounding the beacon is much larger, resulting in a greatly improved image of the FGS. The advantages of using an IR guide star have been described by Sandler, Stahl et al. [1994]. The improvement in sky coverage is dramatic.

7.4.2 Tilt Anisoplanatism

The discussion above has indicated the great importance of tilt anisoplanatism in evaluating the performance of adaptive optics systems using laser beacons. In this section, a method of determining the errors due to tilt anisoplanatism is described, using analytic expressions that are directly related to the basic parameters of adaptive optics systems. The geometry for tilt anisoplanatism is shown in figure 7.34. The angle between the science object and the fixed guide star is θ. This angle is on the order of 10^{-4} rads. The aperture of the telescope is D and the objects being observed are at a zenith angle of ζ. The height at which the two beams no longer overlap is an important parameter and is designated as the beam-clearance height $h_c = D\cos(\zeta)/\theta$. The z coordinate measures distance along the telescope axis. Initially, it is assumed that the three-dimensional turbulence spectrum follows the basic $\kappa^{-11/3}$ power law, with infinite outer scale.

Using Sasiela's method of transverse filter functions, outlined in chapter 3, the two-axis phase variance due to tilt between two objects at infinite range, separated by an angle θ, is

$$\sigma^2 = 0.2073\,k^2 \int_0^\infty dz\, C_N^2(z) \int d\kappa\, \kappa^{-11/3} \times \left[\frac{4J_2(\kappa D/2)}{\kappa D/2}\right]^2 2[1 - \cos(\kappa\theta z \cos(\varphi))] \tag{7.56}$$

After performing the angular integration in the transverse filter plane and converting to angular tilt by multiplying by $(4/kD)^2$, we obtain

$$\sigma_T^2 = \frac{666.9}{D^2} \int_0^\infty dz\, C_N^2(z) \int d\kappa\, \kappa^{-8/3} \times \left[\frac{J_2(\kappa D/2)}{\kappa D/2}\right]^2 [1 - J_0(\kappa\theta z)] \tag{7.57}$$

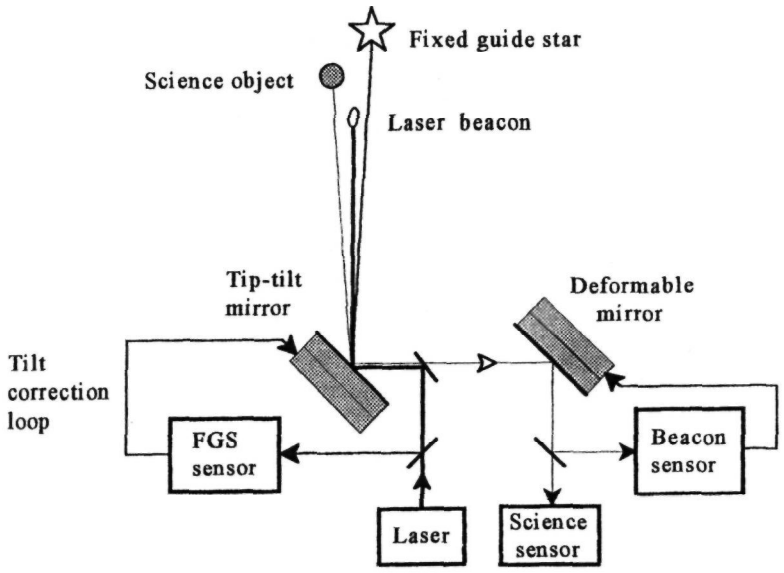


Figure 7.33 Block diagram of telescope pointing system using a fixed guide star (FGS).

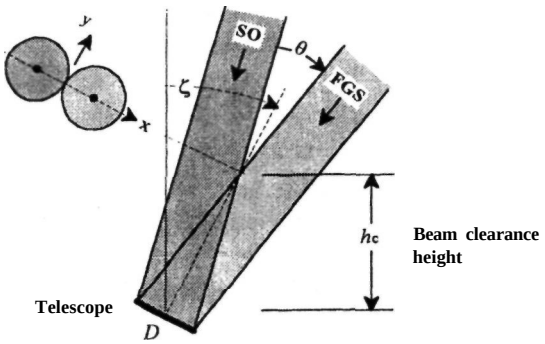


Figure 7.34 Geometry for tilt anisoplanatism between two objects at infinity. FGS, fixed guide star; SO, science object.

The integral in κ-space is evaluated using Mellin transform techniques [Sasiela and Shelton 1993a,b], with the result being expressed as a power series in θ/D. Separate solutions are obtained for $z\cos(\zeta) < h_c$ and $z\cos(\zeta) > h_c$. Above the beam clearance height h_c, the turbulence encountered by the two beams becomes uncorrelated when the separation exceeds the outer scale of turbulence, $\mathcal{L}_0$. The variance between the beams is then twice the variance of a single beam. Below h_c, the tilt variances become more highly correlated as z goes to zero. The value of h_c is usually between 20 and 100 km, so, in practice, there is little turbulence above the beam clearance height.

The anisoplanatic angular tilt error variances (Zernike tilt) produced by turbulence below and above h_c are given by

$$\left.\begin{matrix}\sigma_x^2\\ \sigma_y^2\end{matrix}\right\}_{H-} = \frac{2.67\mu_2^-(h_c)}{D^{1/3}}\left(\frac{\theta}{D}\right)^2\begin{Bmatrix}3\\1\end{Bmatrix} - \frac{3.68\mu_4^-(h_c)}{D^{1/3}}\left(\frac{\theta}{D}\right)^4\begin{Bmatrix}5\\1\end{Bmatrix} + \frac{2.35\mu_{14/3}^-(h_c)}{D^{1/3}}\left(\frac{\theta}{D}\right)^{14/3}\begin{Bmatrix}17/3\\1\end{Bmatrix} + \cdots \tag{7.58}$$

$$\left.\begin{matrix}\sigma_x^2\\ \sigma_y^2\end{matrix}\right\}_{H+} = \frac{6.08\mu_0^+(h_c)}{D^{1/3}}\begin{Bmatrix}1\\1\end{Bmatrix} - \frac{4.04\mu_{-1/3}^+(h_c)}{D^{1/3}}\left(\frac{\theta}{D}\right)^{-1/3}\begin{Bmatrix}0.8\\1.2\end{Bmatrix} + \cdots \tag{7.59}$$

The partial turbulence moments $\mu_m^-(h_c)$ and $\mu_m^+(h_c)$ have been previously defined. Note that h_c itself is also a function of θ.

It is seen that the tilt variances caused by turbulence below h_c are considerably different in the x and y directions, x being the axis of beam displacement. Average values for the single-axis mean-square displacements are given by $(\sigma_x^2 + \sigma_y^2)/2$ The average mean-square values of the main components of tilt anisoplanatism are shown as a function of field angle θ in figure 7.35. The total error obtained by summing these components is also shown. The chart is drawn for good seeing conditions, using a 4-m aperture. The results are independent of wavelength. Use of a different aperture size changes the scale of the abscissa but not the relationship between the various error terms. The atmospheric turbulence profile generally has a consistent shape as a function of altitude, with an overall scaling for turbulence strength. The dependence of the anisoplanatic tilt error terms on the angle θ is determined by the turbulence moments, which also have a consistent shape, with scaling for turbulence strength. These effects considerably simplify the analysis of anisoplanatic tilt errors and allow figure 7.35 to be used as a general model.

7.4.3 Small Angle Approximations for Tilt Anisoplanatism

For small field angles, the beam clearance height h_c is very large and turbulence below the beam crossing height is dominant. For angles up to about 100 μrad (0/40,000) between the science object and the tilt reference, the first term alone of equation (7.58) is a good approximation and the full turbulence moment μ_2 may be used in place of the partial moment. In this region, the total anisoplanatic angular tilt error variance may be approximated as

$$\sigma_T^2 = \sigma_x^2 + \sigma_y^2 = 10.68\mu_2\theta^2 D^{-7/3} \tag{7.60}$$

The average single-axis mean-square angular tilt (in radians squared) is half this value.

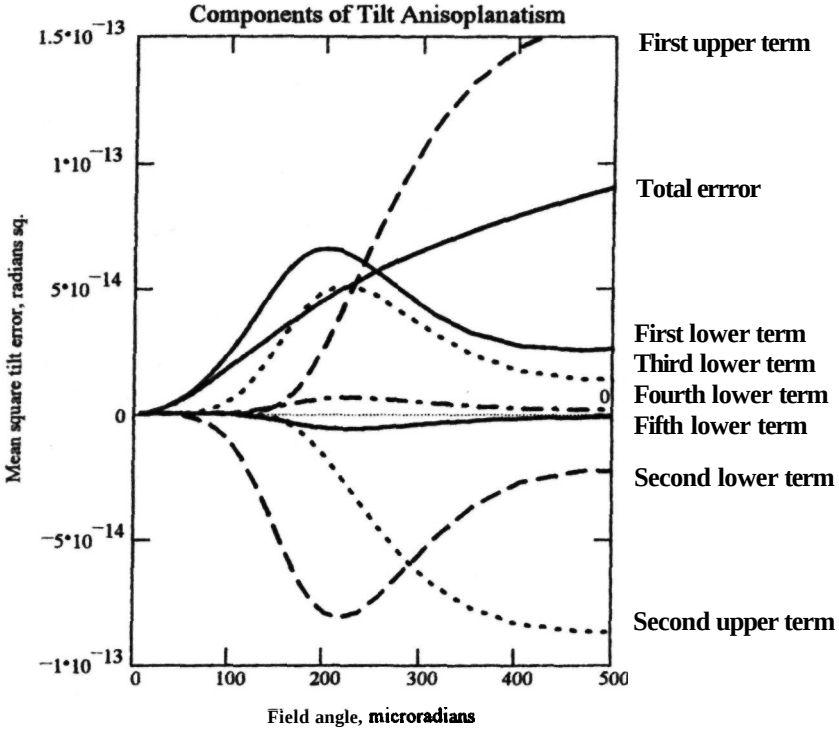
A characteristic angle θ_{TA} been defined by Sasiela and Shelton [1993a] as the angle between two beams at which the wavefront phase error due to to tilt anisoplanatism is 1 rad, so that for a small angle θ, the mean-square wavefront phase error is given by

$$\sigma_{TA}^2 = \left(\frac{kD}{4}\right)^2\sigma_T^2 = \left(\frac{\theta}{\theta_{TA}}\right)^2 \tag{7.61}$$

where

$$\theta_{TA} = \left[0.668\,k^2\mu_2 D^{-1/3}\right]^{-1/2} \tag{7.62}$$

It should be kept in mind that this approximation is valid only for field angles up to about $D/40{,}000$. When observing at IR wavelengths, a larger offset angle can be used between the reference star and the science object, so this simple relationship may no longer hold.



Conditions: Turbulence model HV 20/20, D = **4 m,** ζ = **45 degrees**

Figure 7.35 Main error components of tilt anisoplanatism. The first three components due to turbulence below the beacon (lower terms) are given in equation (7.58), and those due to turbulence above the beacon (upper terms) are given in equation (7.59). HV, Hufnagel-Valley.

7.4.4 Tilt Anisoplanatism at Large Angles

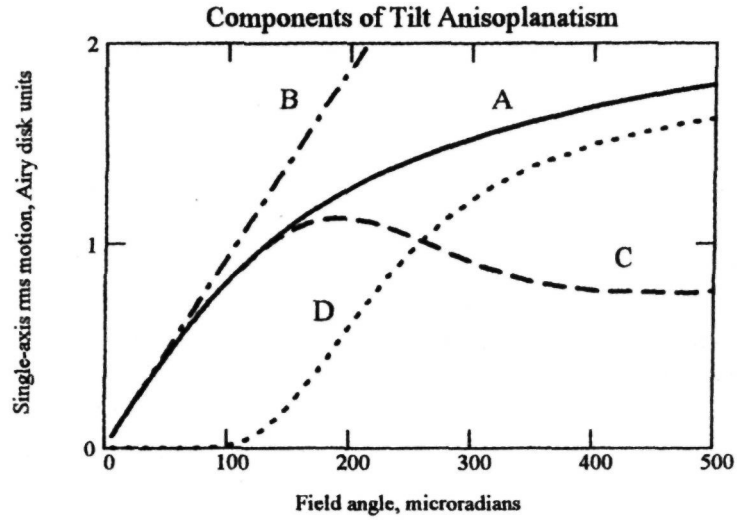
Figure 7.35 shows that higher order terms tend to occur in pairs of similar magnitude but opposite polarity, so care must be taken in truncating the series. For angles greater than $D/40{,}000$, both the second and third terms in equation (7.58) must be included.

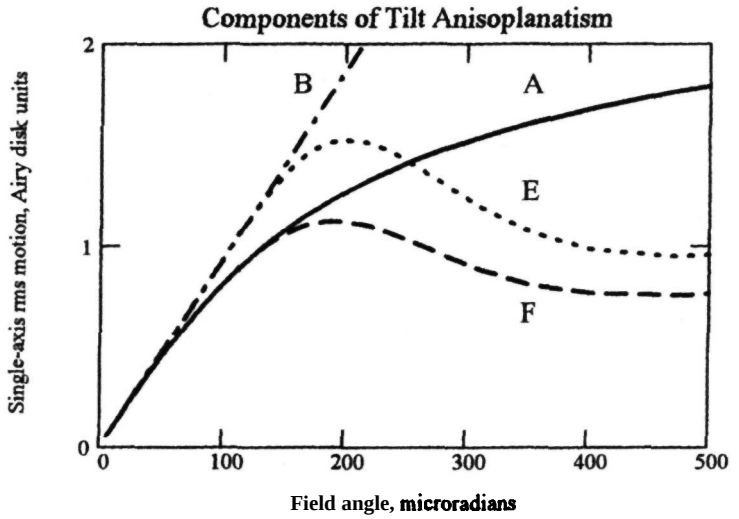
At larger angles, the beam clearance height h_c is reduced to the extent that significant turbulence occurs above the altitude at which the beams are separated. The upper error terms now become important, with the lower contribution gradually diminishing. For values of θ greater than $D/20{,}000$, it is necessary to include the upper terms. Again, note that the first and second upper terms are comparable in magnitude and opposite in polarity, so both must be included. The anisoplanatic tilt errors above and below h_c become equal at an angle of about $D/15{,}000$.

Some insight into the relative importance of anisoplanatic tilt errors can be gained by comparing them with the diffraction-limited resolution of the optical aperture. This ratio shows the increase in long-exposure image size due to tilt anisoplanatism. Figures 7.36 and 7.37 show the angular tilt errors normalized to the radius of the Airy disk, as a function of isoplanatic angle θ. Conditions for good seeing with a 4-m aperture in the visible band are depicted in figure 7.36, the solid curve representing the error computed using the five terms defined in equations (7.58) and (7.59). This curve is not perceptibly changed by the inclusion of additional terms. For these parameters, the normalized angular tilt due to anisoplanatism has a value of unity at an offset angle of $130\,\mu$rad.

The linear approximation of equation (7.60) considerably overestimates the errors due to tilt anisoplanatism at angles greater than $D/40{,}000$. The use of the first lower term in equation (7.59) is not much better. However, as the allowable normalized error due to tilt anisoplanatism is unlikely to be much larger than 1, the single-term approximations are allowable for use at wavelengths in the visible band.

The situation is quite different at longer wavelengths because the Airy disk is larger. In that case, larger values of θ can be used and single-term approximations greatly overestimate the error due

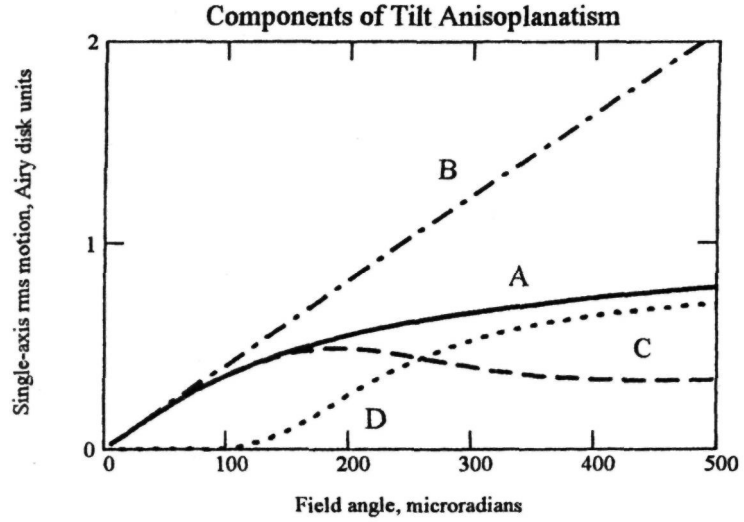


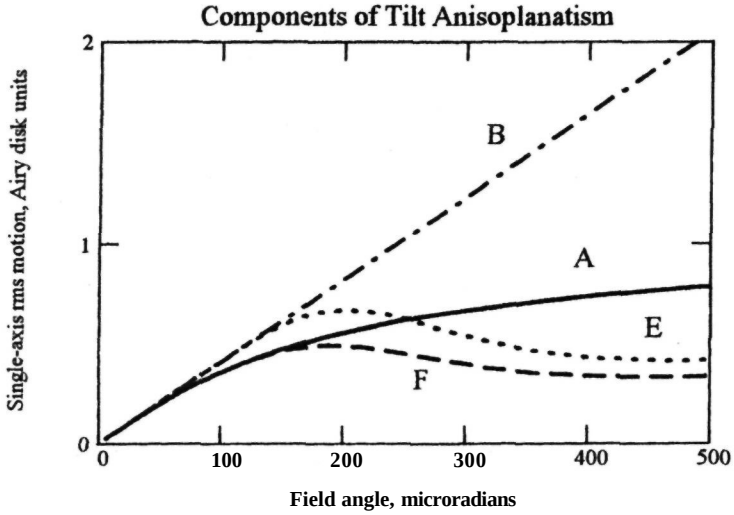


A total error, (eight terms)
B linear approximation, equation (7.60)
C five lower terms
D three upper terms
E first lower term only
F three lower terms only

Conditions: Turbulence model HV 20/ 20, D = 4 m, λ = 0.55 um

Figure 7.36 Angular anisoplanatic tilt errors normalized to radius of Airy disk for a wavelength of 0.55 μm. The components are plotted on two charts for clarity. HV, Hufnagel-Valley.





A total error, (eight terms)
B linear approximation, equation (7.60)
C five lower terms
D three upper terms
E first lower term only
F three lower terms only

Conditions: Turbulence model HV 20/ 20, D = 4 m, λ = 125 μm

Figure 7.37 Angular anisoplanatic tilt errors normalized to radius of Airy disk for a wavelength of 1.25 μm. The components are plotted on two charts for clarity. HV, Hufnagel-Valley.

to tilt anisoplanatism. Figure 7.37 depicts the same conditions as in the previous case except that the wavelength is now $1.25\,\mu$m. It is seen that the normalized anisoplanatic tilt error remains less than 1, even for large offset angles. Neither of the single-term approximations gives useful results at this wavelength, leading to gross over- and underestimates of the error. For systems that must work over a large range of observation wavelengths, it is necessary to use the full expression to determine the maximum value of ϑ. As the anisoplanatic tilt error is the main factor in determining sky coverage, it is essential to use a realistic error model.

7.4.5 Dependence of Tilt Anisoplanatism on Telescope Aperture

Another question of interest is the dependence of tilt anisoplanatism on the size of the telescope aperture. The aperture size is not usually considered as a free parameter; in most observing situations, the maximum available aperture is routinely used. However, it is instructive to see its influence on tilt anisoplanatism, which ultimately determines the sky coverage. The rms value of the anisoplanatic angular tilt error is plotted as a function of field angle ϑ for aperture sizes of 2, 4, 6, and 8 m in figure 7.38. It is seen that at all values of ϑ, the anisoplanatic tilt error is inversely dependent on aperture size, varying as $D^{-1/2}$. The reason is quite clear: for a given field angle, the beam clearance height h_c is directly proportional to D, and a larger value of h_c means that a greater part of the atmosphere is common to the two beams. Hence, the smaller anisoplanatic tilt error.

But, when the anisoplanatic tilt error is normalized to λ/D, as shown in figure 7.39, the net result is that it increases as $D^{1/2}$, because the Airy disk radius is proportional to D^{-1}. Thus, the error due to tilt anisoplanatism becomes a larger factor in the error budget with large apertures.

7.4.6 Effects of the Outer Scale of Turbulence

It has been assumed, up to now, that the turbulence spectrum has an infinite outer scale. It was shown in section 2.4 that when the turbulence spectrum has a finite outer scale, the single-aperture overall tilt component is significantly reduced. Using a von Kantian spectrum with an outer scale of $\mathcal{L}_0$, the variance of the single-aperture Zernike angular tilt component is approximately

$$\sigma_{TS}^2 = \frac{6.08\mu_0}{D^{1/3}}\left[1 - 1.42\left(\frac{D}{\mathcal{L}_0}\right)^{1/3}\right] \tag{7.63}$$

For example, if the outer scale is 10 times larger than the telescope aperture, then the rms value of the overall tilt is only about 60% of that produced with an infinite outer scale.

In the case of tilt anisoplanatism, we are dealing with the difference between two closely spaced beams. Large-scale tilt components have a common effect on these beams, so the outer scale dimension has less influence on the result. Sasiela and Shelton [1993a] show that for small displacement angles, the

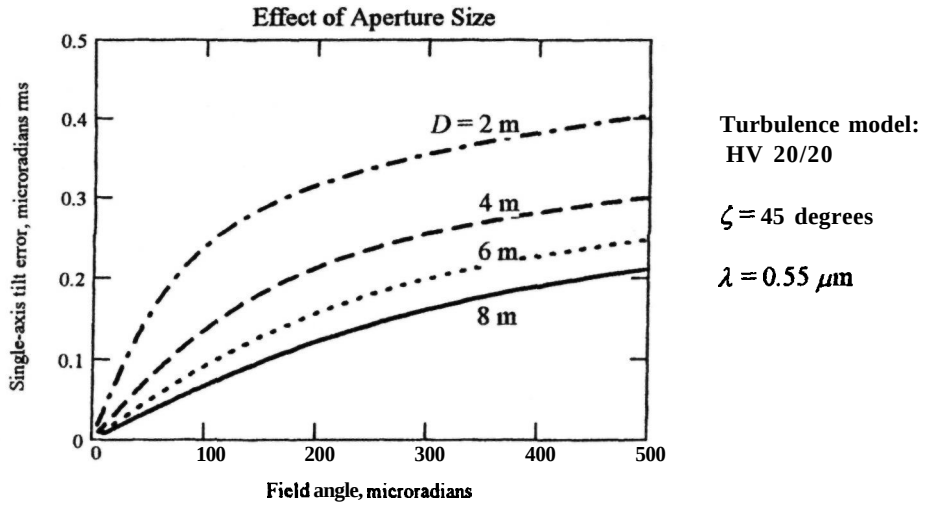


Figure 7.38 Anisoplanatic tilt angular error for apertures of 2, 4, 6, and 8 m. HV, Hufnagel Valley.

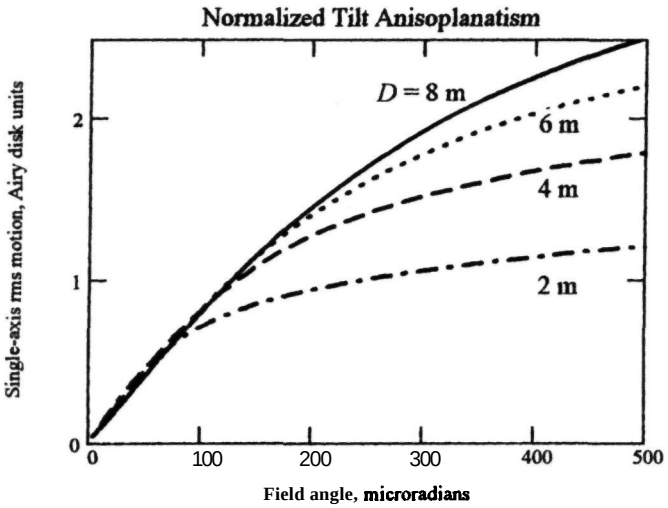


Figure 7.39 Anisoplanatic tilt error, normalized to the Airy disk radius, for apertures of 2, 4, 6, and 8 m. HV, Hufnagel-Valley.

anisoplanatic angular tilt variance with finite outer scale is given by

$$\sigma_{\mathrm{TF}}^2 = \frac{10.68\mu_2}{D^{1/3}}\left(\frac{\theta}{D}\right)^2 \times \left[1 - 20.6\left(\frac{D}{\mathcal{L}_0}\right)^2 + 27.4\left(\frac{D}{\mathcal{L}_0}\right)^{7/3} + \cdots\right] \quad (7.64)$$

In this case, for the same 10:1 ratio of outer scale to aperture, the rms value of the anisoplanatic tilt is 96% of that produced with a infinite outer scale. These results show that outer scale effects are generally negligible for tilt anisoplanatism, except when the outer scale becomes comparable to the size of the aperture.

7.5 Rayleigh Beacon Lasers

7.5.1 Requirements

The process of molecular scattering is not critically dependent on the characteristics of the incident radiation, so the laser requirements are not restrictive and a wide variety of lasers may be used for this application. The dependence of the received signal on the laser parameters can be found from equations (7.1) and (7.2). The number of photons N_p received from a range interval Az may be expressed in terms of the following variables

$$N_p \propto \frac{E\lambda\sigma_B n(z)\Delta z}{z^2} \quad (7.65)$$

where

- E = laser pulse energy, J
- λ = optical wavelength, m
- σ_B = effective backscatter cross-section, m^2
- $n(z)$ = number density of scatterers at range z, m^{-3}
- Az = receiver range gate length, m
- z - range at center of range gate, m

The backscatter cross-section σ_B is proportional to $1/\lambda^4$ and the density of scatterers $n(z)$ is proportional to $\exp(-z/z_0)$. If the receiver range gate is adjusted to cover a range interval at the focus that has an angular diameter of *Act*, then using equation (7.2), we have

$$\frac{\Delta z}{z^2} = \frac{2\Delta\alpha}{D_p} \quad (7.66)$$

Substituting for σ_B, $n(z)$, and Az/z in equation (7.65) and normalizing the received photons to the pulse energy, we obtain an expression for the efficiency of Rayleigh scattering as a function of beacon height and wavelength:

$$\eta_R = \frac{N_p}{E} \propto \frac{\exp(-z/z_0)\Delta\alpha}{\lambda^3 D_n} \quad (7.67)$$

The wavelength dependence is reduced to $X^{\sim}$ because we are counting photons whose energy depends on λ^{-1}.

7.5.2 Lasers for Rayleigh Beacons

Rayleigh (molecular) scattering is most efficient at UV and visible wavelengths between about 0.35 and 0.6 μm, as shown in section 7.2.3. Many different

Table 7.5 Rayleigh Beacon Lasers

Reference	Zollars 1992	Fugate et al. 1994	Sandler 1992b	Fugate et al. 1994	Sandler 1992b
Organization	MIT/LL	PL/SOR	TTC	PL-SOR	TTC
Name	SWAT	Gen I	MACE I	Gen II	MACE II
First operated	1988	1989	1990	1992	
Type	Dye	Cu vapor	Excimer	Cu vapor	YAG
Pump	Flashlamp				
Wavelength, nm	512	511, 578	353	511, 578	532
Average power, W	30-50	75	—	180	
Pulse energy, J	6	0.015	5-10	0.036	2
Repetition rate, Hz	5-10	5000	Single	5000	10(540?)
Pulse width, μs	2	0.05		0.05	
Beam quality	> 20 x DL	30 x DL		30 x DL	1.6 x DL

MIT/LL, Massachusetts Institute of Technology, Lincoln Laboratory; PL/SOR, United States Air Force Phillips Laboratory, Starfire Optical range; TTC, Thermoelectron Technologies Corporation [now ThermoTrex Corporation (1997)]; SWAT, Short-Wavelength Adaptive Techniques; Gen I/II, Generation I/II.
DL, diffraction limit.

types of laser can be used to generate Rayleigh beacons. General requirements include a power output on the order of 100 W, pulse rate of 1000 Hz or more (for continuous closed-loop operation), and good beam quality to produce a small beacon diameter. Some of the lasers that have been used in recent experiments with adaptive optics are listed in table 7.5.

The copper vapor laser has useful characteristics, both for direct use in generating Rayleigh beacons and also as a pump laser for sodium lasers. Characteristics of commercially available copper vapor lasers are listed in table 7.6.

7.6 Sodium Beacon Lasers

7.6.1 Requirements

Lasers for generating sodium beacons have more stringent requirements than those for generating Rayleigh beacons; the requirements involve both the temporal and spectral characteristics of the radiation. Gross limitations on pulse length and repetition rate are easily deduced from the physics and geometry of the scattering layers. The peak power is limited by saturation effects in the sodium layer, as discussed in section 7.2.5. To obtain the most efficient utilization of laser energy, the spectral content of the radiation must closely match the absorption spectrum of sodium. The design of lasers to generate sodium resonance fluorescence therefore requires a detailed knowledge of the physics of the sodium atom.

Considering, first, the macropulse format for sodium beacon lasers, it is found that the pulse energy, pulse width, and repetition rate are determined by two factors:

1. the vertical distribution of scattering species in the atmosphere;
2. the saturation characteristics of the sodium layer.

The laser pulse dimensions associated with these factors are depicted in figure 7.40.

The geometrical factors are easily deduced by reference to the scattering profile shown in figure 7.7. A laser beam tuned to the sodium resonance at 589 nm produces Rayleigh scattering at altitudes up to about 40 km. When using a coaxially projected laser beam with a sodium beacon, the sodium resonance return must be separated from the Rayleigh

Table 7.6 Characteristics of Commercially Available Copper Vapor Lasers[a]

Wavelengths	511 and 578 nm
Single unit average power	10–120 W
Single aperture average power	> 200 W
Peak pulse power	50–500 kW
Pulse duration	10–40 ns
Pulse repetition rate	2–32 kHz
Beam quality	< 10× diffraction limited
Timing jitter	< 1 ns rms

[a]Reference: Lewis et al. 1993. (Data refer to lasers made by Oxford Lasers Ltd, Abingdon Science Park, Abingdon OX14 3YR, U.K.)

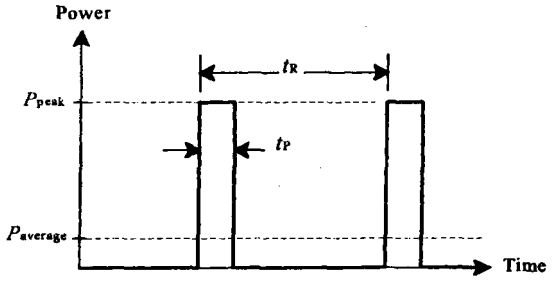


Figure 7.40 Pulse shape constraints for sodium beacon lasers.

scattering that is produced as pulses travel through the atmosphere. This is done primarily by range gating, together with a suitably tailored laser pulse format. The pulse repetition rate is limited by the requirement that scattered light from a pulse at the top of the sodium layer, at about 100 km altitude, should be received at the telescope before the next pulse is transmitted. The spacing between macropulses, t_R, must therefore be sufficient for light to make a round trip of 200 km, which requires a period of 667 μs. The maximum pulse length is governed by the physical gap between the Rayleigh and sodium scattering modes. Rayleigh scattering drops to a low level at about 40 km, while the lowest altitude of the sodium layer is about 80 km. To avoid overlap between the Rayleigh and sodium returns, the spatial extent of the laser pulse must not exceed the physical distance between them, which is about 40 km, corresponding to a maximum pulse length of 133 μs. These results show that in order to separate the sodium return in a coaxial projection system, the maximum pulse repetition rate must not exceed 1500 Hz and the pulse duty cycle must not exceed 20%.

The minimum pulse width, which may, in practice, be much smaller than 133 μs, is determined by the ratio of peak-to-average power in the laser pulses. The peak power is limited by saturation, which occurs when the peak power level exceeds about 50 kW m^{-2}. The average power is determined by the signal-to-noise ratio required by the wavefront sensor, and is typically 5–10 W. These figures give a duty cycle of about 0.03%, corresponding to a minimum macropulse width of 0.2 μs. Lasers specially designed for generation of sodium beacons (table 7.7) fall within these guidelines, having macropulse repetition rates of about 1 kHz and pulse lengths between 0.5 and 150 μs.

With bistatic operation, the sodium and Rayleigh backscatter returns are angularly separated, as discussed in section 7.3.6. Spatial discrimination in the focal plane eliminates the pulse overlap problem, allowing much higher pulse repetition rates with no limitation on the maximum duty cycle. Bistatic operation is therefore suitable for sodium lasers pumped by copper vapor lasers, which have repetition rates of 5 kHz or more. Bistatic operation also allows a continuous-wave laser to be used for generating sodium beacons.

A third method of laser projection, falling somewhere between the coaxial and bistatic modes, is to place the launching aperture in front of the secondary mirror of the telescope. The backscatter from low altitudes is then considerably off-axis at the beacon focal plane, allowing it to be rejected by a field stop.

Table 7.7 Sodium Beacon Lasers

Reference	Humphreys et al. 1991	Jeys et al. 1992	Jeys 1992	Max et al. 1992
Organization	MIT/LL	MIT/LL, NSF, UCh., U.S. Army	MIT/LL, PL/SOR	LLNL, DOE
Year operated	1984	1992	1992	1992
Type	Dye	Nd:YAG sum frequency	Nd:YAG sum frequency	Dye
Pump	Flashlamp	Flashlamp	Flashlamp	Cu vapor laser (AVL1S)
Average power, W	2.4	5	10–20	1500
Repetition rate, Hz	20	10	840	26,000
Pulse energy, mJ	120	500	12-24	38
Macropulse width, μs	4	80	65	0.032"
Micropulse width, ns	—	0.6	0.35	32
Spectral width, GHz	3	1.7	3.0	0.05
Beam quality	4 × DL	2 × DL	3 × DL	1.54 × DL

"Without stretcher.

MIT/LL, Massachusetts Institute of Technology, Lincoln Laboratory; NSF, National Science Foundation; UCh, University of Chicago; PL/SOR, Phillips Laboratory, Starfire Optical Range; LLNL, Lawrence Livermore National Laboratory; DOE, U.S. Department of Energy; AVL1S, Atomic Vapor Laser Isotope Separation Program.

DL, diffraction limit.

This situation has been analyzed in section 7.3.6. The intensity of the unwanted Rayleigh return decreases rapidly toward the center of the image plane, so that the interference is minimal over a small field of view. A variation of this method is to attach the laser launching tube to the side of the main telescope structure.

Considering, now, the laser spectral requirements, the sodium D_2 absorption line has a double-peaked shape, as shown in figure 7.11, caused by the presence of two ground states. The peaks are separated by a frequency of 1.77 GHz and the entire Doppler-broadened, hyperfine structure of the D_2 transition extends over a bandwidth of about 3 GHz.

To obtain the most efficient use of laser energy, its linewidth must be broadened to cover this absorption spectrum, necessitating some form of temporal modulation. For example, if the laser output consists of a train of micropulses, each of width t_o, then the corresponding spectrum is spread over a bandwidth of $1/t_o$. Such an output, depicted in figure 7.41 is obtained by mode locking the laser. In this process, the resonant modes in the laser cavity, which normally have random phases, are forced to have the same phase. If the number of cavity modes is N and the intensity of each mode is I_0, then the total radiance for random phases is NI_0. When the modes are locked, the electric fields add coherently, producing a peak radiance of $N^2 I_0$ The output then takes the form of a train of micropulses of width $2l/Nc$ and spacing $2l/c$, where l is the length of the cavity. The value of N may be 1000 or more, so micropulse widths measured in picoseconds are obtainable. The macropulse generated by the laser may contain several thousand micro-pulses.

For excitation of sodium, Bradley [1992] has found that the optimum micropulse width t_o is in the 0.5–0.7 ns range, producing a transform-limited bandwidth of 1.5–2 GHz, which is less than the full Doppler-broadened spectrum of the D_2 line. The optimum time between pulses is about 2/3 the mean lifetime of the excited state, which for sodium is 16 ns, giving a micropulse rate of 100 MHz. If the pulse rate were constant, then the envelope of the frequency spectrum would contain A' fixed frequency components, each with a very narrow bandwidth determined by the macropulse repetition rate, which does not normally exceed about 1 kHz. Only a small area of the Doppler-broadened absorption spectrum is therefore covered.

To obtain a better match to the atomic absorption spectrum, the pulse output may be phase modulated. The multiline structure of a mode-locked laser can be broadened by using relatively weak modulation, at a fixed frequency, to fill in the gaps between the spectral lines. For the 100-MHz micropulse rate considered by Bradley, the optimum phase-modulation frequency was 30 MHz, with a repetition time of 10 pulse intervals; that is, 100 ns.

In the case of a laser emitting a train of intermediate-duration pulses, such as a copper vapor laser, the pulse spectrum is intrinsically much narrower than the Doppler-broadened sodium D_2 line spectrum. For example, a laser generating pulses of 32 ns width at a repetition rate of 26 kHz produces a spectrum with a width of 30 MHz, densely filled with lines spaced at 26 kHz. The requirement in this case is to broaden the entire spectrum rather than to fill it in, because the lines are already closely spaced. For this purpose, wideband frequency modulation can be used to obtain a better match to the absorption spectrum of the sodium atom.

The interaction of light with sodium atoms depends greatly on the polarization of the incident radiation. The effect of polarization on the optical transitions in sodium is illustrated in the energy-level diagrams shown in figure 7.10. With linear polarization, figure 7.10(A), optical pumping causes a problem when a practical (moderately saturating) level of illumination with a narrow spectrum is used [Morris 1994]. For example, if the laser is tuned to the higher absorption peak, corresponding to the F' = 2 ground state, then only atoms in this state will be excited. Excited atoms decay in 16 ns to both ground states in the ratio 3:5, as described in section 7.2.4. As a result, the population of atoms is rapidly pumped to the F' = 1 state, where they remain inaccessible until a collision occurs. This state has been referred to as a "black hole" by Kibblewhite, Vuilleumier et al. [1994]. In the atmospheric sodium layer, the mean time between atomic collisions is about 100 μs, which also happens to be the longest pulse length that can be used, so atoms pumped to the F' = 1 state will not, on average, be

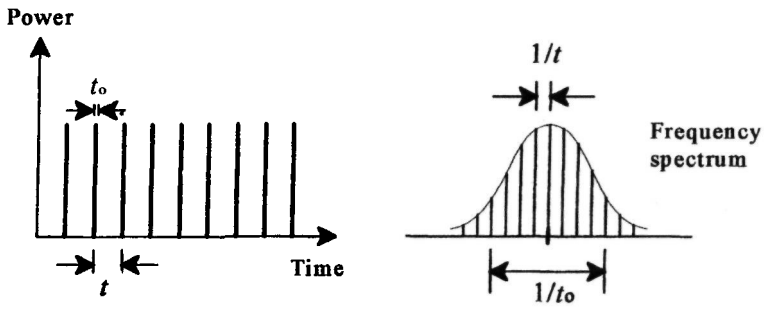


Figure 7.41 Mode-locked laser pulse characteristics.

re-excited during the same laser pulse. Merely lengthening the pulse produces little additional backscatter. When using linear polarization, it is therefore essential to avoid optical pumping by broadening the laser spectrum to cover both of the sodium ground states, which thus requires a spectral width of about 3 GHz.

The situation is quite different when the incident radiation has circular polarization. In this case, when an atom absorbs a photon, it gains one unit of of angular momentum, as indicated in figure 7.10(B). During a long pulse (longer than the spontaneous emission time of 16 ns), the ground-level population shifts toward states having higher M_F values, until it is mostly concentrated in ground state 2,2, marked as "a" in figure 7.10(B). The only permitted optical transition is then to the excited state 3,3, marked as "b." Optical pumping, in this case, is beneficial, as the spontaneous emission from 3,3 to 2,2 is a dipole radiation pattern peaked in the backscatter direction. The "a" to "b" transition is therefore highly efficient. To obtain the most rapid pumping, a wideband pulse is required to cover all Doppler-shifted transition frequencies. Morris [1994] shows that when using the appropriate wideband frequency modulation, circular polarization of the laser produces up to 30% more emission than does linear polarization, with the maximum enhancement ocurring at an irradiance of near 20 $\mathrm{kW\ m^{-2}}$.

7.6.2 Lasers for Sodium Beacons

Two types of laser have been used to generate sodium beacons at 589 nm: (1) tunable dye lasers and (2) solid-state sum-frequency lasers using Nd:YAG. Dye lasers may be pumped by flashlamps or by other lasers, such as copper vapor, argon ion, or Nd:YAG. Sum-frequency Nd:YAG lasers may also be pumped by flashlamps, but higher efficiency is obtained using diode lasers. The macropulse characteristics of each type is determined by the pump laser.

Dye lasers pumped by copper vapor lasers produce short pulses (10–40 ns) at pulse rates of several kilohertz. To obtain an output suitable for the excitation of sodium beacons, these pulses must be stretched in time and broadened in bandwidth. Continuous-wave operation is obtained using argon-ion pump lasers. Solid-state lasers, such as Nd:YAG, have the advantage that by using mode-locking techniques, they can be made to produce their output as a train of very short micropulses, with a spectral distribution matching that of the Doppler-broadened sodium atom.

Characteristics of lasers that have been used or are planned for sodium beacon generation are listed in tables 7.7 and 7.8. The operating principles of the main types will now be described briefly.

7.6.3 Continuous-Wave Dye Lasers

Continuous-wave (CW) lasers may be used for the excitation of sodium beacons in the bistatic configuration, using spatial discrimination to reject the Rayleigh scattering. Because of saturation, the useful CW illumination intensity is limited to about $64\ \mathrm{W\ m^{-2}}$ [Jeys 1991]. Under good seeing conditions, the laser may be focused to produce a spot as small as 0.25 m in diameter. Allowing for atmospheric transmission of 0.6–0.8, the maximum CW laser power that can be used is therefore 4–5 W.

A commercial laser, the Coherent Ring Dye Laser, Model 899-21, is capable of producing about 3 W of power at 589 nm when pumped by a 25-W

Table 7.8 Sodium Beacon Lasers

Reference	Jacobsen et al. 1994	Jacobsen et al. 1994	Friedman et al. 1994	Kibblewhite, Vuilleumier et al. 1994
Organization	CAAO, MMT	CAAO, MMT	LLNL, Lick Observatory	MIT/LL, UCh (ChAOS)
Year operated	1993	1994		
Type	Coherent ring dye	Standing-wave dye	Dye	Nd:YAG sum-frequency
Pump	Argon ion laser, 25 W	Argon ion laser, 13 W	Doubled Nd:YAG	Diode laser
Average power, W	2.7	1.7	30	10
Repetition rate, Hz	CW	CW	10,000	400
Pulse energy, mJ	—	—	3	25
Macropulse width, μs	—	—	0.15	
Micropulse width, ns	—	—	—	0.5
Spectral width, GHz	0.01	0.15	2.0	2

MMT, Multi Mirror Telecope (MT Hopkins, AZ); LLNL, Lawrence Livermore National Laboratory; MIT/LL, Massachusetts Institute of Technology, Lincoln Laboratory; UCh (CHAOS), University of Chicago, Chicago Adaptive Optics System.
CW, continuous wave: CAAO, Center for Astronomical Adaptive Optics, University of Arizona.

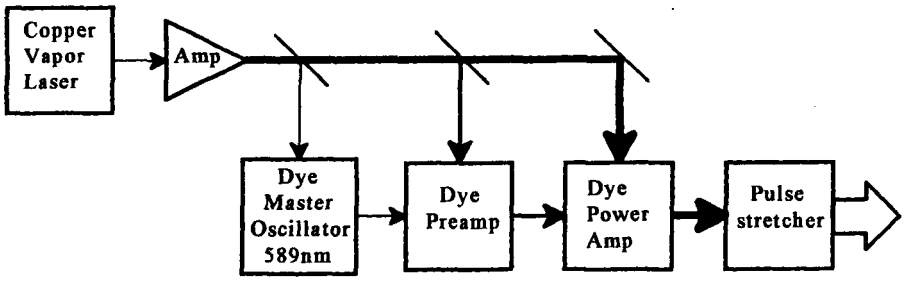


Figure 7.42 Schematic of copper vapor pumped dye laser (at Lawrence Livermore National Laboratory).

Coherent Model I-425 argon ion laser [Jacobsen et al. 1994]. The power consumption of argon ion lasers of this size is on the order of 50 kW. The dye used is Rhodamine 6G in a solution of ethylene glycol.

7.6.4 Copper Vapor Pumped Dye Lasers

This method of pumping has been developed at Lawrence Livermore National Laboratory, where a high-power (1500-W) copper vapor laser was built for the Atomic Vapor Laser Isotope Separation (AVLIS) program. This laser is a fixed installation and can be used only at the Livermore site. A schematic diagram of the system is shown in figure 7.42. It has been used for experiments to determine the behavior of the sodium layer at high power levels [Avicola et al. 1994].

The copper vapor laser is not well suited as a pump for astronomical sodium beacons because of its high pulse rate of 26 kHz and its very short pulse width of 32 ns. To avoid saturation of the sodium layer and to make more efficient use of the laser power, the pulse must be stretched by a factor of 16 and the spectral bandwidth increased by wideband phase modulation, as described in section 7.6.1.

7.6.5 Doubled-YAG Pumped Dye Laser

This type of laser has been specially designed for use on astronomical telescopes [Friedman et al. 1994, 1995] and operates at power levels up to 30 W. The initial application was to generate a sodium beacon for adaptive optics on the 3-m Shane Telescope at Lick Observatory. There are several interesting features in the design, which is shown schematically in figure 7.43. The dye laser is similar to that developed for the copper vapor pump. Solid-state lasers are more suitable for the power levels needed for astronomical sodium beacons, and, in this case, two 70-W, Q-switched, doubled-YAG lasers are employed to pump the dye. Each laser dissipates about 8 kW, which is removed by water cooling. The YAG lasers are themselves pumped with CW flashlamps.

The pulse repetition rate is 10 kHz, with a pulse width of 150 ns. The pulse width is a compromise between the requirement for sodium excitation, which favors a longer pulse to avoid saturation, and the need for efficient conversion in the dye, for which shorter pulses are optimum. Each pump laser head has three fiber outputs, one for the dye master oscillator (DMO), which is located in the equipment

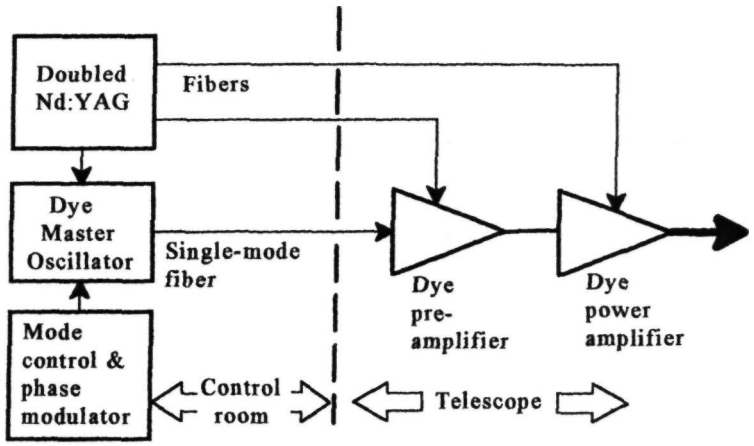


Figure 7.43 Schematic of Nd:YAG pumped dye laser with fiber optic coupling.

room, and one each for the dye preamplifier and power amplifier, which are mounted on the telescope. The DMO consists of the dye laser cavity, single-mode control, frequency control, and phase-modulation functions. Its coherent output is carried on a single-mode fiber, at a power level of a few milliwatts, to the dye amplifiers on the telescope. The pump power is carried on multimode fibers, which are about 70 m in length.

The installation on the telescope consists of three optical tables as shown in figure 7.44. The laser table containing the dye preamplifier, power amplifier, and beam cleanup components is mounted at the lower end of the telescope, adjacent to the primary mirror. A high-speed tip-tilt mirror, driven from the adaptive optics system, stabilizes the laser beam in the optical field of view, compensating for jitter in the outgoing beam plus the residual anisoplanatic wavefront tilt errors due to physical separation between the telescope aperture and the laser projector. The apparent position of the sodium beacon conveys no useful information on where the telescope is pointed, but it is desirable to keep it near the optical axis of the telescope to facilitate measurement of high-order wavefront distortion. There is also provision for a deformable mirror and wavefront sensor on the laser table to allow the laser beam quality to be corrected before projection.

The diagnostics table, containing the beam-pointing and -centering functions and the fast shutter, is mounted about 7 m up the telescope tube. Pointing and centering of the beam is accomplished using two adjustable mirrors in a dogleg configuration. Sensing of the beam position is implemented with two CCD cameras. The status of the laser beam is transmitted to the control computer in the equipment room, where it is visually monitered by the operator. This table also contains the first (negative) component of the afocal laser beam expander and a last shutter to cut off the laser beam in an emergency.

The objective component of the laser projector is mounted near the top of the telescope tube. For the Lick 3 m telescope, the projection aperture is 20 cm, equivalent to about twice the average r_0 value at the site. The entire beacon laser system is controlled and monitored by a Mac Quadra computer using Labview software.

7.6.6 Sum-Frequency Nd:YAG Laser

This type of laser is made possible by a fortunate coincidence of nature. The sum frequency of the two Nd:YAG lines of 1.064 and 1.319 μm produce a wavelength of 0.589 μm, which is exactly the sodium D_2 line. A schematic is shown in figure 7.45. The outputs of two Nd:YAG master oscillators tuned to 1.064 and 1.319 μm are amplified and then superimposed with a dichroic mirror. The beams are then applied to a nonlinear crystal (lithium triborate) which produces the sum frequency corresponding to 0.589 μm.

The solid-state sum-frequency laser is well suited to the task of generating sodium beacons. A duty cycle of 10–15% can be achieved, which is near optimum for this application, allowing macropulse lengths of 100–150 μs at a repetition rate of 1000 Hz. By mode locking the master oscillators, the output is obtained as a train of sub-nanosecond micropulses with a spectral width of 2–3 GHz, matching the Doppler-broadened D_2 line of sodium. These short pulses also produce efficient conversion in the nonlinear crystal.

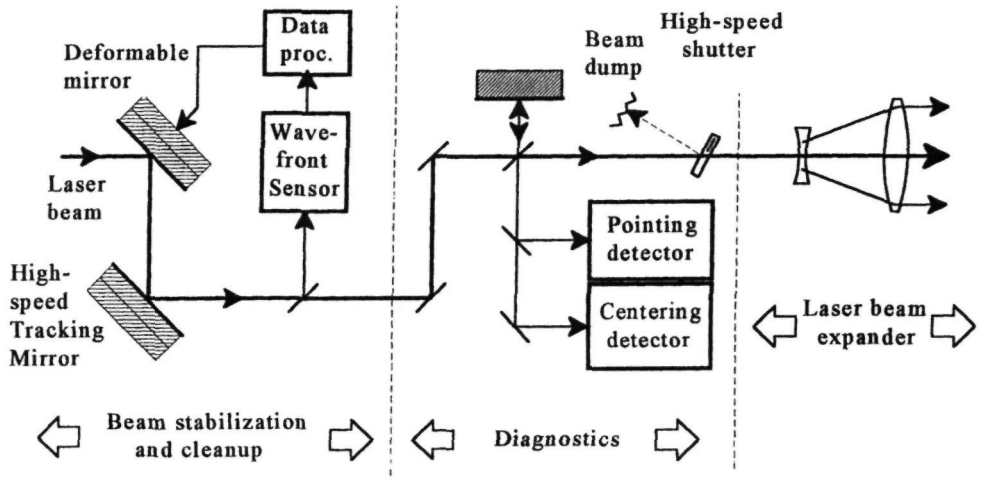


Figure 7.44 Schematic of laser beam control and diagnostic system.

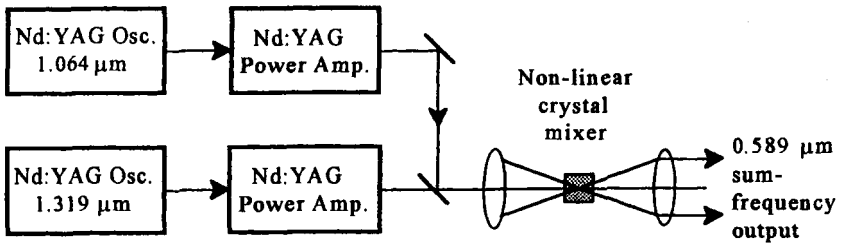


Figure 7.45 Schematic of Nd:YAG sum-frequency laser.

The first laser of this type was built at the MIT Lincoln Laboratory [Jeys et al. 1989]. It was pumped by flashlamp, giving an average power of 5 W at 10-Hz repetition rate, and was employed for sodium layer measurements. A more powerful version, also using flashlamps, producing 10–20 W average power at 840-Hz repetition rate, was built for Phillips Laboratory Starfire Optical Range [Jeys 1992, Jelonek et al. 1992].

For average laser powers of more than 20 W, diode lasers are more attractive than flashlamps for pumping. Diode lasers are efficient, compact, and produce little electrical noise. They also produce their output at near-IR wavelengths for which YAG lasers have strong absorption bands. Kibblewhite, Vuilleumier et al. [1994] have described a 10-W diode-pumped sum-frequency Nd.YAG laser being built by the MIT Lincoln Laboratory for the University of Chicago Adaptive Optics System (ChAOS) program. At the time of writing (1997), diode lasers are expensive and have limited lifetime. However, the reliability of these devices is expected to improve over the next few years.

To summarize the current developments in lasers for sodium beacons, there are two main approaches: (1) dye lasers using solid-state laser pumps; (2) sum-frequency Nd:YAG lasers using diode laser pumps. The trend is to reduce the quantity of hardware that must be mounted on the telescope to the absolute minimum, and to locate power-dissipating components away from the dome. Sum-frequency lasers are well-suited for this application as they are physically compact and have relatively high power efficiency.

7.7 Sodium Beacon Measurements

The characteristics of the mesospheric sodium layer have been explored with lidar techniques since the late 1960s and observations have been made from locations in most parts of the globe. The basic composition and structure of the sodium layer is now well known. When the sodium layer is used to generate laser beacons for astronomy, specific requirements are called upon; these need to be verified in the context of adaptive optics. Wavefront sensors require compact reference sources that emit a large photon flux, to enable measurements to be made at rates of at least several hundred hertz. The laser beacons must have small angular size and good stability. The behavior of the sodium population under strong incident illumination from short laser pulses with high peak power must be verified. Methods must be found to maximize the sodium resonance fluorescence return for a given laser power. An important part of the development of lasers for generating adaptive optics beacons is therefore to verify the expected behavior of the sodium layer. This can, and should, be done before the adaptive optics design is finalized.

Several experimental programs have been conducted over the last few years, specifically to obtain this information and to compile a data base on the properties of the sodium layer that are of particular relevance to adaptive optics. In table 7.9, data from three sources are presented: (1) sodium layer measurements made with the Cedar lidar at the University of Illinois during the 1980s [Gardner et al. 1986, Gardner et al. 1990]; (2) Sodium measurements made with a medium-power (10-W) laser at Phillips Laboratory Starfire Optical Range [Jelonek et al. 1992]; and (3) high-power sodium measurements made with a 1-kW laser at Lawrence Livermore National Laboratory [Avicola et al. 1994].

7.8 Optical Configurations for Beam Sharing

Special care must be taken with the optical train when a laser beacon is projected through an astronomical telescope. The alternative approaches of bistatic and central projection have been discussed in section 7.3.6. The main problem with beam sharing is the very large ratio between the power level of the transmitted laser beam and the radiation received from the reference sources and science object, a ratio that can exceed 10^{12}. The peak and average power levels of the lasers used for adaptive optics are generally not high enough to cause damage to coatings of transmissive components, but high-level radiation must be prevented from reaching sensitive elements,

Table 7.9 Sodium Layer Measurements

Reference	Gardner et al. 1990	Jelonek et al. 1992	Avicola et al. 1994
Organization	UIll	PL/SOR	LLNL
Date of measurements	1980-86	Jun. 1991–Feb. 1992	Aug. 27, 1992
Sodium layer			
Height, km	92 ± 3	95.5 ± 3	
Thickness, km (FWHM)	10	16	
Temperature, K, summer	130-150	—	
winter	210-230	—	
Laser system			
Average power, W	5	10	1100
Pulse energy, mJ	25	12	40
Repetition rate, Hz	200	840	26,000
Pulse width, ns	25	6500/0.35	32
Line width, MHz	1200	3000	2700
Projection aperture, m		0.05 diameter	0.04 x 0.08
Beam divergence, mrad	1		
Peak power, kW	1000		
Polarization		Linear and circular	
Measurements			
Atmospheric transmission			0.6
Coherence length r_0, m			0.11
Spot angular size, μrad		36	22
Spot times λ/D		3.0	1.5
Spot physical size, m		3.25	2.0
Column abundance, m~	$3 - 10 \times 10^{13}$	2 - 3 x 10"	4.1 x 10^{13}
Circular/linear ratio		1.45	

UIll, University of Illinois; PL/SOR, Phillips Laboratory, Starfire Optical Range; LLNL, Lawrence Livermore National Laboratory.

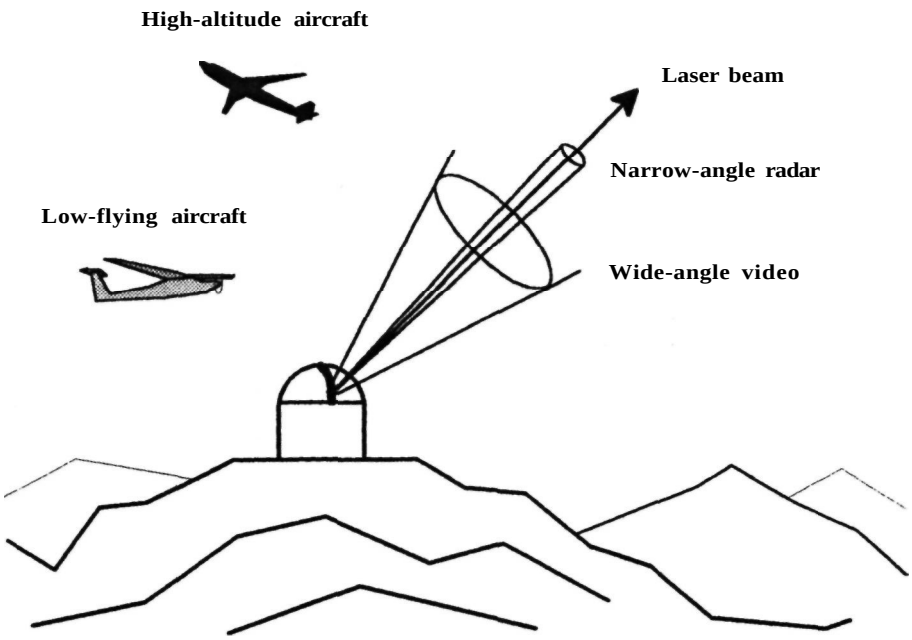


Figure 7.46 Laser beacon safety precautions. The narrow-angle radar detects high-altitude aircraft. A wide-angle video monitor is used to detect low-flying aircraft.

such as detector focal planes, which are easily blinded.

The measures required involve temporal gating with mechanical and electronic shutters, together with spectral filtering. The on/off ratio of solid-state shutters, such as Pockels cells, may be insufficient to prevent light leakage, in which case mechanical back-up is required. When relying on temporal shuttering, care must be taken to eliminate fluorescence in optical components and coatings, as this may persist for some time after a laser pulse.

7.9 Laser Safety Considerations

The lasers required to produce adaptive optics beacons generally launch a beam that exceeds eye safety levels. When using laser beacons, it is necessary to employ a fail-safe method of turning off the laser beam before an aircraft can pass through it. Because of the large range of flight altitudes, it has been found effective to use two detection systems, one for high-altitude aircraft, mostly commercial flights, and one for low-altitude aircraft, mostly privately operated.

A narrow-angle radar aligned with the laser beam has been found to be an effective method for detecting high-altitude aircraft, with the warning system linked to air traffic control. Low-level aircraft are more of a problem because of their high angular velocity and more random flight patterns, necessitating a wide field of view. Human spotters have been employed for this purpose, but their work is expensive and not entirely reliable. The use of a low-light-level, wide-angle television system for aircraft detection has been studied at the University of Chicago [Kibblewhite, Wild et al. 1994]. This system is sensitive enough to detect the lights on all legally equipped aircraft. Fail-safe operation is assured by continuously monitoring stars down to a given magnitude. If a fault occurs and stars are not detected, then the laser system is automatically shut down. The elements of a two-level safety system are depicted in figure 7.46.

An automated optical aircraft spotter to detect aircraft lights at night has been described by Cheselka and Angel [1996].

In addition to real-time safety precautions, an additional constraint may be imposed on the maximum angle at which the laser beam may be pointed away from the zenith. The Federal Aviation Administration (FAA) currently limits this angle to 30°.

8 Wavefront Reconstruction and Control Systems

8.1 Introduction

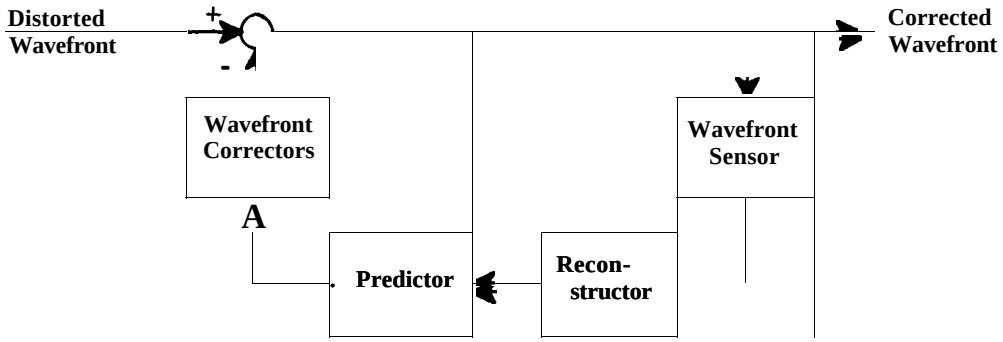
8.1.1 Closing the Loop

The control system is the vital link between the wavefront sensor and the wavefront corrector in all adaptive optics systems. Its function is to convert the wavefront measurements made by the sensor into a set of commands that are applied to the wavefront corrector to satisfy a suitable performance criterion, such as minimizing the residual wavefront error. The main functions in a typical adaptive optics control loop are shown in figure 8.1. The wavefront sensor generates an array of wavefront slope or curvature measurements that are first reconstructed spatially into a continuous two-dimensional representation of the wavefront error. The wavefront predictor minimizes temporal noise by smoothing the data in time and makes the best estimate of the wavefront error for the next correction cycle. The commands to be applied to the wavefront corrector are then computed, taking into account the spatial and temporal characteristics of the wavefront disturbances and the correction system.

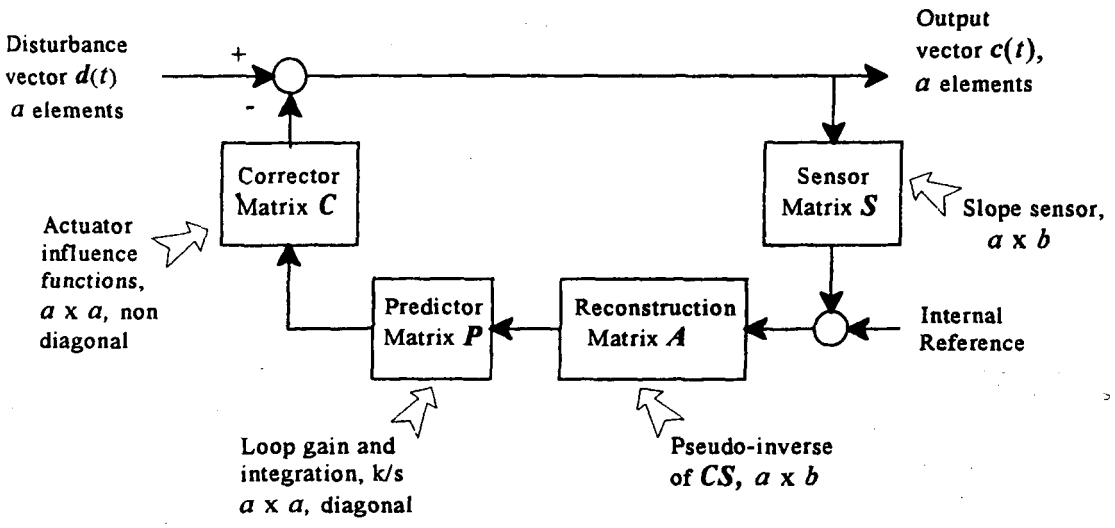
The practical approach to the design of adaptive optics has generally been to divide the system into individual functions that are analyzed and optimized independently. This approach was justified in the early days of adaptive optics when technology was rudimentary. The real-time atmospheric compensator (RTAC) and Compensating Imaging System (CIS), built in the 1970s, employed zonal wavefront sensors and deformable mirrors in which there was a one-to-one relationship between the sensor subapertures and the corrector actuators, enabling the simplest type of wavefront reconstructor to be used. The gain and temporal bandwidths of all control channels were uniform. The only method of real-time optimization available was to adjust the common bandwidth of the feedback loop to minimize the mean-square error in response to variations in the intensity of the reference source and turbulence strength. A simplified model of the reconstruction function, in which the wavefront was represented as an array of points or nodes, connected by linear gradients, was used.

A major advance was made by Wallner [1983], who described the concept of optimal wavefront correction, in which adaptive optics systems are optimized as a whole, taking into account not only all the hardware components, but also the characteristics of the reference source and the wavefront disturbances. Only in this way is it possible to realize the full potential of the technology. Wallner's approach has been expanded by Welsh [1991] to include temporal factors, and by Johnston and Welsh [1994] to include multiconjugate compensation. While fully optimized systems have not yet been realized in practice, a knowledge of the ultimate performance that can be achieved provides a yardstick against which real adaptive optics systems can be judged. The practical goal is to devise hardware and software that approach the performance of ideal systems, but with minimum cost and complexity.

The dominant function in most adaptive optics control systems is wavefront reconstruction, which



(a) Block diagram



(b) Multivariate control system model

Figure 8.1 Main functions in an adaptive optics control loop. The basic building blocks are shown in (a). These functions are modeled by the multivariate control system shown in (b).

involves a large amount of data processing that must be executed in real time. This section contains an overview of the requirements and evolution of wavefront reconstruction and the methods employed to implement it. The principles and practice of wavefront reconstruction are described in more detail in sections 8.2 and 8.3. Methods of wavefront prediction are discussed in section 8.4. Design considerations for the closed-loop control systems used in adaptive optics are outlined in section 8.5. To conclude this chapter, the principles of optimal wavefront correction are described in section 8.6.

8.1.2 Why Is Reconstruction Needed?

Wavefront reconstruction is the process of restoring the absolute phase values that are lost when local wavefront slope or curvature is measured. Most wavefront sensors make their measurements within an array of independent zones covering the optical pupil. If only the local wavefront slope is corrected, the images from each zone would add incoherently. The angular resolution obtainable in that case would be no better than that of a single subaperture, or Xjd. Wavefront reconstruction restores the overall phase relationship between the zones that is essential to obtain the full angular resolution of the aperture,

equal to λ/D. The reconstruction process assembles the individual slope measurements into a continuous three-dimensional representation of the wavefront, each element having the correct absolute phase. When the zones are correctly sized and phased, adaptive optics is capable of producing a potential improvement of D/r_0 in angular resolution.

Many other useful data-processing functions and manipulations can be combined with the basic function of phase restoration. The tasks performed by an adaptive optics wavefront reconstructor are summarized as follows:

1. estimation of the overall wavefront values that best fit the measured wavefront data, using either zonal or modal components;
2. removal of unnecessary components, such as piston, from the reconstructed wavefront;
3. estimation of low-order wavefront modes, such as the overall tilt and defocus components, which can be compensated separately;
4. removal of cross-coupling effects caused by actuator influence functions, which would otherwise cause instability in the feedback loop;
5. improvement of the noise performance by using the statistics of the wavefront being corrected in the reconstruction process;
6. minimization of the error in the reconstructed wavefront by accounting for the signal-to-noise ratio of each gradient input;
7. accommodation of different actuator geometries for the gradient sensor and the corrector.

8.1.3 *Evolution of Real-Time Reconstruction*

The wavefront reconstruction function is a potential bottleneck in large adaptive optics systems using zonal wavefront measurement. Wavefront variations over optical apertures covering several square meters are measured and corrected at high speed using an array of parallel optical channels. At the output of the wavefront sensor, these data are converted to electrical signals that must be processed in real time. The number of arithmetic operations required to perform the reconstruction in each time interval is proportional to the square of the number of independent correction channels. An adaptive optics system with 300 actuators operating at a sampling rate of 1000 Hz requires a reconstructor operating at a data rate of about 100 megaflops. Adaptive optics systems of even modest size consequently require a considerable data-processing capability.

Great efforts were made in the 1980s to improve the speed and capability of reconstructors, but little of this work has been published and no comprehensive review has appeared in the literature. This overview is intended to provide a concise summary of the main developments in wavefront reconstruction as they apply to astronomical adaptive optics. 1 hope that it will serve as an introduction to the material in this chapter, enabling the various approaches to be assessed and placed in perspective.

A two-dimensional wavefront may be reconstructed by summing the gradients over an integration path that includes all subapertures in the array. If there were no noise on the gradient measurements, then all integration paths would give the same result and wavefront reconstruction would be a trivial process. However, the gradient measurements inevitably include shot noise due to finite photon counts, as well as electron noise added in the detection process. If there are more gradient measurements than evaluation points or "nodes" in the network, as is usually the case, then it is possible to use the redundant information to reduce the noise on the estimated wavefront. The criterion normally used is to minimize the mean-square error between the reconstructed wavefront and the measured gradients; that is, to find the wavefront that best fits the measurements in the least-squares sense. Mathematically, this is achieved by solving a set of simultaneous equations, or, equivalently, by inverting the matrix describing the relation between the wavefront gradients and the wavefront nodes. The reconstruction process therefore reduces to a matrix multiplication.

The development of wavefront reconstructors for adaptive optics has always been dominated by the need to process large quantities of data in real time. The input data (wavefront slopes) and the output data (actuator drive commands) are both analog quantities, so it is logical to use analog processing. The first real-time wavefront reconstructors, used in adaptive optics systems designed in the 1970s, used parallel analog processors to perform the calculations. The reconstruction problem was reduced to manageable proportions by using simple reconstruction networks that could be defined by sparse matrices, enabling the use of iterative processing. Digital computation was primitive at that time.

A major breakthrough in reconstruction technology occurred in the 1980s when fast and inexpensive digital multiplier-accumulator chips became available, leading to the development of "brute force" reconstructors that used parallel data processors to perform the matrix multiplication operation. Since then, the technology of wavefront reconstruction has improved in two ways: (1) the speed of digital processors has increased to the point where multiplication of large matrices, corresponding to apertures containing about 250 subapertures, can be accomplished in less than 100 ms; and (2) the reconstruction process itself has been optimized to include characteristics of the adaptive optics system and of the atmospheric turbulence, improving the accuracy of compensation.

Although "brute force" wavefront reconstruction using parallel digital processors is feasible, it is still a very expensive solution, especially in systems that

have a large number of subapertures for which the reconstructor can be a major item in the budget. There is still a need for "economical" reconstructors that provide the flexibility and performance of full matrix multiplication, without the current high cost. It is possible that the continuing increase in speed (clock rate) of digital processors will gradually reduce the cost and complexity of parallel processors. There are, however, other useful approaches to wavefront reconstruction, including iterative and hybrid processors. Most adaptive optics systems require digital-to-analog conversion to drive the wavefront corrector, so hybrid processors (using both digital and analog components) may be an economical solution in some systems.

8.1.4 Aspects of Wavefront Reconstruction

To get a clearer picture of the various issues and approaches that are involved in wavefront reconstruction, it is useful to consider three different aspects of the technology:

1. the functional models employed, which define the operating principle and overall capabilities of the reconstructor;
2. the methods of calculation, which define the mathematical operations involved;
3. the practical realizations that have been employed in adaptive optical systems.

The main ideas that have been used during the development of wavefront reconstruction for adaptive optics are summarized in table 8.1.

8.1.5 Basic Models

The wavefront reconstruction process was initially formulated to reduce data from static shearing interferograms, using a simple network in which the wavefront is evaluated at a rectangular array of single points or *nodes* that are connected by gradient vectors [Rimmer 1974]. The gradients were obtained by scanning laser interferograms manually. An iterative (relaxation) method was used to calculate the wavefront. The first real-time wavefront reconstructor [Hardy et al. 1974, 1977] employed a similar reconstruction model, using analog processing to achieve the necessary speed. The adaptive optics system used a shearing interferometer in which the wavefront slope measurements were averaged over finite areas of the pupil, together with a deformable mirror that had an influence function also extending over a finite area. In spite of these approximations, the simple reconstruction model (shown as network 1 in figure 8.2), using an array of nodes connected by slope vectors, gave satisfactory results. In this model, the x-axis and y-axis gradients are measured in two separate sets of subapertures, displaced from each other.

With the development of the Shack-Hartmann sensor, a different sensor geometry was needed because the x-axis and y-axis gradients are measured in the same subapertures. This geometry is shown as network 2 in figure 8.2. The gradient measurements can be resolved into two 45° components, connecting diagonal nodes. It is seen that this leads to two separate, interlaced networks. Zonal slope measurements do not provide the data to connect these networks, so the data must be inferred from a priori information.

In network 3, also shown in figure 8.2, the measurement subapertures are centered on the nodes rather than being displaced between them, as in the other two networks. This arrangement is not often used. It makes an adaptive optics system more difficult to calibrate, because deflecting or "poking" one node (actuator) produces no gradient in the corresponding sensor subaperture.

None of the above networks provides an optimal solution to the reconstruction problem. Parameters such as the structure function of the turbulence-degraded wavefront, the noise statistics of the gradient measurements, and the response functions of the

Table 8.1 Overview of Real-Time Wavefront Reconstruction Techniques

Reconstruction Models	Methods of Calculation	Practical Realizations
‣ Simplified network models Nos 1, 2, and 3 (figure 8.2)	‣ Iterative methods Jacobi Gauss-Seidel SOR"	• Iterative techniques Analog Digital Hybrid
Partially optimized reconstructors	Exponential	• Serial computation
	Matrix inversion	• Parallel processor (multiplier accumulator)
‣ Optimal reconstructor		

" SOR, successive overrelaxation.

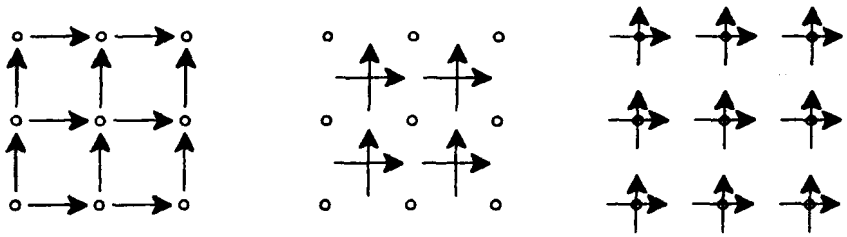


Figure 8.2 Three models for the reconstruction of wavefront points from gradient measurements.

wavefront corrector can all be used to improve the estimate of the wavefront. The advantages of optimal wavefront estimation were first pointed out by Wallner [1982, 1983]. The optimal reconstruction process may be defined as a matrix relating every point in the wavefront to every measurement made by the wavefront sensor. This is the most general type of wavefront reconstructor and is the preferred approach, even though not all of the collateral functions may be implemented. Quantities, such as the wavefront coherence length and noise covariances, change with time, making it desirable to update the reconstruction matrix in real time according to operating conditions. Adaptive reconstruction systems using dynamic optimization represent the state of the art in this technology.

8.1.6 Methods of Calculation

Iterative methods are useful for solving sparse arrays. A local algorithm is applied sequentially to each node of the array, the process being repeated until the solution settles to within the required tolerance. The Jacobi and Gauss-Seidel methods are similar, the latter converging more rapidly because the solution is updated after each application. The successive overrelaxation (SOR) method is more optimal because its rate of convergence can be adjusted by a relaxation parameter. The exponential reconstructor (section 8.2.3) is a completely different iterative approach, consisting of two steps. First, the overall structure of the array is determined by combining the gradient measurements to enlarge the subapertures by powers of two, after which the details are reconstructed by halving the dimensions at each step. The main advantage of iterative methods is that a solution can usually be obtained with fewer arithmetical operations than by an equivalent matrix multiplication. A disadvantage of iterative methods is their lack of flexibility in tailoring the algorithm to accommodate different conditions.

Matrix inversion is the most rigorous method of wavefront reconstruction, in which a least-squares (or other) solution is obtained by direct calculation. The matrix relates the measured gradients to the wavefront nodes. To determine the nodal values, the matrix must be inverted. This is not a trivial operation, as it usually involves extraction of the overall phase (piston) and tilt components, as described in section 8.2.2. To obtain an optimal solution, the matrix may also include additional factors, such as wavefront and noise covariances, as well as the influence function of the actuators on the wavefront corrector. The inverted matrix is stored for use, as required, to reconstruct the measured gradients. Several matrices covering different conditions may be inverted and stored, it being relatively easy to switch between them in the data-processing operation.

8.1.7 Practical Realizations

As mentioned above, the first real-time wavefront reconstructor used in adaptive optics employed an analog processor, implementing an algorithm similar to the Jacobi method. Using an electrical network, the rate of iteration was extremely high, resulting in a settling time of a few microseconds. This type of reconstructor employs local algorithms, which have a limited capability for optimization, although it is possible to include functions covering a moderate area around each node, such as the influence function of the wavefront corrector.

All of the iterative methods may be implemented digitally, offering somewhat greater flexibility than that with analog implementation. The saving in data-processing operations is most apparent with simple reconstruction algorithms, such as those of networks 1, 2, and 3. While it is possible to use iteration to solve matrices of any complexity, there is no practical advantage in doing so when the algorithm involves a large number of nodes, corresponding to a dense matrix, as occurs with optimal reconstructors.

Hybrid reconstructors combine the best features of analog and digital techniques and may be applicable where digital-to-analog conversion is necessary in any case; for example, in the drive circuits for a

wavefront compensator. In such cases, the data Formatting and multiplication functions are best performed digitally, while the summation at each node is accomplished analogically.

With the availability of high-speed parallel processors, it has become possible to multiply the dense matrices, typical of optimal reconstruction, in real time. The standard approach is to use a parallel array of multiplier-accumulators, which are available as special-purpose integrated circuits. These devices operate at clock rates as high as 30 MHz. The inverted matrix elements that define the "weights" to be used in the multiplication process are stored in a random-access memory (RAM). To implement the adaptive feature of modern reconstructors, several different (inverted) matrices may be computed in advance and stored ready for use, to be accessed as required in the reconstruction process.

8.1.8 Reconstruction Errors

The wavefront reconstructor is only one element in the chain of operations that determines the wavefront error of an adaptive optics system. The optimization process involves the performance of the entire system, so, with a fully optimized system, errors are normally contributed by several components. Wavefront reconstruction does not normally produce wavefront errors directly, but contributes indirectly in two ways: (1) the wavefront error due to noise on the gradient measurements is modified by the reconstruction process; and (2) the time required to perform the reconstruction contributes to the time delay error in the system. The reconstruction time can usually be made a small part of the total delay time or latency in the system. However, the measurement noise is usually a major contributor to the total error budget, so the noise factor or error propagation of wavefront reconstructors is of considerable interest.

The error propagation of a reconstructor can be determined only when it is modeled realistically in a system context. Using simple stand-alone reconstruction models, such as networks 1, 2, and 3, it is found that the noise propagation depends not only on the size of the network, as expected, but also appears to depend on the exact configuration used, being greater than unity for network 2 and less than unity for the others. Wallner [1982, 1983] showed that this apparent difference disappears when the reconstruction process is realistically modeled. It has been shown from basic principles and by simulation that the error propagation factor varies as $\sqrt{N^2}$, where N^2 is the total number of nodes in the array. For the array sizes used in astronomical adaptive optics (up to 1000 actators), the error propagation for slope sensing systems is less than unity.

8.2 Principles of Wavefront Reconstruction

8.2.1 Simplified Network Models

Wavefront reconstruction is basically a spatial integration process that converts an array of independent slope or gradient measurements into a smoothly varying three-dimensional surface defining the wavefront errors to be corrected. Wavefront gradient measurements are inevitably perturbed by random noise, because of the quantum nature of light and the addition of electrons in the detection process. Because there are multiple gradient paths between any two points on the wavefront, there is no unique wavefront that satisfies all the measured gradients exactly, so a statistical solution is employed. The usual criterion is to minimize the mean-square error between the reconstructed wavefront and the individual gradient measurements.

Wavefront reconstruction is required in all adaptive optics systems in which the wavefront sensor makes zonal measurements. The geometrical configuration of the reconstructor (that is, the relation between the nodes at which the wavefront is evaluated and the measurement zones) depends on the type of sensor used. Gradient sensors such as the shearing interferometer and the Shack-Hartmann sensor use different detector configurations. Reconstructor configurations for these sensors, showing the relation between the subapertures and the measured slopes, are shown in figures 8.3-8.5.

In network 1, the x-axis and y-axis gradients are displaced from each other, enabling all nodes to be interconnected into a single network. The x-axis and y-axis gradient measurements are usually made with two sets of contiguous zones, requiring two detector arrays, as shown in figure 8.3. The performance of this reconstructor was analyzed by Hudgin [1977a]. Network 1 is compatible with both the shearing interferometer and the Shack-Hartmann sensor. It has been used in the Compensated Imaging System (CIS), the Atmospheric Compensation Experiment (ACE), and the Wavefront Control Experiment (WCE), all of which used shearing interferometers and which have been tested on astronomical telescopes (see chapter 10). It was also used in the Massachusetts Institute of Technology (MIT) Lincoln Laboratory's Short Wavelength Adaptive Techniques (SWAT) system, which employed a Shack-Hartmann sensor [Barclay et al. 1992].

The second configuration, network 2, shown in figure 8.4, is used primarily with Shack-Hartmann sensors in which the x-axis and y-axis gradients are measured in identical subapertures. This reconstructor has been analyzed by Fried [1977]. Although it operates with a single detector array, this configuration is composed of two independent

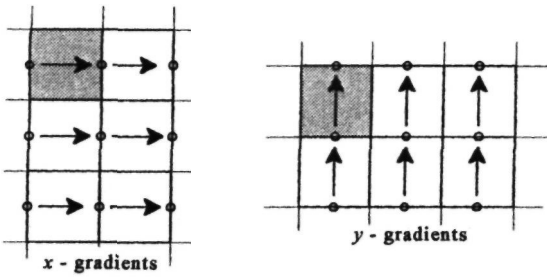


Figure 8.3 Reconstruction network 1 using displaced subapertures for *x*- and *y*-gradient measurements.

grids [Herrmann 1980], as can be seen when the measured gradients are resolved into 45° components, connecting diagonal nodes. The presence of two separate grids necessitates some means of establishing the relation between them. If a displacement exists between the two grids, a checkerboard wavefront pattern is produced, to which the wavefront sensor is insensitive. This is not a trivial matter and the error produced can be difficult to eliminate.

In the third configuration, network 3, figure 8.5, *x*-axis and *y*-axis gradients are, again, measured in the same subapertures, but, in this case, the subapertures are centered on the nodes rather than located between them [Southwell 1980]. In network 3, the gradient connecting two nodes is the average of the two gradients measured in the zones centered on those nodes. Thus, all nodes are connected by gradients into a single network. This arrangement can be considered a modification of that shown in figure 8.3, with the subapertures displaced by half the zone spacing so that *x* and *y* subapertures coincide. There is no record of this configuration having been used in practice, although it appears to be suitable for Shack-Hartmann sensors. It allows the use of a single detector array without producing the unsensed checkerboard mode.

8.2.2 Matrix Representation

The process of wavefront reconstruction and the derivation of the methods used to implement it are most easily explained using a simplified model, as shown in figure 8.6. The wavefront is represented as an array of discrete points or *nodes* on a uniform grid, and the measured gradients are assumed to define the differences between these nodes in the *x* and *y* axes. This model does not account for the fact that, in practical adaptive optics systems, the wavefront measurements are made over discrete contiguous zones in the optical aperture, or the fact that the corrections are likewise made using actuators with influence functions extending over finite areas of the aperture. In addition, the simplified model does not account for the statistics of the wavefront disturbance, such as its spatial frequency spectrum. A more comprehensive approach to the wavefront reconstruction function is described in section 8.6. In the present section, the simplified reconstruction model is analyzed, keeping in mind its limitations.

In figure 8.6 the *nodes* at which the wavefront is evaluated are labeled $w_{i,j}$ where $i = 1, 2 \cdots N$ is the index for the *x* axis and $j = 1, 2 \cdots N$ is the index for the *y* axis. A square array is assumed, having a total of N^2 nodes. The measured wavefront slopes in the *x* axis are labeled $s^x_{i,j}$ and those in the *y* axis are labeled $s^y_{i,j}$ where the indexes refer to the node immediately to the left of, or below, the slope. In the square array, there are $N(N-1)$ *x* slopes and $N(N-1)$ *y* slopes. Use of the term *phase* is deliberately avoided in the context of, wavefront reconstruction, because the wavefront w is defined in terms of optical path difference (meters) and the wavefront sl
that are measured independently of wavelength.

From figure 8.6, it is seen that the relation between the wavefront slopes and node values is

$$\begin{aligned} s^x_{11} &= w_{21} - w_{11} & s^y_{11} &= w_{12} - w_{11} \\ s^x_{21} &= w_{31} - w_{21} & s^y_{12} &= w_{13} - w_{12} \\ s^x_{31} &= w_{41} - w_{31}, & s^y_{13} &= w_{14} - w_{13} \\ \text{etc.} & & \text{etc.} & \end{aligned} \tag{8.1}$$

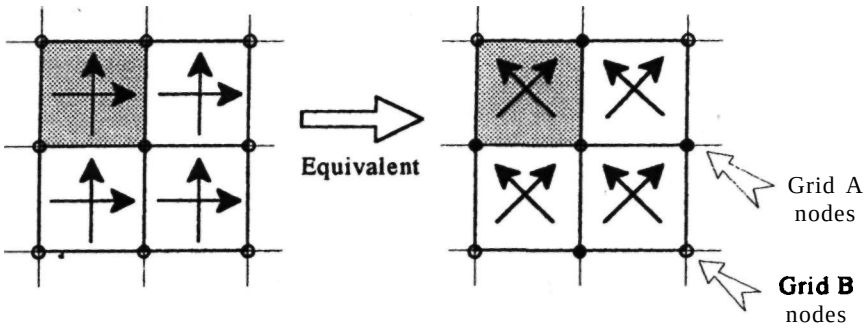


Figure 8.4 Reconstruction network 2, using common subapertures for *x*- and *y*-gradient measurements. This results in two unconnected nodal grids.

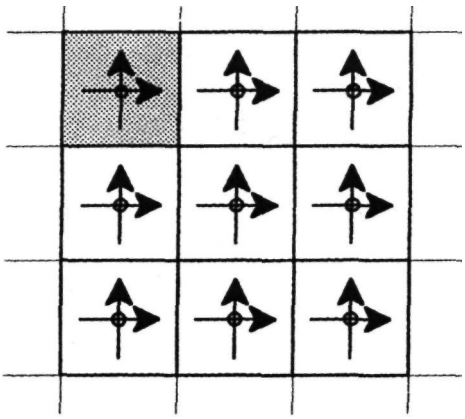


Figure 8.5 Reconstruction network 3, using common subapertures for gradient sensing, centered on the actuators.

This is a set of $2N(N-1)$ equations for N^2 unknowns (the wavefront node values). The number of equations exceeds the number of unknowns so the solution is overdetermined.

A set of simultaneous linear equations can be expressed compactly in matrix form. Representing the gradient or slope measurements as the vector s and the wavefront as the vector **w**, the relation becomes

$$\mathbf{s} = \mathbf{A}\mathbf{w} + \mathbf{n} \tag{8.2}$$

where

A = matrix that depends on the geometry of the nodes and gradients,
n = noise added in the measurement process

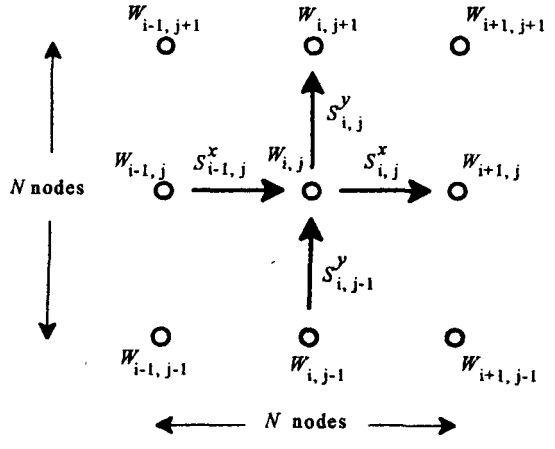


Figure 8.6 Simplified reconstruction model and nomenclature.

The "hat" or circumflex indicates estimated or measured values.

For a square wavefront array using network 1, using the head-to-tail configuration of measured slopes (figure 8.3), the total number of nodes is $A = N^2$. The number of *x* gradients, s^x, is $N(N-1)$ and the number of *y* gradients, s^y, is also $(N-1)N$, for a total of $B = 2N(N-1)$ The number of gradients, *B*, is greater than the number of nodes, *A*, for all values of N greater than 2. The matrix A is rectangular with dimensions *A* x *B* and the matrix values are [Herrmann 1980]

$$\begin{aligned} &\mathbf{A}[q + (p-1)(N-1)q + (p-1)N] = -1 \\ &\mathbf{A}[q + (p-1)(N-1), \ 1 + q + (p-1)N] = 1 \\ &\quad \text{for } p = 1 \ldots N, \ q = 1 \ldots N-1 \end{aligned} \tag{8.3}$$

$$\begin{aligned} &\mathrm{A}[r + B/2, \ r] = -1 \\ &\mathbf{A}[r + B/2, \ r + N] = 1 \\ &\quad \text{for } r = 1 \ldots B/2 \end{aligned} \tag{8.4}$$

All remaining elements have a value of zero.

For a square wavefront array of $A = N^2$ nodes, using network 2 (figure 8.4), the number of *x* gradients, s^x, and *y* gradients, s^y, is each $(N-1)^2$, for a total of $B = 2(N-1)^2$. Here, *B* is greater than *A* for values of N greater than 3. For this configuration, the dimensions of the matrix A are *A* x *B* and the matrix values are

$$\begin{aligned} &\mathbf{A}[m, n] = 1 \\ &\mathbf{A}[m+1, n) = -1 \\ &\mathbf{A}[m, n + B] = 1 \\ &\mathbf{A}[m+1, n + B] = -1 \\ &\mathbf{A}[m, n + NB] = -1 \\ &\mathbf{A}[m+1, n + NB] = 1 \\ &\mathbf{A}[m, n + NB + B] = -1 \\ &\mathbf{A}[m+1, n + NB+B] = -1 \end{aligned} \tag{8.5}$$

where

$$\begin{aligned} m &= 1 + 2(q-1) + 2(p-1)(N-X) \\ n &= [q + (p-1)N]B \end{aligned} \tag{8.6}$$

for

$$\begin{aligned} p &= 1 \ldots N-1 \\ q &= 1 \ldots N-1 \end{aligned}$$

and all remaining elements of the array have a value of zero.

Given A and s, it is required to find the wavefront vector **w** that minimizes the mean-square error $(\mathbf{A}\mathbf{w} - \hat{\mathbf{s}})^2$. It is assumed that the noise is random, uncorrelated, and equal on all slope measurements. The classic Gauss solution for a nonsingular matrix is found by multiplying each side of equation (8.2) by the transpose of matrix A and then inverting the resulting square matrix A A, giving

$$\hat{\mathbf{w}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\hat{\mathbf{s}} \tag{8.7}$$

The mean-square error of this solution is

$$\sigma_w^2 = \sigma_n^2 Tr(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1} \tag{8.8}$$

where

mean-square noise on each slope measurement
Tr = trace of the matrix

In the present case, the matrix A is singular because the slope measurements s cannot detect the piston component of **w**, which is consequently undetermined. Alternative solutions must therefore be found. One method is to use the generalized inverse of A, designated A®, giving the solution

$$\hat{\mathbf{w}} = \mathbf{A}^{\oplus}\hat{\mathbf{s}} \tag{8.9}$$

Another approach, described by Herrmann [1980], is to use a minimum-norm solution in which the wavefront is forced to have a mean value of zero. This can be achieved by adding an extra row of 1's to A to form the extended matrix $\mathbf{A_E}$, and adding an extra element (zero) to the vector s. The absolute phase of a wavefront is irrelevant in adaptive optics and, in any case, the minimum-norm solution is desirable since it minimizes the excursion of the wavefront corrector. Using the extended matrix, the solution giving the least-squares error is

$$\begin{aligned}\hat{\mathbf{w}} &= (\mathbf{A}_E^{\mathrm{T}}\mathbf{A}_E)^{-1}\mathbf{A}_E^{\mathrm{T}}\hat{\mathbf{s}} \\ &= \mathbf{B}\hat{\mathbf{s}}\end{aligned} \tag{8.10}$$

The mean-square wavefront error is given by

$$\sigma_w^2 = \frac{\sigma_n^2}{N^2}\sum_i\sum_j \mathbf{B}_{i,j}^2 \tag{8.11}$$

where

σ_n^2 = mean-square error on the slope measurements
$\mathbf{B}_{i,j}^2$ = squares of all elements of the B matrix

Practical methods of implementing the wavefront reconstruction process outlined above are described in section 8.3.

A similar process can be used to reconstruct the modal components of a wavefront, such as overall tilt, defocus, or astigmatism. These signals can then be used to control appropriate devices to null the error. Modal reconstruction is described in section 8.2.6.

When using zonal reconstruction, it is usually necessary to extract modal components, such as overall x and y tilt, from the wavefront and to reconstruct a tilt-removed wavefront. The tilt components may be extracted directly from the gradient array before reconstruction, by averaging the x-tilt and y-tilt measurements, or may be found from the reconstructed wavefront, using a matrix that weights each wavefront value according to its distance from the x and y axes. For a square aperture, the overall x- and y-axis angular tilts are

$$\begin{aligned}T_x &= \frac{1}{d}\sum_{\substack{p=-N/2\\p\neq 0}}^{N/2}\sum_{q=-N/2}^{N/2}\frac{w_{p,q}}{p} \\ T_y &= \frac{1}{d}\sum_{p=-N/2}^{N/2}\sum_{\substack{q=-N/2\\q\neq 0}}^{N/2}\frac{w_{p,q}}{q}\end{aligned} \tag{8.12}$$

where

d = node spacing
p, q = node indices in the x and y axes, respectively

A matrix can be constructed that will produce the tilt-removed wavefront $\mathbf{w_{TR}}$ to be applied to the deformable mirror directly, and will simultaneously produce the x- and y-tilt signals to be applied to the fast-steering mirror [Sasiela and Mooney 1985]. This matrix is constructed as follows. First, a matrix E is defined with dimensions 2 x N^2, which extracts the overall x and y tilts, $\mathbf{T_x}$ and $\mathbf{T_y}$, from the reconstructed wavefront. These tilts may be combined into a single matrix, T:

$$\mathbf{T} = \begin{bmatrix}\mathbf{T}_x \\ \mathbf{T}_y\end{bmatrix} = \mathbf{E}\hat{\mathbf{w}} \tag{8.13}$$

Next, the D matrix, which produces the tilt-removed wavefront and extracts the tilt signals from the full reconstructed wavefront, is specified:

$$\begin{bmatrix}\hat{\mathbf{w}}_{\mathrm{TR}} \\ \mathbf{T}\end{bmatrix} = \mathbf{D}\begin{bmatrix}\hat{\mathbf{w}} \\ \mathbf{T}\end{bmatrix} \tag{8.14}$$

The full reconstructed wavefront **w** and the overall tilts T are obtained from the gradients in the usual way:

$$\begin{bmatrix}\hat{\mathbf{w}} \\ \mathbf{T}\end{bmatrix} = \begin{bmatrix}\mathbf{B} \\ \mathbf{E}\,\mathbf{B}\end{bmatrix}\hat{\mathbf{s}} \tag{8.15}$$

The matrices are then combined to give

$$\begin{bmatrix}\hat{\mathbf{w}}_{\mathrm{TR}} \\ \mathbf{T}\end{bmatrix} = \mathbf{D}\begin{bmatrix}\hat{\mathbf{w}} \\ \mathbf{T}\end{bmatrix} = \mathbf{D}\begin{bmatrix}\mathbf{B} \\ \mathbf{E}\,\mathbf{B}\end{bmatrix}\hat{\mathbf{s}} = \mathbf{H}\hat{\mathbf{s}} \tag{8.16}$$

In this way, a single matrix multiplication of the measured gradients by H yields the required tilt-removed array $\mathbf{w_{TR}}$, as well as the x- and y-tilt components T.

The reconstruction problem in adaptive optics is usually more complicated than that shown in the simple model of figure 8.6. For example, the noise on the gradients may vary and have statistical correlations. Also, the wavefront structure function and coherence length are often known. If the noise statistics are represented by the noise covariance matrix $\mathbf{C}_N$, then the wavefront may be found using the Gauss-Markov estimate:

$$\hat{\mathbf{w}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}_N^{-1}\hat{\mathbf{s}} \tag{8.17}$$

If the spatial structure of the wavefront is represented by covariance matrix Cw, then the best solution is the Wiener estimate, given by

$$\hat{\mathbf{w}} = \left(\mathbf{C}_{\mathrm{W}}^{-1} + \mathbf{A}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{C}_N^{-1}\hat{\mathbf{s}} \tag{8.18}$$

This matrix can be reduced to a single matrix to implement the reconstruction. The mean-square error in the reconstruction process is found from

$$\sigma_w^2 = \sigma_n^2 Tr\left(\mathbf{C}_{\mathrm{W}}^{-1} + \mathbf{A}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathbf{A}\right)^{-1} \tag{8.19;}$$

8.2.3 Iterative Solutions

Wavefront reconstruction involves a large amount of computation, and real-time reconstruction requires very high data-processing rates. The first practical real-time reconstructors, devised in the 1970s, were based on iterative algorithms, implemented either analogically or digitally. Iterative methods generally involve less computation than direct solutions using matrix equations, but are limited to solving the simplified reconstruction problems described above and do not always converge to an exact solution. Digital processors are now capable of performing large matrix multiplications using arrays of multiplier-accumulators (the "brute force" approach), allowing much greater flexibility in the control algorithms used in adaptive optics. To make the coverage of wavefront reconstruction techniques as complete as possible, a brief account of iterative methods is offered here. Further information on iterative solutions may be found in Young [1972].

The iterative approach to wavefront reconstruction applies a local algorithm to each node of the network in turn, iterating this process until the wavefront values settle to within a given tolerance. The basic algorithm for network 1 may be stated as follows

$$w_{j,k} = \frac{1}{g_{j,k}}\left(w_{j,k}^{S} + s_{j,k}^{S}\right) \tag{8.20}$$

where

$w_{j,k}$ = wavefront value at node j, k
$g_{j\,k}$ = number of neighboring wavefront values (2, 3, or 4)
$w_{j,k}^{\mathrm{S}}$ = sum of the $g_{j,k}$ nearest wavefront values
$s_{j,k}^{\mathrm{S}}$ = sum of the $g_{j\,k}$ connecting slope values

Each iteration consists of N^2 calculations (one for each node).

There are three ways to implement the iteration process. In the first method, the $w_{j\,k}$ values are held until all N^2 nodes have been calculated, whereupon they are all updated simultaneously. This procedure is known as the Jacobi method; it converges at a slow rate. To improve the rate of convergence, the $w_{j,k}$ values may be updated immediately and used in each succeeding calculation. This is known as the Gauss-Seidel method. For both of these methods, the wavefront value at node j, k at the $(m+1)$th iteration is given by

$$w_{j,k}^{(m+1)} = \frac{1}{g_{j,k}}\left(w_{j,k}^{\mathrm{S}(m)} + s_{j,k}^{\mathrm{S}(m)}\right) \tag{8.21}$$

In equation (8.21), the value of $w_{j,k}^{(m+1)}$ is recalculated for each iteration and its previous value is only used to establish the sum of the neighboring nodes. The rate of convergence may be improved further if the value of the wavefront at each node in the previous iteration is taken directly into account. In effect, calculated wavefront values are "remembered" from one iteration to the next. The decay factor is established by the relaxation parameter ω. This technique is known as successive overrelaxation (SOR) and is described by the equation

$$w_{j,k}^{(m+1)} = \frac{1}{g_{j,k}}\left[w_{j,k}^{(m)} + \omega\left(w_{j,k}^{\mathrm{S}(m)} - w_{j,k}^{(m)} + s_{j,k}^{\mathrm{S}(m)}\right)\right] \tag{8.22}$$

The rate of convergence depends on ω. When $\omega = 1$, this equation reduces to equation (8.21). The optimal value of ω for this type of matrix is given by

$$\omega = \frac{2}{1 + \sin[\pi/(N+1)]} \tag{8.23}$$

Iterative solutions have been implemented using both digital and analog computers. The analog approach using a resistor network is described in section 8.3.1.

A totally different type of recursive algorithm is the exponential reconstructor, developed by D. Ehn at Itek Optical Systems in the 1980s. It provides an approximate least-squares solution to the slope reconstruction problem for network 1, requiring many fewer arithmetic operations than a full matrix inversion of the same-size array. The number of multiply–add operations in the exponential algorithm is proportional to the number of nodes in the array, whereas with full matrix inversion the number of multiply–adds is proportional to the square of the number of nodes. The price paid for this large increase in efficiency is that the exponential reconstructor has a slightly higher noise propagation.

The exponential reconstructor algorithm operates on a square grid of nodes that covers the optical aperture, which may be of arbitrary shape. The underlying idea of the algorithm is first to simplify the network by eliminating nodes in successive steps until it is easily solved by explicit calculation, and then to build the network back to full resolution iteratively, using the measured slope data to fill in the nodes that were eliminated. The first operation is called *decimation* and consists of weighting and combining the input data along several paths to produce estimates of the phase differences at twice the nodal spacing in both axes, thereby reducing the number of nodes by a factor of four at each iteration. The decimation operation is halted when a simple

array with a nontrivial solution emerges, such as 3 x 3 or 4 x 4. Such an array is readily solved by direct calculation, establishing the overall shape of the wavefront. The *build* operation then estimates the wavefront values at intermediate nodes by weighting and averaging the input slope data, increasing the number of nodes by a factor of four at each iteration until the array is, again, fully populated.

Using this method, the number of multiply–add operations required to reconstruct a 9 x 9 array is reduced from 12,000 (for full matrix multiplication) to 5000, while for a 17 x 17 array the corresponding reduction is from 157,000 to 18,000. The noise propagation for the exponential reconstructor is about 10-15% higher than that for full matrix multiplication.

8.2.4 Error Propagation of Simple Networks

Two performance factors are of interest in wavefront reconstruction. The first of these is the accuracy of reconstruction; that is, how closely the original wavefront can be recovered from an array of noise-free gradients. The second factor is the error propagation, which is the ratio of the error in the reconstructed wavefront to the errors in the input gradients, assumed to be random and independent. Using the simplified reconstruction model, in which the wavefront is sampled at discrete nodes and the gradients are measured between those nodes, the basic equations developed above are all exact solutions. The error propagation for the simplified reconstructor models does appear to vary considerably, although these differences tend to disappear when all factors are taken into account, as explained later. It should be noted that the adaptive optics *fitting error*, which depends on the wavefront sampling density (node spacing), is independent of the reconstructor configuration and is evaluated separately.

The error propagation of network 1 has been analyzed by Hudgin [1977a], that of network 2 by Fried [1977], and that of network 3 by Southwell [1980]. Performance comparisons of these configurations have been made by Herrmann [1980], Southwell [1980], and Wallner [1983].

If the noise on all gradient inputs is uncorrelated with a mean-square value of σ_n^2 then, using equation (8.11), the error in the reconstructed wavefront may be expressed as

$$\sigma_w^2 = \frac{\sigma_n^2}{N^2} \sum_i \sum_j \mathbf{B}_{i,j}^2 = \sigma_n^2 E \tag{8.24}$$

The error propagation factor E is the ratio of the mean-square error of the reconstructed wavefront to that of the input gradients. It depends on the geometry of the reconstructor and is typically on the order of unity. Using the simplified reconstructor model, in which the wavefront is defined at an array of points or nodes, Herrmann [1980] has calculated the error propagation of networks 1 and 2 for various array sizes, as shown in table 8.2.

The error propagation of network 2 is about twice that of network 1, but of more interest is the fact that the error propagation of both networks is only weakly dependent on the number of wavefront nodes. For large arrays, Noll [1978] found that this dependence has the form

$$E = a + b \ln(N^2) \tag{8.25}$$

where a and b depend on the geometry of the array and N^2 is the total number of nodes.

The reason for the large apparent difference in error propagation between the network configurations has been examined by Wallner [1982, 1983]. He found that when a realistic adaptive optics system model is used, the differences in performance between these configurations disappear. Wallner analyzed a general model in which the wavefront slope is measured and corrected over finite subapertures, using an optimal reconstructor and control law based on a knowledge of the structure function of the wavefront. In this general case, the residual

Table 8.2 Error Propagation of Wavefront Reconstructors Using Gradient Inputs (Simplified Model)

Array Size	Error Propagation	
	Network 1	Network 2
2 x 2	0.31	0.50
3 x 3	0.43	0.99
4 x 4	0.50	1.15
5 x 5	0.55	1.23
6 x 6	0.58	1.29
7 x 7	0.62	1.33
9 x 9	0.66	1.39
12 x 12	0.69	1.41

error depended on the total number of gradient measurements made over the aperture, as expected, but not on their geometry. The difference in error propagation between networks 1 and 2 only appears when using simplified reconstruction models, in which the wavefront is determined at arrays of single points (nodes) that are connected by linear phase differences.

8.2.5 Applications of the Curl Operator

The configuration of network 1 allows the closure gradients to be estimated around each group of four nodes, as shown in figure 8.7. These closure gradients or "curls" can be extracted from the array of gradient measurements by means of a discrete curl operator in the form of a matrix, leading to some interesting possibilities. To start with, a knowledge of the curl components enables the noise on the gradient measurements to be determined irrespective of the wavefront being measured. If the curl matrix is Q and the gradient array is g, then, assuming the noise on each gradient is statistically independent, the curl values are given by

$$\mathsf{q} = \mathsf{Qg} \tag{8.26}$$

The matrix Q has been defined by Herrmann [1980] as follows

$$\mathbf{Q}[p, q + N - 1] = -1$$
$$\text{for } p = 1 \ldots (N-1)^2, \qquad q = 1 \ldots (N-1)^2 \tag{8.27}$$

$$\mathbf{Q}[q + (p-1)(N-1)M/2 + q + (p-1)N] = -1$$
$$Q[q + (p-1)(N-1),\ 1 + M/2 + q + (p-1)N] = 1$$
$$\text{for } p = 1 \ldots (N-1),\ q = 1 \ldots (N-1) \tag{8.28}$$

The size of this matrix is $(N-1)^2 \times M$ (curls x gradients), where N^2 is the number of nodes and $M = 2N(N-1)$ is the number of gradients.

The relation between the noise variances of the curl and the gradients is

$$\sigma_q^2 = 4\sigma_g^2 \tag{8.29}$$

Thus, by obtaining several samples of curl, either from a single array or from successive frames, or both, the mean-square noise on the measured gradients may be found.

If the gradient measurements were noise free, the curl would be zero. In this case, a least-squares solution to the wavefront reconstruction problem would be unnecessary. The gradient vectors could be added directly over a single integration path, considerably simplifying the computation. Herrmann [1980] has suggested that this result could also be achieved by subtracting the closure or "solenoidal" error, determined by the curl operation, from the gradient measurements. He shows that the curl-free gradients are given by

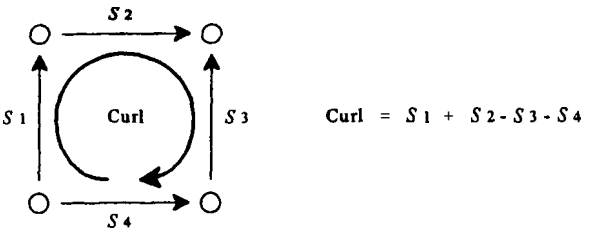


Figure 8.7 Definition of curl in a reconstruction network.

$$\mathbf{g}_c = (1 - \mathbf{Q}^+\mathbf{Q})\mathbf{g} \tag{8.30}$$

The reconstructed wavefront is then obtained from the curl-free gradients by making $N^2 - 1$ additions (one for each node, with one node set to zero). The total number of arithmetic operations in this process is therefore [(curls) x (gradients)] + [nodes], or $[\{N-1)^2 \times 2N(N-1)] + [N^2 - 1]$ When N is large, the number of operations tends to increase as which is the same as that for a conventional matrix inversion process. The curl-free reconstructor does not therefore appear to reduce the amount of data to be processed.

8.2.6 Modal Estimation

Zonal wavefront gradient data may be reconstructed into a modal estimate of the wavefront, for use in adaptive optics systems using modal control or correction. Modal operation is appropriate for low-order compensation because whole-aperture functions, such as tilt, defocus, and astigmatism, can be implemented more smoothly than with zonal correction. The principles of modal estimation have been described by Cubalchini [1979] and Southwell [1980, 1982].

A wavefront may be generally defined in modal terms, using a set of whole-aperture basis functions or modes, as

$$W(x, y) = \sum_{k=1}^{M} a_k Z_k(x, y) \tag{8.31}$$

where values of a_k are the coefficients of the basis functions $Z_k(x, y)$.The set of modes is assumed to be linearly independent; that is, the value of each coefficient a_k is independent of all the others. The wavefront gradients measured in the x and y axes are

$$g^x(x, y) = \frac{\partial}{\partial x} W(x, y)$$
$$g^y(x, y) = \frac{\partial}{\partial y} W(x, y) \tag{8.32}$$

In zonal wavefront sensors, the gradient measurements are made over discrete areas and therefore represent a finite set of gradient samples. The gradients at location $i = (x_i, y_i)$ may then be expressed as

$$g_i^x = \frac{\partial}{\partial x} W_i = \sum_{k=1}^{M} a_k \frac{\partial}{\partial x} Z_{ki}$$
$$g_i^y = \frac{\partial}{\partial y} W_i = \sum_{k=1}^{M} a_k \frac{\partial}{\partial y} Z_{ki} \qquad (8.33)$$

This equation models the relation between the measured gradients g_i and the modal coefficients a_k. The modal estimation process requires each of the M expansion coefficients a_k to be determined from the gradient measurements.

Equation (8.33) may be written in matrix form, relating the gradient vector g to the mode coefficient vector a, as follows

$$\mathrm{g} = \mathrm{D}\mathrm{a} \qquad (8.34)$$

where D is a matrix containing the derivatives of the basis functions in equation (8.33) and g contains both x- and y-gradient measurements. It is desired to find the values of the coefficients a, which define the magnitudes of the modal components. This is done in the usual way by finding the inverse of matrix D and multiplying it by the gradient vector g. The result is

$$\hat{\mathbf{a}} = \mathbf{B}\hat{\mathbf{g}}$$
$$\text{where} \quad \mathbf{B} = \left(\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{D}^T \qquad (8.35)$$

In the case of most interest, when the basis functions Z_k *axe* Zernike polynomials, the required derivatives can be found explicitly. An interesting insight into this operation was provided by Southwell [1982], who noted that since the derivative operation reduces the order of a polynomial, each scalar component of the gradient of a Zernike polynomial can be expanded as the sum of lower order Zernike polynomials, as shown in table 8.3. The matrix B is constructed in this way, containing both x and y derivatives.

The number of modes, M, is usually determined by the residual error requirement. With a knowledge of the wavefront statistics and the basis functions Z_k, the fitting error can be determined as a function of M. Residual errors for Kolmogorov wavefronts corrected with Zernike polynomials are given in section 3.5. In practical cases, the number of modes employed may be much smaller than the number required for good correction. For example, wavefront correction may be limited to the first six Zernike modes, including x and y tilts, defocus, and astigmatism, but excluding coma, trifoil, and spherical aberration. In this case, Southwell [1982] has shown that it is possible to design a matrix in which lower order modes partly compensate for higher order aberrations. This is a standard technique in optical design, known as "aberration balancing." In the context of adaptive optics, it implies a solution that minimizes the mean-square wavefront error for a given number of correction modes, rather than merely correcting specific modal components.

The coefficients of the Zernike modes are determined from equation (8.35), which may be expressed in terms of the x and y gradients as follows

$$\hat{\mathbf{a}} = \mathbf{B}^x\hat{\mathbf{g}}^x + \mathbf{B}^y\hat{\mathbf{g}}^y \qquad (8.36)$$

Using the aberration-balancing approach for 11 Zernike terms, the rows of the $\mathbf{B}^x$ and $\mathbf{B}^y$ matrices are given by the factors b^x and b^y in table 8.4. The columns correspond to the x and y gradients in each zone of the aperture; these are stacked end to end as indicated in figure 8.8. Each element of the matrix, $b_{k,j}$, is obtained from the expression in the table by evaluating $Z_k(x_j, y_j)$. Using this approach, the overall tilt correctors remove the tilt component of coma, and the focus corrector partly compensates for spherical aberration. From the narrow viewpoint of control system design, this cross-coupling of modes may appear detrimental, but in the wider context of max-

Table 8.3 Zernike Polynomials and their Gradients

k		Z_k	$Z_k^x = \partial Z_k/\partial x$	$Z_k^y = \partial Z_k/\partial y$
1	Piston	1		
2	x Tilt	2v	$2Z_1$	
3	y Tilt	$2v$		$2Z_1$
4	Defocus	$\sqrt{3}(2r^2 - 1)$	$2\sqrt{3}Z_2$	$2\sqrt{3}Z_3$
5	0° Astigmatism	$\sqrt{6}(x^2 - y^2)$	$\sqrt{6}Z_2$	-V6Z3
6	45° Astigmatism	$\sqrt{6}(2xy)$	$\sqrt{6}Z_3$	$\sqrt{6}Z_2$
7	x Coma	$\sqrt{8}(3r^2 - 2)$*	$\sqrt{8}Z_1 + 2\sqrt{6}Z_4 + 2\sqrt{3}Z_5$	$2\sqrt{3}Z_6$
8	y Coma	V8(3r^2 - 2)y	$2\sqrt{3}Z_6$	$\sqrt{8}Z_1 + 2\sqrt{6}Z_4 - 2\sqrt{3}Z_5$
9	x Trifoil	$\sqrt{8}(4x^2 - 3r^2)x$	$2\sqrt{3}Z_5$	$-2\sqrt{3}Z_6$
10	y Trifoil	$\sqrt{8}(3r^{\epsilon} - 4y^2)y$	$2\sqrt{3}Z_6$	$2\sqrt{3}Z_5$
11	Spherical aberraration	$\sqrt{5}(6r^4 - 6r^2 + 1)$ where r = x + y^{-} < 1	$2\sqrt{5}Z_2 + 2\sqrt{10}Z_7$	$2\sqrt{5}Z_3 + 2\sqrt{10}Z_8$

Table 8.4 Components of The Matrices B^x and B^y, for Optimal Zernike Mode Extraction Using 11 Modes

k	Mode	Z_k	b_k^x (divide each term by π)	b_k^y (divide each term by π)
I	Piston	1		
2	x Tilt	$2x$	$Z_1/2 - Z_4/4\sqrt{3} - Z_5/4\sqrt{6}$	$-Z_6/4\sqrt{6}$
3	y Tilt	$2y$	$-Z_6/4\sqrt{6}$	$Z_1/2 - Z_4/4\sqrt{3} + Z_5/4\sqrt{6}$
4	Defocus	$\sqrt{3}(2r^2 - 1)$	$Z_2/4\sqrt{3} - Z_7/4\sqrt{6}$	$Z_3/4\sqrt{3} - Z_8/4\sqrt{6}$
5	0° Astigmatism	$\sqrt{6}(x^2 - y^2)$	$Z_2/2\sqrt{6}$	$-Z_3/2\sqrt{6}$
6	45° Astigmatism	$\sqrt{6}(2xy)$	$Z_3/2\sqrt{6}$	$Z_2/2\sqrt{6}$
7	x Coma	$\sqrt{8}(3r^2 - 2)x$	$Z_4/4\sqrt{6} + Z_5/8\sqrt{3}$	$Z_6/8\sqrt{3}$
8	y Coma	$\sqrt{8}(3r^2 - 2)y$	$Z_6/8\sqrt{3}$	$Z_4/4\sqrt{6} - Z_5/8\sqrt{3}$
9	x Trifoil	$\sqrt{8}(4x^2 - 3r^2)x$	$Z_5/4\sqrt{3}$	$-Z_6/4\sqrt{3}$
10	y Trifoil	$\sqrt{8}(3r^2 - 4y^2)y$	$Z_6/4\sqrt{3}$	$Z_5/4\sqrt{3}$
11	Spherical aberration	$\sqrt{5}(6r^4 - 6r^2 + 1)$	$Z_7/4\sqrt{10}$	$Z_8/4\sqrt{10}$

imizing image quality, it is recognized to be beneficial.

A different approach to modal wavefront reconstruction, using Fourier transforms, has been described by Freischlad and Koliopoulos [1985]. This process is equivalent to a filtering operation in the spatial frequency domain, following which zonal actuator commands may be retrieved. If the basis functions are defined as a set of complex exponentials $Z_{j,k}(x, y)$, where

$$Z_{j,k}(x, y) = \frac{1}{N}\exp[(2\pi i/N)(jx + ky)] \quad \text{for } 0 \le j, k \le N - 1 \tag{8.37}$$

then the expansion of the wavefront in these basis functions is identical to the inverse discrete Fourier transform of the coefficients $a_{j,k}$; that is,

$$W(x, y) = \sum_{j,k=1}^{N=1} a_{j,k} Z_{j,k}(x, y) = \mathcal{DF}^{-1}(a_{j,k}) \tag{8.38}$$

The goal of the reconstruction process is to find the expansion coefficients $a_{j,k}$ that correspond to the measured wavefront gradients $g(x,y)$. To find the expansion coefficients, the gradients measured by the wavefront sensor are, first, Fourier transformed into the spatial frequency domain, and then multiplied by a filter function corresponding to the least-squares fit. The wavefront is then retrieved using the inverse discrete Fourier transform, equation (8.38). The error propagation of this process is comparable to that of zonal reconstructors, except that most of the error occurs in sharp peaks at the edges and corners of the aperture. This is because of the fact that there are only $(N - 1)N$ x-gradients and $N(N - 1)$ y-gradients in an array of $N \times N$ nodes, so that missing data must be synthesized. An advantage of this approach is that the number of arithmetic operations is approximately proportional to N^2, the number of nodes, whereas matrix multiplication requires N^4 operations.

8.2.7 Optimal Wavefront Reconstruction

The approach to wavefront reconstruction outlined above has been used in the design of most existing adaptive optics systems. While it has given service-

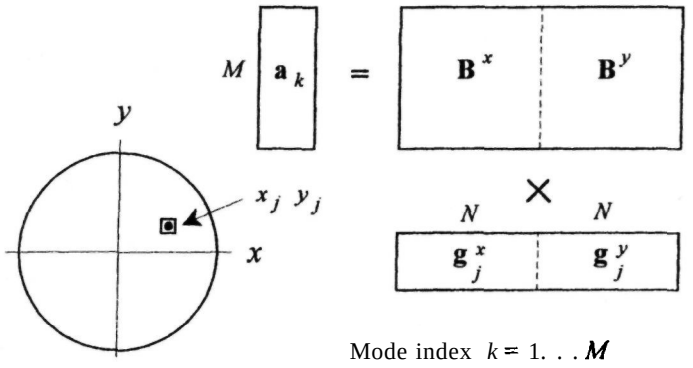


Figure 8.8 Zernike mode extraction matrix.

able results, it is based on a simplified reconstruction model that does not account for many characteristics of actual systems. In particular, it has the following limitations:

1. Wavefront measurements are treated as point-to-point differences rather than averages over finite subaperture areas.
2. Estimates of the reconstructed wavefront are obtained as an array of points, with no guidance on how to interpolate between them.
3. The statistics of the wavefront being estimated are not used.

Wallner (1983) has suggested a more comprehensive approach, in which the wavefront reconstructor is generalized into an optimum estimator that takes into account the characteristics of the wavefront sensor and the wavefront corrector, as well as the spatial and temporal statistics of the turbulence. The reconstruction process is considered to be an intrinsic part of the overall control law, rather than an inconvenient operation made necessary by the use of wavefront slope sensors. The control law takes into account all the relevant parameters of the adaptive optics system, both external and internal, and can be optimized to maximize any desired performance criterion.

The system model used in this enlightened approach to the design of adaptive optics systems is described briefly as follows. The wavefront sensor produces an array of wavefront measurements s_n, that are unavoidably contaminated with noise. The measurement array is multiplied by a control matrix $\mathbf{M}_{jn}$ to produce an array of actuator commands c_i that are applied to the wavefront corrector. The process may be optimized in a number of ways. Two results of interest are: (1) Given the actuator response functions $r_j(\mathbf{x})$, it is possible to find the optimal matrix $\mathbf{M}^*_{jn}$ that will minimize the error criterion (for example, least-mean-square error) for a given number of degrees of freedom. (2) Given only the number of degrees of freedom, it is possible to find the optimal actuator response function $r_i^*(\mathbf{x})$ that will minimize the error criterion. The principles of optimal wavefront correction are summarized in section 8.6.

8.3 Practical Reconstructors

8.3.1 Analog Network

The analog reconstructor using a resistor network was the first practical solution to the problem of real-time wavefront reconstruction [Hardy et al. 1977]. The network determines the wavefront values at an array of nodes, using the differences between these nodes as the input data. The geometry is shown in figure 8.9. The wavefront is represented by a rectangular array of nodes $W_{x,y}$, which are connected by wavefront slopes or gradients $S^x_{x,y}$ and $S^y_{x,y}$. The analog network is based on the proposition that in any group of five nodes, the wavefront value at the center node, $W_{j,k}$, is equal to the average value of the four adjacent nodes, plus the average of the four connecting gradients. The relation is formally stated as

$$W_{j,k} = \tfrac{1}{4}\left[W_{j-1,k} + W_{j+1,k} + W_{j,k-1} + W_{j,k+1} + S^x_{j-1,k} - S^x_{j+1,k} + S^y_{j,k-1} - S^y_{j,k+1}\right] \tag{8.39}$$

This algorithm was derived heuristically for wavefront reconstruction in adaptive optics [Hardy 1975a]. It has been proved to give the least-mean-square solution in the case of noisy gradient inputs [Hudgin 1977a].

The equation is implemented in the analog reconstructor as shown in figure 8.10. The nodes are connected by a network of equal resistors that produce the average value of the four adjacent W values. A current proportional to the algebraic sum of the four surrounding gradients is injected into each node. At the edges and corners of the array, where there are only three or two adjacent nodes, the proportional constant is modified accordingly. One node is normally grounded (clamped to zero) to serve as a reference. If $S^x_{j,k}$ and $S^y_{j,k}$ are wavefront gradients in the x and y directions (meters per subaperture) and k_g is the scale factor (volts per meter) then the currents corresponding to each gradient measurement are

$$i^x_{j,k} = \frac{k_g}{R}\left(S^x_{j,k}\right)$$
$$i^y_{j,k} = \frac{k_g}{R}\left(S^y_{j,k}\right) \tag{8.40}$$

The net current into each node is the resultant of four gradients

$$I_{j,k} = \frac{1}{4}\left[i^x_{j-1,k} - i^x_{j,k} + i^y_{j,k-1} - i^y_{j,k+1}\right]$$
$$I_{j+1,k} = \frac{1}{4}\left[i^x_{j,k} - i^x_{j+1,k} + i^y_{j+1,k-1} - i^y_{j+1,k+1}\right], \tag{8.41}$$
$$\text{etc.}$$

The current associated with any subaperture gradient measurement must flow only through the single resistor corresponding to that subaperture. This is accomplished by injecting the current into one node and extracting it from the opposite end of the appropriate resistor. With a given set of input currents containing random uncorrelated noise, the node voltages settle at values satisfying equation (8.39), minimizing the mean-square value of the noise over the array. A change in any input current propagates through the network, which settles at a new value satisfying the equation at all nodes. The settling time of the network depends on the time constant defined by the network resistance R between nodes and the shunt capacitance C of each node to ground. The performance of analog reconstruction networks is analyzed in the following section.

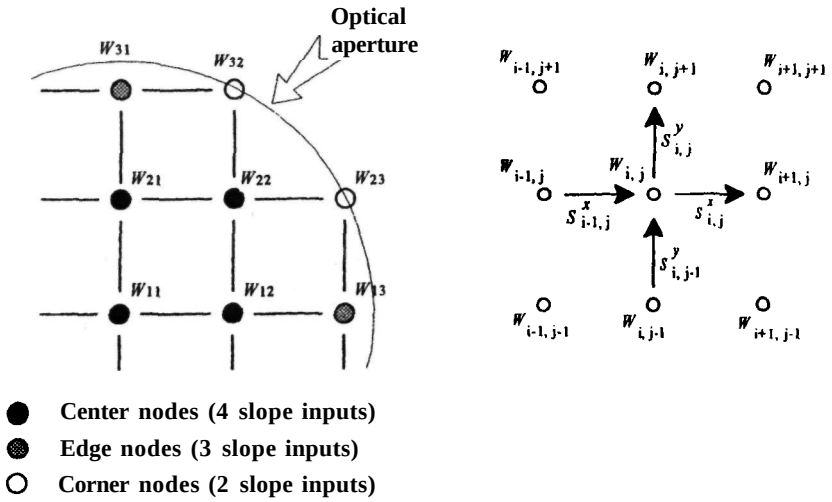


Figure 8.9 Iterative reconstruction algorithm for network 1.

Analog reconstructors of this type were used successfully in the 21-actuator RTAC system, built in 1973, and in the 168-actuator CIS, first operated in 1982. The advantages of analog reconstructors are that they are extremely fast (the settling time for a 32 x 32 array is typically less than 10 μs) and the hardware is relatively simple, the number of components being directly proportional to the number of nodes. The perceived problem with these devices is the basic limitation of all analog computers: restricted dynamic range and accuracy. Analog devices are typically limited by noise to a dynamic range of 10^3–10^4 (60–80 dB) and are susceptible to offsets due to temperature and other environmental changes. These problems can be overcome by suitable design; for example using a self-calibration system to eliminate offsets. Studies have shown that it is feasible to design analog networks with adequate performance for adaptive optics systems as large as 10,000 actuators.

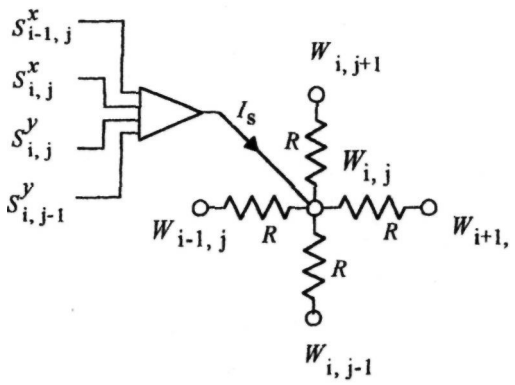


Figure 8.10 Analog implementation of iterative reconstruction algorithm.

A block diagram of the hardware required to implement an analog reconstructor is shown in figure 8.11. The input consists of serially multiplexed x-axis and y-axis gradient data from the wavefront sensor, which has been conditioned to eliminate offsets and to provide the correct scaling, as described in section 5.3.6. Each frame of conditioned slope data is stored in a RAM, where it can be accessed by the reconstructor, which requires algebraic summation of gradients in groups of four. Two memory banks are normally used, one being updated while the other is accessed by the reconstructor. The memory may be organized and updated in sections, corresponding to multiple readout ports in the wavefront sensor. The wavefront data are weighted and added in groups of four (two S^x gradients and two S^y gradients) using a multiplier-accumulator (MAC). In the iterative algorithm used by the analog reconstructor, each node requires only four multiply–adds, compared with N^2 in the case of matrix multiplication. The MACs can therefore be multiplexed to serve several nodes, a convenient arrangement being to have one adder per row or column of the network. The number of MACs required for the analog reconstructor is therefore much smaller than that necessary for full matrix multiplication.

The digital outputs of the MACs, comprising the four-element slope sums, are then converted to analog and distributed to the current drivers. Each node has a dedicated current driver, which is basically an operational amplifier configured to operate as a high-impedance source. The outputs of the current drivers are connected to the nodes of the resistor matrix. The reconstructed wavefront value is represented by the voltage at each node of the network.

8.3.2 Hybrid Reconstructors

Analog reconstructors are not restricted to the simplified configuration described in the preceding text. In principle, reconstructors of any complexity may be realized using analog or a combination of analog and digital (hybrid) techniques. This possibility has received little or no attention in the literature, so a brief discussion is presented here.

The basic elements of all wavefront reconstructors using zonal slope measurements are:

1. an array of nodes representing the actuators of the correction device, which may have any configuration, not necessarily regular or symmetrical;
2. a means of weighting and distributing the zonal slope measurements to the nodes;
3. a means of summing the signals at each node.

These elements describe the process defined in the basic linear reconstruction equation (8.31). In the analog reconstructor, items 1 and 3 are implemented by a passive network interconnecting all nodes according to their geometry. Item 2 is implemented by active devices that inject a current into each node proportional to the wavefront difference between that node and its neighbors.

In the simplified network using localized slopes and zones, only the four nearest neighbors are used for item 2. This is not a basic restriction and there is no limit to the nodal distribution of each measured slope. In this way, factors such as the actuator influence functions, noise covariances, and wavefront statistics can be accommodated in the analog reconstruction process by weighting and combining the slope measurements, as is done in the general matrix equation. The advantage of the analog system is that the basic process of adding the subaperture gradients over the two-dimensional array is accomplished by the passive network. This eliminates a considerable part of the data processing that is required in matrix multiplication systems, especially for large arrays.

A major concern about the use of analog reconstructors is whether analog (A) devices can be justified in otherwise exclusively digital (D) processors. The need for D/A and A/D data conversion would appear to outweigh any possible advantages. However, a suitable location does exist. The output devices used in adaptive optics, typically deformable mirrors, are driven analogically, requiring D/A conversion of the actuator drive signals. This suggests that a suitable (and economical) location for an analog wavefront reconstructor would be in combination with the wavefront correction driver system.

In practice, the reconstructor system would be hybrid, using both digital and analog components. A possible scheme is shown in figure 8.11. Wavefront slope data are downloaded from the bus and held in a register, where they are accessed by the multiplier-accumulator and weighted using the constants stored in the RAM. These functions are similar to those in an all-digital matrix multiplication reconstructor. The difference is that the number of multiply-add operations required in the hybrid system is typically much smaller than for an all-digital system. The node signals are converted from D to A and held in sample-and-hold circuits at the input to the current amplifier serving each node. The MACs and D-to-A converters may be multiplexed to serve a number of current amplifiers, depending on the data rates involved. The current amplifiers inject current into the summing network, producing the required voltages at each node; these voltages are fed directly to the wavefront corrector drivers.

It is of interest to compare the data-processing requirements of a hybrid reconstructor of this type with that of an all-digital matrix multiplication system. For *an N x N* array, full matrix multiplication requires a total of $2N(N - 1) \times N^2$ multiply–add operations, irrespective of the number of additional covariance functions involved. (These functions change the matrix elements, especially those off-diagonal, but do not change the size of the matrix.) In the case of analog reconstruction, the number of multiply-adds required is $M \times N$, where M is the number of slopes input to each node. The minimum value of M for a simple reconstructor is 4. The additional number required depends on the area covered by the covariance functions, in units of actuator spacing, d. For example, the reconstructor may be designed to account for the actuator influence functions. These usually have significant value over a radius of $2d$, which amounts to an area covering 20 nodes. In this case, M will have a value of 20. If the reconstructor weights slope measurements according to their signal-to-noise ratio, this will only affect individual inputs, with no influence on M. For array sizes of practical interest (5 x 5 or over), the analog reconstructor will, in most cases, require many fewer multiply-add operations than a matrix multiplication reconstructor of similar performance. Only when the additional covariance functions cover a large area (that is, when M approaches the value of N^2) will the analog reconstructor require as much hardware as a matrix multiplier.

8.3.3 Performance of Analog Reconstructors

The performance factor of prime interest in all reconstructors is the error propagation (that is, the ratio of mean-square error on the output wavefront to that on the input gradients), and how it varies with the size of the wavefront array. Analysis of the basic reconstruction process (section 8.2.4) has shown that the error propagation factor is ideally about 0.4 for small arrays, and it increases as the logarithm of the number of nodes. This performance is achieved with a full matrix inversion, which requires approxi-

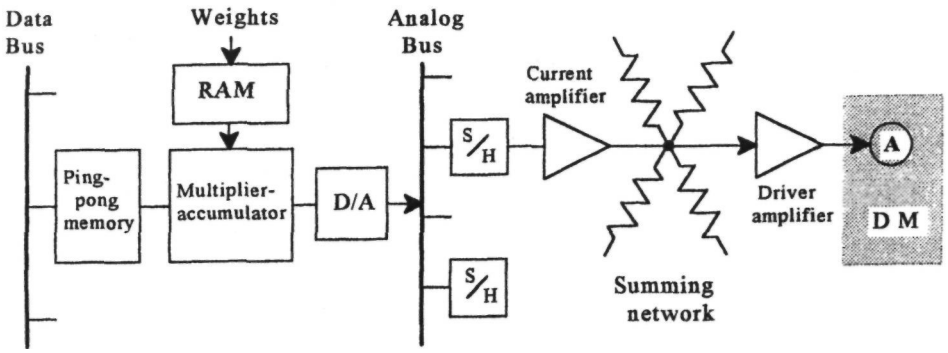


Figure 8.11 Block diagram of hybrid reconstructor combined with actuator drivers. RAM, random-access memory; D/A, digital-to-analog converter; S/H, sample-and-hold circuit; A, actuator; DM, deformable mirror.

mately N^4 arithmetic operations for an $N \times N$ array of nodes. In the case of a digital processor, the computation speed is set by the clock rate and the number of parallel processors available. To achieve high data rates, a considerable amount of hardware may be necessary.

In the case of analog devices, there is usually a more flexible tradeoff between speed and performance, with the performance degrading gradually with bandwidth. It is of some interest to see how analog reconstructors compare with digital processors in this regard. An incentive for this investigation is the possible use of solid-state or liquid crystal wavefront correctors that may employ 10^4 or more actuators, for which digital processing would be very expensive. As mentioned above, an economical solution in this case would be to combine an analog reconstructor with the correction device.

An ideal analog processor (using network 1) solves the simplified reconstruction problem exactly, producing an output wavefront that minimizes the mean-square error between the wavefront and the measured gradients [Hudgin 1977a]. An ideal processor consists of a pure conductance network with no parasitic reactances, and a linear current driver with no offset.

Practical analog reconstructors are susceptible to two errors: (1) temporal error due to shunt capacitance from each node to ground, which causes the network to require a finite time to settle to the final solution; and (2) offsets due to leakage currents in each node, which distort the reconstructed wavefront.

The temporal errors in analog wavefront reconstructors using the resistor network were analyzed by Menikoff [1987]. Assuming that the wavefront data have a Kolmogorov spectrum, he showed that the convergence error, defined as the difference between the reconstructed wavefront at time t and the final value at $t = \infty$, is given by

$$\sigma_c = 0.146\left(\frac{RC}{t}\right)^{1/2}\left(\frac{d}{r_0}\right)^{5/6} N^{11/6} \times \exp\left[-\frac{t}{2RC}\left(\frac{2\pi}{N}\right)^2\right] \tag{8.42}$$

where

σ_c = convergence error at time t, in waves rms at wavelength λ
t = time after inputs are applied, s
R = resistance between nodes, Si
C = capacitance of each node to ground, F
d — subaperture dimension, m
r_0 = turbulence parameter, meters at wavelength λ
N = number of nodes per side ($N \times N$ network)

Convergence errors for reconstructor arrays of 10×10, 32 x 32, and 100 x 100 are shown in figure 8.12, using network parameters of $R = 1000\,\Omega$, $C = 20$ pF, with $d/r_0 = 1$. It is seen that even with these conservative parameters, the convergence time for a 1000-element array is less than 10 μs. The temporal error is therefore negligible in analog reconstructors employed in astronomical adaptive optics.

The error due to offsets in the current injected into the nodes of the network may be estimated as follows. If the gradients are all zero, then the input currents should likewise all be zero and the network will settle to a null (flat) condition. The currents are generated by operational amplifiers, which are liable to offsets due mainly to temperature variations. The offsets are therefore likely to be correlated, which will produce the largest error on the reconstructed wavefront. If the current injected into each node is /, then the voltage at node j, k is

$$V_{j,k} = \tfrac{1}{4}\left[V_{j-1,k} + V_{j+1,k} + V_{j,k-1} + V_{j,k+1} + RI\right] \tag{8.43}$$

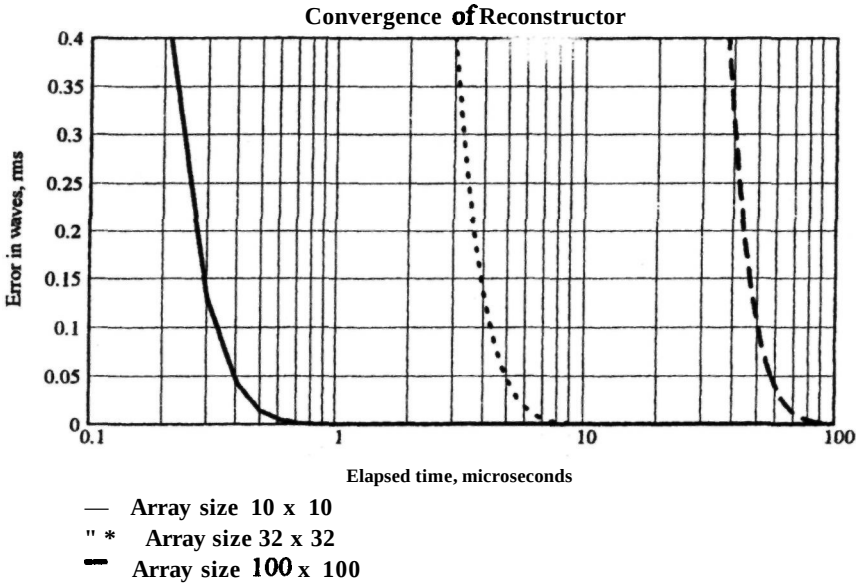


Figure 8.12 Convergence errors for an iterative reconstructor. These curves show the error versus settling time for a high-speed analog reconstructor using a resistance network, as a function of the array size. The internode resistance is 1000 ohms and the shunt capacitance is 20 pF per node. The input wavefront has a Kolmogorov spectrum and is constant during reconstruction. The value of d/r_0 is 1.

The system can be modeled as continuous Laplacian. For a circularly symmetrical aperture, it is convenient to use polar coordinates. The surface profile produced by a constant current per unit area may be defined in terms of the radius r by

$$U(r) = \tfrac{1}{4}r^2 + A\ \ln(r) + B \tag{8.44}$$

To determine the value of A, the condition is imposed that no current flows across the boundary of the aperture; that is,

$$\frac{\partial U}{\partial r} = 0 \qquad \text{when } r = 1 \tag{8.45}$$

This condition implies that $A = -1/2$. There is a singularity at the origin, so, to determine the B term, it is necessary to specify a minimum radius $r_{\min}$, which will serve as the reference for measuring U. The inner ring of radius r can be considered to be connected to ground, draining all of the current injected into the nodes. Then,

$$B = -\tfrac{1}{4}r_{\min}^2 + \frac{1}{2}\ln(r_{\min}) \tag{8.46}$$

Substituting A and B in equation (8.44), the current profile is

$$U(r) = \tfrac{1}{4}(r^2 - r_{\min}^2) - \tfrac{1}{2}\ln\left(\frac{r}{r_{\min}}\right) \tag{8.47}$$

If there are N nodes across the diameter, then the voltage as a function of radius is given by

$$\begin{aligned} V(r) &= \tfrac{\pi}{4}N^2\tfrac{1}{4}RIU(r) \\ &= \tfrac{\pi}{16}N^2RI\left[\tfrac{1}{4}(r^2 - r_{\min}^2) - \tfrac{1}{2}\ln\left(\frac{r}{r_{\min}}\right)\right] \end{aligned} \tag{8.48}$$

This function is plotted against radius in figure 8.13 for $r_{\min}$ — 0.05. It shows the shape of wavefront distortion produced by equal offset currents of / (amperes) in each node. The maximum voltage excursion at the outer edge is $V_o = 0.248N^2RI$.

If necessary, the current drivers may be calibrated to trim their offset current to zero. This can be done automatically using a regular cycle of operations. The procedure is very simple: (1) All gradient inputs are set to zero. (2) The node voltages are compared to zero using comparator devices. (3) The binary output of the comparator is used to increment or decrement a digitally controlled calibration current into the node, updating at each clock pulse until the offset is within the resolution of the comparator. (4) After a predetermined number of cycles, the final calibration word is saved and the current is maintained until the next calibration cycle. Normal operation then resumes.

Calibration of an analog reconstructor normally takes place when the equipment is first switched on, and at subsequent intervals, depending on the operational environment. A study of this technique at Itek

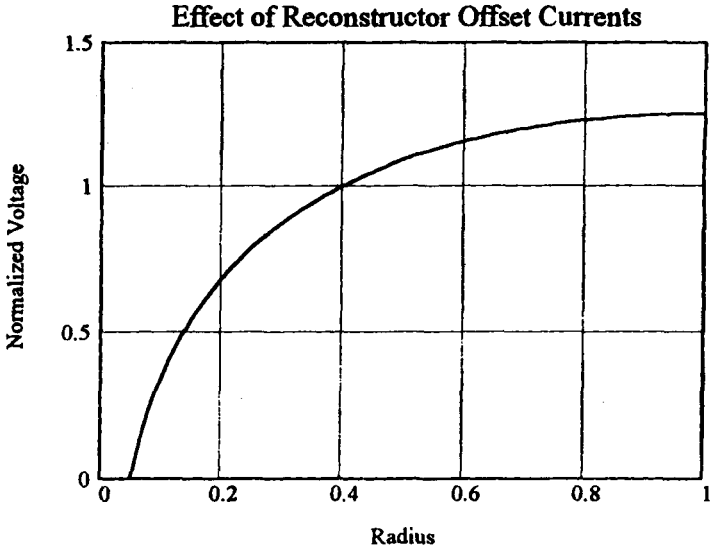


Figure 8.13 Voltage profile produced by uniform offset currents in an analog reconstructor. To obtain the actual offset voltage, the normalized scale should be multiplied by $(\pi/16)(N^2)RI$, where N is the number of nodes across the diameter, R is the internode resistance, and / is the offset current; $r_{min} = 0.05$.

Optical Systems in 1987 showed that sparse calibration of one node in each 5×5 block is an effective and economical procedure.

8.3.4 Matrix Multiplication

Matrix multiplication is the most direct and flexible method of wavefront reconstruction in adaptive optics. This method implements the process defined in equation (8.10), in which the required wavefront values are obtained by weighting and summing the measured gradients. The operation is formally described as

$$w_i = \sum_{k=1}^{k=B} b_{i,k} s_k \qquad (8.49)$$

where

- w_i = elements of the wavefront array W, which is of order A
- s_k = elements of the gradient array S, which is of order B
- $b_{i.k}$ = elements of the matrix B, which is of order $A \times B$.

Implementation of this equation requires $A \times$ multipiy-add operations.

This is evidently a general solution because the wavefront value at every node contains a contribution from every measured gradient. As there are no restrictions on the weights that can be used, equation (8.49) must include every possible linear solution, including those using optimal reconstruction, as described in section 8.6. The process is applicable to zonal or modal systems, using any criterion for minimizing the wavefront error. Computation is performed digitally so there is no limit to the precision that can be achieved.

The major difference between matrix multiplication and iterative solutions is in the complexity of the algorithms used. Iterative solutions employ simple algorithms that require many repetitions to converge to a solution, whereas the matrix solution employs a complex algorithm in return for a single iteration. There is a continuum of tradeoffs between these two extremes, but only the two limiting cases have any practical value. Simple algorithms for which iterative solutions are most useful are found in control matrices with diagonal structure, representing reconstruction schemes in which there is little coupling between nodes. The other extreme is represented by dense control matrices with many off-diagonal elements, typical of optimal reconstruction algorithms, for which matrix multiplication is best suited.

The required weighting functions are held in the control matrix B, which is computed a priori and stored in memory. It is also possible to change the parameters of the control law in real time by updating the elements of B in response to changes in the reference source brightness or turbulence conditions.

The main problem in implementing wavefront reconstructors using matrix multiplication is the large amount of data to be handled in a limited time. In a system with N x N actuators, the dimension A has a value of N^2 and B is approximately $2(N - 1)^2$, so the matrix multiplication involves $2N^2(N \sim \backslash)^2$ multiply–add operations. For a 16×16 wavefront array, this requires 115,200 multiply-adds, which typically must be completed in less than $100\,\mu s$. This task is usually accomplished using special-purpose parallel processors.

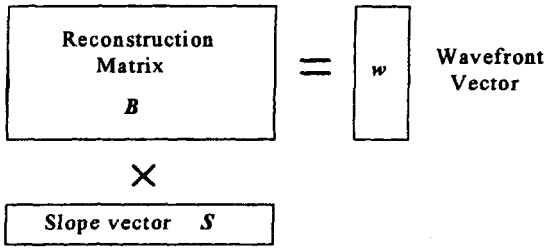
The standard approach is to use a parallel array of multiplier-accumulator circuits (MAC) as described by Sasiela and Mooney [1985]. The operation is shown schematically in figure 8.14. The matrix B has A rows corresponding to the wavefront elements w, and B columns corresponding to the gradients s. The wavefront values w are obtained by multiplying each gradient s by the corresponding element of B and accumulating the result. The simplest scheme is to assign one MAC to each wavefront node, as shown. In this case, each MAC performs B operations for each set of gradient measurements, multiplying each value of s by the corresponding element in its assigned row of B, and summing the result.

In general, the number of MACs required depends on their clock speed C, the required reconstruction time T, and the number of nodes N. For a wavefront reconstruction system with N x N nodes, the minimum number of MACs required is

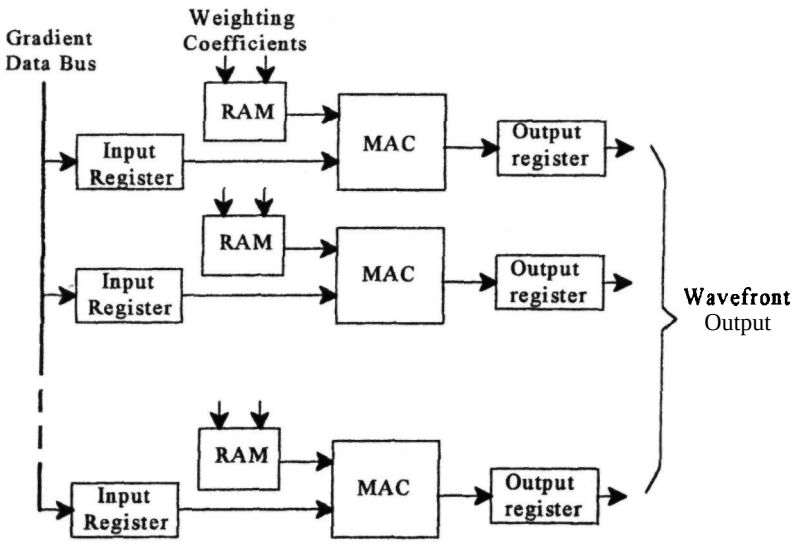
$$M = \frac{AB}{CT} = \frac{2N^2(N-1)^2}{CT} \tag{8.50}$$

The reconstruction time must allow for overhead operations, such as unloading data from the output registers. Currently available multiplier-accumulator devices operate at clock rates of about 30 MHz. For a net reconstruction time of $T = 1 \times 10^{-4}$ seconds, using one 30-MHz MAC per node, the maximum array size is $N = 39$ nodes and the number of MACs required would be $N^2 = 1521$. With smaller arrays, the MACs may be shared between nodes. For example, a 16 x 16 array would require only 43 MACs, with each servicing six nodes.

In a "brute force" reconstructor of this type, the number of multiply–add operations is proportional to the square of the number of nodes in the array, or N^4. For large arrays, a considerable amount of hardware is required. Consequently, there is a strong incentive to find methods of simplifying the opera-



(a) Matrix operations



(b) Block diagram

Figure 8.14 Wavefront reconstruction using matrix multiplication.

tion; basically, this involves reducing the power-law dependence of the number of multiply–add operations. One approach, the exponential reconstruction algorithm, is described in section 8.2.3. Another approach is to use the wavefront statistics (structure function) to reduce the influence of gradients far from the node being computed. Small values can then be truncated with minimal loss of accuracy. The use of wavefront statistics is one of the elements of optimal wavefront reconstruction described in section 8.6.

8.3.5 Calibration of Actuator Response Functions

An advantage of the matrix multiplication approach to wavefront reconstruction is its flexibility in modifying the control law to accommodate variations in operating conditions and hardware status. A useful procedure is to self-calibrate the actuator response functions, which allows the actual influence function of each actuator (rather than an estimate or average) to be used in the reconstruction process. Calibration is implemented by driving or "poking" each actuator, in turn, when the adaptive optics system is operating with a clean input wavefront, usually generated by an internal (self-test) reference source. The wavefront sensor responses for each actuator are then separable and can be saved. The resulting data are then used to establish the elements of the reconstruction matrix.

8.4 Wavefront Prediction

8.4.1 Strategies for Wavefront Prediction

Time delays are inevitable in adaptive optics systems, partly because of the finite integration time required by the wavefront sensor to collect photons from the reference source, to which must be added pure (non-integrating) time delays, such as those due to reading out a charge-coupled device (CCD) detector, processing the wavefront data, and setting the wavefront corrector. Time delays between the measurement and correction of a wavefront produce a *temporal* or *prediction* error, the magnitude of which depends on the dynamics of the turbulence. The contributing factors include transportation of turbulent layers by winds, the wind vectors often being different for layers at different heights, together with the motion or evolution of the turbulence itself. The correlation of atmospherically distorted wavefronts decays with time, the time constant being typically between 1 and 10 ms, depending on the turbulence strength, wind speed, and wavelength.

In principle, the effects of time delays can be compensated by using a prediction scheme to anticipate the corrections that will be required in the next cycle. Wavefront prediction may be based on data measured in previous sampling periods and on data from surrounding subapertures, as well as data from the subaperture in question. There are three main prediction strategies:

1. Zero-order prediction, in which the current value of the wavefront is used as the best estimate for the next period. This is the optimum strategy for the random walk scenario.
2. First-order or linear prediction, based on extrapolation from recent samples. This is the optimum strategy for a signal contaminated with white noise; for adaptive optics, the predicted wavefront may be a function of both time and position.
3. Nonlinear prediction, as implemented by a neural network, which is appropriate for a chaotic process.

Most closed-loop adaptive optics systems use the zero-order prediction strategy, generally implemented in the form of an integrator that holds the current correction, usually with a decay factor, until the next measurement is available. This method requires the wavefront sampling interval to be considerably shorter than the decay time of the turbulence.

Two methods of wavefront prediction beyond simple integration or time smoothing have been investigated for use in adaptive optics. The first of these is linear prediction, using measurement of the wind vector that transports turbulence across the telescope aperture. A "traveling wave predictor," designed to propagate wavefront measurements across the aperture with the correct speed and direction, is described in the following section. Linear prediction is appropriate for processes that can be modeled as band-limited white noise, of which Kolmogorov turbulence is an example. It could be of considerable value at good astronomical sites, where the turbulence occurs mostly in well-defined layers in the vicinity of the tropopause.

Nonlinear prediction for adaptive optics has been investigated in the form of neural networks. Jorgenson and Aitken [1992, 1993] state that there is evidence that atmospheric turbulence sometimes behaves like a chaotic process, in which case a nonlinear predictor, such as a neural network, would be the appropriate choice. They found, however, that with severe turbulence, the largest gain was obtained in going from a conventional zero-order predictor to a first-order (linear) predictor, with only a small further gain for the neural network. With low turbulence, there was little difference between the three methods.

8.4.2 Traveling Wave Predictor

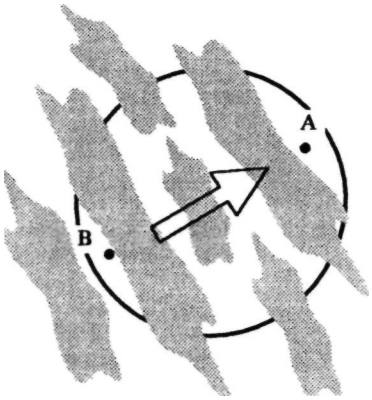
A linear predictor of this type was proposed for use in the Compensated Imaging System in 1978. The

CIS employed a 1.6-m optical telescope and was designed for imaging of low-Earth-orbit satellites, which typically have angular velocities of 10-25 mrad s when viewed from Earth's surface. The large slew rate required high sampling rates, with a compensation bandwidth approaching 1 kHz. Because of the high slew rate, the turbulence structure appears to move across the telescope aperture with consistent speed and direction. The traveling wave predictor was designed to propagate wavefront measurements across the aperture at a given speed and direction, with a decay factor to take account of the decorrelation with time. Further measurements were added to the traveling wave in its journey across the aperture to improve its precision so that the average compensation was enhanced. Analysis predicted a modest improvement in image resolution, but this could not be verified because the traveling wave predictor was never reduced to practice.

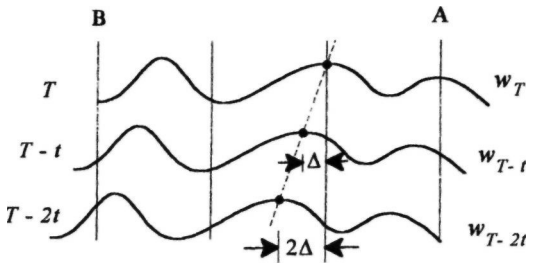
There is experimental evidence that atmospheric turbulence occurs in well-defined layers traveling with consistent speed and direction. For example, data obtained by Vernin and Roddier [1973] shows two layers of turbulence at 5 km and 10 km, traveling at velocities of 7 and 27 $m\,s^{-1}$, respectively. The structure and motion of turbulence may be visualized directly by looking at a star through a telescope (even a small one), using a knife-edge at the focus, as described by Babcock (see chapter 1). When the turbulence structure persists for a significant fraction of its journey across the telescope aperture, it becomes possible to use spatiotemporal wavefront reconstruction, in which wavefront measurements are propagated across the aperture to make a better estimate of the wavefront.

The principle of the traveling wave predictor is shown in figure 8.15. Wavefront disturbances traversing the telescope aperture pass across the subapertures of the wavefront sensor with a velocity vector determined by the wind speed and direction. In normal operation, each wavefront measurement is independent, using whatever photons are collected during the detector integration time. When the wind velocity is high, the integration time must be kept short to minimize the measurement error due to motion of the turbulence. If the turbulence structure remains "frozen" while it traverses the aperture, then many additional samples may be obtained by combining previous measurements, with the appropriate spatial shift. This process increases the effective exposure time and consequently improves the signal-to-noise ratio of the wavefront measurement.

Subapertures on the lee side of the telescope aperture, *A*, have the most favorable location, because the turbulence can be tracked across the full diameter, producing a large increase in the effective exposure time. At the least favorable location *B*, on the windward edge of the aperture, the traveling wave predictor produces no improvement. The average distance of all points within an aperture, measured from one edge in one direction, is $0.5D$ for a square aperture and $\pi D/8 = 0.393D$ for a circular aperture, so in either case a significant increase in the average signal-to-noise ratio is obtained, resulting in lower wavefront measurement errors. To implement a system of this kind, it is necessary to have a priori knowledge of the wind vector. In practice, a decay or weighting factor is used to take account of the decorrelation of the wavefront structure with time, reducing the contribution of earlier measurements.



(1) Turbulence motion across the telescope aperture



(2) Wavefront profiles along the motion vector

Figure 8.15 Principle of the traveling wave predictor. Wavefront measurements made along the motion vector at times $T - t$, $T - 2t$, etc. are shifted by A, 2A, etc., and included in the current measurement.

The wavefront estimate w_T^x made at location x and time T, using a traveling wave predictor, is formally described as

$$\hat{w}_T^x = a_0 w_T^x + a_1 w_{T-t}^{x-\Delta} + a_2 w_{T-2t}^{x-2\Delta} + \cdots + a_m w_{T-mt}^{x-m\Delta} = \sum_{n=0}^{m} a_n w_{T-nt}^{x-n\Delta} \tag{8.51}$$

where

a = decay factor
w_T^x = measured wavefront at location x and time T
t = measurement interval
A = spatial wavefront shift between measurements = vt
v = effective wind velocity
m = number of measurements used for estimate

The improvement in effective exposure time, brought about by the use of traveling wave prediction, is easily estimated for the ideal case of frozen turbulence, when there is essentially no change in the structure as it traverses the telescope aperture. If d is the subaperture dimension at the primary mirror, then the time required for the turbulence to traverse one subaperture is d/v seconds. It is easily shown that the *average* distance of points within a circle of diameter D from the edge (in one axis) is $\pi D/8 \simeq 0.39D$. For a system with N subapertures across the diameter, the wavefront can therefore be measured over an average distance of $0.39N$ subapertures. The effective exposure time t_e is therefore

$$\begin{aligned} t_e &= \text{average number of subapertures used} \\ &\quad \times \text{ time to traverse one subaperture} \\ &= 0.39N \times \frac{d}{v} \end{aligned} \tag{8.52}$$

The maximum exposure time without a traveling wave propagator is approximately equal to the measurement interval t, so the improvement factor F_{TWP} is given by

$$F_{TWP} = \frac{t_e}{t} = \frac{0.39Nd}{vt} \tag{8.53}$$

For a typical case with $N = 10$, $d = 0.2$ m, $v = 10\,\text{ms}^{-1}$, and $t = 0.002$s, F_{TWP} has a value of 39. The corresponding improvement in wavefront sensor signal-to-noise ratio is $\sqrt{F_{TWP}}$. Under realistic conditions, the actual improvement would be less than this, but the example shows the potential of this prediction technique.

8.5 Adaptive Optics Control Systems

8.5.1 *Introduction*

The control system is the vital link between the wavefront sensor and the wavefront corrector in an adaptive optics system. There are several ways in which this control function can be implemented. The most common is a negative feedback loop in which the incoming wavefront is first compensated by the wavefront corrector. The residual error after compensation is then measured and a proportional signal is fed back to the wavefront corrector, with the correct polarity to reduce the error. With this arrangement, the wavefront sensor need only detect small deviations away from the null. Another arrangement is to use an open-loop or feed-forward configuration in which the wavefront is measured before it is corrected; this requires precise measurement capability over a large dynamic range. Each of these systems has its advantages and disadvantages. Closed-loop control systems do not require precise calibration (except to establish the null), but their temporal frequency response is limited by the need to avoid oscillation in the feedback loop. Open-loop control systems are free of this limitation, but, in this case, it is necessary for the sensor and corrector to have a linear response over the entire dynamic range of the signal to be compensated, a requirement that is difficult to achieve in practice.

Adaptive optics control systems can be based on either zonal or modal wavefront compensation; the standard approach is to use modal compensation for low-order errors, such as overall tilt, and zonal compensation for the higher order wavefront errors. The control algorithms employed in adaptive optics usually have some capability for updating their parameters in real time, to accommodate changes in operating conditions. While most control systems employ linear operation, some experiments have been made with neural networks, which employ nonlinear algorithms. These alternative approaches are discussed briefly later in this section.

The theory of automatic control is well established and standard design procedures are adequate for the design of adaptive optics systems. From the control system viewpoint, adaptive optics wavefront compensation systems may be treated as linear, time-invariant feedback systems. To provide the context for this section, the basic equations used in the analysis of linear control systems will be reviewed briefly. Their application to adaptive optics control systems will then be described.

When the data are continuous with time, they may be analyzed in the time (or frequency) domain using Laplace transforms. The Laplace transform of a function of time $f(t)$ is given by

$$\mathcal{L}[f(t)] = \int_0^\infty dt f(t) e^{-st} = F(s)$$

where the complex variable $s = a + j\omega$. The Laplace transform $F(s)$ is a continuous function of time, making it suitable for use with analog systems. The transfer function of a system may be defined as

$$G(s) = \frac{\mathcal{L}[c(t)]}{\mathcal{L}[u(t)]}$$

where $\mathcal{L}[c(t)] = C(s)$ is the system output produced by the input $\mathcal{L}[u(t)] = U(s)$. The value of the Laplace transform is that it enables the response of systems composed of a number of interconnected subsystems to be determined readily. For example, if a number of subsystems with transfer functions G|(s),

$G_2(s)..G_n(s)$ are connected in series, then the output $C(s)$ is given by the product of the input $U(s)$ and the individual transfer functions:

$$C(s) = G_1(s)G_2(s)\cdots G_n(s)U(s)$$

For subsystems connected in parallel, the individual transfer functions are added.

When the data are sampled in time, which is the usual case for digital processing, control system analysis is facilitated by using the Z-transform, which is a generalized version of the $\mathcal{L}$-transform. The variable z is defined as $z = e^{sT}$, where T is the sampling period, and the Z-transform of a time-sampled function $u^*(t)$ is given by

$$Z[u(t)] = \mathcal{L}[u^*(t)]_{e^{sT}=z} = \sum_{n=0}^{\infty} u(nT)z^{-n}$$

These transforms enable the performance of single-loop control systems to be determined from a knowledge of the transfer functions of the components in the feedback loop.

Most adaptive optics systems employ a closed-loop control system to minimize the wavefront error. An adaptive optics control loop may be represented as shown in figure 8.16(a). The quantity being controlled is the shape of the output wavefront. This is usually a vector quantity, but for the present purpose we consider it to be a single scalar value. The basic function of the system is to control the shape of the output wavefront $C(s)$ in sympathy with the reference signal $R(s)$, in the presence of wavefront disturbances $W(s)$. In the adaptive optics context, the transfer function $H(s)$ represents the wavefront sensor, the output of which is compared with the reference signal $R(s)$ to produce the error signal $E(s)$. The error is multiplied by the transfer function $G(s)$, which, includes the wavefront processor, reconstructor, predictor, and wavefront corrector, to produce a signal $B(s)$ that is added to the input wavefront. Ideally, $B(s)$ should be exactly equal and opposite to the shape of the input wavefront $W(s)$, so the output $C(s)$ will exactly follow the value of the reference $R(s)$. If this is not the case, an error signal $E(s)$ is generated, driving $B(s)$ until it again neutralizes $W(s)$.

In adaptive optics systems, the objective is to maintain a flat wavefront at the output, so that the controlled values of $C(s)$ and the reference $R(s)$ are both zero. The control loop can then be considerably simplified, as shown in figure 8.16(b). All of the components in the feedback loop are now included in the transfer function $G(s)$, which is designed to make $B(s)$ follow the input wavefront disturbances $W(s)$ as closely as possible. Adaptive optics systems usually consist of many parallel feedback loops of this type. Initially, each loop will be considered to operate independently. The effect of coupling between the loops is considered later.

The performance of the system in figure 8.16(b) is completely specified by $G(s)$, which contains the cascaded transfer functions of all elements in the loop. The relation between the input and output is

$$E(s) = W(s)\frac{1}{1+G(s)} \tag{8.55}$$

Thus, to minimize the wavefront error, the transfer function $G(s)$ should be as large as possible.

Feedback systems are classified broadly into three categories, depending on the dominant temporal response in the loop. If G is independent of time (that is, constant with temporal frequency), the servo is type 0, in which the error is $E_0(s) = W(s)/(1+G)$. The error is minimized by using a large value of G, but there is always a finite

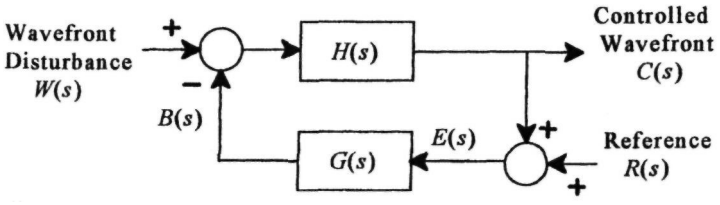


Figure 8.16(a) Conventional representation of adaptive optics control system.

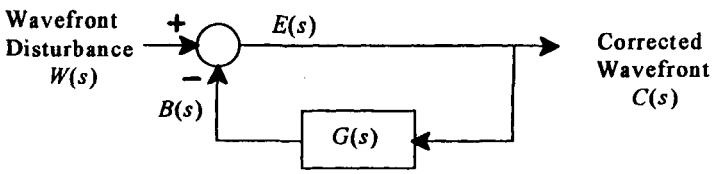


Figure 8.16(b) Simplified adaptive optics control loop.

error, proportional to the wavefront $W(s)$, even with a steady-state input.

To eliminate the steady-state error, a single integration may be included in the loop, producing a type-1 servo. A constant error then produces a constant rate of change of the correction function $B(s)$. With a constant input, $B(s)$ will increase or decrease until it is equal to $W(s)$, at which point the error is reduced to zero and the integration stops, maintaining $B(s)$ at the required value. If $W(s)$ changes, then it is obvious that the ability of $B(s)$ to follow these changes depends on the time constant of the integrator, which therefore dominates the temporal response of the adaptive optics system. Closed-loop adaptive optics systems normally use type-1 servos.

If two integrations are included in the loop, the system is known as a type-2 servo. In this case, a constant error produces a constant acceleration in the correction $B(s)$. The error is then reduced to zero, even when the input wavefront changes at a constant rate. Type-2 servos are difficult to implement because of the stability problem. Each integration can produce a phase shift of up to 90°, and if the gain in the loop is greater than unity when the phase shift reaches 180° then oscillation will result. The components in the feedback path of adaptive optics systems contribute additional phase shifts that make the application of type-2 servos questionable for this application.

Adaptive optics systems employ a number of parallel feedback loops, typically two modal (whole-aperture) paths for compensating x and y tilt, plus a large number of identical paths for zonal compensation of the wavefront. Strictly speaking, such systems cannot be represented completely using transform theory because of interactions between the loops. However, if the multiple loops are made to operate independently, the design of control systems for adaptive optics is considerably simplified. In this case, the design problem boils down to two loops: one for overall tilt control and one for control of the wavefront zones.

For complete parametric control of factors varying in real time, the state variable approach is appropriate. The state of a system is defined as the minimal amount of information required to determine the future response from a given input. This information takes the form of a vector x having n variables, where n is the number of initial conditions defining the state at time t_0. Instead of considering only one independent variable, as in the case of transform theory, the state variable approach considers n variables. To characterize such a system, the transfer function is replaced by n simultaneous first-order differential equations, relating the input, output, and state variables. The rate of change (denoted by the overdot) of the state vector x may be expressed in matrix form as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{8.56}$$

where u is an m-dimensional vector representing the inputs to the system, A is an n x n matrix and B is an n x m matrix. If x represents the initial value of the state vector, and u represents the controls applied to the system over a time period, then equation (8.56) defines the trajectory of x in the state space as a function of time.

8.5.2 Closed-Loop Control Systems

The specific characteristics of the feedback paths employed in adaptive optics systems will now be examined in more detail. The implications of using a wavefront sensor that produces output data at discrete time intervals is considered first. Periodic sampling of the detector output at time intervals t_f generates multiple sidebands at frequencies $\pm n\omega_f$ where $\omega_f = 2n/tr$, as shown in figure 8.17. To avoid overlap, the spectrum of the signal to be controlled should not exceed $\omega_f/2$. The sidebands produce unwanted high-frequency interference on the control signal, so it is necessary to remove them by spectral filtering. In analog systems, this is done with a low-pass filter. The digital equivalent is a "hold" circuit, which reconstructs an approximation to the smoothed turbulence spectrum. Zero-order hold circuits, which simply maintain each value of the signal until the next sample arrives, are commonly used for this purpose. The transfer function of a zero-order hold circuit with sampling interval T is

$$G_{0h}(s) = \frac{1 - e^{-sT}}{s} \tag{8.57}$$

The operation of a sample-and-hold circuit is depicted in figure 8.18. For $sT \ll 1$ (that is, when the sampling time T is much smaller than the period of the signal), a zero-order hold may be approxi-

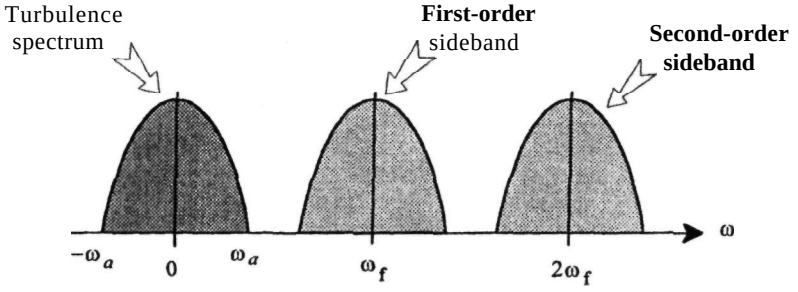


Figure 8.17 Frequency spectrum produced by sampling.

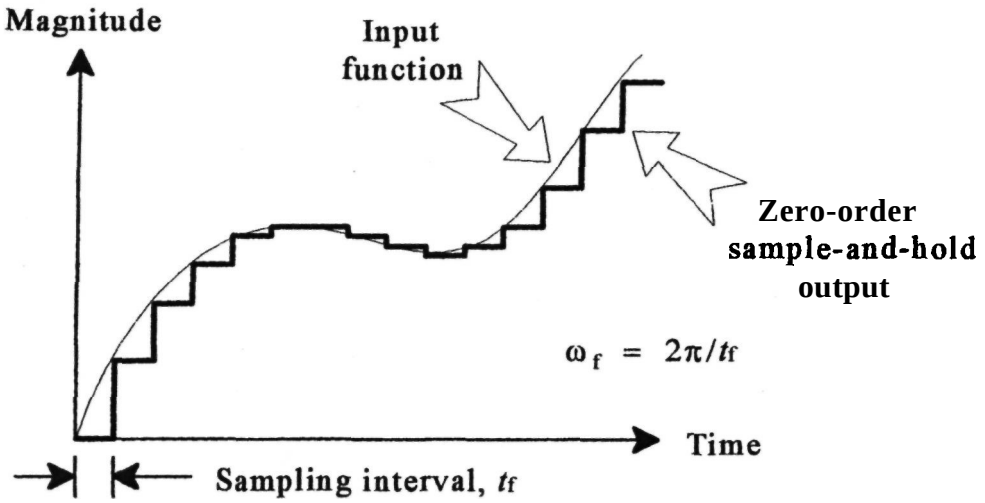


Figure 8.18 Operation of a zero-order sample-and-hold circuit.

mated as a simple *RC* low-pass filter with $T = RC$, producing a phase shift of $\phi = -\tan^{-1}\omega T$.

The effects of sampling on the control signal are diminished when the sampling frequency is much higher than the control bandwidth. In this case, the control loop can be evaluated as a continuous process rather than as a discretely sampled process, providing that all phase shifts produced in filtering out the sampling frequency are properly accounted for. This approach greatly simplifies the control system analysis and is used in the following discussion.

The functions in a typical adaptive optics system are depicted in figure 8.19. The feedback loop includes both optical and electronic functions. Although there may be several components in the optical path, they can be grouped into two categories: those before the beam splitter and those after. Components before the beam splitter are common to both the control path and the imaging path. "Common path" errors in these components are detected by the wavefront sensor and are consequently corrected by the feedback loop. This category includes optical aberrations in the overall tilt corrector (fast-steering mirror) and the wavefront corrector (deformable mirror). However, optical aberrations occurring in or after the beam splitter, whether in the control loop or the imaging path, produce differences between the sensor and image and therefore appear as errors in the optical output. In most cases, it is possible to calibrate the wavefront

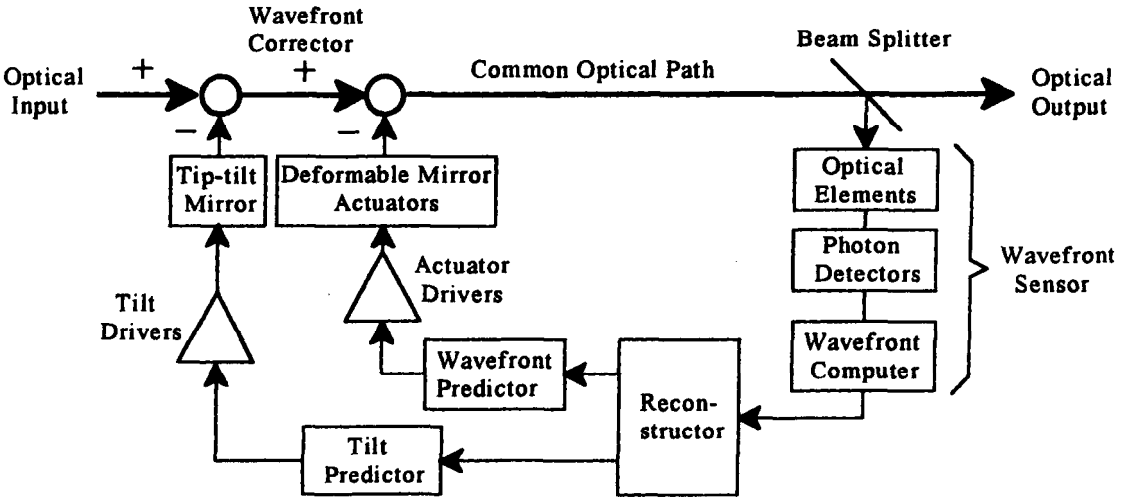


Figure 8.19 Components of an adaptive optics control loop.

sensor to eliminate its internal optical and alignment errors. Uncorrectable errors include not only those outside the loop, but also wavefront errors within the loop that are of higher spatial frequency than the actuator spacing cutoff. Such errors are not covered by the following analysis, but must be included in the overall error budget when the system performance is evaluated (see chapter 9).

Most adaptive optics systems employ separate modal correction of angle-of-arrival variations, using a two-axis fast-steering mirror. The drive signals for this corrector are separated out in the reconstruction process, as indicated in figure 8.19. The wavefront correction loop actually consists of a number of parallel channels (several hundred in large systems), one for each actuator in the wavefront corrector. The control functions in each loop may be modeled as shown in figure 8.20. The transfer function of each component will now be described.

Parameters of the wavefront sensor are defined as follows:

Transducer gain, g\, volts per meter

Exposure time, t_e, seconds

Readout time, t_s, seconds

Frame (sampling) period, t_f, seconds

Wavefront processing time, t_c, seconds

From the control system viewpoint, wavefront sensors consist of three main components: an optical "front end" that determines the transducer gain, a photon detector that determines the temporal response, and a digital processor that contributes a time delay. The transducer gain is specified as the voltage produced by a gradient of 1 meter per subaperture. (The actual gradients only amount to micrometers per subaperture.) In the case of a Shack-Hartmann sensor, an output of V volts per radian is equivalent to a transducer gain of $g_1 = V/d$ volts per meter, where d (meters) is the subaperture size. The time line for a frame-transfer CCD is shown in figure 8.21. The whole CCD is exposed for the time t_e required to collect the necessary number of photons from the reference source. At the end of this period, the accumulated charges are transferred rapidly to a shielded readout array, from which they are scanned out in sequence, converted to digital form, and held in memory.

The time t_s required to read each frame of CCD data determines the maximum sampling rate of the wavefront sensor. The frame (sampling) period t_f must equal or exceed the greater of the read time t_s or the exposure time t_e. (Ideally, if $t_e > t_s$, the read time should be increased to match the exposure time, in order to minimize the read noise.) The exposure time t_e is a true integration for the whole array, but the readout time t_s is a pure time delay, because the data are unchanged during the readout process.

The CCD data are then transferred to the wavefront processor, which computes the gradients corresponding to the intensity data. In principle, some of the time delay could be saved if pipeline processing was used to compute the gradients as they are read out from the CCD array. However, in Hartmann-type sensors, the data required to compute the gradients in each subaperture are spread over at least two lines of the scan, so the sample-and-hold function is not eliminated entirely. Using batch processing with a wavefront computation time of t_c, as shown in figure 8.21, the wavefront sensor model for each channel consists of a transducer gain of g_1,

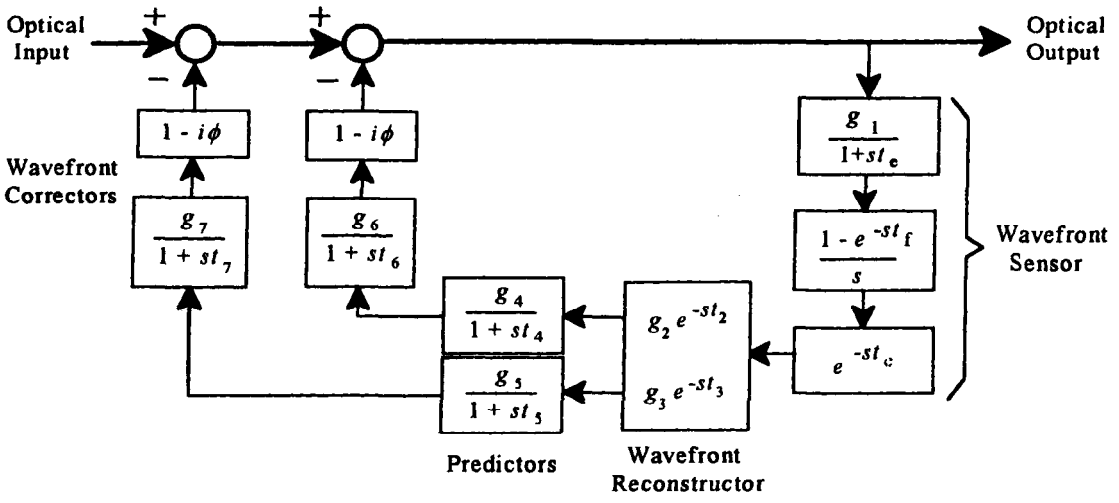


Figure 8.20 Transfer functions of an adaptive optics control loop.

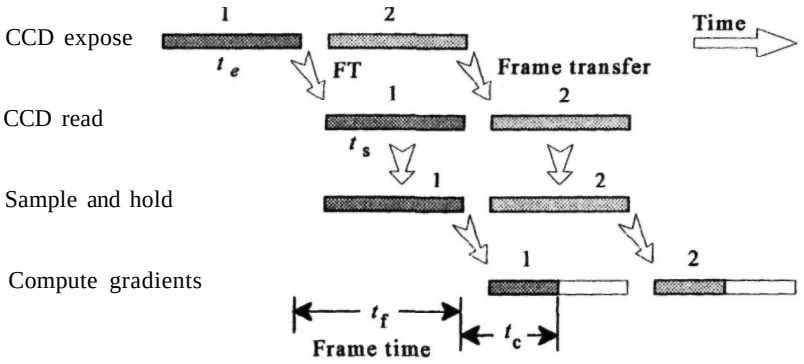


Figure 8.21 Time line for wavefront sensing using a frame-transfer CCD.

an integration over time t_e, a sample-and-hold of period t_f, and a computation time delay of t_c. The transfer function of the wavefront sensor is then

$$G_1(s) = \frac{g_1}{1 + st_e} \frac{(1 - e^{-st_f})}{s} e^{-st_c} \tag{8.58}$$

The reconstructor is usually a massively parallel processor that accepts two vectors from the wavefront sensor representing the x- and y-axis gradients. Its output is a vector representing the estimated wavefront, together with the overall x- and y-axis wavefront tilts. It is simply modeled with a gain of g_2 for each wavefront correction channel and a gain of g_3 for each tilt correction channel. The delay times are t_2 and t_3, respectively. The reconstructor transfer functions are

Wavefront correction:

$$G_2(s) = g_2 e^{-st_2} \tag{8.59}$$

Tilt correction:

$$G_3(s) = g_3 e^{-st_3} \tag{8.60}$$

Most adaptive optics systems employ zero-order prediction, implemented as a "leaky" integrator, with time constant t_4 for wavefront correction and t_5 for tilt correction. These time constants are generally the largest in the control loops and dominate their temporal response. The predictor gains g_4 and g_5 likewise dominate the gains in the control loops. The predictor transfer functions are

Wavefront correction:

$$G_4(s) = \frac{g_4}{1 + st_4} \tag{8.61}$$

Tilt correction:

$$G_5(s) = \frac{g_5}{1 + st_5} \tag{8.62}$$

The wavefront correctors themselves are at the end of the chain, converting the electrical signals provided by the control loop into a physical or optical pathlength change. Each actuator-driver combination is modeled as a subsystem with three basic parameters: the transducer gain g_6 or g_7, the temporal bandwidth f_6 or f_7, and the hysteresis phase angle fa. The relation between the 3-dB bandwidth (the frequency at which power is reduced to half, equivalent to a voltage reduction of $1/\sqrt{2}$) and the time constant is $f_{3dB} = 1/2\pi t$.

Hysteresis is present in many types of actuator, especially piezoelectric, and can have a significant effect on performance. It affects both the modulus and the phase of the response. Hysteresis may be modeled as a phase shift that is independent of frequency. A typical hysteresis loop is similar to the Lissajous figure produced by plotting two cyclic waveforms on orthogonal axes. When the two functions are in phase, the resulting plot is a straight line at an angle of 45°, corresponding to zero hysteresis. If a phase shift exists between the functions, the line splits into a loop in which the horizontal (or vertical) separation at the center is proportional to the sine of the phase angle, as shown in figure 8.22. When the two functions are at 90°, the figure becomes a circle. Using this approach, hysteresis may be approximately modeled as a phase delay equal to $fa = \sin^{-1}(a/b)$, where a is the horizontal separation of the hysteresis loop and b is the peak-to-peak amplitude. The transfer function of the wavefront corrector is modeled as

Wavefront correction:

$$G_6(s) = \frac{g_6}{1 + st_6}(1 - i\phi_h) \tag{8.63}$$

Tilt correction:

$$G_7(s) = \frac{g_7}{1 + st_7}(1 - i\phi_h) \tag{8.64}$$

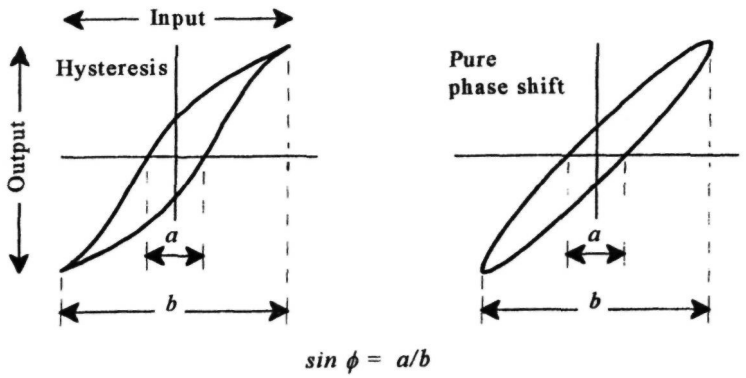


Figure 8.22 Estimation of phase shift due to hysteresis.

In addition to these parameters, nonlinearity of the wavefront corrector response can be corrected using a lookup table.

To obtain the overall transfer function of the adaptive optics control loop, the individual functions are cascaded. The overall transfer functions of the wavefront correction and tilt correction loops are then

$$G_{\text{wave}}(s) = G_1(s)G_2(s)G_4(s)G_6(s) \tag{8.65}$$

$$G_{\text{tilt}}(s) = G_1(s)G_3(s)G_5(s)G_7(s) \tag{8.66}$$

If type-1 servos are used, the dominant integrations in the loops are those of the predictor, the time constants of which (t_4 and t_5) are usually adaptively controlled to optimize performance. In the context of adaptive optics, this integration may be regarded as a zero-order predictor, in which the current value of the wavefront correction is the best estimate for the next correction. Methods of prediction have been described in section 8.4.

In the design of adaptive optics systems, two main characteristics of the control loop are of interest: (1) the gain as a function of frequency; and (2) the stability of the system. In each case it is necessary to determine the frequency response, which is found by evaluating the transfer function $G(s)$ along the $j\omega$ axis. The reduction in wavefront error due to the adaptive optics control loop is then determined using equation (8.55). This is just one of the many factors determining the overall residual error in adaptive optics systems. Optimization of system performance is covered in chapter 9. The stability of closed-loop control systems is discussed in the following section.

8.5.3 Stability Criteria

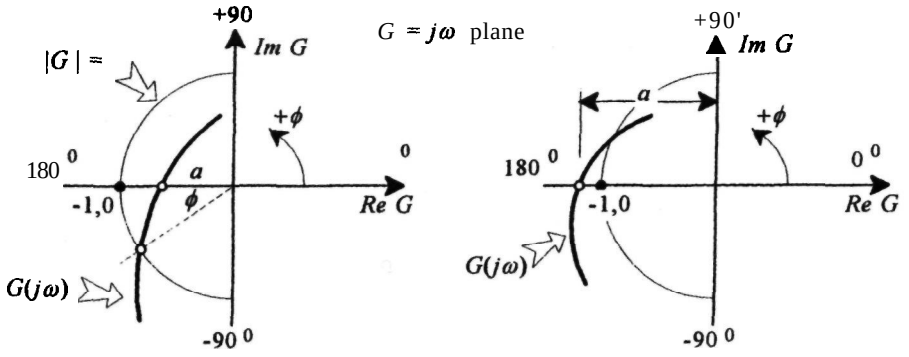
The evaluation of control system stability is a vital part of the design process. Adaptive optics systems are multiple-input multiple-output (MIMO) systems with coupling between the channels. Although such systems have been studied [Maciejowski 1989], reliable stability criteria have yet to be developed. For adaptive optics applications that contain many identical parallel feedback loops, it has been found effective to base the design on a single feedback channel, which is replicated as necessary. The justification for this approach is that in a properly designed adaptive optics system, the zones or modes operate independently; in any case, the simpler approach works well in practice. It is necessary, however, to guard against unwanted coupling between channels, caused by misalignment between the sensor and corrector subapertures. Significant alignment errors may invalidate the control system design. Subaperture alignment tolerances and methods of verification are discussed in section 8.5.4.

Several methods have been devised to verify the stability of single feedback loops, including the Nyquist criterion, the Bode plot, and the log-magnitude versus phase-angle plot, from which the Nichols chart is derived. All of these methods require knowledge of the frequency response of the control system, which is obtained from the transfer function $G(s)$ by evaluating it on the $j\omega$ axis. The output of a system with transfer function $G(s)$, for an input of unit amplitude and frequency ω is

$$C(t) = |G(j\omega)| \sin[\omega t + \angle G(j\omega)] \tag{8.67}$$

The frequency response of a real system may be measured directly and compared with the values predicted from the design in equation (8.67), allowing the performance and stability of the control system to be verified experimentally.

The Nyquist criterion is evaluated from the open-loop response $G(j\omega)$ of a system by plotting the real part versus the imaginary part of $G(j\omega)$ for values of the parameter ω in the range $0 < \omega < \infty$. The system is closed-loop stable if its Nyquist plot does not pass through or encircle the $-1, JO$ point in the G plane. The Nyquist criterion is illustrated in figure 8.23. The semicircle represents a gain of 1 at phase angles between +90° and -90°. The response curve shown crosses the negative real axis (where the phase shift is 180°) at a value of a, and crosses the unity-gain envelope at an angle of ϕ. The system represented by this



(a) Stable system (b) Unstable system

Figure 8.23 Nyquist plots for stable and unstable feedback systems. The criterion for stability is that the plot of $G(j\omega)$ in the complex plane must not enclose the point $(-1, 0)$. The stable system in (a) has a gain margin of $1/a$ and a phase margin of ϕ. The system in (b) is unstable because the gain a is greater than unity at the frequency at which the phase shift is 180°, resulting in positive feedback.

curve is therefore stable because it is well inside the $-1, j0$ point. The *gain margin* is defined as the additional gain, at a frequency ω where the curve crosses the negative real axis, required to make the gain unity; that is, $1/a$. The *phase margin* is the additional (negative) phase shift required to make the phase angle 180° at the frequency where the gain is unity; that is ϕ. These parameters represent the stability margin built into the system; typical values are a gain margin of 6 dB $(a = 0.5)$ and a phase margin of 30°.

The criterion for stability can be expressed in other terms that are more directly related to measurable quantities. For example, the transfer function of any first-order system can be factored into the form

$$G(s) = \frac{K(1 + sT_a)(1 + sT_b)\cdots}{s(1 + sT_1)(1 + sT_2)\cdots} \tag{8.68}$$

The zeros of this function are at $s = -1/T_a, -1/T_b$, and so on, and the poles are $s = -1/T_1, -1/T_2$ and so on. Higher order systems have additional complex poles, but the expression is similar in form. The log magnitude of the frequency response is

$$\begin{aligned}\log|G(j\omega)| = {} & \log K + \log|1 + j\omega T_a| \\ & + \log|1 + j\omega T_b| + \cdots \\ & - \log|\omega| - \log|1 + j\omega T_1| \\ & - \log|1 + j\omega T_2| - \cdots\end{aligned} \tag{8.69}$$

and the phase angle is

$$\begin{aligned}\angle G(j\omega) = {} & \tan^{-1}\omega T_a + \tan^{-1}\omega T_b + \cdots \\ & - 90° - \tan^{-1}\omega T_1 - \tan^{-1}\omega T_2 - \cdots\end{aligned} \tag{8.70}$$

These functions can be used to determine the stability of a feedback control system and also to indicate the compensation needed to modify the gain and phase margins.

The Bode plot is a graphical method of designing closed-loop systems without resorting to extensive calculations. There is a direct relation between the plotted values of magnitude and phase and the components in the feedback loop. It consists of companion plots of $20\log_{10}|G(j\omega)|$ and $\angle G(j\omega)$. Because of the logarithmic scale, the magnitude values of each component are added, making the contribution of each element apparent. A Bode plot of the function $1/(1 + j\omega T)$ is shown in figure 8.24. There is a "corner frequency" at $\omega = 1/T$, equivalent to the pole in the corresponding expression for $G(s)$ in equation (8.62). Above this frequency, the gain falls at an asymptotic rate of 20 dB per decade; that is, as $1/\omega$, corresponding to a single "lag" or *RC* time constant. At the corner frequency, the gain has dropped exactly 3 dB and the phase shift is —45°.

The Bode plot for the inverse function $(1 + j\omega T)$, corresponding to a zero in $G(s)$, is similar in form but is inverted about the zero-gain and zero-phase angle axes; that is, the gain increases at 20 dB per decade and the phase shift is +45°. The effect of additional terms on the frequency response is therefore easy to plot on the chart, using the corner frequencies and adding or subtracting slopes of 20 dB per decade and phase shifts as required. The plot is also generated easily by exact calculation on a digital computer.

The closed-loop stability of a system is readily assessed from a Bode plot of the open-loop transfer function. The Bode plot of a stable feedback system is shown in figure 8.25, and the stability criterion is depicted in figure 8.26. Two critical points are determined: (1) the frequency ω_1 at which the gain plot passes through OdB (unity gain); and (2) the fre-

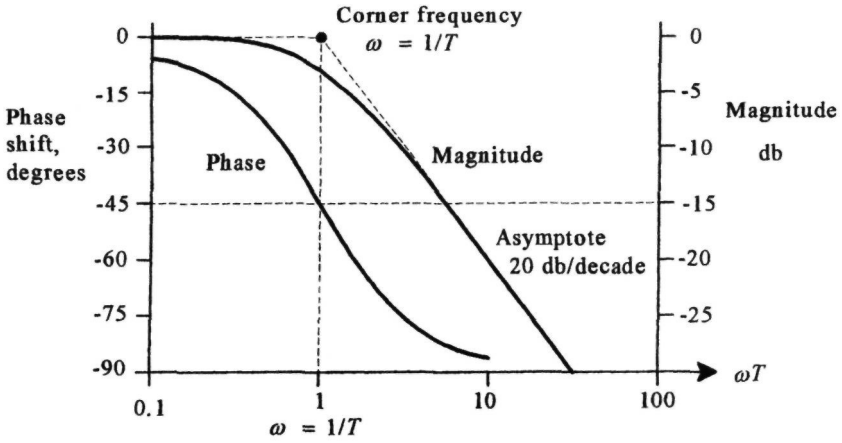


Figure 8.24 Bode plots (log magnitude and phase) of the function $1/(1 + j\omega T)$. At the corner frequency $\omega = 1/T$, the magnitude is -3 dB and phase angle is -45°. For this function, with a single time constant, the gain falls off asymptotically at 20 dB per decade and the phase approaches —90°.

quency ω_2 at which the phase-angle plot passes through -180°. If $\omega_1 < \omega_2$, then the phase lag must be less than 180° at the point of unity gain, so the system is stable. However, if $\omega_1 > \omega_2$, the gain is greater than unity at the point where the phase lag is 180°, so the system is unstable. The gain margin is the absolute value of the gain at ω_2, and the phase margin is 180° less the phase angle at ω_1.

A different type of chart is obtained when the log magnitude of $G(j\omega)$ is plotted directly against the phase angle $\angle G(j\omega)$. In this case, there is no separate frequency axis; each point on the curve describes the magnitude and phase of $G(j\omega)$ at a specific frequency. A log-magnitude versus phase-angle plot of this type is shown in figure 8.27. The chart is centered on the point OdB gain and —180° phase shift. A

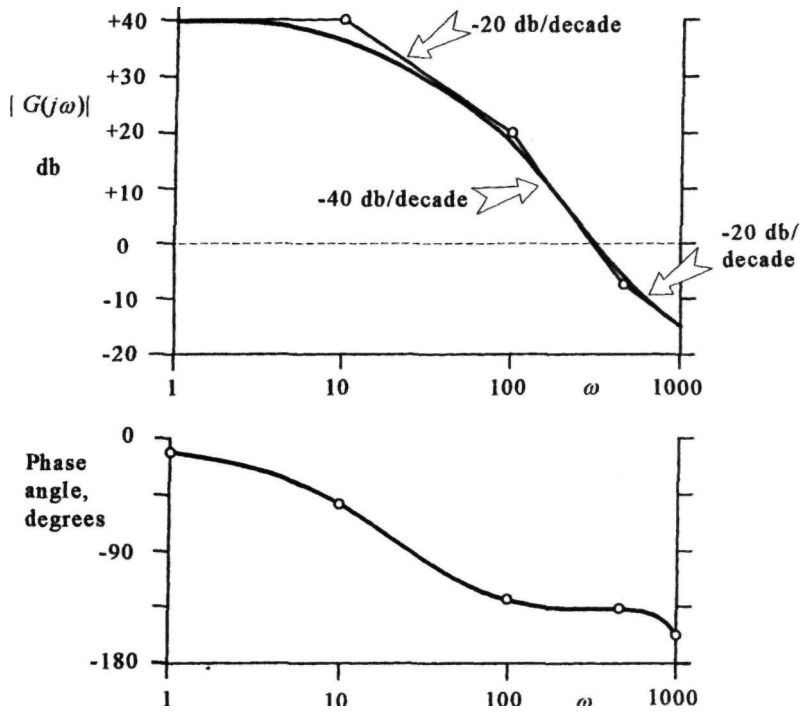


Figure 8.25 Typical Bode plot of a stable feedback system.

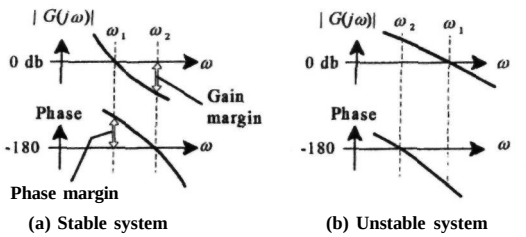


Figure 8.26 Stability criterion for the Bode plot.

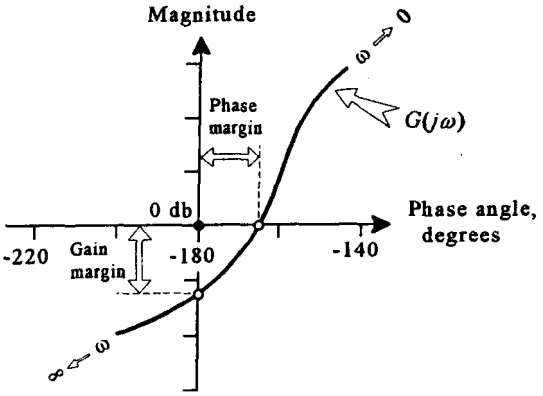


Figure 8.27 Plot of log magnitude versus phase angle in the $G(j\omega)$ plane for a system that is stable when the loop is closed. If the $G(j\omega)$ function passed to the left of the (OdB, -180°) point, the system would be unstable closed-loop.

typical system trajectory is shown, with the gain falling and the phase angle increasing as the frequency goes from zero to infinity. For stability, the plot must pass to the right of the 0 dB, -180° point, as indicated in the diagram. The gain and phase margins are the intercepts on the two axes. If the system trajectory passes through or to the left of the 0 dB, —180° point, then the system will be unstable in closed-loop operation.

The Nichols chart is a modification of the log-magnitude phase-angle plot, having an additional set of axes called the closed-loop axes. For a point x in the log-magnitude phase-angle plot, the Nichols chart provides a set of nonrectangular coordinates giving the values of $20\log_{10}\left|\frac{x}{1+x}\right|$ and $\angle\left(\frac{x}{1+x}\right)$. Thus, if an open-loop frequency response is plotted using the rectangular coordinates (the same curve as that described in the previous paragraph), the new coordinates enable the values of the closed-loop response $20\log_{10}\left|\frac{G(j\omega)}{1+G(j\omega)}\right|$ and $\angle\left(\frac{G(j\omega)}{1+G(j\omega)}\right)$ to be read off on the closed-loop axes.

In the adaptive optics control model, the transfer function $G(j\omega)$ is considered to be in the feedback path and not in the forward path, as is assumed in the Nichols chart (see figures 8.16(a) and (b)). The "closed-loop response" on the Nichols chart is actually the ratio of the correction function $B(j\omega)$ to the wavefront disturbance $W(j\omega)$ in the adaptive optics system. Ideally, this ratio should be near unity, corresponding to a ratio $E(j\omega)/W(j\omega)$ of near zero. These considerations have no effect on the criterion for closed-loop stability, which is based on the rectangular coordinates on the chart.

8.5.4 *Subaperture Alignment*

In zonal adaptive optics systems, it is essential for the subapertures in the sensor and corrector to be aligned accurately. The wavefront sensor and deformable mirror (or other compensation device) are usually separated by an optical path that may be several meters in length. Optical misalignment between corresponding subapertures of the sensor and corrector is caused by physical displacement of optical or structural components caused by temperature or attitude variations. Misalignment inevitably produces cross-coupling between the control channels, reducing performance and ultimately causing instability in the feedback loop.

Two main steps can be taken to ensure correct alignment: (1) the optomechanical design must take into account the full operating environment of the adaptive optics system; and (2) provision should be made to check the optical alignment before, or even during, operation, preferably with an automated routine that intrudes as little as possible on normal observations.

In the previous section, it was pointed out that the conventional tests for the stability of feedback systems, based on the Nyquist plot, were formulated for single feedback loops. These stability tests are valid for the MIMO control systems employed in adaptive optics, only when the optical path is perfectly aligned. It is therefore of considerable interest to determine the amount of misalignment that can be tolerated before significant performance degradation or instability results. For design purposes, an approximate value is sufficient; for example, it is usually sufficient to know whether the allowable misregistration is 1%, 10%, or 50% of the subaperture size.

The answer can be obtained by modeling a single actuator channel in one dimension, as depicted in figure 8.28. Two slope-sensing subapertures of width d are located symmetrically on each side of the actuator center, using the arrangement of network 1. The actuator is assumed to have a triangular profile with base $2d$ and height a. The slope signals s_1 and s_2 are proportional to the difference in actuator deflection across each subaperture. When the system is in perfect alignment, the slope signals are

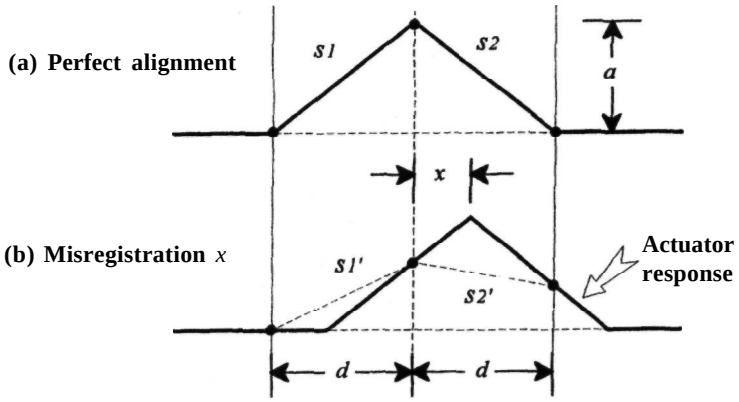


Figure 8.28 Sensor and actuator alignment model.

$$s_1 = Ka$$
$$s_2 = -Ka \tag{8.71}$$

where K is a constant of proportionality. The feedback path corresponding to this model is shown in figure 8.29. The sensor outputs are combined and inverted in polarity to provide negative feedback, so that any deflection of the actuator is countered by a restoring signal. The actuator deflection is then

$$A = -6(5, -s_2) \tag{8.72}$$

where b is the actuator sensitivity.

If the actuator is now misaligned laterally by distance x, the slope signals become

$$s_1' = Ka\left(1 - \frac{x}{d}\right)$$
$$s_2' = -Ka\left(1 - \frac{2x}{d}\right) \tag{8.73}$$

The restoring signal is then

$$-(s_1 - s_2) = -Ka\left(2 - \frac{3x}{d}\right) \tag{8.74}$$

which falls to zero when $x = 2d/3$; that is, when the misalignment is two-thirds of the subaperture spacing. At larger values of x, the feedback becomes positive and the system will be unstable. Although the control loop is modeled crudely in this example, the result is similar to that reported by Merritt et al. [1992], who found that a realistic computer model of an adaptive optics control loop became unstable when the misregistration was 0.62 subapertures.

Of course, one would not want an operational adaptive optics system to get anywhere near the point of instability; the performance would be degraded long before that state was reached. The main effect of small misalignments is to produce cross-coupling between correction channels, which reduces the feedback loop gain and degrades the temporal response. Equation (8.74) shows that a misregistration of 1/10 subaperture reduces the gain of the feedback loop to about 85% of its maximum value, while a misregistration of 1/5 subaperture reduces the gain to 70%. An alignment tolerance of 1/10 subaperture appears to be a reasonable and conservative design goal for practical systems.

When a Shack-Hartmann sensor is used, the subaperture alignment referred to in the preceding paragraph relates the lenslet array of the sensor to the actuator locations on the deformable mirror. Lenslet sizes typically range between 50 and 500 μm, so the required alignment precision can be as small as 5 μm referred to the wavefront sensor. In addition, the detector array in the sensor must be aligned accurately to the lenslets. Because of the small size of the pixels used in CCD cameras, typically about 25 x 25 μm, the alignment requirement in this case is on the order of 1 μm.

The next question to be addressed is how to verify that the subapertures are properly aligned. Merritt et al. [1992] investigated several tests that could be made on the feedback loop transfer function to warn of misalignment and impending instability. They found that in the case of MIMO control systems, tests based on the transfer function were not reliable indicators of system misalignment and that instability could suddenly appear without warning. These results indicate that the best approach is to measure the subaperture alignment directly, which can be done in a straightforward way, rather than

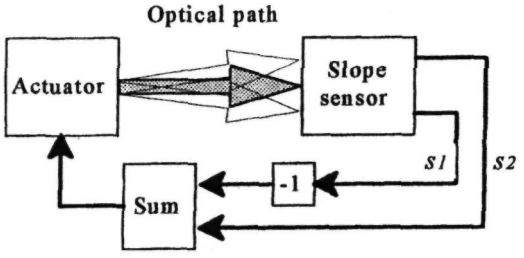


Figure 8.29 Feedback path for alignment model.

to rely on secondary indicators such as the loop transfer function.

The question of alignment is closely related to the actuator calibration, discussed in section 8.3.5. In principle, the relative alignment of sensor and corrector can be accommodated by a calibration procedure in which each actuator is "poked," in turn, and the corresponding wavefront sensor response is recorded. This procedure determines the elements of the matrix A, as shown in section 8.2.2. The control matrix required for the adaptive optics control loop is the inverse of A, which must be computed and stored in memory for use in the reconstruction process. If this process was implemented each time the system was calibrated, then no mechanical adjustment would be necessary, providing that the alignment errors were within certain gross limits. Small alignment errors would be accommodated simply by changing all the elements in the reconstruction matrix. This approach would be suitable for a system that could be relied upon to maintain its alignment during normal observation periods, the matrix being updated just before each session.

In cases where the subaperture alignment is implemented mechanically, it is only necessary to take sufficient data to establish the overall alignment of the sensor and corrector. Assuming that the misalignment consists only of rigid-body motions, there are four degrees of freedom: lateral x and y shifts, rotation angle θ, and radial shift or magnification m. These elements can be measured by "poking" only two actuators, although more accuracy can be obtained by measuring several pairs and averaging the results. The alignment process is illustrated in figure 8.30. The relative displacement of the corrector and sensor are measured at two diametrically opposite locations a and b on the corrector, at radius r, corresponding to two actuator positions. If these displacements are xa, ya, xb, and yb, then the four elements of misalignment are

$$\begin{aligned} \text{Overall } x \text{ shift} &= \tfrac{1}{2}(xa + xb) \\ \text{Overall } y \text{ shift} &= \tfrac{1}{2}(ya + yb) \\ \text{Rotation angle } \theta &= \tfrac{1}{2r}(xa - xb) \\ \text{Magnification } m &= \tfrac{1}{2r}(ya - yb) \end{aligned} \tag{8.75}$$

The displacements xa, ya, and so on, are obtained from the outputs $sx1$, $sx2$, and so on, of the slope-sensor subapertures adjoining the poked actuator, as shown in figure 8.30:

Network 1

$$\begin{aligned} xa &= K(sx2 + sx1) \\ ya &= K(sy2 + sy1), \\ &\text{etc.} \end{aligned} \tag{8.76}$$

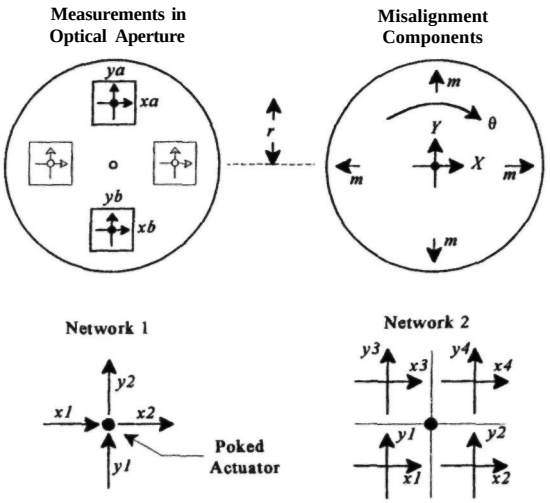


Figure 8.30 Optical alignment of wavefront sensor and corrector using measurements from poked actuators.

Network 2

$$\begin{aligned} xa &= K(sx2 + sx4 + sx1 + sx3) \\ ya &= K(sy3 + sy4 + sy1 + sy2), \\ &\text{etc.} \end{aligned} \tag{8.77}$$

To eliminate any errors due to the wavefront sensor calibration (factor K), the process can be made iterative, driving the alignment errors to zero by using successively smaller adjustment steps.

8.5.5 Modal Control Systems

The control system may be implemented quite independently of the type of wavefront sensor and corrector being used in the adaptive optics system. For practical reasons, most wavefront sensors and correctors are zonal devices consisting of two-dimensional arrays of subapertures, usually with uniform spacing. The wavefront is therefore sampled and corrected at a fixed spatial frequency. However, this does not constrain the data processing that takes place between the sensor and corrector, which may be organized in any convenient way. Modal processing has some compatibility with the wavefront structure produced by atmospheric turbulence, making a near-optimum control system relatively easy to implement.

The phase variations produced by atmospheric turbulence contain spatial frequencies distributed over a wide interval from the inner to the outer scales, a range of over 1000 to 1. The outer scale of turbulence is comparable to the apertures of large astronomical telescopes. Atmospheric turbulence therefore produces spatial wavefront modes varying from less than 1 cycle across the aperture (piston and tilt) to about 1000 cycles.

In the control system connecting the sensor to the corrector, this wavefront may be represented as uniformly spaced samples (zonal processing), or as a series of modes of ascending order, such as the Zernike polynomials, or the related Karhunen-Loeve modes. A randomly distorted wavefront may be equally well represented in either zonal or modal form. The difference between zonal and modal representations appears when the dynamics of the wavefront are taken into account. The two-dimensional spatial frequency power spectrum due to Kolmogorov turbulence has a characteristic $f^{-8/3}$ dependency. When frozen turbulence is transported at a fixed velocity by wind, the temporal power spectrum has the same shape. The lower order (spatial) components of the wavefront are therefore associated with lower temporal frequencies. Random noise is always present in the wavefront sensor output, especially with faint astronomical sources. In a modal control system, in which the spatial modes are separated, it is therefore possible to optimize the bandwidth of each mode independently.

The optimization procedure for each mode is basically the same as that for a zonal system, in which the sum of the mean-square errors due to photon noise and to temporal bandwidth (servo lag) is minimized by adjustment of the servo bandwidth:

$$\sigma^2 = \sum \sigma_i^2(f_i) \tag{8.78}$$

where

$$\sigma_i^2 = [\sigma_i^2(f_i)]_{\text{noise}} + [\sigma_i^2(f_i)]servo \tag{8.79}$$

and

i = mode number

Gaffard and Ledanois [1991] and Gendron [1993] have analyzed optimal modal control systems based on Zernike or Karhunen-Loeve modes in which the temporal bandwidth of each mode is adjusted according to the signal-to-noise ratio of the wavefront sensor, to minimize the mean-square error. Implicit in this approach is the need to stop driving a mode, by reducing the bandwidth to zero, when it no longer makes a useful contribution to the wavefront correction. A block diagram of a multiple-bandwidth modal control system is shown in figure 8.31. A control system of this type has been employed successfully in the "Come-On Plus" adaptive optics system developed by a team headed by Office National d'Etudes et de Recherches Aerospatiales (ONERA), France and the European Southern Observatory (ESO).

The use of multiple control bandwidths in modal control systems for adaptive optics has been investigated by Ellerbroek et al. [1994]. They modeled a system in which the basis of the modes was optimized, as well as the bandwidth of each mode. The performance of a multimode control system, in which the bandwidth of each mode was optimized for maximum signal-to-noise ratio, was compared with that of a control system employing just two fixed bandwidths, one of which was zero. The latter can be regarded as a "reduced range" single-bandwidth system in which higher order modes are successively dropped as the sensor noise increases, as shown in figure 8.32. The performance of a conventional single-bandwidth system was used as a reference.

As expected, the multiple-bandwidth modal system had appreciably better performance than a conventional single-bandwidth system, especially at low signal-to-noise ratios. But, it was found, somewhat surprisingly, that the performance of the single-bandwidth reduced-range modal control system was very close to that of the full multiple-bandwidth system. The control system shown in figure 8.32 using a single optimized bandwidth f_1 is far simpler to implement than the system of figure 8.31 which employs multiple bandwidths, each of which must be optimized in real time.

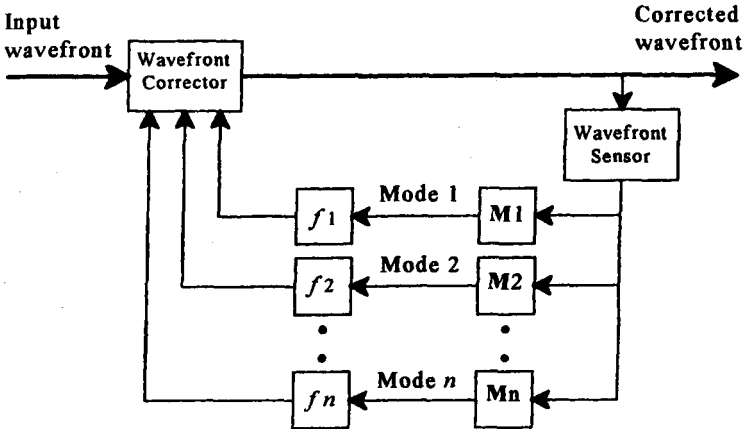


Figure 8.31 Principle of multiple-bandwidth modal compensation. Orthogonal modes are extracted by the matrices M1, M2 etc. The bandwidth of each mode is optimized individually in real time by adjusting/1, $f2$, etc., to minimize the error in that mode.

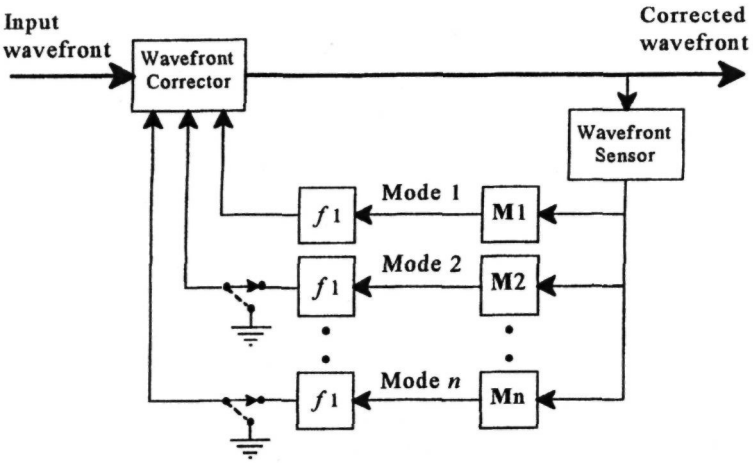


Figure 8.32 Reduced-range, single-bandwidth modal compensation. All modes use the same optimized bandwidth. Higher-order correction modes are simply set to zero if they are unable to reduce the residual error because of poor signal-to-noise ratio.

8.5.6 Open-Loop Operation

It is not always possible to implement closed-loop control of adaptive optics systems. The main exception occurs in the case of a pulsed reference source (laser beacon), when the pulse rate of the laser is comparable to, or longer than, the change time of the atmospheric turbulence. The bandwidth of a closed-loop control system would then be much lower than that of the turbulence, and compensation would be ineffective. The solution is to use an open-loop or "feed-forward" control system, also known as a "go-to" servo system, in which each wavefront measurement is applied directly to the corrector with the minimum delay. When the measurement interval is much longer than the atmospheric change time, the corrections are uncorrelated; each measurement and correction cycle is a separate operation and no information can be obtained from past history.

Open-loop operation has both advantages and disadvantages compared with closed-loop operation. Its main advantage is that the latency or time delay between measurement and correction is usually less than that of a closed-loop system because there are no constraints due to feedback loop instability. The main disadvantage of open-loop operation is that it is necessary to measure the *uncorrected* wavefront over its full dynamic range, and the combined accuracy of the sensor and corrector (including any latency) must equal the specified error requirement over this range. One may recall that in closed-loop operation, the wavefront sensor is required only to measure *residual* wavefront errors (that is, changes that occur over short measurement intervals) over a small dynamic range, and that, in this case, the accuracy of compensation is determined mainly by the stability of the "null" of the wavefront sensor.

To achieve similar performance, open-loop operation places more stringent requirements on both the wavefront sensor and the corrector than does closed-loop operation.

The main requirements for open-loop operation are:

1. The wavefront sensor must be capable of measuring the full wavefront excursion to be corrected by the adaptive optics system.
2. The wavefront correction produced by the deformable mirror or other compensator must follow the input wavefront accurately. This can be achieved by calibrating the response; for example, by using a lookup table.

The wavefront correction range required by an open-loop adaptive optics system may be calculated as follows. Overall tilt is usually measured and corrected separately, so it must be subtracted from the local tilts measured in each subaperture by the wavefront sensor. The tilt error due to turbulence is given in section 3.4.6, equation (3.59). The rms angular error in one axis over a subaperture of diameter d, when the overall tilt is corrected over an aperture D, is

$$\sigma_\alpha = 0.426 \frac{\lambda}{r_0^{5/6}} \left[\frac{1}{d^{1/3}} - \frac{1}{D^{1/3}} \right]^{1/2} \tag{8.80}$$

To determine the peak value of the tilt, a value of 2.5a may be used, corresponding to a probability of .995 that it will not be exceeded. For the typical conditions of d = 0.2m, r_0 = 0.18m at $X = 0.589\mu$m (sodium beacon), and D = 4m, the peak wavefront slope over each subaperture is $\pm 2.73\mu$rad = ±0.93 waves per subaperture.

Reference to Section 5.3 and figure 5.11 shows that the linear response of a Shack-Hartmann sensor using a bicell in each axis is limited to about ±0.5 waves per subaperture (at the sensor wavelength). When using reference sources at visible wavelengths, the simple form of a Shack-Hartmann sensor therefore has insufficient dynamic range for use with open-loop adaptive optics systems. To obtain the necessary

linear range, a detector array with 4×4 pixels per subaperture would be required, as described in section 5.3.1. When using a shearing interferometer (section 5.4), a large linear measurement range can be obtained by reducing the shear distance.

8.6 Optimal Wavefront Correction

8.6.1 Introduction

Up to this point, each component of an adaptive optics system has been considered as a separate entity, with its parameters being optimized individually without regard to the operation of the system as a whole. To maximize its performance over the expected range of operating conditions, an adaptive optics system must be considered in its entirety, from the distant reference source to the focal plane instrumentation that collects the scientific data. The analytical models derived for each major function not only must represent their individual performance accurately, but also must be combined so that all interactions are accounted for. Only by treating wavefront estimation and correction as a unified process can the performance of the whole system be maximized.

A major step in this direction was made by Wallner [1982, 1983], who, by using realistic models of the wavefront sensor and wavefront corrector, developed an approach to optimal wavefront correction that could be applied to any type of adaptive optics system. This analysis enables a control matrix to be defined for any given set of components, in order to optimize a given criterion, such as mean-square wavefront error, in the presence of an input wavefront specified by any structure function. The formalism used by Wallner has been extended by Welsh and Gardner [1989a, 1991] to include additional parameters, such as time delays, anisoplanatism, and multiple turbulent layers.

In addition to its value for optimizing system design, this analysis fulfills another very important function: it predicts the "best possible" performance that can be achieved with adaptive optics for any given set of conditions. The basic performance curves are easily scaled for any aperture size, turbulence strength, and photon flux, providing a standard against which the performance of practical systems may be judged. The derivation of the optimal control law for adaptive optics is outlined in the following section.

8.6.2 System Model

The adaptive optics system model being considered consists of the basic elements shown in figure 8.33. The optical aperture is described by weighting function $W_A(\mathbf{x})$, proportional to the intensity of the light entering the system, where x is a two-dimensional position vector in the aperture plane. It is convenient to normalize this weighting function so that

$$\int d\mathbf{x}\, W_A(\mathbf{x}) = 1 \tag{8.81}$$

where the integration is over the entire aperture plane. The wavefront sensor and the wavefront corrector are located at conjugate images of the aperture, so that positions in these planes are also described by the vector x.

The phase or optical path difference of the incoming wavefront is specified as $\Psi(\mathbf{x}, /)$, a random process of position x and time *t*. The use of the term "phase" in this context is purely for convenience, because the wavefront is actually measured and corrected in terms of its optical path differences (meters), independent of wavelength, and not in terms of phase angle. The absolute optical phase is of no interest when imaging with a single aperture, so the phase $\phi(\mathbf{x}, t)$ is defined with the aperture average removed:

$$\phi(\mathbf{x}, t) = \Psi(\mathbf{x}, t) - \int d\mathbf{x}\, W_A(\mathbf{x}')\Psi(\mathbf{x}', t) \tag{8.82}$$

The wavefront sensor measures the weighted average of the wavefront slope $\phi(\mathbf{x}, ()$ over an array of subapertures, adding random noise in the process. The output of the nth sensor at time t is

$$s_n(t) = \int d\mathbf{x}\, W_{sn}(\mathbf{x})\phi_n^s(\mathbf{x}, t) + \alpha_n(t) \tag{8.83}$$

where

$W_{sn}(\mathbf{x})$ = subaperture weighting function for the nth slope sensor

$\phi_n^s(\mathbf{x}, I)$ — average slope of the wavefront in the measurement direction of the nth slope sensor

$\alpha_n(t)$ = slope measurement error for the nth sensor at time *t*, independent of $\mathbf{tf}(\mathbf{x}, t)$.

The sensor output signal can be expressed in terms of the wavefront phase by integrating the first term of equation (8.83) by parts:

$$s_n(t) = \int d\mathbf{x}\left[-W_n^s(\mathbf{x})\phi(\mathbf{x}, t)\right] + \alpha_n(t) \tag{8.84}$$

where $W_n^s(\mathbf{x})$ is the derivative of $W_{sn}(\mathbf{x})$ in the measurement direction of the nth sensor.

The control law generates a command for each actuator of the wavefront corrector, based on all of the sensor outputs. Using a linear control law, the command to the *j*th actuator is

$$c_j(t) = \sum_n \mathbf{M}_{jn} s_n(t) \tag{8.85}$$

where $\mathbf{M}_{jn}$ is the weighting of the nth sensor in they'th actuator command.

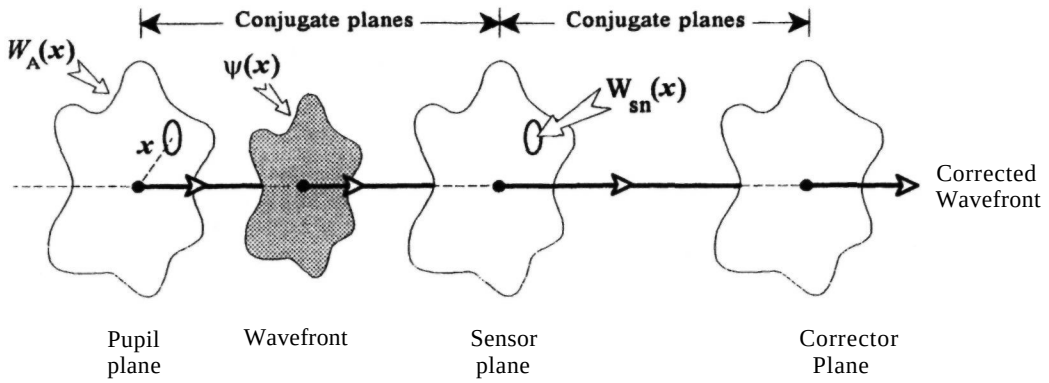


Figure 8.33 Model for optimal wavefront correction.

The total wavefront correction, assuming linear superposition of the actuator responses (and no time delay in the corrector itself), is then

$$\hat{\phi}(\mathbf{x}, t) = \sum_j c_j(t) r_j(\mathbf{x}) \tag{8.86}$$

where $r_j(\mathbf{x})$ is the response of the jth actuator to a unit command.

8.6.3 Evaluation of Residual Error

Using equations (8.85) and (8.86), the residual wavefront error after correction may be expressed as

$$\begin{aligned}\varepsilon(\mathbf{x}, t, \tau) &= \hat{\phi}(\mathbf{x}, t-\tau) - \phi(\mathbf{x}, t) \\ &= \sum_j r_j(\mathbf{x}) \sum_n \mathbf{M}_{jn} s_n(t-\tau) - \phi(\mathbf{x}, t)\end{aligned} \tag{8.87}$$

where T is the time delay between wavefront sensing and correction. If $\phi(\mathbf{x}, t)$ and $\alpha_n(t)$ are random processes that are stationary in time over the observation periods, then the residual error can be written as a function of the time delay only:

$$\langle \varepsilon(\tau)^2 \rangle = \int d\,\mathbf{x}\, W_A(\mathbf{x}) \langle \varepsilon^2(\mathbf{x}, \tau) \rangle \tag{8.88}$$

The average mean-square error is then

$$\begin{aligned}\langle \varepsilon(\tau)^2 \rangle = &\sum_j \sum_{j'} \sum_n \sum_{n'} \mathbf{M}_{jn} \mathbf{M}_{j'n'} \mathbf{S}_{nn'} \mathbf{R}_{jj'} \\ &- 2 \sum_j \sum_n \mathbf{M}_{jn} \mathbf{A}_{jn}(\tau) + \langle \varepsilon_0^2 \rangle\end{aligned} \tag{8.89}$$

where $\mathbf{S}_{nn'}$ is the matrix involving the product of two sensor functions:

$$\begin{aligned}\mathbf{S}_{nn'} &= \langle s_n(t) s_{n'}(t) \rangle \\ &= \int d\,\mathbf{x}' \int d\,\mathbf{x}''\, W_n^s(\mathbf{x}') W_{n'}^s(\mathbf{x}'') \langle \phi(\mathbf{x}', t) \phi(\mathbf{x}'', t) \rangle \\ &\quad + \langle \alpha_{n'}(t) \alpha_{n'}(t) \rangle\end{aligned} \tag{8.90}$$

$\mathbf{R}_{ii'}$ is the matrix involving the product of two actuator response functions:

$$\mathbf{R}_{jj'} = \int d\,\mathbf{x}\, W_A(\mathbf{x}) r_j(\mathbf{x}) r_{j'}(\mathbf{x}) \tag{8.91}$$

$\mathbf{A}_{jn}(\tau)$ is the matrix involving products of the sensor and actuator response functions:

$$\begin{aligned}\mathbf{A}_{jn}(\tau) = &\int d\,\mathbf{x}\, W_A(\mathbf{x}) r_j(\mathbf{x}) \langle s_n(t-\tau) \phi(\mathbf{x}, t) \rangle \\ &- \int d\,\mathbf{x} \int \mathrm{d}\,\mathbf{x}'\, W_A(\mathbf{x}) r_j(\mathbf{x})\, W_n^s(\mathbf{x}') \\ &\langle \phi(\mathbf{x}, t) \phi(\mathbf{x}', t-\tau) \rangle\end{aligned} \tag{8.92}$$

The average mean-square uncorrected error is

$$\langle \varepsilon_0^2 \rangle = \int d\,\mathbf{x}\, W_A(\mathbf{x}) \langle \phi^2(\mathbf{x}) \rangle \tag{8.93}$$

To optimize system performance, it is necessary to choose a set of weights in the control matrix $\mathbf{M}_{jn}$ that will minimize the mean-square error. This is achieved by differentiating equation (8.89) with respect to $\mathbf{M}_{jn}$ and equating the result to zero. The resulting optimal control matrix $\mathbf{M}_{jn}^*$ is given by Wallner [1983] as

$$\mathbf{M}_{jn}^*(\tau) = \mathbf{R}_{jj'}^{-1} \mathbf{A}_{j'n'}(\tau) \mathbf{S}_{n'n}^{-1} \tag{8.94}$$

Because of the need to invert the R and S matrices, the question of their singularity arises. Wallner states that the R matrix will not be singular unless there are redundant actuators that allow the same correction to be made with different commands. This problem may be solved either by eliminating the redundant functions, or by using the generalized inverse of the matrix. The S matrix will not be singular, even with redundant sensors, if the noise is independent on each measurement, which is always the case in real systems.

The average mean-square residual error of the optimized system is found by substituting equation (8.94) in equation (8.89):

$$\left\langle \varepsilon(\tau)_{\text{opt}}^2 \right\rangle = \langle \varepsilon_0^2 \rangle - \mathbf{R}_{jj'}^{-1} \mathbf{A}_{j'n'}(\tau) \mathbf{S}_{n'n}^{-1} \mathbf{A}_{jn}(\tau) \tag{8.95}$$

8.6.4 Optical Transfer Function

The mean-square error enables the Strehl ratio of the optimized adaptive optics system to be determined, providing that the phase error is less than about 2 rad rms (see section 4.3.3). As discussed in chapter 4, to obtain a more complete understanding of the effect of wavefront correction on the structure of images, it is necessary to obtain the optical transfer function (OTF) of the system. Welsh and Gardner [1989a] have derived an expression for the OTF of a compensated system in terms of the parameters defined above. Starting with the complex-amplitude field $E(\mathbf{x})$ in the optical aperture, produced by a far-field point source, the OTF can be written as the convolution of $E(\mathbf{x})$ with its complex conjugate $E^*(\mathbf{x})$:

$$S(\boldsymbol{\rho}) = \frac{\int d\,\mathbf{x}\, W_A(\mathbf{x}) E(\mathbf{x}) W_A^*(\mathbf{x}-\boldsymbol{\rho}) E^*(\mathbf{x}-\boldsymbol{\rho})}{\int d\,\mathbf{x} |W_A(\mathbf{x}) E(\mathbf{x})|^2} \tag{8.96}$$

where x and p are two-dimensional position vectors in the aperture plane. If the optical aperture is in the near field of the turbulence, which is usually the case for astronomical adaptive optics, then amplitude variations due to interference may be ignored and the absolute magnitude of the field $|E(\mathbf{x})|$ may be set to unity. The ensemble average OTF *(S(p))* may then be written in terms of the wavefront phase error $\varepsilon(\mathbf{x}, /, \mathrm{T})$, assuming it to be a Gaussian, zero-mean, random process:

$$\langle S(\boldsymbol{\rho})\rangle = \frac{\int d\,\mathbf{x}\, W_A(\mathbf{x}) W_A^*(\mathbf{x}-\boldsymbol{\rho}) \exp\left\{-\frac{1}{2}\langle[\varepsilon(\mathbf{x}, t, \tau) - \varepsilon(\mathbf{x}-\boldsymbol{\rho}, t, \tau)]^2\rangle\right\}}{\int d\,\mathbf{x} |W_A(\mathbf{x})|^2} \tag{8.97}$$

Substituting for the wavefront phase error e(x, /, r) equation (8.97) can be expanded to give the OTF of the system after correction

$$\begin{aligned}\langle S(\boldsymbol{\rho})\rangle = {} & \frac{\exp\left\{-\frac{1}{2}\langle[\phi(\mathbf{x}, t) - \phi(\mathbf{x}-\boldsymbol{\rho}, t)]^2\rangle\right\}}{\int d\,\mathbf{x} |W_A(\mathbf{x})|^2} \\ & \times \int d\,\mathbf{x}\, W_A(\mathbf{x}) W_A^*(\mathbf{x}-\boldsymbol{\rho}) \\ & \times \exp\left\{-\frac{1}{2}\sum_j \sum_i [r_j(\mathbf{x}) - r_j(\mathbf{x}-\boldsymbol{\rho})] \right. \\ & \times [r_i(\mathbf{x}) - r_i(\mathbf{x}-\boldsymbol{\rho})]\mathbf{C}_{ji} \\ & + \sum_j [r_j(\mathbf{x}) - r_j(\mathbf{x}-\boldsymbol{\rho})]\langle c_j(t-\tau)[\phi(\mathbf{x}, t) \\ & \left. - \phi(\mathbf{x}-\boldsymbol{\rho}, t)]\rangle\right\}\end{aligned} \tag{8.98}$$

where

$$\begin{aligned}\mathbf{C}_{ji} &= \langle c_j(t) c_i(t)\rangle \\ &= \sum_n \sum_m \mathbf{M}_{jn} \mathbf{M}_{im} \mathbf{S}_{nm}\end{aligned} \tag{8.99}$$

8.6.5 Wavefront Structure Function

The expressions in equations (8.95) and (8.98) enable the optimal performance of an arbitrary system with arbitrary wavefront statistics to be determined. In the case of atmospheric turbulence, the statistics of the input wavefront can be represented by a wavefront structure function:

$$D(\mathbf{x}, \mathbf{x}', \tau) = \langle[\Psi(\mathbf{x}, t) - \Psi(\mathbf{x}', t-\tau)]^2\rangle \tag{8.100}$$

The phase correlations required to compute S, A, and *(el)* can be found using this equation. Following Wallner [1983], the expressions are

$$\begin{aligned}\mathbf{S}_{nn'} = {} & -\tfrac{1}{2}\int d\,\mathbf{x}' \int d\,\mathbf{x}''\, W_n^s(\mathbf{x}') W_{n'}^s(\mathbf{x}'') D(\mathbf{x}', \mathbf{x}'', 0) \\ & + \langle \alpha_n(t) \alpha_{n'}(t)\rangle\end{aligned} \tag{8.101}$$

$$\begin{aligned}\mathbf{A}_{jn}(\tau) = {} & -\int d\mathbf{x} \int d\mathbf{x}'\, W_A(\mathbf{x}) r_j(\mathbf{x}) W_n^s(\mathbf{x}')\left[-\tfrac{1}{2} D(\mathbf{x}, \mathbf{x}', \tau)\right. \\ & \left. + g(\mathbf{x}', \tau)\right]\end{aligned} \tag{8.102}$$

where

$$g(\mathbf{x}, \tau) = \tfrac{1}{2}\int d\mathbf{x}'\, W_A(\mathbf{x}') D(\mathbf{x}, \mathbf{x}', \tau) \tag{8.103}$$

$$\begin{aligned}\langle \varepsilon_0^2\rangle &= \int d\,\mathbf{x}\, W_A(\mathbf{x}) \langle \phi^2(\mathbf{x})\rangle \\ &= \tfrac{1}{2}\int d\,\mathbf{x}'' \int d\,\mathbf{x}'''\, W_A(\mathbf{x}'') W_A(\mathbf{x}''') D(\mathbf{x}'', \mathbf{x}''', \tau)\end{aligned} \tag{8.104}$$

The corresponding expression for the OTF in terms of the wavefront structure function, as derived by Welsh and Gardner [1989a], is

$$\begin{aligned}\langle S(\boldsymbol{\rho})\rangle = {} & \frac{\exp\left\{-\frac{1}{2}[D(\mathbf{x}, \mathbf{x}-\boldsymbol{\rho}, 0)]\right\}}{\int d\mathbf{x} |W_A(\mathbf{x})|^2} \int d\mathbf{x}\, W_A(\mathbf{x}) W_A^*(\mathbf{x}-\boldsymbol{\rho}) \\ & \times \exp\left\{-\frac{1}{2}\sum_j \sum_i [r_j(\mathbf{x}) - r_j(\mathbf{x}-\boldsymbol{\rho})][r_i(\mathbf{x}) \right. \\ & - r_i(\mathbf{x}-\boldsymbol{\rho})]\mathbf{C}_{ji} + \sum_j [r_j(\mathbf{x}) - r_j(\mathbf{x}-\boldsymbol{\rho})] \\ & \times \tfrac{1}{2}\sum_n \mathbf{M}_{jn} \int d\mathbf{x}' W_n^s(\mathbf{x}')[D(\mathbf{x}, \mathbf{x}', \tau) \\ & \left. - D(\mathbf{x}-\boldsymbol{\rho}, \mathbf{x}', \tau)]\right\}\end{aligned} \tag{8.105}$$

8.6.6 Measurement Noise

The limiting performance of most astronomical adaptive optics systems is determined by the shot noise resulting from relatively small photon counts in the wavefront sensor. The shot noise produces random, zero-mean errors in the wavefront slope measurements, modeled as the function $\alpha_n(t)$ in the

above equations, for subaperture n. These noise errors are spatially and temporally uncorrelated. Orthogonal wavefront measurements made simultaneously in the same subaperture (as in a Hartmann sensor) also have uncorrelated noise errors [Wallner 1983]. For a white-noise spectrum, the noise correlation function may be written

$$\langle \alpha_n(t)\alpha_{n'}(t')\rangle = \sigma_n^2 k_{nn'}\delta(t-t') \tag{8.106}$$

where

σ_n = mean-square slope error in sensor subaperture n
$k_{nn'}$ = $\delta_{nn'}\cos\theta$
θ - angle between the slope measurement directions of the nth and n'th sensors
$\delta_{nn'}$ = 1 if the nth and n'th subapertures coincide, 0 otherwise
$\delta(t)$ = Dirac delta function

The value of σ_n^2 depends on the photon flux received from the reference source and on the configuration of the wavefront sensor.

8.6.7 Optimal Performance Calculation

The theory outlined above establishes the optimal performance of adaptive optics systems. The results are presented conveniently in the form of charts plotting the normalized residual error against the normalized density of photons received from the reference source. The normalization accounts for factors such as subaperture size d, turbulence parameter r_0 and wavefront sensor efficiency K_s. The results are then scaled easily to adaptive optics systems of any size. These ideal performance curves provide a standard against which the performance of real systems can be judged.

Residual wavefront error for an optimized system is calculated using equation (8.95). The elements required to make this computation will now be summarized. The structure function of turbulence-degraded wavefronts following Kolmogorov's law is defined as

$$D(\mathbf{x},\mathbf{x}') = 6.884\left|\frac{(\mathbf{x}-\mathbf{x}')}{r_0}\right|^{5/3} \tag{8.107}$$

where r_0 is the coherence length of the turbulence. The mean-square uncorrected wavefront error produced by this turbulence over a square aperture of dimensions A x A (section 3.3) is

$$\langle \varepsilon_0^2\rangle = 1.31\left(\frac{A}{r_0}\right)^{5/3} \quad \text{rad}^2 \tag{8.108}$$

The mean-square angular excursion of the wavefront in one axis due to atmospheric turbulence over a square aperture of d x d meters at a wavelength of X meters, from equation (3.59), is

$$\sigma_\alpha^2 = 0.18 d^{-1/3} r_0^{-5/3}\lambda^2 \quad \text{rad}^2 \text{ of angle}$$

To convert this angle into radians squared of phase per meter squared at wavelength X, the expression is multiplied by $(2\pi/\lambda)^2$ to obtain

$$\sigma_c^2 = 7.11\, d^{-1/3} r_0^{-5/3} \quad \text{rad}^2 \text{ of phase per m}^2 \tag{8.109}$$

With all wavefront sensors, the mean-square slope error due to photon noise is inversely proportional to the number of detected photons used in the measurement. For the ideal case of a point source reference and the subaperture size matched to the atmospheric coherence length r_0, the photon noise error may be expressed in the form

$$\sigma_n^2 = \frac{K_s}{H_s t d^2} \tag{8.110}$$

where

K_s = a constant depending on the wavefront sensor characteristics, rad^2 m^{-2}
H_s = irradiance from the reference source, photoelectrons per meter squared per second
t = exposure time,
d = subaperture dimension, m

The response functions of most zonal deformable mirrors using continuous faceplates are approximately Gaussian in form, enabling a simple model to be used:

$$r_j(x,y) = a_m \exp -\left[\frac{(x-x_j)^2+(y-y_j)^2}{r_m}\right] \tag{8.111}$$

where

x, y = coordinates in the plane of the mirror,
a_m = a constant defining the actuator sensitivity,
x_j, y_j = specification of the actuator locations
r_m = radius at which the response is $1/e$

These equations, together with equation (8.95), enable the residual mean-square error to be plotted against the photoelectron density, $H_s t$, for systems using an optimal control law to connect the wavefront sensor and corrector. Wallner [1982, 1983] found that when optimal control is used, the geometrical arrangement of the sensors and actuators has a negligible effect on the system performance, reinforcing the fact that this is truly a fundamental approach to the design of adaptive optics systems.

8.6.8 Normalized Adaptive Optics Performance

Normalized performance curves for adaptive optics systems with actuator arrays of 2 x 2 through 6×6 are shown in figure 8.34. This chart plots the normalized mean-square error $A^{-5/3} r_0^{-5/3}\sigma_c^2$ in radians squared of wavefront phase, against the normalized photoelectron density $A^{11/3} r_0^{-5/3} k^{-1} H_s t$, where A is

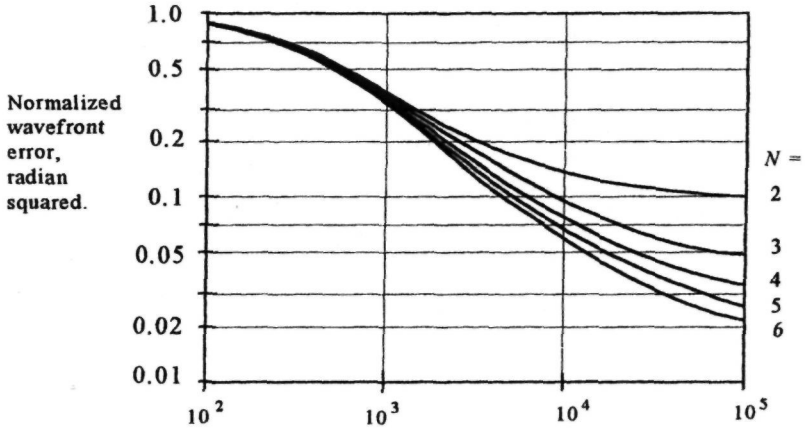


Figure 8.34 Normalized performance of optimized adaptive optics systems [Wallner 1983]. At low photon densities there is insufficient information for the optimal control law to make an estimate of the wavefront, so the system error approaches that of the wavefront itself. At high photon densities, the fitting error is dominant and the performance is determined mainly by the number of actuators, $N \times N$. Practical adaptive optics systems normally operate between these two extremes.

the side of the full aperture (square in this case) and r_0 is the turbulence coherence length. The photoelectron density is directly related to the brightness of the reference source.

Figure 8.34 shows three main regimes of operation:

1. At low photon densities, the errors due to measurement noise become comparable to the phase excursions of the uncorrected wavefront. The optimal control law then reduces its estimate to near zero, resulting in no correction, so the wavefront error of the system approaches that of the uncorrected wavefront. Under these conditions, the number of actuators becomes irrelevant because they are not utilized.
2. At very high photon densities, the slope measurement error becomes negligible. In the system modeled, the wavefront fitting error becomes dominant, with the result that the performance is limited mainly by the number of actuators.
3. Adaptive optics systems usually operate in the intermediate regime, where the overall performance is determined by many factors, including measurement noise, fitting errors, time delays, and anisoplanatism. In this operating environment, the balancing of errors to achieve optimal performance requires a detailed knowledge of all facets of adaptive optics systems, involving both fundamental limitations and practical issues, as well as the all-important turbulence environment.

In the design of real adaptive optics systems, many compromises have to be made and the actual performance usually falls short of the ideal. Multidimensional tradeoffs must be made between system parameters in order to maximize performance over the operating envelope. A systematic approach to the design and optimization of adaptive optics systems is described in chapter 9.

9 Adaptive Optics Performance Analysis and Optimization

9.1 Introduction

9.1.1 Objectives

The literature on adaptive optics contains many different approaches to the design of wavefront compensation systems. This situation is typical of a rapidly evolving technology. Each new idea is tested and evaluated over a period of time; eventually, the technology matures and crystallizes into well-defined components and system configurations. Adaptive optics is far from reaching that state of maturity. New ideas on wavefront sensing, wavefront correction, and control systems are being formulated and tested. However, all systems are constrained by the same basic principles, and a systematic design procedure may be followed, whatever the state of the technology.

The objective of this chapter is to provide the tools needed to analyze the performance of adaptive optics systems using both natural stars and laser beacons, and to optimize their major parameters. Because there are many different philosophies and applications of astronomical adaptive optics, it is not intended to derive specific designs, but rather to illustrate a general procedure that should be widely applicable.

9.1.2 The Design Process

The process of design and optimization involves a large number of parameters, covering both the adaptive optics system itself and the external environmental conditions. The quantities r_0, θ_0, and f_0 that describe the spatial and temporal characteristics of the turbulence, optimistically referred to as "constants," usually change during extended observing periods. To maintain optimal performance, an adaptive optics system not only must compensate random wavefront errors, but also must be capable of adjusting its parameters to accommodate varying conditions. For any adaptive optics system to deliver its peak performance, it must be kept well tuned to the current atmospheric conditions.

The design of an adaptive optics system should start with a set of technical requirements that specify the scientific purpose and overall performance objectives of the complete compensated telescope. These specifications should include at least the following items:

- the seeing qualities of the observing site, including statistical data over extended time periods;
- the characteristics of the telescope, including any constraints on locating the adaptive optics;
- the scientific goals of the installation, including the type of object and field of view to be observed, and the information to be gathered;
- the scientific devices used for capturing data (imaging camera, spectrograph, etc.);
- the spectral band(s) of the observations;
- the sky coverage required;

- the availability and characteristics of reference sources;
- the overall performance required in terms of angular resolution, image contrast, or Strehl ratio.

The scientific objectives are usually tempered by a cost budget that must not be exceeded.

Ideally, the basic system architecture and major components should be selected purely on technical merit or on a combination of performance and cost. In the real world, the situation is usually far more complex. Specific hardware items, such as wavefront sensors, deformable mirrors, or lasers, may be available or be specified in advance, placing restrictions on the system configuration. Scientific tasks cannot be foreseen years in advance and the hardware should therefore include the flexibility needed to try out new ideas and to exploit special opportunities. The budget may be inadequate to satisfy all of the technical requirements. Making the compromises necessary to allow the project to proceed can be the most challenging task in the design of an adaptive optics system.

In adaptive optics systems, it is particularly important to avoid "peak designs," which offer high performance over a narrow range of conditions, but are unduly sensitive to variations in external conditions or the physical alignment of the hardware. With so many uncontrolled variables, it is essential to use robust optimizations that are tolerant to a wide range of conditions. To achieve a durable design, it is necessary to explore the effect of changes in many parameters, for which a graphical approach is highly efficient. The optimization procedure outlined in sections 9.5 and 9.6 therefore employs multiple tradeoff charts that give a synoptic view of performance, rather than the conventional approach of finding the peak of a single comprehensive analytic function.

The desire to maximize performance should be tempered by the realization that because so many factors influence the performance of a compensated telescope, including purely practical housekeeping items—such as the cleanliness and alignment of the optics, there is a threshold below which small performance differences become lost in the noise.

9.1.3 Operating Modes for Adaptive Optics

One of the first decisions to be made in the design of an adaptive optics system is to select the operating mode(s) that will be used for astronomical observations. These operating modes are distinguished by the type and arrangement of the reference source in relation to the science object. A natural guide star is always required, but the compensation strategy depends on the angular offset between the star and the science object. There are three basic modes of operation:

1. Natural reference sources.
2. Laser beacon(s) with uncompensated natural guide star.
3. Laser beacon(s) with compensated natural guide star.

The arrangement of the reference source(s) in relation to the science object, in the telescope field of view, is depicted in figure 9.1 for each mode of operation.

1. Natural Reference Sources

This mode may be divided into (a) self-referencing operation, in which the science object itself functions as the reference source, and (b) offset referencing, in which a nearby (natural) star provides the reference. A single reference source is used to compensate both the wavefront and the image motion.

The self-referencing mode may be used when the science object is brighter than 10th to 14th magnitude, depending on the observing wavelength. Low-order compensation at IR wavelengths has been obtained with natural stars down to 15th visual magnitude, as reported by Roddier et al. [1994]. The compensated area has a radius of only 2 or 3 arc seconds at visible wavelengths, but grows as the 6/5 power of the observing wavelength. Self-referencing is a suitable mode for adaptive optics in the search for dim companions, in which the parent star acts as the reference.

When the science object is too faint for self-referencing, offset referencing may be used if a suitable guide star occurs within the isoplanatic area. The sky coverage can be maximized by using multispectral operation, observing at IR wavelengths with an offset visible or IR reference star. Although this mode of operation offers larger sky coverage than self-referencing, it is still less than 1%.

2. Laser Beacon with Uncompensated Guide Star

The use of a laser beacon as the reference source enables faint objects to be compensated by adaptive optics, at the expense of greater hardware complexity. The laser beacon must be directed within the isoplanatic angle of the science object at the observation wavelength but can provide only short-exposure (wavefront) correction. A fixed (natural) guide star is required to stabilize the image during long exposures, so that sky coverage, again, depends on the distribution of stars. For tilt sensing, stars of m_V = 16 to 20 can be used, greatly improving the sky coverage compared with mode 1. The precision of overall tilt correction depends not only on the brightness of the star, but also on its apparent size. The angular size of an uncompensated guide star is enlarged by atmo-

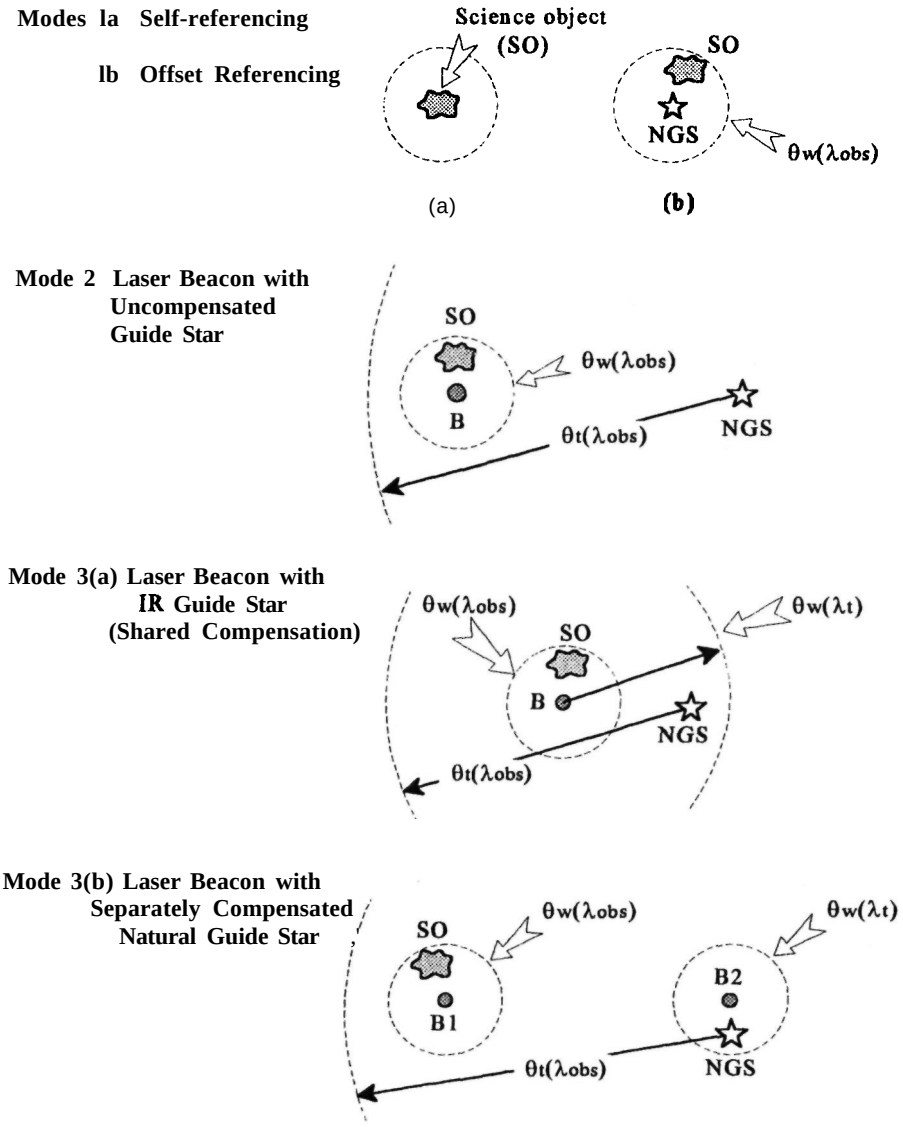


Figure **9.1** Operating modes for adaptive optics using natural guide stars and laser beacons. This figure shows the geometrical relation of the reference sources and the iosplanatic angles, which are designated as follows: SO, science object; B1 and B2, laser beacons; NGS, natural guide star; $\theta_w(\lambda_{obs})$, isoplanatic angle for wavefront compensation at observation wavelength; $\theta_t(\lambda_{obs})$, isoplanatic angle for tilt compensation at observation wavelength; $\theta_w(\lambda_t)$, isoplanatic angle for wavefront compensation at tilt sensor wavelength.

spheric turbulence, necessitating a larger photon count to achieve a given tilt accuracy than if it were diffraction limited. This requirement limits the sky coverage in this mode to about 1% in the visible and 10% at $1.6\mu m$. (This limitation is removed in mode 3 by compensating the natural guide star.)

There are several options for implementing the laser beacons. To start with, there is the choice between beacons generated by Rayleigh scattering within the atmosphere, or beacons generated in the sodium layer above the atmosphere. In each case, there are several methods of launching the laser beam: (a) through the telescope optical system; (b) through separate optics on the same mount as the telescope; or (c) through a separate launching system (bistatic operation).

3. Laser Beacon with Compensated Guide Star

Mode 3 includes the same laser beacon options as mode 2, with the additional capability of compensating the natural guide star to increase sky coverage. There are two methods of achieving this compensation. The first method, known as "shared compensation," uses a natural IR guide star that is within the large isoplanatic angle of the primary laser beacon at IR wavelengths, while the second method employs a separate adaptive optics system to compensate the natural star. In both cases, the compensation concentrates the light from a fixed reference star into a smaller and brighter image that enables fainter stars to be used, thereby increasing the sky coverage. Sandler, Stahl et al. [1994] have suggested that many science objects may contain IR stars bright enough to function as the tilt reference source.

The ultimate tilt stabilization system employs separate compensation of the natural guide star, using dual adaptive optics, as suggested by Rigaut and Gendron [1992]. in addition to the primary laser beacon system that compensates the wavefront of the science object, a secondary laser beacon system is employed to compensate the natural star's disk, so that the centroids of even dim stars can be measured with precision. This approach gives more flexibility in the choice of natural guide stars than does mode 2; it maximizes the capabilities of adaptive optics systems for astronomy, at the expense of additional complexity.

Each of these operating modes requires a different error model and a different procedure for optimizing the performance. The error components associated with each mode of operation are summarized in table 9.1

Once the overall performance requirements and basic configuration have been decided, the detailed design process may proceed. This phase is the main subject of this chapter. The objective is to establish an adaptive optics design that satisfies the technical requirements defined above in the most economical way, using a process of system modeling and optimization. The design process can be clarified by separating the system parameters into four main categories as follows:

1. Fixed parameters of the installation. These include (a) environmental conditions at the observing site, such as the expected ranges of r_0 and θ_0, wind velocities, and ambient temperature range; (b) characteristics of the telescope, the optical path, and the scientific sensors; and (c) availability of reference sources, whether natural stars or laser beacons using Rayleigh scatter or sodium resonance.
2. Performance goals of the observing system, such as sky coverage, spectral bands, compensated field of view, and the degree of compensation (or residual error) required. These goals may have to be modified during the design process, for technical or economic reasons.
3. Adaptive optics characteristics to be determined by the optimization procedure, which would normally include the number of degrees of freedom, the types of wavefront sensor and corrector to be used, the wavefront sampling rate, the servo bandwidth, and the beacon laser requirements (power, beam quality, and pulse rate).
4. Adjustable parameters of the adaptive optics that can be optimized in real time, while observing, to maximize performance under varying environmental conditions. Such parameters include the photon integration time of the wavefront sensor, the reconstructor weights, and the servo bandwidth of the correction system.

The criteria for optimization are important factors in overall system design. Factors involving image quality and their relation to residual wavefront and tilt errors have been discussed in chapter 4. A true end-to-end optimization procedure must include the science instrumentation, taking into account its spatial resolution and detector noise characteristics.

9.7.4 *Performance Estimation*

The performance of adaptive optical systems is determined using analytical models that define the wavefront errors produced by each component or function of the system. The effects of atmospheric turbulence are described using the statistical models developed in chapter 3. Analytical models defining the errors in the wavefront sensor and wavefront correctors have been developed in chapters 5 and 6. These analytical models are summarized and put into a convenient form for system analysis in sections 9.2, 9.3, and 9.4. Using this approach, the external conditions and the adaptive optics characteristics are defined in terms of a relatively small number of parameters, whose effect on system performance is readily apparent, allowing the design to be optimized in a straightforward way. It should be kept in mind that this gives a somewhat idealized view of system operation. In reality, the parameters employed are approximations and the statistics may not be stationary, so the practical results will sometimes vary from the predictions. However, experience with several operating adaptive optics systems has shown that relatively simple analytical models are sufficient to allow these systems to be optimized and their average performance to be predicted.

The major sources of error in adaptive optics systems are summarized in figure 9.2. Error sources are conveniently divided into two main groups: (1) External factors, which include the structure and dynamics of the atmosphere, and the characteristics of the star or beacon used as the wavefront sensor

Table 9.1 Basic Operating Modes for Adaptive Optics

	Self Referencing Mode 1(a)	Offset Referencing Mode 1(b)	Laser Beacon Modes 2 and 3
Tilt reference →	Science object	Natural star	Natural star
Wavefront Reference →	Science object	Natural star	Laser beacon
Error sources			
Overall tilt:			
Photon error	o[a]	o[a]	•
Temporal error	o[a]	0*	•
Offset (aniso) error	—	•	•
Wavefront:			
Photon error	•	•	•
Temporal error	•	•	•
Fitting error	•	•	«
Offset (aniso) error	—	•	o[b]
Cone (aniso) error	—	—	•
Atmospheric dispersion error	o[c]	o[c]	o[c]

•, Primary error source; o, secondary error source.
" Overall tilt errors due to photon noise and time delays should be much smaller than the corresponding wavefront errors, if the same reference source is used.
[b] This error will be small when the laser beacon is within the isoplanatic angle of the science object.
' Dispersion errors occur if the spectral bandwidths or mean wavelengths of the reference and observations are different, as in multispectrai operation.

reference; and (2) Instrumental factors or limitations of the adaptive optics system itself. Figure 9.2 shows, in broad outline, the relation between these factors and the errors that they generate.

The normal procedure in performance evaluation is to define a set of input conditions that include the reference source characteristics, the atmospheric turbulence parameters, and the major specifications of the adaptive optics system, such as the telescope aperture, subaperture size, wavefront sensor characteristics, type of deformable mirror, and observing wavelength. Using each set of input data, the wavefront error variances from each component of the adaptive optics system are computed.

If the individual error sources are random and uncorrelated, the residual wavefront phase error is given by the sum of their variances:

$$\sigma^2_{\text{phase}} = \sum \sigma_i^2 \tag{9.1}$$

In practice, there are correlations between some errors, so the straightforward sum of the variances may lead to an overestimate of the total residual error. This is not necessarily a disadvantage in system design, because it tends to compensate for unrecognized errors; the resulting performance predictions are often more realistic. Note that equation (9.1) covers only the random errors in the system; systematic errors, such as those due to offsets in the wavefront sensor, will normally be removed by calibration.

Although astronomical imaging systems operate with incoherent light, the wavefront errors are usually expressed in terms of phase errors at a given mean wavelength. Most astronomical adaptive optics systems are designed to measure and correct optical path differences, allowing wideband operation. However, because image formation is basically an interferometric process, the phase relationships at each wavelength are of prime concern, hence the use of optical phase in the analytical expressions.

From the total wavefront error variance, the Strehl ratio representing the normalized peak intensity of a compensated point source is computed from the relation

$$S = \exp(-\sigma^2_{\text{phase}}) \tag{9.2}$$

While the Strehl ratio is a useful measure of performance for some applications of adaptive optics, such as laser beam compensation, it does not completely represent the performance of imaging systems, for which the signal-to-background ratio is more important. The structure of compensated images is discussed in chapter 4, where it is shown that partially compensated images consist of a diffraction-limited core surrounded by a much larger halo. The peak contrast ratio between the core and the halo may be the critical factor for some imaging tasks. The subject of suitable criteria for the evaluation

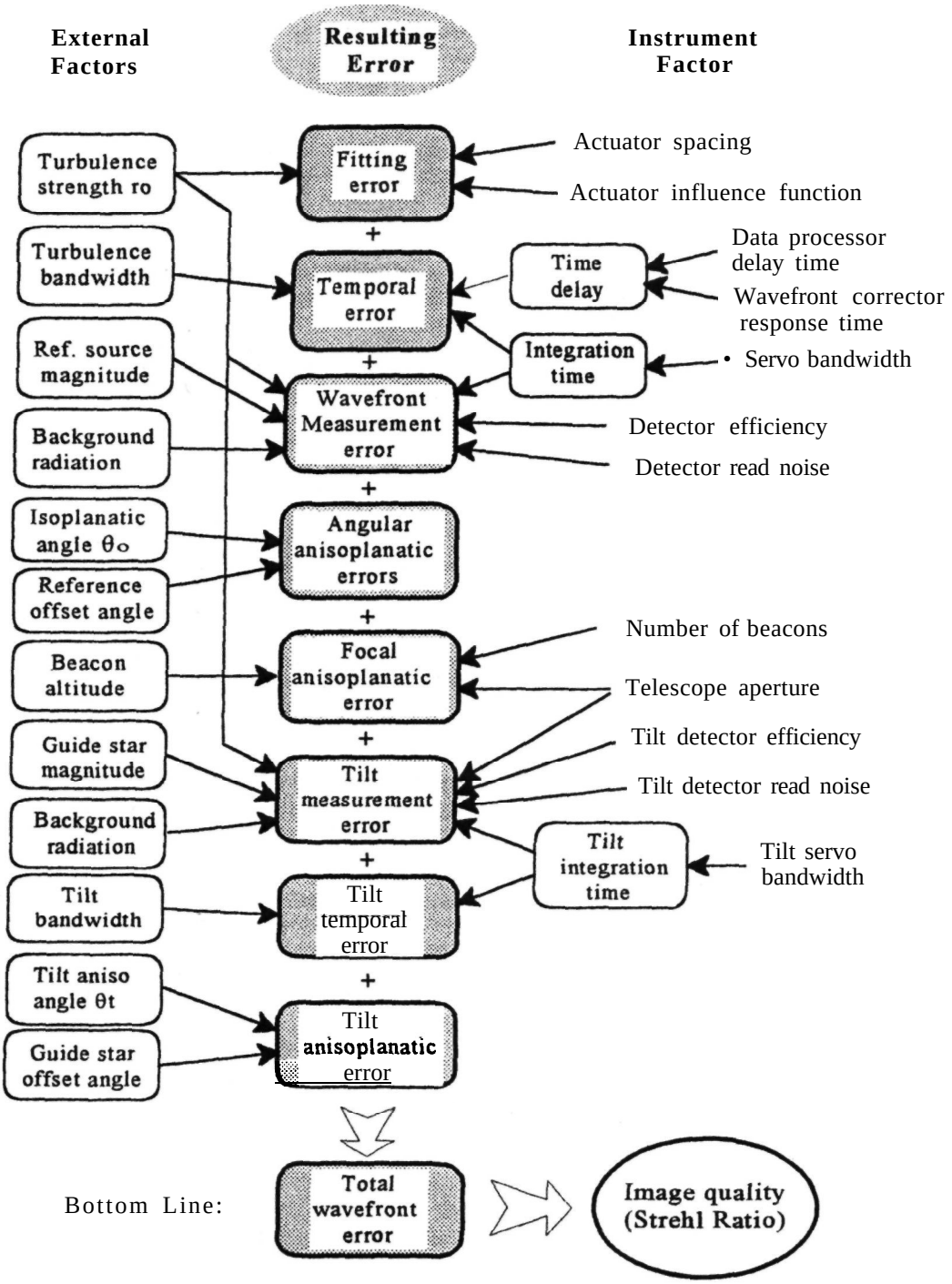


Figure 9.2 Main sources of wavefront error in adaptive optics.

and optimization of adaptive optics systems is also treated in chapter 4.

The image structure may also be analyzed in terms of the optical transfer function (OTF) of the compensated telescope. The OTF is obtained from the structure functions of the errors defined in the preceding text. The transfer function of the compensated system is

$$\begin{aligned} S(r) &= B(r) \cdot T(r) \\ &= e^{-0.5D_s(r)} T(r) \end{aligned} \tag{9.3}$$

where

$B(r)$ = transfer function of the compensated wavefront
$T(r)$ = telescope transfer function
$D_s(r)$ = structure function of the compensated wavefront, which is equal to the sum of the structure functions of the individual components:

$$D_s(r) = D_f(r) + D_p(r) + D_m(r) + D_a(r) \tag{9.4}$$

The point spread function is obtained by the inverse Fourier transform of the transfer function $S(r)$:

$$P(x, y) = \int_{-\infty}^{+\infty}\int S(r) e^{-ir(x+y)}\, dx\, dy \tag{9.5}$$

In the following sections, the wavefront error models and other computations required to calculate the performance of adaptive optics systems using natural stars and laser beacons are described.

9.1.5 The Optimization Process

If a performance criterion, such as the total wavefront error, can be expressed analytically in terms of the system parameters, then the optimum value of each parameter may be found using the standard approach of differentiating the performance function with respect to that parameter and setting the derivative to zero. To maximize the overall performance, all discretionary parameters must be optimized simultaneously. This is achieved conventionally by establishing the relation between individual parameters in order to reduce the analytical expression that defines system performance, until it is a function of a single variable, from which all other parameters are then derived.

In using these mathematical procedures, it is important to keep track of the sensitivity of the solution to variations in individual parameters. Robust adaptive optics systems must accommodate naturally occurring changes in the external (atmospheric) conditions, as well as parameter shifts due to temperature, gravity, and so on, within the instrument itself. A flat peak to the optimization function is desirable, indicating an ability to accommodate large parameter variations that could lead to simplifications in the system design. On the other hand, if the optimization function has a sharp peak, indicating excessive sensitivity, then the design should be reviewed to ensure that the tolerances on the parameter involved can be met under practical conditions. If not, it may be necessary to provide local adaptive control to reduce this sensitivity.

An example of this procedure is the optimization of integration time in adaptive optics systems, which depends on the brightness of the reference source and on the temporal bandwidth of the turbulence, driven by the notoriously capricious atmospheric wind profile. Increasing the integration time allows more photons to be collected from the reference source, thereby reducing the wavefront error due to random noise. However, the resulting increase in time delay between the measurement and correction of wavefronts produces a larger temporal error. The optimum exposure time required to minimize the sum of these errors depends on the wind velocity, which may vary considerably over short periods. One solution is to use an auxiliary control loop that optimizes the integration time continuously, based on measurements of the reference source brightness and effective wind velocity. Such adaptive parameter control systems are discussed in section 9.7.

The overall error of an adaptive optics system is determined by a large number of interacting parameters, so the optimization process must be global.

The availability of fast desktop computers greatly facilitates this approach. It is no longer necessary to reduce the performance of complex multivariate systems, such as adaptive optics, to a single analytical model, a process that usually involves approximations and omissions in order to make the process tractable. Using current software, it is easy to set up individual performance models that include all the relevant factors, to produce numerical solutions that are displayed graphically. In this way, families of curves can be produced interactively, enabling the effect of parameter changes to be assessed rapidly. This process gives much-needed visibility to the sensitivity of overall performance to individual parameters. Single-model optimization is useful to obtain first-order results, but a comprehensive optimization procedure should include a review of the behavior of all relevant parameters; this is now readily accomplished with numerical methods. The process of system optimization is outlined in sections 9.5 and 9.6. Examples of the broad approach to the optimization of integration time and subaperture size are shown in figures 9.24 and 9.26..

9.2 Atmospheric Turbulence Summary

9.2.1 Turbulence Parameters Used in Adaptive Optics

To facilitate the design of adaptive optics systems, the atmospheric characteristics derived in chapter 3

are expressed as a compact set of equations defining the disturbances to be compensated in the optical path. Integration of the $C_N^2(h)$ profile yields the zero-order turbulence moment μ_0, from which is computed the coherence length r_0, usually specified at a standard wavelength of 0.5 μm. The five-thirds moment of the turbulence distribution, $\mu_{5/3}$, yields the angular correlation of the turbulence structure, expressed as the isoplanatic angle θ_0. The $C_N^2(h)$ profile and the wind profile $v(h)$ together yield the velocity moment $v_{5/3}$, from which the temporal power spectrum of the turbulence is computed, yielding the temporal bandwidth f_0 or equivalently, the turbulence time constant, $\tau_0 \approx 1/2\pi f_0$. In a similar way, the second-order wind velocity moment v_2 is used to determine the overall tilt bandwidth f_T of the turbulence.

From these basic descriptors of the atmospheric turbulence, the following key parameters of an adaptive optics system are derived:

1. The coherence length r_0 determines the number of degrees of freedom (or number of subapertures) required for the wavefront corrector, this number being proportional to $(D/r_0)^2$, where D is the telescope aperture. The error in correcting the wavefront is proportional to $(d/r_0)^{5/6}$, where d is the subaperture size.
2. The angle-of-arrival variations and the wavefront phase excursions over the telescope aperture are also computed from the ratio $(D/r_0)^{5/6}$, thereby determining the angular dynamic range requirement for the fast-tracking mirror and the drive distances for the wavefront corrector.
3. The vertical structure of the turbulence provides information for the optimum location of the wavefront corrector, while the angle θ_0 determines the anisoplanatic errors.
4. The turbulence bandwidth f_0 and the time constant τ_0 control the time-dependent errors in an adaptive optics system and therefore determine the temporal bandwidth of the wavefront correction loop and the allowable time delays in the compensation system.
5. The tilt bandwidth f_T similarly determines the bandwidth of the tilt correction loop.

The derivation of these descriptors from atmospheric data and their use in computing wavefront errors will now be reviewed.

9.2.2 Turbulence Coherence Length r_0

The value of the turbulence parameter r_0 produced by any $C_N^2(h)$ profile may be found from equation (3.51), restated here for convenience:

$$r_0 = \left[0.423k^2 \sec\zeta \int_0^H C_N^2(h)dh\right]^{-3/5} \qquad (9.6)$$

The integral of the refractive index structure function C_N^2 is found either from direct measurements made at the observing site, or from a model of the turbulence structure such as those described in section 3.3.4. The parameter r_0 is usually specified as the value at the zenith at a wavelength of 0.5 μm; it is often given as a range of values in terms of the probability of occurrence.

In general, the wavelength at which the science observations are made is different from the wavelength at which the turbulence is measured, and neither of them is centered at 0.50 μm, the wavelength at which the turbulence parameter r_0 is normally specified. The turbulence parameter also depends on the zenith angle of observation. The effective value of the turbulence parameter r_0' at the sensor wavelength λ_0' micrometers is given by

$$r_0' = r_0(\lambda_0'/0.5)^{6/5}\cos^{3/5}(\zeta) \qquad (9.7)$$

where ζ is the zenith angle.

Taking into account the turbulence conditions at the site, the operating wavelengths, and the zenith angles for the adaptive optics, a working value for r_0 can be established, at which the adaptive optics system must deliver a specified image quality. The value of r_0 at the sensor wavelength controls the number of degrees of freedom required in the wavefront corrector.

9.2.3 Angle of Arrival

The wavefront tilt due to atmospheric turbulence is also determined by r_0. The angle-of-arrival variance, specified as the mean-square value over a circular aperture of diameter D for each axis, is given by Greenwood and Fried [1976] as

$$\sigma_\alpha^2 = 0.184\left(\frac{D}{r_0}\right)^{5/3}\left(\frac{\lambda_0}{D}\right)^2 \quad \text{rad}^2 \qquad (9.8)$$

For a Gaussian distribution, the probability is 99.4% that the value will be within ±2.5 standard deviations. The peak excursion of the angle of arrival for this probability is

$$\alpha_{PK} = \pm 1.07\left(\frac{D}{r_0}\right)^{5/6}\left(\frac{\lambda_0}{D}\right) \quad \text{rad} \qquad (9.9)$$

When using a tip-tilt mirror for compensation the required lilt angle is half of this value. Note that because r_0 is proportional to $\lambda^{6/5}$, the angle-of-arrival variations are independent of the observing wavelength.

9.2.4 Peak Wavefront Excursion and Drive Distances

The single-point wavefront excursion with respect to the average phase over a circular aperture of diameter D, expressed as the average mean-square phase at wavelength λ_0, is given by

$$\sigma^2_{T,W} = 1.075\left(\frac{D}{r_0}\right)^{5/3} \quad \text{rad} \quad \text{(overall tilt included)} \tag{9.10}$$

$$\sigma^2_{W} = 0.141\left(\frac{D}{r_0}\right)^{5/3} \quad \text{rad}^2 \quad \text{(overall tilt removed)} \tag{9.11}$$

When averaged over finite subapertures, the wavefront excursion is reduced slightly, so that using the single-point values gives a conservative estimate of the correction range required. These equations show that the excursion required by the wavefront corrector is reduced by a factor of 7(1.075/0.141) = 2.76 when the overall tilt (angle-of-arrival) component is removed with a separate tilt compensator. The peak wavefront excursions are obtained by multiplying the standard deviation by the appropriate factor. For 99.4% probability, the deviation is $\pm 2.5\sigma$, giving the following expression for peak wavefront excursion, in meters, assuming overall tilt removal:

$$W_{PK} = \pm 0.149\lambda_0\left(\frac{D}{r_0}\right)^{5/6} \quad \text{meters} \tag{9.12}$$

When using a deformable mirror for correction, the mirror excursions will be half of this value. In this equation, λ_0 is the wavelength at which r_0 is specified, normally 0.5μm. Because of the negligible dispersion of the atmosphere over the wavelengths of interest, the wavefront excursion is essentially independent of the observing wavelength.

9.2.5 Power Spectrum of Turbulence

The temporal power spectrum of atmospheric turbulence is controlled by wind velocities. The atmosphere is modeled as a series of thin lavers, each characterized by a turbulence strength $C_N^2(\Delta z)$ and wind vector v. It is assumed that the turbulence structure does not change over time scales of less than a few seconds, so that temporal effects are entirely due to transport by the wind. Taking advantage of the fact that the C_N^2 values add linearly, the total turbulence and the resulting power spectral density is found by summing the contributions from each layer. At locations such as mountain-top observatories, the turbulent layers may be well separated and few in number, and so may be treated independently. In general, the atmospheric turbulence and wind profiles are modeled analytically and integrated over the optical path.

For adaptive optics system design, the basic turbulence spectra required are:

- wavefront phase averaged over a subaperture *d*, with respect to the mean phase over the full aperture of diameter *D*;
- wavefront phase averaged over a subaperture *d*, with overall tilt and mean phase removed over the full aperture of diameter *D*;
- wavefront tilt, defined as the average gradient over an aperture of diameter *D*.

It has been found that these spectra can be modeled, with analytic expressions using relatively simple power laws, over the whole frequency range of atmospheric turbulence.

Turbulence may be described using the spatiotemporal phase structure function (see section 3.2.3) given by

$$D(\mathbf{r},\tau) = \langle[\phi(\mathbf{x},t) - \phi(\mathbf{x}+\mathbf{r},t)][\phi(\mathbf{x},t+\tau) - \phi(\mathbf{x}+\mathbf{r},t+\tau)]\rangle \tag{9.13}$$

where x and r are the spatial coordinates, and t and τ represent time. When T = 0, this reduces to the familiar spatial phase structure function. The phase-difference power spectrum $W_\phi(\mathbf{r},f)$ is the Fourier transform of $D(\mathbf{r}, r)$. For plane-wave propagation with a circular aperture of diameter D, the phase-difference power spectrum is

$$W_\phi(\mathbf{r},f) = 2.079\left(\frac{f}{f_0}\right)^{-8/3}\frac{1}{f_0}\left(\frac{D}{r_0}\right)^{5/3}\sin^2\left(\frac{f}{f_0}\frac{r}{D}\right) \tag{9.14}$$

where f_0 is a characteristic frequency equal to v/nD , and v is the turbulence-weighted wind velocity over the optical path, defined by

$$v^{5/3} = \frac{\int C_N^2(\mathbf{z})v^{5/3}(\mathbf{z})\,dz}{\int C_N^2(\mathbf{z})\,dz} = \frac{v_{5/3}}{\mu_0} \tag{9.15}$$

In their classic paper on turbulence power spectra, Greenwood and Fried [1976] show that the required turbulence power spectra for phase and tilt can be expressed as

$$F_\phi(f) = -\tfrac{1}{2}\int W_\phi(\mathbf{r},f)T_\phi(\mathbf{r})\,d\mathbf{r}$$
$$F_\alpha(f) = -\tfrac{1}{2}\int W_\alpha(\mathbf{r},f)T_\alpha(\mathbf{r})\,d\mathbf{r} \tag{9.16}$$

where $T_\phi(\mathbf{r})$ and $T_\alpha(\mathbf{r})$ are transfer functions that define the exact conditions for the required spectra, specifically whether the wavefront phase is referenced to average phase or average tilt. These transfer functions consist of overlap integrals in the optical aperture and subaperture, and account for the auto- and cross-covariances of the wavefront phase and tilt, for the full aperture D and a subaperture d, located anywhere within D.

The total power in each spectrum, which gives the variance of the phase or tilt angle, is given by

$$\sigma^2_\phi = -\tfrac{1}{2}\int D_\phi(\mathbf{r})T_\phi(\mathbf{r})\,d\mathbf{r}$$
$$\sigma^2_\alpha = -\tfrac{1}{2}\int D_\phi(\mathbf{r})T_\alpha(\mathbf{r})\,d\mathbf{r} \tag{9.17}$$

These integrals were evaluated numerically by Greenwood and Fried, yielding results that will now be summarized.

For wavefront phase, the power spectrum contains three basic power laws, which are depicted in figure 9.3 using a representative frequency scale with arbitrary units on the ordinate. The first power law, $F1(f)$, is the two-dimensional Kolmogorov phase spectrum, representing the absolute phase at a single point. This spectrum forms a high-frequency asymptote, defined as follows, in units of radians squared of phase per hertz:

$$F1(f) = 0.0326 \sec(\zeta) k^2 v_{5/3} f^{-8/3} \tag{9.18}$$

In optical imaging systems, the absolute phase is irrelevant and the wavefront is referenced to the mean phase over the aperture. This is equivalent to removing the piston component of the wavefront and it results in a spectrum having a —2/3 power law at low frequencies. This section of the total spectrum is defined by

$$F2(f) = 1.29\left(\frac{r_1}{D}\right)^2 \sec(\zeta) k^2 D^2 v_{-1/3} f^{-2/3} \tag{9.19}$$

In this case, with overall tilt included, the spectral power depends on the position of the sampling point within the full aperture, specified by the ratio (r_1/D). It is highest at the edge of the aperture $(r_1/D = 0.5)$, because here the tilt excursions are largest. At the center of the aperture, there is no phase excursion due to tilt, so the spectrum for this location becomes the same as that for a tilt-removed wavefront.

If the overall tilt is removed over the aperture D, the power spectrum of the wavefront phase follows an $f^{4/3}$ law at low frequencies, considerably reducing the wavefront excursion. This power law, in radians squared of phase per hertz, is given by

$$F3(f) = 0.132 \sec(\zeta) k^2 D^4 \mu_0^{12/5} (v_{5/3})^{-7/5} f^{4/3} \tag{9.20}$$

The large drop in the spectrum at low frequencies produces a commensurate decrease in the required stroke of the deformable mirror, which, of course, is the main reason for using a separate corrector for overall tilt. The transition frequency f_c between the 4/3 and —8/3 power laws is

$$f_c = 0.705 D^{-1} \mu_0^{-3/5} (v_{5/3})^{3/5} = 0.705 v/D \tag{9.21}$$

where v is the turbulence-weighted wind velocity defined in equation (9.15).

The combination of these three power laws produces a sharply peaked power spectrum for wavefront phase. The steep cutoff is a favorable factor in the design of adaptive optics systems, limiting the bandwidth requirement of the servo system. When wavefront phase is averaged over a finite aperture, spatial averaging causes the power spectrum to roll off at a faster rate at high frequencies. The larger the aperture, the lower the frequency at which the roll-off occurs. The rolloff is modeled by multiplying the spectrum by an attenuation factor of the form

$$\gamma = a - b\frac{d}{v} f \qquad 1 \geq \gamma \geq 0 \tag{9.22}$$

where

a and b = constants,
d = subaperture diameter
v = weighted wind velocity and
f = temporal frequency

The values given by Greenwood and Fried for the phase spectrum $F1(f)$ are $a = 1.26$ and $b = 2.24$.

The power spectrum of wavefront tilt has been investigated by Tyler [1994a]. The tilt over a given aperture may be defined in different ways, such as the average wavefront gradient, which is called G-*tilt*, or as the normal to the plane that minimizes the wavefront distortion, which is known as *Z-tilt* because it is is equivalent to the two tilt terms in the Zernike polynomial expansion (see section 3.5). For imaging purposes, we are usually concerned with stabilizing the center of intensity of an image, which is equivalent to minimizing the G-tilt. The standard method of wavefront measurement, using either a Shack-Hartmann or shearing interferometer sensor, is also based on detecting the center of intensity of the reference source image, so this is a consistent approach.

The power laws governing the frequency spectrum of wavefront tilt are shown in figure 9.4. At low frequencies, the power spectral density, expressed in radians squared of single-axis tilt per hertz, is

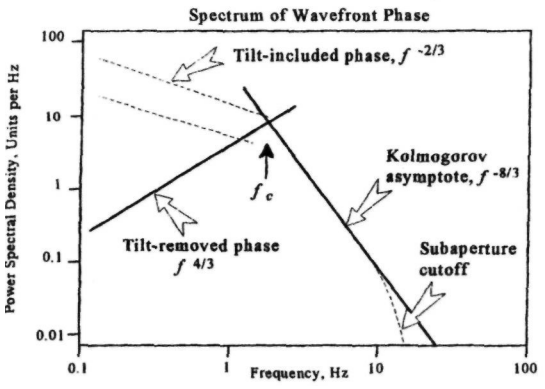


Figure 9.3 Power-law dependencies of the spectrum of wavefront phase variations due to atmospheric turbulence. At low frequencies, the phase is dominated by tilt and varies as the —2/3 power of frequency. If overall tilt is removed, the phase spectrum at low frequencies follows a 4/3 power law. At middle frequencies, the basic two-dimensional Kolmogorov phase spectrum is dominant, producing a —8/3 power law. At high frequencies, the phase spectrum is cut off by subaperture averaging and drops rapidly.

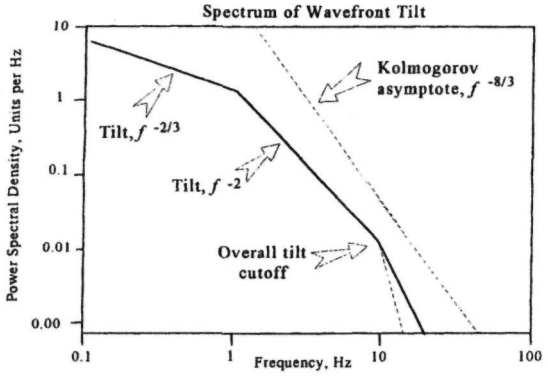


Figure 9.4 Power-law dependencies of the spectrum of wavefront tilt. At low frequencies, the tilt spectrum is proportional to the —2/3 power of frequency, which becomes an inverse-square dependency at intermediate frequencies. When tilt is averaged over an aperture, the overall tilt spectrum is cut off at a frequency determined by the diameter of the full aperture and the wind velocity. For gradient tilt, the spectrum falls off as the —11/3 power of frequency.

$$F4(f) = 1.29\sec(\zeta)v_{-1/3}f^{-2/3} \qquad (9.23)$$

This has the same $f^{-2/3}$ dependency as $F2(f)$, with the difference that the angular tilt spectrum is independent of wavelength and aperture diameter. At intermediate frequencies, the spectrum falls off at a higher rate, f . There is little difference between the two definitions of tilt in these ranges. However, at a frequency of approximately $f = v/D$, where v is the turbulence-weighted wind velocity, spatial averaging over the aperture causes a further break in the power law, and from here the two types of tilt fall off at different rates: $f^{-11/3}$ in the case of G-tilt and $f^{-17/3}$ for Z-tilt. At this point, the spectral density is so low that the difference has little practical significance.

Turbulence power spectra for specific atmospheric conditions are shown in figures 9.5 and 9.6, using both logarithmic and linear plots. The effect of varying wind velocity and turbulence strength at a constant wavelength is shown in figure 9.5, using turbulence-weighted wind velocities of 10, 15, 20, and 25 $\mathrm{m\,s^{-1}}$. Turbulence strength normally increases with wind velocity and this is modeled in the figure. The spectra are sharply peaked at frequencies between about 1 and 10 Hz, the cutoff frequency increasing with wind velocity as expected. The high frequency response is inversely proportional to the size of the subapertures, so that for a given wind velocity v, the frequency response is reduced for larger subapertures. The power and bandwidth of the spectrum both decrease considerably at IR wavelengths.

The power spectrum of tilt, shown in figure 9.6 for the same conditions as figure 9.5, behaves in a different way. To start with, the tilt angle is independent of wavelength. At low frequencies, the power spectral density is inversely proportional to the 2/3 power of frequency. The high frequency cutoff is governed by the telescope aperture and wind velocity. The larger the aperture, the lower the overall tilt bandwidth. For the relatively benign conditions at good observing sites, the tilt spectrum drops very rapidly between about 1 and 10 Hz, frequencies at which the phase spectrum has its peak response.

The use of logarithmic scales for these power spectra clearly shows the underlying power laws on which they are based, but it distorts the spectral power distribution. To show the spectral power correctly, in proportion to the area under the curve, linear scales are necessary. These linear plots reveal the true characteristics of the wavefront phase and tilt spectra, suggesting that the models may be further simplified by using only a single dominant power law. Temporal errors caused by finite servo bandwidth may then be determined directly by comparing the bandwidth of the turbulence with that of the adaptive optics servo. This approach was first suggested by Greenwood [1977] and is discussed further in section 9.4.3.

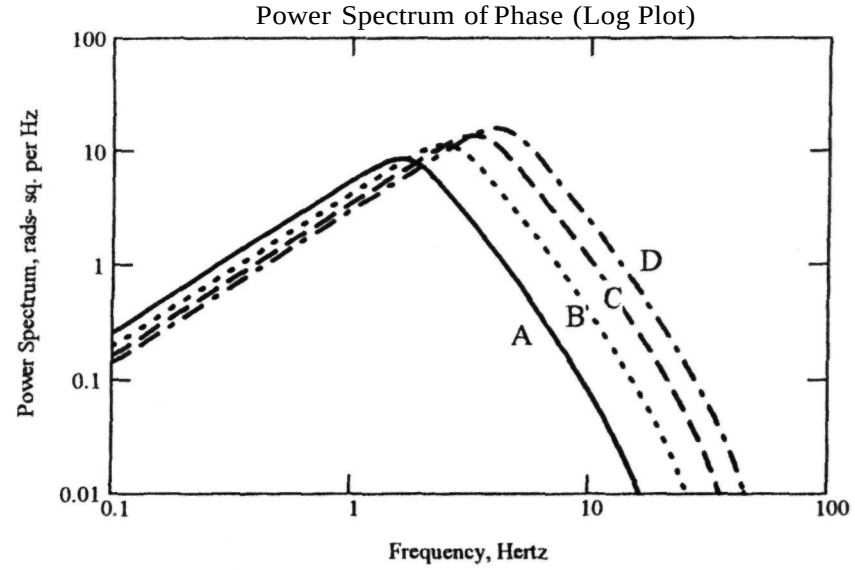
9.3 External Sources of Wavefront Error

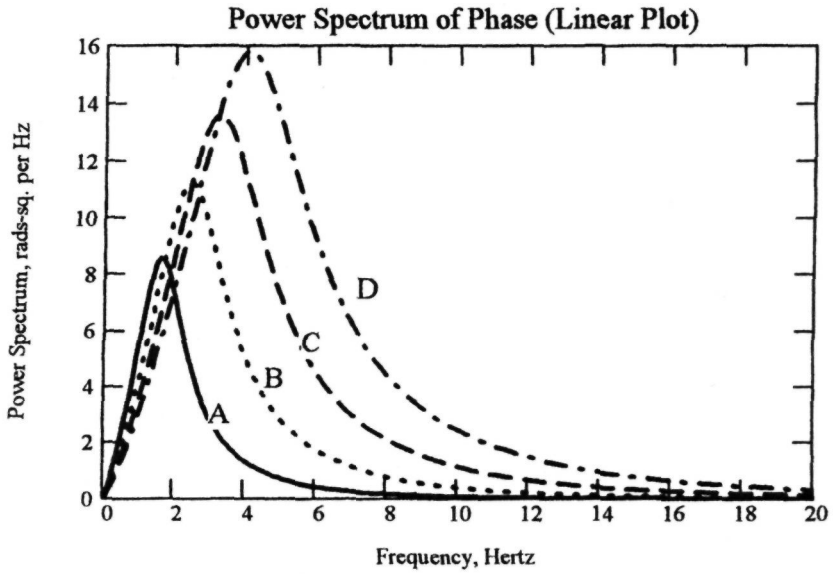
9.3.1 Angular Anisoplanatism

The isoplanatic angle is an extremely important parameter in adaptive optics, as it governs the field of view over which wavefront compensation is possible using a single corrector. Wavefront measurements made in a specific direction through distributed turbulence lose their validity as the angle between the measurement and the observation increases. The isoplanatic angle θ_0 is defined as the angle at which the ensemble average mean-square error between the measured and observed wavefronts is 1 rad^2. The isoplanatic angle depends on the vertical structure of the turbulence, as given by the C_N^2 profile, weighted by the 5/3 power of the range from the telescope. It is computed from the following relation, derived in section 3.7, valid when $D \gg r_0$:

$$\theta_0 = \left[2.91k^2(\sec\zeta)^{8/3}\mu_{5/3}\right]^{-3/5} \qquad (9.24)$$

The weighted turbulence profile $\mu_{5/3}$, known as the *five-thirds order turbulence moment*, has been defined in section 3.3.6. Typical profiles are shown in figure 9.7. As a result of this weighting, the isoplanatic angle is determined mainly by high-altitude turbulence, especially that occurring at the tropopause. Turbulence near the ground, even when severe, has little effect on the isoplanatic angle.





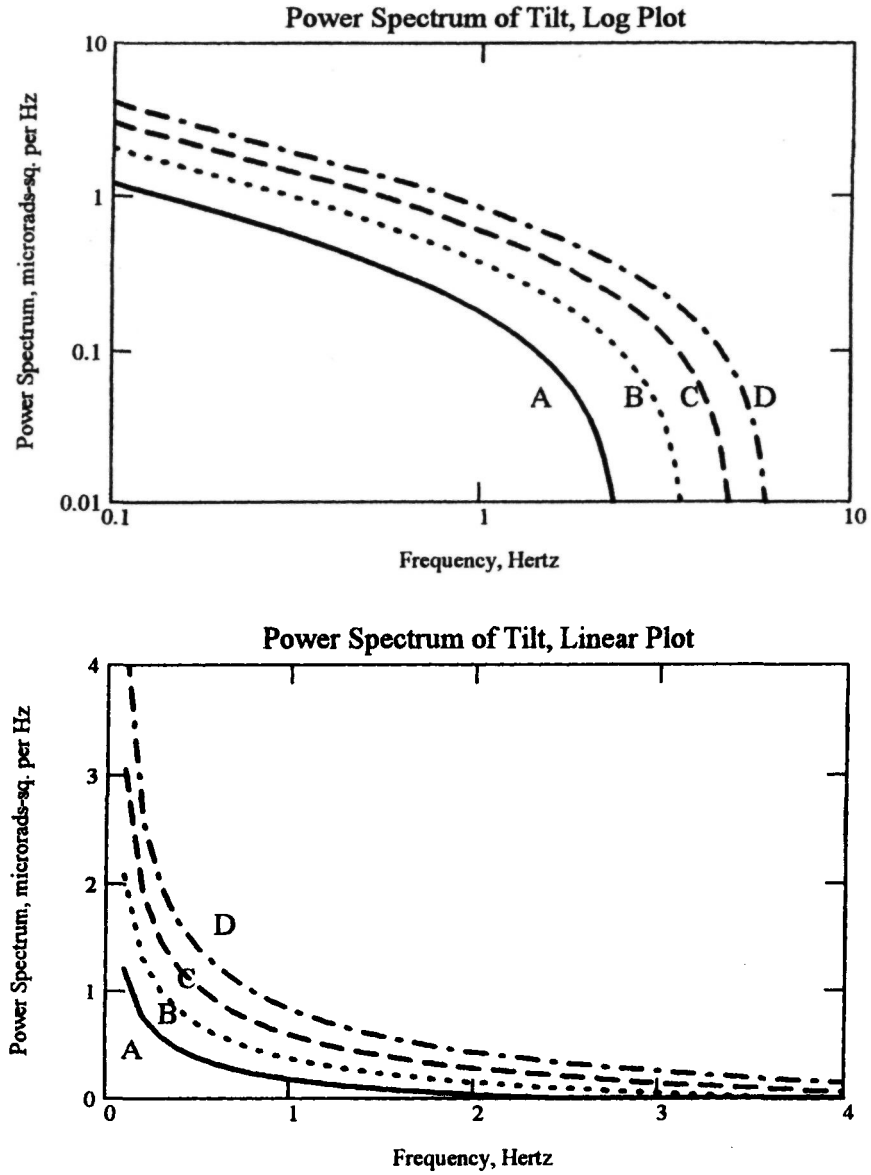
A wind velocity 10 m/s, r_0 - 0.2 m at λ = 0.5 μm

B wind velocity 15 m/s, r_0 — 0.13 m

C wind velocity 20 ra/s, r_0 = 0.1 m

D wind velocity 25 m/s, r_0 = 0.08 m

Figure 9.5 Power spectra of wavefront (phase) error due to atmospheric turbulence, as a function of wind speed and r_0. The conditions are: aperture 4m, wavelength 0.7 μm, zenith angle 45°, subaperture size 0.25 m.



A wind velocity 10 m/s, r_0 = 0.2 m at λ = 0.5 μm

B wind velocity 15 m/s, r_0 = 0.13 m

C wind velocity 20 m/s, r_0 = 0.1 m

D wind velocity 25 m/s, r_0 = 0.08 m

Figure 9.6 Power spectrum of single-axis overall tilt due to atmospheric turbulence, as a function of turbulence-weighted wind speed and r_0 (measured at the zenith). The aperture is 4 m and the zenith angle is 45°. This power spectrum is independent of wavelength.

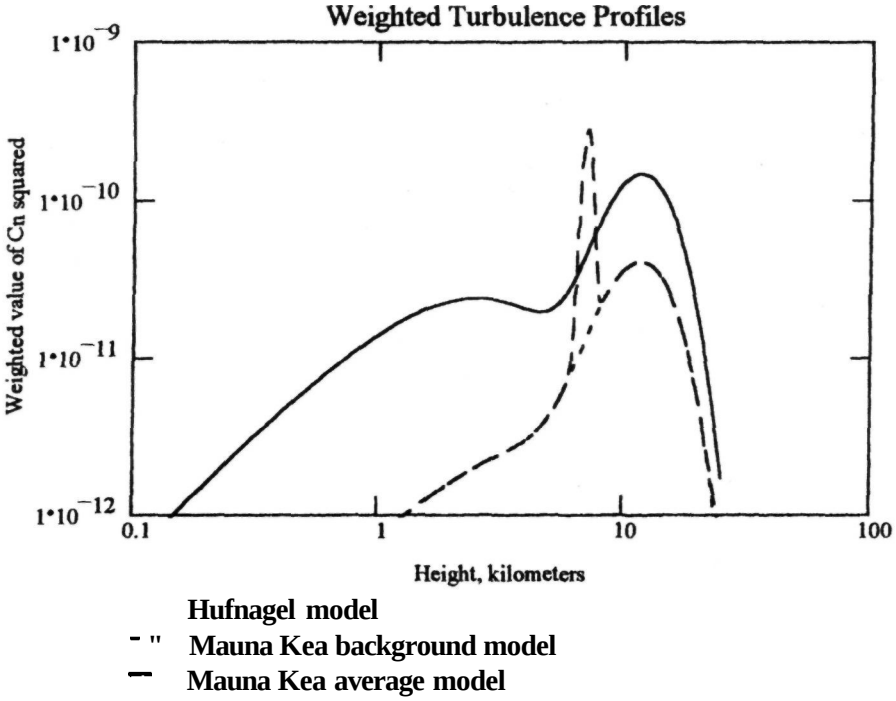


Figure 9.7 Turbulence profiles weighted for angular anisoplanatism. To show the relative contribution of turbulence at each height to the anisoplanatic error, values of C_N^2 are weighted by the 5/3 power of the height. The dominant contributions come from turbulence in the region of 10 km.

As discussed in section 3.7, it is often convenient to define the isoplanatic angle in terms of r_0 and the mean turbulence height $\bar{h}$, using the expression

$$\theta_0 = 0.314 \frac{r_0}{\bar{h}} \cos\zeta \tag{9.25}$$

where ζ is the zenith angle and the mean turbulence height h is defined as

$$\bar{h} = \left[\frac{\int dh\, h^{5/3} C_N^2(h)}{\int dh\, C_N^2(h)} \right]^{3/5} = \left(\frac{\mu_{5/3}}{\mu_0} \right)^{3/5} \tag{9.26}$$

Typical values of h are 1500 m for daytime turbulence and 3000 m for night-time operation.

Even with a fixed value of A, the isoplanatic angle varies with both wavelength and zenith angle. The turbulence parameter r_0 is normally specified at $\lambda_0 = 0.5\,\mu$m for vertical propagation, but its actual value varies as $\lambda^{6/5} \cos^{3/5}(\zeta)$. The isoplanatic angle at wavelength λ and zenith angle ζ is then

$$\theta_0(\lambda, \zeta) = 0.314 \frac{r_0}{\bar{h}} \left(\frac{\lambda}{\lambda_0} \right)^{6/5} \cos^{8/5}(\zeta) \tag{9.27}$$

The mean-square error due to anisoplanatism at any angle θ may be expressed as

$$\sigma_A^2 = \left(\frac{\theta}{\theta_0} \right)^{5/3} \tag{9.28}$$

9.3.2 Focal Anisoplanatism

The use of laser beacons in adaptive optics systems produces the following types of wavefront error:

1. *Beacon geometry.* The physical location of laser beacons within the atmosphere means that atmospheric turbulence is sampled only partially by the beacon light. The errors due to focal anisoplanatism with single and multiple beacons are addressed in this section.
2. *Image stabilization error.* Laser beacons do not allow the measurement of the absolute position of a source, for which a fixed reference (such as a natural star) is required. Due to the scarcity of suitable stars, the angular separation between the science object and the fixed reference may be considerable, resulting in tilt anisoplanatic errors. These errors are addressed in section 9.4.6.

The wavefront measurement errors produced by the geometry of laser beacons have been analyzed in section 7.3. These errors are now put into a suitable form for use in the adaptive optics performance model. In the general case of multiple laser beacons, the wavefront errors associated with the beacon geometry are:

- focal anisoplanatism, σ_B^2;
- conic tilt (the tilt component of focal anisoplanatism), σ_{CT}^2;
- beacon position uncertainty, CT£.

The total wavefront error due to beacon geometry, from equation (7.37), is

$$\sigma_{MB}^2 = \sigma_B^2 + E(\sigma_{CT}^2) + \sigma_P^2 \tag{9.29}$$

where E is an error propagation factor dependent on the correlation between the tilt components of multiple beacons. Values of E have been given in table 7.4. For a single beacon, E is zero, while for array sizes of interest, E may be approximated as $0.1N^{5/3}$. It was shown in section 7.3.5 that the error due to uncertainty in the beacon position σ_P^2 is about one-tenth of the conic tilt error CTCT for Rayleigh beacons, and between one-tenth and one-third of the conic tilt error for sodium beacons. On this basis, the beacon position error may be neglected for first-order analysis. Using equations (7.24), (7.39), and (9.35), the wavefront error due to multiple beacons may then be expressed in terms of system parameters as

$$\begin{aligned}\sigma_{MB}^2 &= \left(\frac{D}{N d_0}\right)^{5/3} + E\left(\frac{D}{N d_1}\right)^{5/3} \\ &\approx D^{5/3}\left[\left(\frac{1}{N d_0}\right)^{5/3} + \left(\frac{1}{10 d_1}\right)^{5/3}\right]\end{aligned} \tag{9.30}$$

where

D = telescope aperture
N = number of beacons in one dimension
d_0 = focal anisoplanatism parameter defined by equation (7.36), values of which are given in table 7.2
d_1 = tilt anisoplanatism parameter defined by equation (7.40), values of which are given in table 7.3

9.3.3 Effects of Atmospheric Dispersion

In section 3.3, it was shown that the refractivity of Earth's atmosphere varies significantly with altitude and wavelength. The refractivity, expressed in parts per million, has a value of close to 280 at sea level for $X = 0.5\ \mu$m, decreasing to about 95 at an altitude of 10 km. The exponential decrease in refractivity with altitude causes incoming light rays, at angles away from the zenith, to follow a curved path through the atmosphere, changing the apparent direction of celestial objects. This effect, in itself, has no significance for adaptive optics. It is only the dispersive component of the refractivity that causes errors. The dispersion causes rays at different wavelengths to traverse different paths through the atmosphere, producing four distinct effects, which are identified as follows:

1. angular dispersion,
2. chromatic path-length error,
3. dispersion displacement error,
4. multispectral error.

The wavefront errors produced by these effects will now be described.

Angular Dispersion

Angular dispersion is the result of the average refraction of light as it passes through the atmosphere, producing an effect found in all ground-based telescopes, whether compensated or not. Incoming rays from a distant object are bent by atmospheric refraction through an angle depending on the zenith angle of observation. At the telescope focal plane, the image is dispersed in a vertical plane by an amount proportional to the spectral bandwidth and the zenith angle. For high-resolution imaging over broad spectral bands, it is essential to compensate the angular dispersion, using prisms specially designed for the purpose.

The angular deviation of a ray passing through Earth's atmosphere is given by

$$\Delta z = -N_S(\lambda)\tan(\zeta)\frac{\rho(h)}{\rho_S} \tag{9.31}$$

where

N_S = refractivity of the atmosphere at standard temperature and pressure at wavelength λ, an expression for which is given in equation (3.16)
ζ = zenith angle of observation
$\rho(h)$ = density of the atmosphere at height h
ρ_S = standard density of the atmosphere

The effect of angular dispersion as a function of zenith angle and wavelength is illustrated in figure 9.8, showing the image spread over a wavelength range of 0.4–2.2 μm. To provide a reference for comparison, the Airy disk of a 1 m aperture telescope is shown to scale at each wavelength. It can be seen that dispersive effects are negligible at wavelengths longer than 1.25 μm, except at zenith angles of 60° or more. However, at visible wavelengths, angular dispersion exceeds the Airy disk diameter even at small zenith angles, making correction essential for high-resolution observations through the atmosphere.

In adaptive optics systems, it is necessary not only to correct the imaging path for angular dispersion, but also to correct the wavefront measurement path, if a broadband reference source, such as a natural

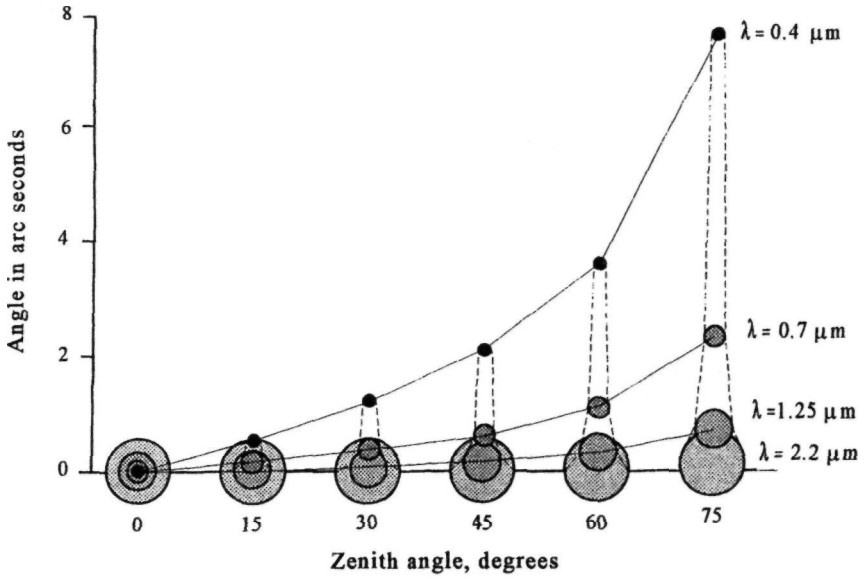


Figure 9.8 Angular dispersion of a star image, caused by the density profile of the atmosphere. The wavelength range shown is 0.4–$0.2\,\mu$m. The Airy disk of a 1-m aperture telescope is shown to scale at each wavelength. At wavelengths longer than $1.25\,\mu$m, dispersion effects are small compared with the Airy disk, except at very large zenith angles. At visible wavelengths, however, angular dispersion exceeds the Airy disk diameter, even at small zenith angles, necessitating prismatic correction for white light.

star, is used. If this were not done, the performance of the wavefront sensor would be compromised by the angular spread of the reference image, reducing the signal-to-noise ratio of the wavefront measurement.

Angular dispersion may be corrected using a Risley prism, which consists of two rotating prisms located between the deformable mirror and the wavefront sensor. In the null position, the prism axes are oriented at 180°, so their effects cancel. The magnitude of the correction is controlled by differential rotation, while the orientation of the correction is controlled by the common rotation of the prisms. By suitable choice of glasses, these devices can be made to operate over a wide spectral band [Wallner and Wetherall 1980].

Chromatic Error

Rays of different wavelengths traversing the same atmospheric path travel at different velocities, depending on the refractivity of air at each wavelength. The optical path length then appears to be different at each wavelength, causing chromatic errors. When a deformable mirror is used to compensate turbulence, it makes the same path length correction at all wavelengths, so the chromatic error is not compensated. Note that the chromatic error applies to a single atmospheric path: if the dispersion causes the rays to traverse different regions of the atmosphere, then an additional error due to dispersion displacement (defined later in this section) is incurred. The chromatic error has been analyzed by Wallner [1977], who expressed it in the form

$$\sigma_{ch}^2 = \sigma_u^2 \varepsilon(\lambda) \tag{9.32}$$

where

σ_u = uncorrected wavefront error due to turbulence
= $0.026(D/r_0)^{5/3}$ waves squared

and

$$\varepsilon(\lambda) = \text{chromatic error coefficient} = \frac{\int_0^\infty d\lambda\, I(\lambda) N^2(\lambda)}{N^2(\lambda_0)} \tag{9.33}$$

where

$I(\lambda)$ = effective intensity distribution of the source,
$N(\lambda)$ = refractivity at wavelength λ and
λ_0 = mean wavelength.

For operation at visible wavelengths, the value of $\varepsilon(\lambda)$ is approximately 1.0×10^{-4}. The chromatic error for a 4-m compensated telescope is then about 1/30

wave, with an r_0 value of 0.1 m. At longer wavelengths, the error will be even less. In most cases, an error of this magnitude can be ignored.

Dispersion Displacement Error

Dispersion error is another result of the bending of incoming light rays by atmospheric refraction. The angular dispersion causes rays of different wavelengths that were coincident at the top of the atmosphere to be displaced, so that they enter the telescope aperture at different locations. Spectral dispersion produces errors in adaptive optics systems in two different ways: (1) errors are caused by using a finite bandwidth for either the reference source or the science observations, even if they are centered at the same wavelength; and (2) errors are produced by using different center wavelengths for the reference and science observations, even when each is monochromatic. The second situation is known as multispectral operation and the errors incurred are discussed in section 9.3.4.

The dispersion displacement error with finite spectral bandwidth is due to the fact that each location on the telescope aperture receives rays that arrived from the same source but which traversed different atmospheric paths. The correlation between the path-length errors decreases as the spectral bandwidth increases, making exact compensation impossible. The error due to this effect has also been analyzed by Wallner [1977]. Figure 9.9 shows the relevant geometry. The vertical displacement of a ray due to atmospheric dispersion can be expressed as

$$\Delta b_0 = [N_s(\lambda) - N_s(\lambda')]\sec(\zeta)\tan(\zeta)(P_0/g\rho_s) \quad (9.34)$$

where

$N_s(\lambda)$ and $N_s(\lambda')$ = refractivities at wavelengths A and X'', respectively
ζ = true zenith angle of the ray,
P_0 = atmospheric pressure at the telescope,
g = acceleration of gravity
Ps = density of air for standard conditions

The term $P_0/g\rho_S$ can be interpreted as the effective depth of the atmosphere at the telescope site. Under standard conditions at sea level, it has a value of 8432 m.

If the adaptive optical system makes the best achromatic correction to rays of all colors, then there will remain a residual mean-square error given by

$$\sigma_D^2 = \frac{1}{2}\sec^{8/3}(\zeta)\tan^{5/3}(\zeta)\left[\frac{P_0}{g\rho_S}\right]^{5/3} \times G(\lambda) \times H(h) \quad (9.35)$$

where

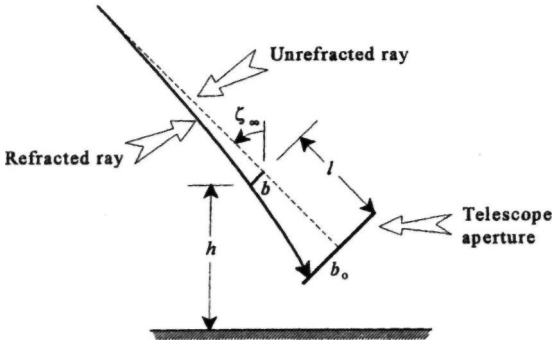


Figure 9.9 Geometry of ray displacement due to atmospheric refraction.

$$G(\lambda) = \int_0^\infty d\lambda\, d\lambda' [2I_A(\lambda) - I_B(\lambda')] \times I_B(\lambda')|N(\lambda) - N(\lambda')|^{5/3} \quad (9.36)$$

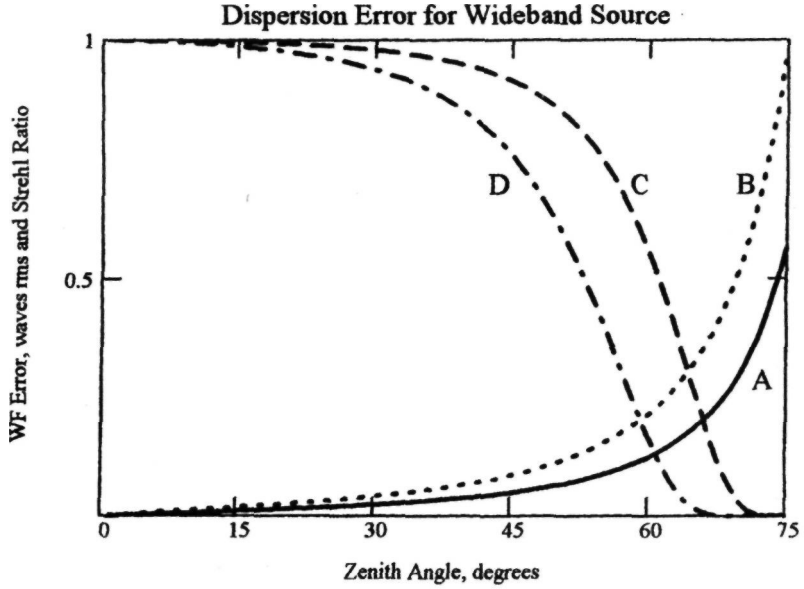
$$H(h) = \int_0^\infty dh \frac{2.91}{\lambda_0^2} C_N^2(h)\left|1 - \frac{P(h)}{P_0}\right|^{5/3} \quad (9.37)$$

and

$I_A(\lambda)$ = spectral intensity distribution of the observed object
$I_B(\lambda')$ = spectral intensity distribution of the reference source
λ_0 = mean reference wavelength
$C_N^2(h)$ = refractive index structure parameter, $m^{-2/3}$

The derivation of equation (9.35) assumes that the aperture is larger than than the ray separations due to dispersion, which is generally true for astronomical telescopes with apertures over about 1 m. When $I_A(\lambda)$ and $I_B(\lambda')$ are both centered in the visible band at wavelengths between 0.4 and 0.75 μm, using a 5500° blackbody source distribution through 1 atmosphere, and with detectors having an S-20 photocathode response, the spectral integral $G(\lambda)$ has the value of 4.19 x 10^{-10}. The height integral $H(h)$ depends mainly on the atmospheric turbulence profile specified by $C_N^2(h)$. An exponential turbulence distribution with a scale height of 3200 m, corresponding to an r_0 value of 0.1 m, gives $H(h) = 1.3$ waves squared $m^{-5/3}$ at a reference wavelength of 0.55 μm. When the low-altitude turbulence is reduced by half and a tropopause layer having equal strength is added at 15 km, for the same value of r_0, then $H(h)$ = 3.95 waves squared $m^{-5/3}$.

The dispersion error for these two atmospheric turbulence profiles is plotted in figure 9.10 as a function of zenith angle. For zenith angles up to about 30°, the wavefront error is below 1/20 wave and therefore tolerable, but it grows rapidly at larger



A wavefront error, no tropopause layer
B wavefront error with tropopause layer
C Strehl ratio, no tropopause layer
D Strehl ratio with tropopause layer

Figure 9.10 Wavefront error caused by spectral dispersion when using a wideband reference source, such as a natural star. Rays at different wavelengths traverse separate paths through the atmosphere, resulting in a measurement error that depends on the turbulence structure and bandwidth. This chart is drawn for moderate turbulence, using a spectral band of 0.4–0.75 μm, similar to solar radiation.

angles, especially when a tropopause layer is present. Because this error is produced over a long atmospheric path, it is not possible to compensate it completely with a single dispersion correction device. Wallner [1977] has shown that a lateral dispersion corrector, which must be located before the wavefront compensation device, is capable of reducing the dispersion error to about one-third.

9.3.4 *Multispectral Error*

This error is encountered when the wavefront sensor operates at a different wavelength from the science observations, often the case in astronomical adaptive optics. The rationale for using multispectral operation is that atmospheric turbulence degrades observations much less in the IR windows between 1 and 4 μm than at visible wavelengths, enabling simpler adaptive optics to be used, while the most efficient, low-noise detectors for wavefront sensing operate at wavelengths below 1 μm. Adaptive optics systems used for observations at IR wavelengths consequently employ wavefront sensors operating at shorter wavelengths. This is true for systems using either natural stars or laser beacons as reference sources. With laser beacons, Rayleigh scatter is produced at all wavelengths, but is most efficient at ultraviolet (UV) wavelengths, while the alternative approach using sodium resonance fluorescence produces yellow light at a fixed wavelength of 0.589 μm.

To make multispectral operation work, it is necessary for the adaptive optics to measure and correct the turbulence over exactly the same atmospheric path as that taken by the science observations. If the dispersion of air was zero (that is, the refractivity was constant with wavelength), then the optical paths would be identical at all wavelengths and multispectral operation would incur no additional errors. In fact, dispersion causes rays at different wavelengths to become separated by distances comparable with r_0, resulting in large errors; great care is therefore necessary in the design and operation of adaptive optics systems using multispectral operation.

The ray separation at the telescope aperture in the vertical plane, caused by dispersion, is easily computed using equation (9.34). For example, using wavelengths of 0.5 and 2.2 μm, at a zenith angle of 45°, the sensing and observing beams would be displaced by 0.066 m. The effect of this ray separation on the wavefront error has been analyzed by Wallner [1984]. The random optical path difference (OPD) caused by turbulence between two rays at wavelengths λ_1 and λ_2 depends on their separation, but not on their absolute positions. Rays from an object at height H travel through the atmosphere to the telescope aperture, arriving at points separated by vector y. Each ray is displaced in the vertical plane by dispersion, according to equation (9.34). The resulting wavefront error can be expressed in terms of the structure functions of the optical path differences between the two rays as follows:

$$\sigma_{\text{opd}}^2 = D(0, \lambda_1, \lambda_2) - \int_{-\infty}^{\infty} dy\, A(\mathbf{y}) \times \left[\tfrac{1}{2}D(\mathbf{y}, \lambda_1, \lambda_2) + \tfrac{1}{2}D(\mathbf{y}, \lambda_2, \lambda_1) - D(\mathbf{y}, \lambda_1, \lambda_1)\right] \tag{9.38}$$

where $A(y)$ is a weighting function equal to the overlap area of two circular apertures of radius a displaced by $y = |\mathbf{y}|$:

$$A(y) = \frac{2}{\pi}\left\{\cos^{-1}(y/2a) - (y/2a)[1 - (y/2a)^2]^{1/2}\right\} \qquad 0 < y < 2a$$
$$= 0 \qquad y \geq 2a \tag{9.39}$$

For Kolmogorov turbulence, the structure functions are given by

$$D(\mathbf{y}, \lambda_1, \lambda_2) = 2.914 \sec\zeta \int_0^H dh\, C_N^2(h) |s(\mathbf{y}, \lambda_1, \lambda_2, h)|^{5/3} \tag{9.40}$$

where $s(\mathbf{y}, \lambda_1, \lambda_2, h)$ is the separation at height h in the atmosphere, of rays of wavelength λ_1 and λ_2, originating from an object at height H and reaching coordinate v at the telescope aperture. Substituting equation (9.40) in equation (9.38), the following expression for the mean-square error in terms of the separation of the rays is obtained:

$$\sigma_{\text{opd}}^2 = 2.914 \sec\zeta \int_0^H dh\, C_N^2(h) \left\{ |s(0, \lambda_1, \lambda_2, h)|^{5/3} - \frac{1}{a}\int_{-\infty}^{\infty} dy\, A(\mathbf{y}) \left[\tfrac{1}{2}|s(\mathbf{y}, \lambda_1, \lambda_2, h)|^{5/3} + \tfrac{1}{2}|s(\mathbf{y}, \lambda_2, \lambda_1, h)|^{5/3} - |s(\mathbf{y}, \lambda_1, \lambda_1, h)|^{5/3}\right] \right\} \tag{9.41}$$

The ray separation is the sum of two components: (1) the geometrical separation of rays traveling in a straight line between the object and the telescope aperture; and (2) the separation due to atmospheric dispersion. The total separation is then

$$s(\mathbf{y}, \lambda_1, \lambda_2, h) = \left(1 - \frac{h}{H}\right)\mathbf{y} + [b_0(\lambda_1) - b_0(\lambda_2)] \left\{1 - \frac{P(h)}{P_0} - \left[1 - \frac{P(H)}{P_0}\right]\frac{h}{H}\right\} \tag{9.42}$$

where $[b_0(\lambda_1) - b_0(\lambda_2)]$ is the vertical separation due to dispersion, and $P(h)$ and $P(H)$ are the atmospheric pressures at heights h and H, respectively.

For astronomical objects, H is set to a height at which the atmospheric pressure $P(H)$ is effectively zero. The effect of multispectral operation on telescope performance is best illustrated by converting the OPD error given by equation (9.41) into the Strehl ratio at the observing wavelength.

Figure 9.11 shows the multispectral Strehl ratio at fixed sensor wavelengths of 0.4, 0.589, and 1.25 μm, as a function of observation wavelength for zenith angles of 45 and 60°. The Hufnagel-Valley (HV) turbulence model is used to compute the wavefront error, with an r_0 value of 0.1 m, defined at a wavelength of 0.5 μm and at sea level. These charts illustrate the useful range of observation wavelengths obtained for each sensor or beacon wavelength. For zenith angles less than 45°, the loss due to dispersion is relatively small, except at wavelengths below 0.5 μm. However, at 60° zenith angle, dispersion can produce significant losses, especially with sensor wavelengths below 0.5 μm. The narrowness of the peaks for short sensing wavelengths are striking. For observation wavelengths in the visible spectrum, it is clear that the sensor wavelength should be as close as possible to the observation wavelength. For observing at IR wavelengths, considerably greater latitude is allowed in the choice of sensor wavelength, although, again, the use of sensor wavelengths below 0.5μm should be avoided. The rise in Strehl ratio at longer wavelengths is due to the fact that a fixed OPD error becomes a smaller phase error as the wavelength increases.

The main conclusions to be drawn from this analysis of multispectral operation are:

1. The best compensation of observation wavelengths over 0.5 μm is obtained by using a sensor wavelength greater than 0.5 μm.
2. Multispectral operation with either sensor or observation wavelengths below 0.5 μm is of questionable value at large zenith angles because of the rapid increase in dispersion.

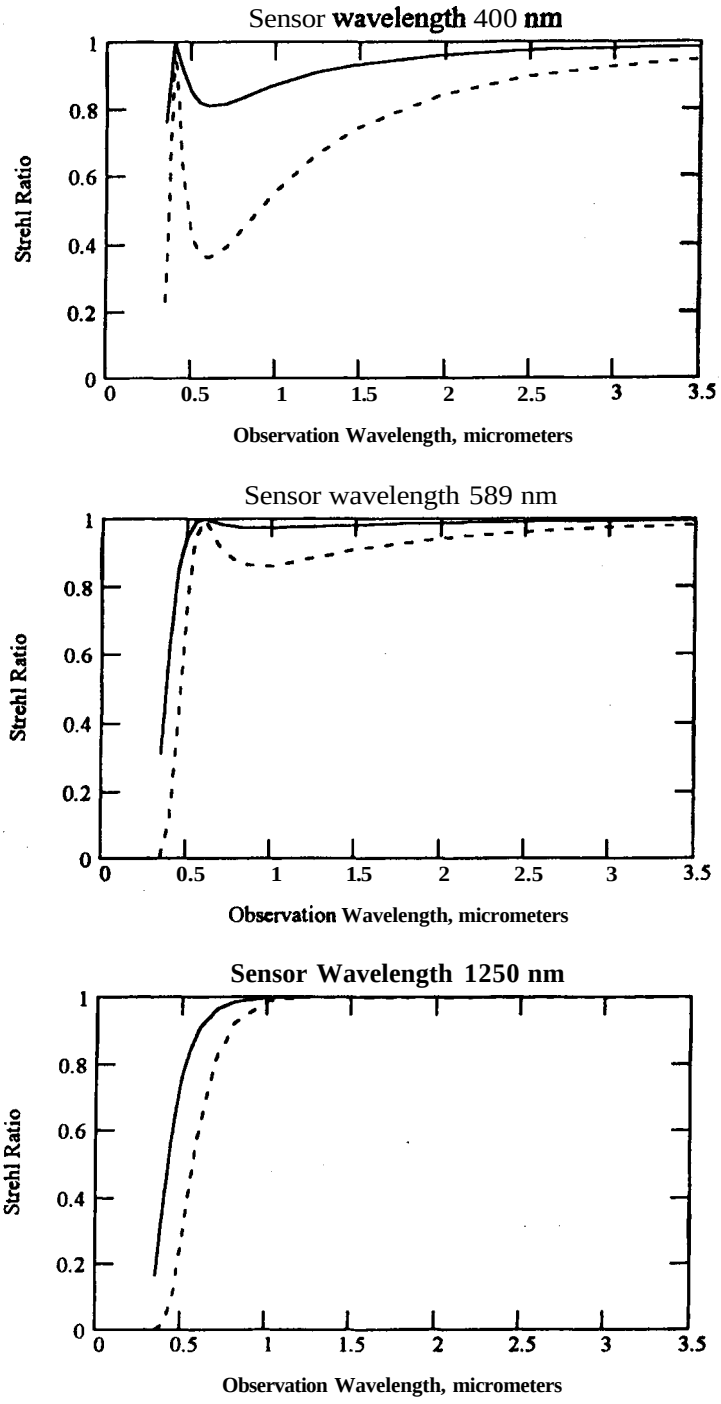


Figure 9.11 Effect of atmospheric dispersion in multispectral operation. These charts show the reduction in Strehl ratio caused by optical path separation in the atmosphere when different wavelengths are used for wavefront sensing and observation. The telescope has an aperture of 4 m and is situated at an altitude of 3000 m. The HV turbulence model is used, with $r_0 = 0.1$ m at $\lambda = 0.5\mu$m, at sea level.

9.4 Instrumental Error Sources

The main errors contributed by the components within an adaptive optics system are the measurement error of the wavefront sensor, the error propagation of the reconstruction process, temporal errors due to time delays and the limited bandwidth of the feedback loop, and the wavefront fitting error associated with the deformable mirror or other correction device. The analytical models required to compute the instrumental errors are summarized in the following sections.

9.4.1 Wavefront Measurement Errors

The measurement error of all wavefront sensors depends largely on two characteristics of the reference source: its brightness, which determines the signal-to-noise ratio of the detected signal, and its angular size. A third factor is the density of sampling points on the wavefront, which determines the wavefront fitting error. The fitting error is usually accounted for in the wavefront corrector, on the assumption that the wavefront sensor has at least as many degrees of freedom as the corrector, which is true in most cases.

The sensor signal-to-noise ratio is calculated from the photon counts produced by the reference source and sky background, together with the electron noise contributed by the photon detector. In the performance models, it is convenient to separate the natural parameters and constants from the discretionary parameters that will be optimized. It is also necessary to specify the available photon flux at the entrance to the wavefront sensor, rather than at the photon detector, because the optical efficiency of wavefront sensors varies greatly (see section 9.4.2) and the decision on which configuration to use is part of the optimization process.

Shack-Hartmann Sensor

Expressions for the measurement errors of the Shack-Hartmann sensor and shearing interferometer were developed in sections 5.3 and 5.4. For a Shack-Hartmann sensor using a quadrant detector, the single-axis mean-square error in radians squared of phase per subaperture, from equation (5.16), is

$$\sigma_{MH}^2 = \left[\frac{\pi^2 K_q}{4(SNR)}\right]^2 \left[1.5^2 + \left(\theta \frac{d}{\lambda_S}\right)^2\right] \qquad d < r_0$$

$$= \left[\frac{\pi^2 K_q}{4(SNR)}\right]^2 \left[\left(1.5 \frac{d}{r_0}\right)^2 + \left(\theta \frac{d}{\lambda_S}\right)^2\right] \qquad d > r_0 \tag{9.43}$$

where

K_q - loss factor due to the gap between quadrant detector elements (= 1.3-1.5)

r_0 = turbulence parameter at sensor wavelength

θ = angular subtense of reference source, rad

At large photon counts, the signal-to-noise ratio SNR is proportional to $\sqrt{n_p}$. The wavefront measurement error increases significantly when an extended reference source is used; for the same photon count, a 10-μrad source has an error 1.6 times larger than that for a point source.

Shearing interferometer

The wavefront measurement error for a lateral shearing interferometer using a near-optimum detection system with sine/cosine weighting, as described in section 5.4, is given by

$$\sigma_{MS}^2 = \frac{2}{\gamma^2 (SNR)^2} \left(\frac{d}{s}\right)^2 \tag{9.44}$$

where

σ_{MS}^2 = single-axis mean-square phase error in radians squared per subaperture

γ - fringe contrast (modulation)

d = subaperture size

s = shear, expressed as units of length in the aperture.

When a four-bin detection system is used with a shearing interferometer having either temporal or spatial modulation, the factor of 2 in the numerator is replaced by $(\pi/2)^2$. The measurement error is inversely proportional to the shear, so that, in general, the largest possible value of shear should be used. However, there are two limitations: (1) when the shear distance becomes comparable to the turbulence coherence length at the sensing wavelength, the correlation between the sheared wavefronts decreases, resulting in a smaller value of γ; and (2) if the shear distance is sufficient to resolve the reference source (that is, if the reference is no longer effectively a point source), then, again, the value of γ decreases. In the second case, γ is equal to the coherence function of the source. It is therefore necessary to find the optimum value of shear that maximizes the shear-gamma product $s\gamma$. For a point source, the shear s_0 should initially be equal to the smaller of either the subaperture size d or the value of r_0. With extended sources, the optimum value of shear is approximated by

$$s_{opt} = \left[\left(\frac{1}{s_0}\right)^2 + \left(\frac{2\theta}{\lambda}\right)^2\right]^{-1/2} \tag{9.45}$$

where θ is the angular subtense of the reference source, assumed to be a uniform disk.

It is then necessary to determine the value of γ. If the wavefront slope were constant over areas equal to the subaperture size plus the shear distance, then the phase difference due to shear would be uniform in each subaperture and the fringe contrast, γ, would be

unity. Variations of wavefront phase over each subaperture of area $A = d^2$ will reduce the value of y. If the wavefront at position x is represented by a Gaussian random variable $\phi(\mathbf{x})$ and the shear is s, then the square of the contrast y may be expressed as

$$\langle \gamma^2 \rangle = \frac{1}{d^4} \int_A \int_{A'} dA\, dA' \exp\left\{ -\tfrac{1}{2} \left\langle [\phi(\mathbf{x}+\mathbf{s}) - \phi(\mathbf{x}) - \phi(\mathbf{x}'+\mathbf{s}) + \phi(\mathbf{x}')]^2 \right\rangle \right\} \tag{9.46}$$

The variance of the wavefront phase distribution is described by a structure function

$$D(\mathbf{s}) = \langle [\phi(\mathbf{x}+\mathbf{s}) - \phi(\mathbf{x})]^2 \tag{9.47}$$

Substituting equation (9.47) into the expansion of equation (9.46) gives

$$\langle \gamma^2 \rangle = \frac{1}{d^4} \int_A \int_{A'} dA\, dA' \exp\left\{ -D(\mathbf{s}) - D(\mathbf{x}-\mathbf{x}') + \tfrac{1}{2}D(\mathbf{x}-\mathbf{x}'+\mathbf{s}) + \tfrac{1}{2}D(\mathbf{x}-\mathbf{x}'-\mathbf{s}) \right\} \tag{9.48}$$

For Kolmogorov turbulence, the structure function is given by

$$D(s) = 6.88\left(\frac{s}{r_0}\right)^{5/3}$$

The value of y^2 has been evaluated numerically by Wallner (private communication), who developed the following empirical expression that fits the data to about 5%. The expression includes the effects of averaging the wavefront over the sensor subapertures (d/r_0 term) and also the decorrelation of the wavefront with shear (d/s term):

$$\gamma^2 = \exp\left[\frac{-2\left(\frac{d}{r_0}\right)^{5/3}}{1.33 + \left(\frac{d}{s}\right)^{1.37}} \right] \tag{9.49}$$

The wavefront measurement error for a lateral shear interferometer is inversely proportional to the shear. Using a point reference source, the shear distance can approach the subaperture size ($d/s \rightarrow 1$), giving optimal performance. With an extended reference source, such as a laser beacon, the shear must be reduced according to equation (5.47), resulting in a larger error.

9.4.2 Comparison of Wavefront Slope Sensors

The relative merits of the shearing interferometer (SI) and Shack-Hartmann (SH) wavefront slope sensors have been hotly debated ever since they were first developed in the early 1970s. In this section, it is shown that the performance of all wavefront sensors is dominated by the efficiency of the photon detection system. Practical considerations, such as using an optical system with high transmission and a detector with high quantum efficiency, are far more important than the theoretical differences between sensors.

The first shearing interferometers used for adaptive optics employed diffraction gratings to generate orthogonally sheared replicas of the input beam. The interference patterns representing x and y wavefront gradients were modulated by rotating the gratings, producing sinusoidal intensity signals of constant frequency, the phase of which is proportional to the wavefront gradient in each subaperture. This arrangement has several desirable features:

1. The wavefront slope in each subaperture is a linear function of the temporal phase of the detected signal, which is easily measured.
2. The shear is proportional to wavelength, making the sensor almost achromatic.
3. The shear is easily changed by operating at a different radius on the grating disk.
4. All subapertures share a common optical system, eliminating the need for precise optical calibration.
5. Temporal modulation allows the phase of the detected signals to be measured accurately, without the need for individual detector calibration, even in the presence of spatial intensity variations over the pupil.

An additional advantage of the SI is that field stops can be added at the focal planes to reduce the field of view to a small area, enabling the sensor to operate with bright backgrounds or with extended sources, such as the Sun.

Variable shear is a useful feature of shearing interferometers that does not show up directly in performance comparisons. With this capability, the sensitivity and dynamic range of an SI can be adjusted rapidly to accommodate large wavefront errors that would normally saturate the wavefront sensor during initial convergence of the adaptive optics system. When the adaptive optics has converged and the residual wavefront error is small, the shear may be increased to improve the measurement precision. By controlling the shear value in real time so that it is inversely proportional to the residual wavefront error, the sensitivity of an SI can be maintained at its optimum value through a wide range of operating conditions, especially in the presence of large changes in the strength of turbulence. In comparison, changing the dynamic range of an SH sensor is much more difficult, requiring additional detector pixels and/or defocusing the spots.

The temporally modulated SI has two significant disadvantages. While giving optimum performance for continuous, point reference sources, it is less efficient with extended sources because of the need to reduce the shear to maintain fringe visibility; this leads inevitably to a loss of precision. It is also incap-

able of handling pulsed reference sources (except at frequencies much higher than the modulation frequency) and so is not suitable for use with most laser beacons.

Later versions of the SI, described in section 5.4, employ spatial modulation and are therefore compatible with pulsed beacons. An SI employing a single two-dimensional transmission grating, from which both *x*- and *y*-wavefront slopes are obtained, looks very much like an SH sensor. The optical grating in the spatially modulated version of the SI performs a similar function to the lenslet array in the SH sensor, but has the advantage that the shear, or sensitivity, is easily varied. As pointed out in section 5.2, both of these sensors use similar optical principles, the differences being mostly in the details of the implemention. Practical considerations often dictate the final choice.

The SH sensors have no moving parts and do not usually employ temporal modulation, so they are suitable for either pulsed or continuous reference sources. Because the aperture is subdivided into a large number of separate imaging channels, SH sensors require calibration of the optics and detectors. There is no practical method of controlling the dynamic range of a SH sensor to accommodate changing operating conditions, as there is with a variable-shear interferometer. In fact, as described in section 5.3.1, the dynamic range of slope measurement in SH sensors is fixed by the number of pixels in each subaperture, which is determined by the detector hardware and is not easily changed. Using larger detector arrays to increase the dynamic range is expensive and requires tradeoffs involving readout speed and detector noise. The dynamic range of shearing interferometers, on the other hand, is easily controlled by varying the shear, without any effect on the size of the detector arrays.

But, the most important factor in determining the performance of wavefront sensors has not yet been discussed: intrinsic differences between the SI and the SH sensors are small compared with the great variations encountered in the photon detectors on which they depend. *The performance of any wavefront sensor is dominated totally by the characteristics of its photon detector, and the associated optical path.*

The most important properties of the detectors used in wavefront sensors for astronomy are (1) the photon detection probability at the sensing wavelength (quantum efficiency); and (2) the noise added in the readout process. Three basic types of detector are used in astronomical adaptive optics:

- Photocathodes followed by electron multiplication, for which quantum efficiency is relatively low (10%) at visible wavelengths, but for which readout noise is negligible;
- Silicon charge-coupled device (CCD) arrays, having high quantum efficiency (80-90%) at visible and near-IR wavelengths, but with significant readout noise (2-10 electrons per pixel) at normal scanning rates;
- Avalanche photodiodes (APDs) operating in the Geiger mode, which have detection efficiencies (quantum efficiency x detection probability) approaching 70% in the visible band, but which are difficult to assemble in arrays because of crosstalk and uniformity problems; cooling reduces dark noise to negligible proportions.

Wavefront sensors with more than about 20 subapertures usually employ scanned arrays, such as CCDs. The available size of such arrays is a fundamental consideration in the system design, because it determines the number of pixels available per subaperture and the pixel readout rate required to obtain a given wavefront sampling frequency, which, in turn, determines the noise added to each pixel in the readout process.

With noise-free detectors, better sensor performance can usually be obtained by using more pixels per subaperture, but when readout or amplifier noise is present (and it is exceedingly difficult to avoid), there is a penalty for using each additional pixel, which may outweigh any advantage gained. It is therefore vitally important to understand how specific detector configurations influence the performance of each type of wavefront sensor. The key parameter in comparing wavefront sensor performance is the signal-to-noise ratio. The number of counted photoelectrons, n_c , is equal to $n_p b \eta p_d$, where n_p is the average number of photons available to the wavefront sensor per subaperture measurement, b is the fraction of light used in making wavefront tilt measurements on each axis, η is the quantum efficiency of the detectors, and p_d is the probability of detection. To standardize the comparison between different wavefront sensor configurations, it is assumed that CCD detectors are used throughout, so that ηp_d is constant. The signal-to-noise ratio for each tilt measurement axis is then

$$SNR = \frac{bn'}{\sqrt{bn' + me^2}} \tag{9.50}$$

where n = available photon count per subaperture $= n_p \eta p_d$. The fraction of light, b, used in each measurement may be due to intensity division using a beam splitter, to spatial division using a fixed grating, or to time division using a chopper. Similarly, the detector pixel readouts, m, in each measurement may occur in either space or time. The effects of detector noise are best conveyed through specific examples. The performance of four types of shearing interferometer will be compared to four types of Shack-Hartmann sensor, using realistic models and taking care to account properly for every photon.

Four detector configurations used in shearing interferometers are shown in figure 9.12. The first two employ temporal modulation. The optical

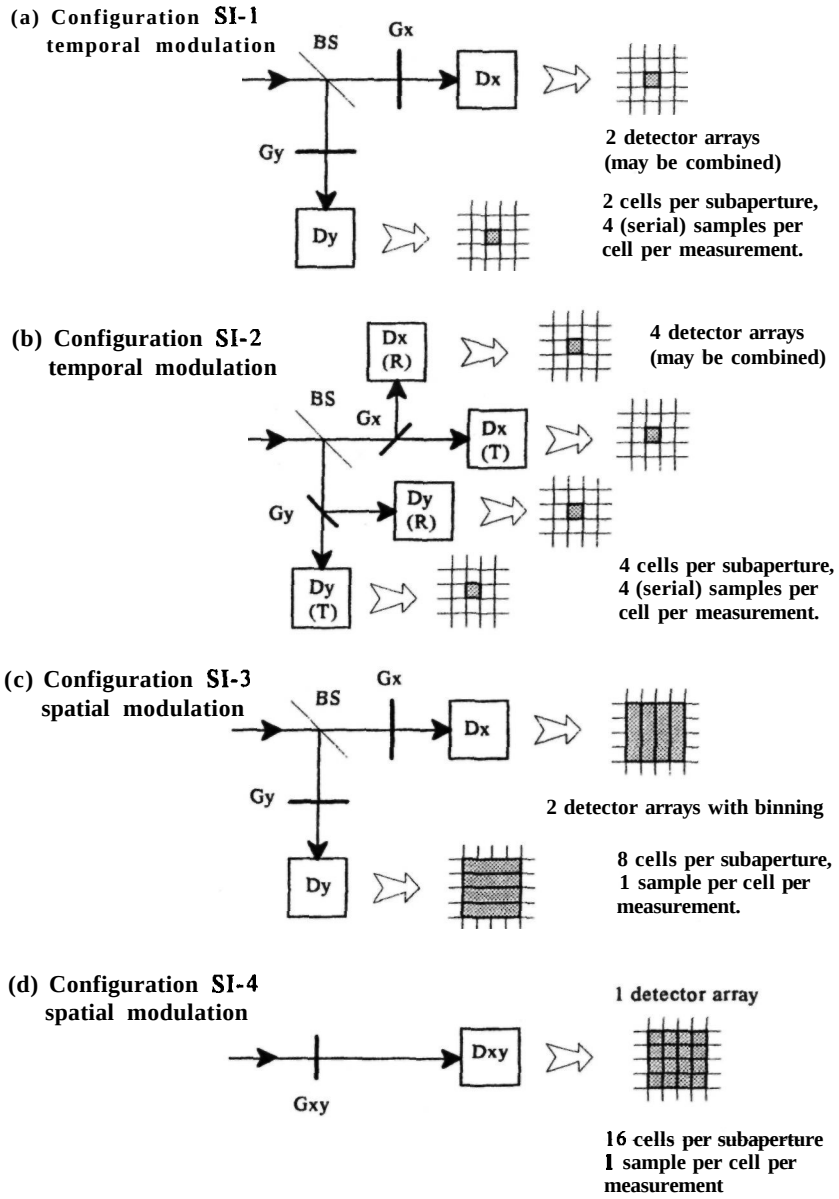


Figure 9.12 Shearing interferometer detector configurations. The shaded areas represent one subaperture of the wavefront sensor.

beam is split into two paths by the beam splitter BS. Rotating gratings Gx and *Gy* in the *x*-slope and *y*-slope paths produce modulated, sheared signals that are detected by arrays *Dx* and *Dy* that contain one detector pixel for each subaperture. In configuration SI-1, only the light transmitted by the gratings is measured. When using a Ronchi grating, the average light reaching each detector is therefore 1/2 x 1/2 = 1/4 of the input intensity. To measure the phase of the temporal modulation produced by grating rotation, at least three, and conventionally four, samples are required per cycle. For the present comparison, it is assumed that phase is computed using a four-bin algorithm (see section 5.4.3), requiring the single detector to be read out four times in sequence for each measurement. For this configuration, therefore, $b = 1/4$ and $m = 4$ in equation (9.50).

The second shearing interferometer configuration, SI-2, is similar, except that the radial gratings have reflecting lines and a second set of detectors is used,

so that none of the incident light is lost. In this case, half of the incident light is used for slope measurement on each axis, so $b = 1/2$, and because there are now two pixels per subaperture, the number of readouts per measurement is $m = 8$.

The third and fourth shearing interferometer configurations depict spatially modulated sensors, in which the gratings are stationary. The configuration shown in SI-3 splits the beam into separate x and y paths, each with a 4 x 4 pixel detector array. Using phase gratings with a transmission of near unity, half the light reaches each detector, so $b \approx 1/2$. The wavefront phase is measured using a four-bin algorithm requiring four pixels per subaperture, so $m = 4$. The pixels in the direction orthogonal to the shear are summed on the CCD chips before readout, in a process known as "binning," to eliminate any additional noise.

In configuration SI-4, a single two-dimensional (crossed) phase grating and a single detector array is used. All of the light reaches the detector array, so $b = 1$. Although each wavefront slope measurement requires only four samples, the full 4×4 array of pixels must be read out each time because the detector is shared between the x and y channels, so it is not possible to perform on-chip binning. The value of m in this case is 16. These signal-to-noise ratio factors are summarized in table 9.2, along with similar data for four types of Shack-Hartmann sensor.

Detector configurations for Shack-Hartmann sensors are shown in figure 9.13. The simplest of these, SH-1, employs a quad cell with 2×2 pixels for each subaperture, giving $b = 1$ and $m = 4$. To improve dynamic range (see section 5.3.1), the configuration SH-2 employs a 4 x 4 array in each subaperture, resulting in $b = 1$ and $m = 16$. As with the corresponding shearing interferometer configuration, on-chip binning is not possible because the detector array is shared between the x and y channels: the one-dimensional summing necessary to compute the x and y slopes must therefore be performed after all pixels are read out. As a consequence of the large value of m, the performance of the 4 x 4 configuration is extremely sensitive to pixel readout noise.

The last two Shack-Hartmann configurations split the optical path into two sections, with separate x and y detector arrays. The reason for using this configuration, as will become apparent, is that better performance can be obtained with noisy detectors, by using on-chip binning. The configuration shown in SH-3 employs two-element arrays (bi-cells) in each x and y subaperture, with on-chip binning, giving $6 = 1/2$ and $m = 2$. To obtain a larger dynamic range, configuration SH-4 employs four-element subaperture arrays, also with on-chip binning, giving $b = 1/2$ and $m = 4$.

Performance factors of these eight different wavefront sensor configurations, computed using equations (9.43) through (9.50), are compared in figures 9.14 through 9.16. The shear used in the shearing interferometers has been adjusted to the optimum value. In each case, the wavefront measurement error due to photon and detector noise is expressed as a Strehl ratio using the relation $5 = \exp(-\sigma^2)$

The conditions are:

Wavelength	$\lambda = 0.55\,\mu$m
Turbulence parameter	$r_0 = 0.15$ m
Subaperture size	$d = 0.2$ m
Detector readout noise	$e = 2$ electrons per pixel
Reference source size	$6 = 0$–$20\,\mu$rad
Available photon count	$N' = 10$-500 counts per subaperture

In figure 9.14, the Strehl ratio for each type of sensor is shown as a function of available photon count per subaperture, N', for two reference source sizes, 0 and $10\,\mu$rad. In figure 9.15, Strehl ratios are plotted against the reference source angular diameter, 6, for two values of available photon count, 100 and

Table 9.2 Wavefront Sensor Detector Configurations

Sensor Type	Intensity Ratio b to x or y Detector	Subaperture Array Size (pixels)	Readouts per Measurement	Detector Noise Factor, m
Shearing interferometers				
SI-1 temporal, 2 arrays	0.25	1	4	4
SI-2 temporal, 4 arrays	0.5	2	4	8
SI-3 spatial, 2 arrays	0.5	1 x 4	1	4
SI-4 spatial, 1 array	1.0	4 x 4	1	16
Shack-Hartmann sensors				
SH-1 one array, 2×2	1.0	2 x 2	1	4
SH-2 one array, 4×4	1.0	4 x 4	1	16
SH-3 two arrays, 1 x 2	0.5	1 x 2	1	2
SH-4 two arrays, 1 x 4	0.5	1 x 4	1	4

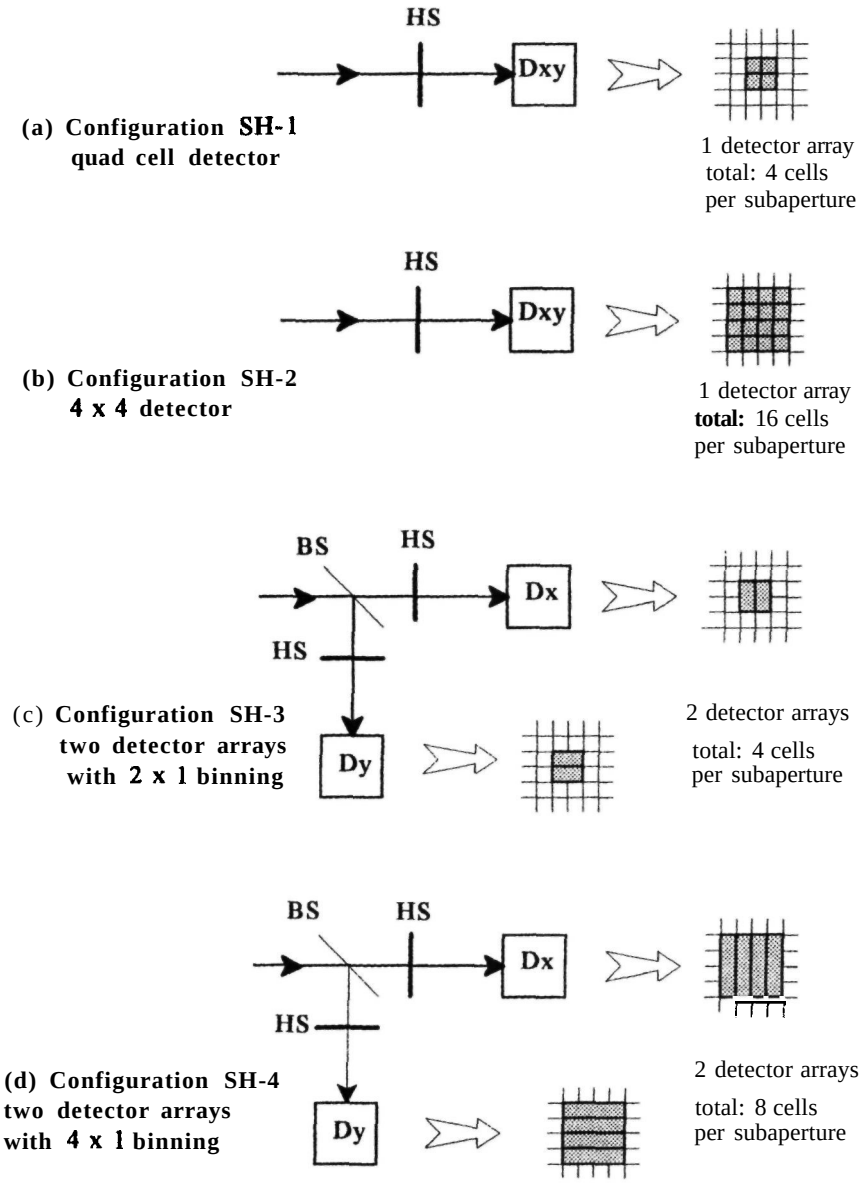
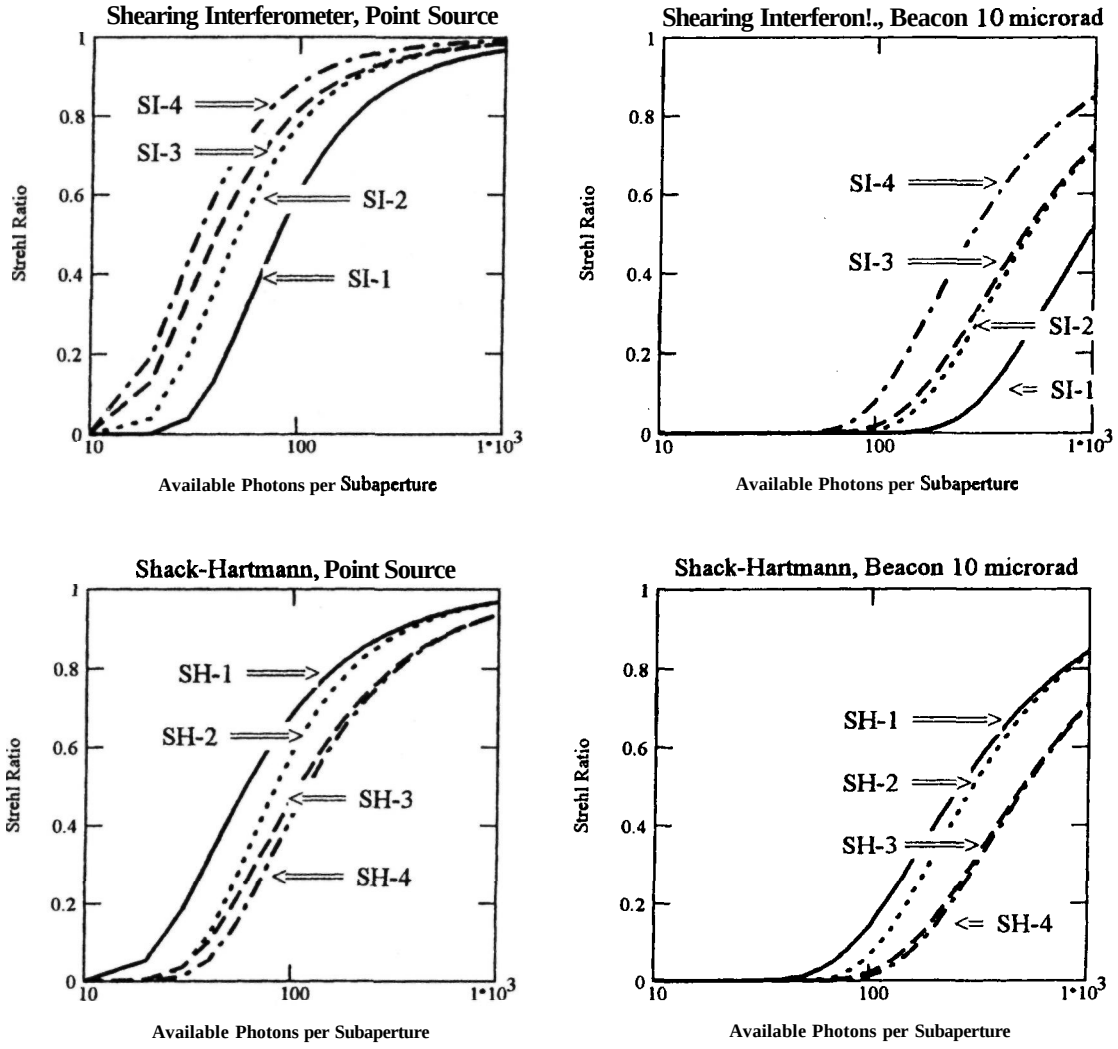


Figure 9.13 Shack-Hartmann sensor detector configurations. The shaded areas represent one subaperture of the wavefront sensor. HS, Hartmann screen.

500. These charts show that the performance differences due to detector configuration within each type of sensor are far greater than any intrinsic difference between the shearing interferometer and the Shack-Hartmann sensors themselves.

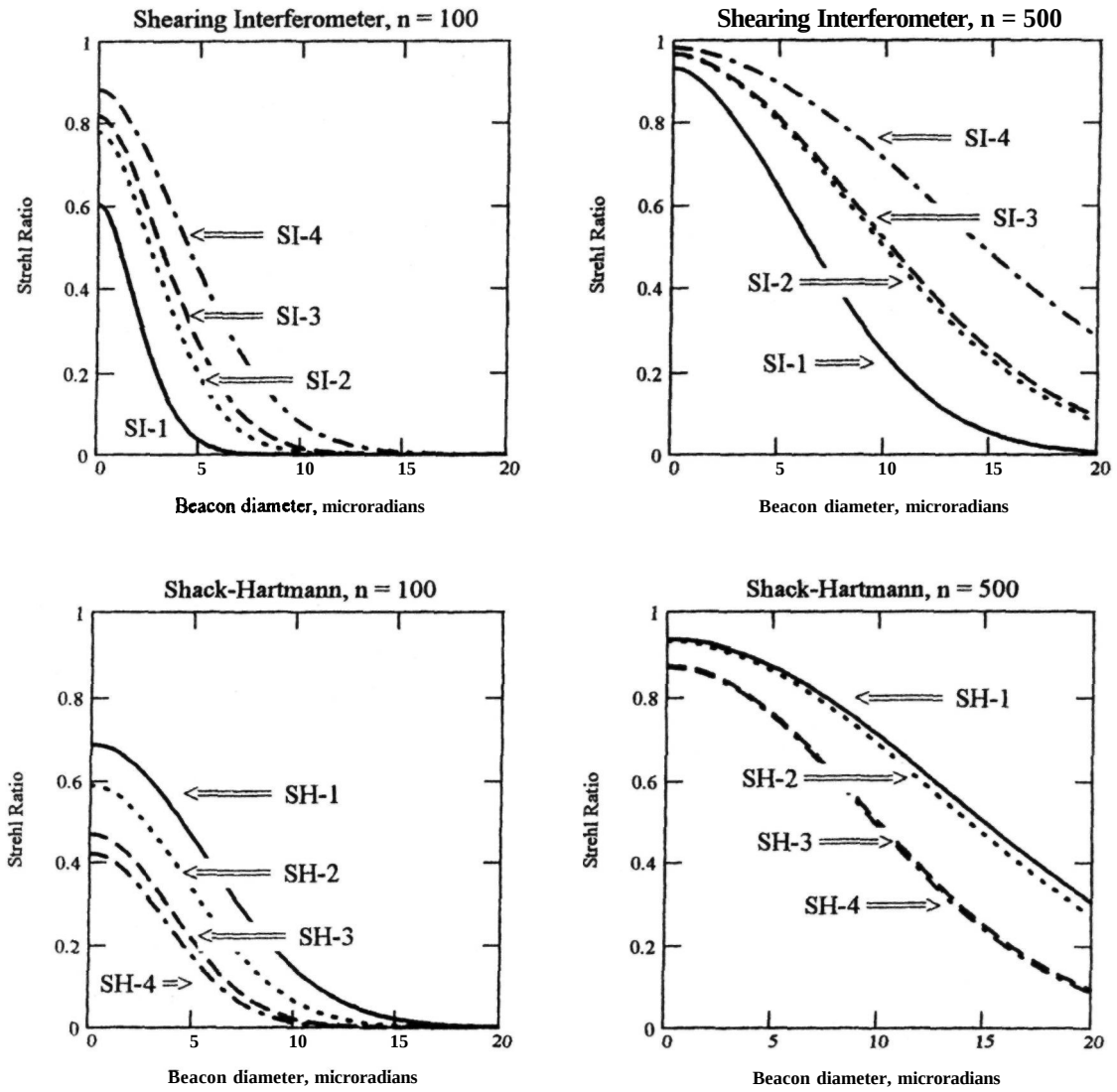
The best-performing shearing interferometer for the specified detector readout noise of two electrons per pixel is SI-4, a spatially modulated configuration using a single detector array. The best-performing Shack-Hartmann sensor for the same conditions is the simplest configuration, SH-1, which uses a single array of quad cells. This sensor has a linear tilt-measurement range of less than ± 1 wave per subaperture. A larger dynamic range is obtained with SH-2, which employs 4 x 4 pixels per subaperture. Shack-Hartmann configurations SH-3 and SH-4 are useful when the detector and background noise is high, over 20 electrons rms per pixel; this can occur when high readout rates are employed.

Figure 9.16 shows a direct comparison of the two "best" sensors of each type when the available photon count and reference size are varied. There is



Conditions: sensor wavelength λ = 0.55 μm, turbulence parameter r_0 = 0.15 m subaperture size d = 0.2 m, detector read noise = 2 e per pixel.

Figure 9.14 Wavefront sensor performance comparison 1: photon count. Measurement error (Strehl ratio) as a function of the available photon count per subaperture at the entrance of the wavefront sensor, for reference source diameters of 0 and 10 μrad. These charts take into account the specific optical and detector configurations used in each type of sensor.



Conditions: sensor wavelength λ = 0.55 μm, turbulence parameter r_0 = 0.15 m subaperture size d = 0.2 m, detector read noise = 2 e per pixel.

Figure 9.15 Wavefront sensor performance comparison 2: beacon size. Effect of beacon angular size on measurement error (Strehl ratio), for available photon counts of 100 and 500 per subaperture per measurement.

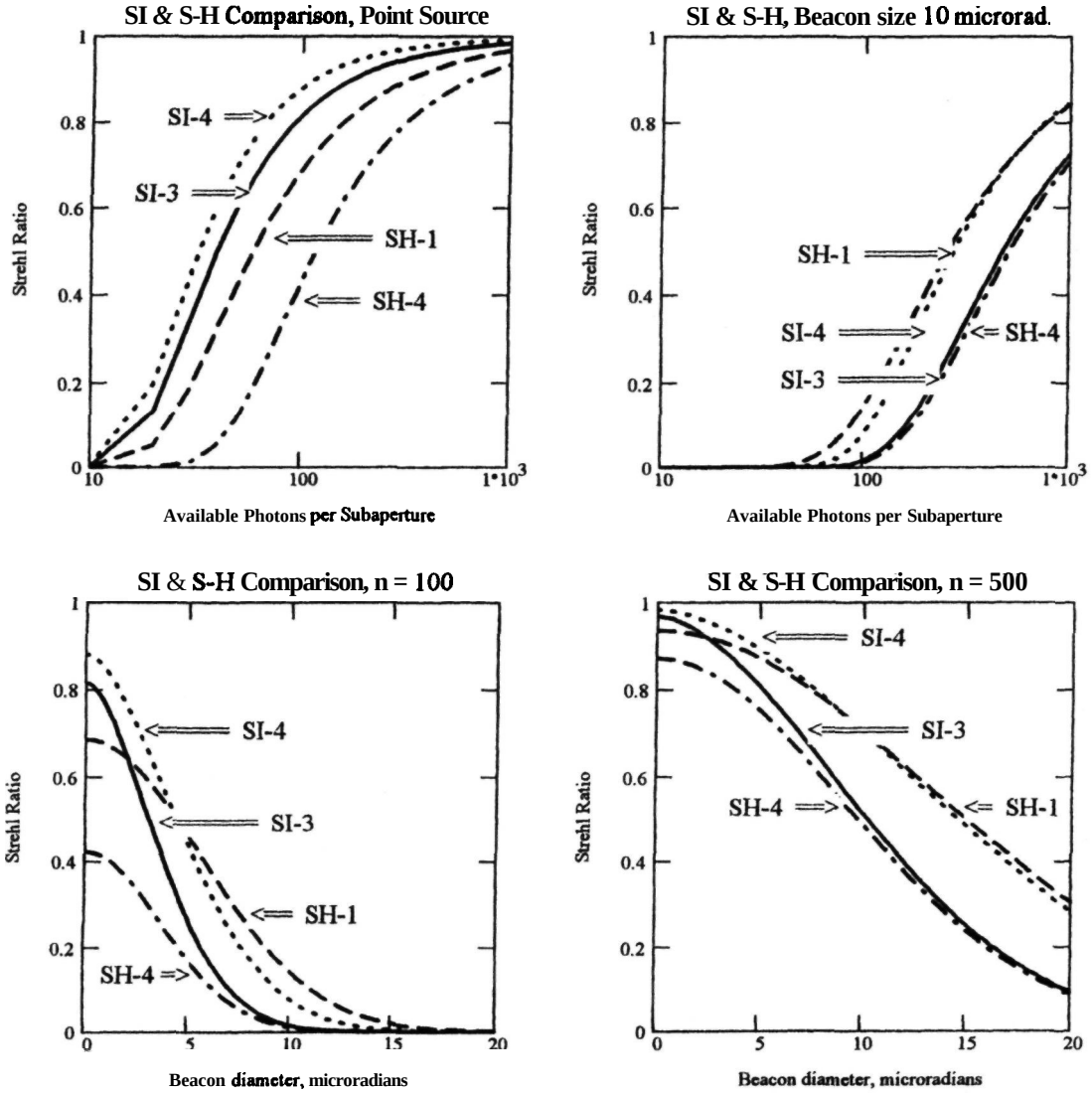


Figure 9.16 Wavefront sensor performance comparison 3: SI and SH. Direct comparison of shearing interferometer and Shack-Hartmann sensors under identical operating conditions.

little difference at large photon counts. For photon-starved conditions ($N' < 100$), the shearing interferometer gives slightly better performance with unresolved reference sources, but when the source size exceeds about 10 μrad, the Shack-Hartmann has a smaller random measurement error. The performance of the SI is degraded with large reference sources because the shear must be reduced to maintain contrast. Optimizing the shear-gamma product, as outlined earlier, is not sufficient to compensate this loss of performance. Adaptive optics systems employing laser beacons must be capable of operating with reference source diameters in the range of 5-10 μrad. For these relatively small sources, either type of wavefront sensor is suitable.

The importance of CCD read noise on the performance of Shack-Hartmann sensors is brought into sharp focus in figure 9.17, which shows the Strehl ratio of three SH configurations as a function of available photon count per subaperture, with detector read noise levels of 0, 5, and 10 electrons rms per pixel. The difference in performance, especially at photon counts below 500, is startling. These curves show, beyond any doubt, that detector efficiency is the most important factor in wavefront sensor performance, far outweighing any differences between the types of sensor.

9.4.3 Temporal Errors

Temporal errors are caused by the inability of adaptive optics systems to respond instantly to changes in the wavefront. Any time delay between the measurement and correction of a wavefront disturbance results in a temporal error. The amount of error depends on the relation between the dynamics of the turbulence and the response time of the adaptive optics system. The contributing factors include transportation of turbulent layers by winds (the wind vectors usually vary with height), together with the motion or evolution of the turbulence itself. In adaptive compensation systems, the main time delay is usually the integration time of the servo control loop, during which the wavefront sensor collects photons from the reference source. To this must be added pure time delays, such as those due to reading out a CCD detector and processing the wavefront data.

In principle, the effects of time delays can be compensated by predicting what the wavefront structure will be at the time it is compensated. The correlation of atmospherically distorted wavefronts decays with time, the time constant typically being between 1 and 10 ms, depending on the turbulence strength, wind speed, and wavelength. The simplest method of prediction is simply to use the current value of the wavefront, with a decay factor, as the best estimate of the next sample. This is the method generally employed in adaptive optics systems, using a (temporal) integrator or low-pass filter. It implies a wavefront sampling rate considerably higher than the integration time, in order to ensure the stability of the feedback loop.

Power spectra for the wavefront phase and overall tilt errors due to atmospheric turbulence have been developed in section 9.2.5. To reduce these wavefront disturbances to a negligible level, the temporal response of the compensation system must be comparable to the turbulence spectrum. The frequency response of a closed-loop adaptive optics system is limited primarily by the sampling rate of the wavefront sensor, which is typically between 1 and 10 kHz, being usually determined by the frame rate of the photon detector. To ensure the stability of the feedback loop, the closed-loop bandwidth cannot exceed about 1/6 of the sampling rate and is more often about 1/10. The servo bandwidth determines the effective integration time during which photons are collected from the reference source. The optimum integration time (minimizing the total error) is found by trading off the measurement error, due to the collection of a finite number of photons, against the temporal error, due to a finite bandwidth. Most astronomical adaptive optics systems are photon starved, so their optimal temporal bandwidth is usually lower than that of the wavefront disturbances, resulting in significant errors. In addition, there are always fixed delays in the feedback loop caused by data transfer and processing times. In order to optimize the integration time, it is necessary to determine the temporal errors caused by (1) finite correction bandwidth and (2) pure time delays.

Errors Due to Correction Bandwidth

Greenwood [1977] solved the first problem by finding the residual wavefront error in terms of the ratio of the effective bandwidth of the turbulence to that of the correction servo. If the servo response is defined by the complex function $G(f)$, then the power rejected by the filter may be expressed as the response $|1 - G(f)|^2$For an input with power spectrum $F_\phi(f)$, the rejected or uncorrected power is

$$\sigma_R^2 = \int_0^\infty |1 - G(f)|^2 F_\phi(f) df \qquad (9.51)$$

For a servo with a single time constant $t = RC$, where R is the resistance in ohms and C is the capacitance in farads, the response function is

$$G(f) = \frac{1}{1 + i\left(\frac{f}{f_c}\right)} \qquad (9.52)$$

where $f_C = 1/2\pi RC$.

Using the asymptotic turbulence power spectrum, equation (9.18), for $F_\phi(f)$,and performing the integration of equation (9.51), the wavefront temporal error may be expressed in the form

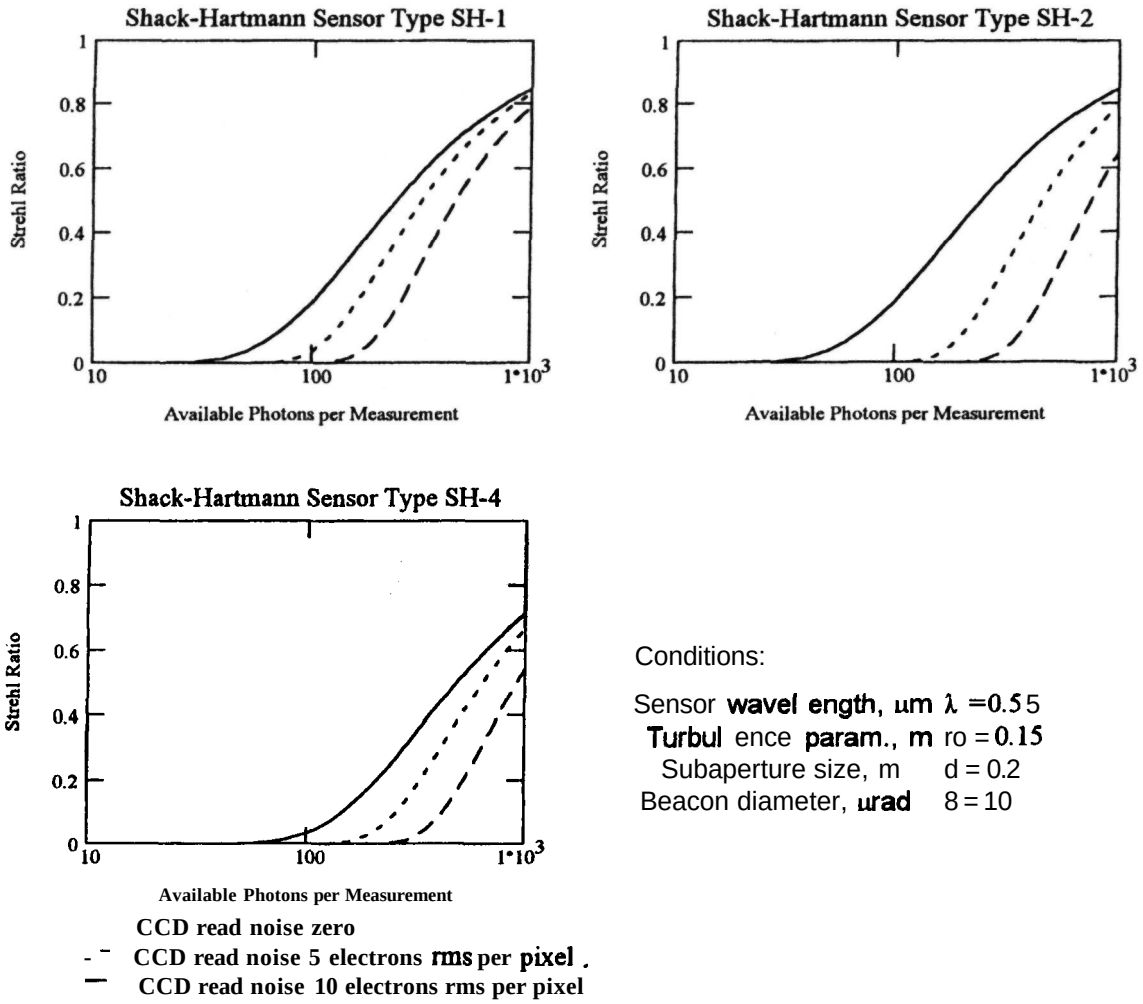


Figure 9.17 Effect of detector read noise on the performance of three configurations of the Shack-Hartmann wavefront sensor, as a function of the available photon count per subaperture at the input of the sensor. The performance is based on the wavefront slope measurement error per axis, converted to Strehl ratio. Configuration SH-1 employs 2×2 detector arrays for each subaperture, SH-2 employs 4×4 detector arrays, and SH-4 employs two 1 x 4 arrys with binning, as shown in figure 9.13. With no photon noise, SH-1 and SH-2 have similar performance, but the additional pixels in SH-2 degrade its performance when read noise is present. The SH-4 has worse performance, in all cases, because the available photon flux is shared between two detector arrays.

$$\sigma_{TR}^2 = \kappa\left(\frac{f_G}{f_S}\right)^{5/3} \tag{9.53}$$

where f_S is the servo bandwidth of the system. In the case of a simple RC network with the bandwidth defined at the half-power point, the value of κ is 1. At the other extreme, a servo bandwidth having an infinitely sharp cutoff at f_S produces a κ-value of 0.191.

The parameter f_G is a characteristic frequency of the atmospheric turbulence, known as the Greenwood frequency, and is given by

$$f_G = \left[0.102\,k^2 \sec(\zeta) \int_0^\infty C_N^2(h) v^{5/3}(h)\, dh\right]^{3/5} \tag{9.54}$$

The value of f_G is determined from the turbulence and wind profiles of the atmosphere, enabling the error induced by any servo bandwidth to be calculated. In the special case of a single turbulent layer with wind velocity v, the Greenwood frequency may be expressed as

$$f_G = 0.427\frac{v}{r_0} \tag{9.55}$$

Several simplifying assumptions have been made in deriving these equations. The temporal frequency spectrum is assumed to have a constant —8/3 power law, ignoring the low-frequency breaks due to the removal of average phase and tilt. This is justified on the basis that these breaks occur at frequencies much lower than those used for controlling the subapertures. The expression for temporal wavefront error using the Greenwood frequency [equation (9.53)] does not account for time delays in the feedback loop, nor for the averaging effects of finite subaperture size. In other words, it assumes that the wavefront is measured at a single instant of time, in subapertures comparable in size to r_0. The wavefront error caused by time delays can be estimated separately and added to the bandwidth error, as discussed later. The effect of using subapertures significantly larger than r_0 is to reduce the temporal component of the error and to increase the fitting error.

Equation (9.53) therefore tends to overestimate the temporal error when $d/r_0 > 1$.

Errors Due to Pure Time Delay

Pure time delays produce different effects from those of an integrator or low-pass filter; the phase shift is linear with frequency and the spectral content is unchanged by the delay. Such propagation delays occur when data are temporarily stored during conversion from one format to another (analog to digital or parallel to serial), or during processing in a serial computer. In an adaptive optics system, a fixed time delay in the correction signal produces a phase shift that increases with frequency, progressively degrading the compensation of high-frequency wavefront errors. Errors due to time delays in the feedback loop must be added to those produced by the bandwidth limitation previously discussed.

Fried [1990] and Karr [1991] independently analyzed the effect of pure time delays in an adaptive optics feedback loop. They considered the case of a turbulence-degraded wavefront with Kolmogorov statistics, measured at some instant of time and corrected after a delay of τ_S, the correction being held until the next update. They showed that the error can be expressed as

$$\sigma_{TD}^2 = \left(\frac{\tau_S}{\tau_0}\right)^{5/3} \tag{9.56}$$

where τ_0 is a time constant related to the Greenwood frequency by $\tau_0 = 0.134/f_G$. The wavefront error due to a time delay τ_S is then

$$\sigma_{TD}^2 = 28.4(\tau_S f_G)^{5/3} \tag{9.57}$$

Welsh [1991] and Harrington and Welsh [1994] have developed a frequency-domain analysis that predicts the performance of an adaptive optical system that has an arbitrary temporal response. This analysis takes into account the removal of aperture-averaged phase and tilt, as well as the spatial and temporal averaging due to subaperture size and the finite integration time of the wavefront sensor. They use a von Karman spectrum for the index of refraction fluctuations, rather than the Kolmogorov spectrum used by Greenwood and Fried. The main difference between these spectra occurs at low frequencies, where the von Karman spectrum flattens out, generally reducing the temporal errors in comparison with the Kolmogorov spectrum, which continues to rise as the frequency is reduced. Use of the von Karman spectrum in the model reduces the predicted errors for delay times normally encountered in adaptive optics systems by about 30%. Harrington and Welsh found that accounting for overall piston and tilt removal has no significant effect at delay times of $\tau_S < 0.1D/v$, which is the case for most adaptive optics systems.

As the time delay between measurement and correction increases, the measured wavefront data becomes less descriptive of the actual wavefront, especially at high frequencies. Under these conditions, the best correction is therefore obtained by discarding the high frequencies. This is exactly the effect of finite integration time in the wavefront sensor, which acts as a low-pass temporal filter, removing some of the high-frequency power in the turbulence spectrum.

Time Lines for Adaptive Optics

The time lines of adaptive optics compensation are determined largely by the system architecture, and, in particular, by whether serial or parallel data processing is employed. The first large-aperture adaptive optics system (the CIS) employed parallel analog processing to achieve the required bandwidth. Current systems employ digital processing exclusively, using a combination of serial and parallel operations to minimize the processing time. The longest time delays usually occur during photon detection and wavefront reconstruction.

Time lines are shown in figures 9.18-9.20 for three types of adaptive optics system. Figure 9.18 shows the time line for a wavefront sensor using an array of photon-counting detectors, the outputs of which are read out in parallel. A sensor of this type using photomultipliers was employed in Itek's CIS [Hardy 1993]. Avalanche photodiodes are now preferred for photon counting in wavefront sensors because of their higher quantum efficiency and similar freedom from noise [Graves et al. 1994, Roddier et at. 1994]. Photon-counting detectors such as APDs have the advantage that that there is no readout noise penalty, as there is with CCDs, so the counts may be accumulated over periods of any length. This gives photon-counting detectors great flexibility, as they are suitable for continuous reference sources (which necessitate high readout rate) and also for pulsed sources, such as laser beacons, in which the readout is gated and synchronized to the pulse rate.

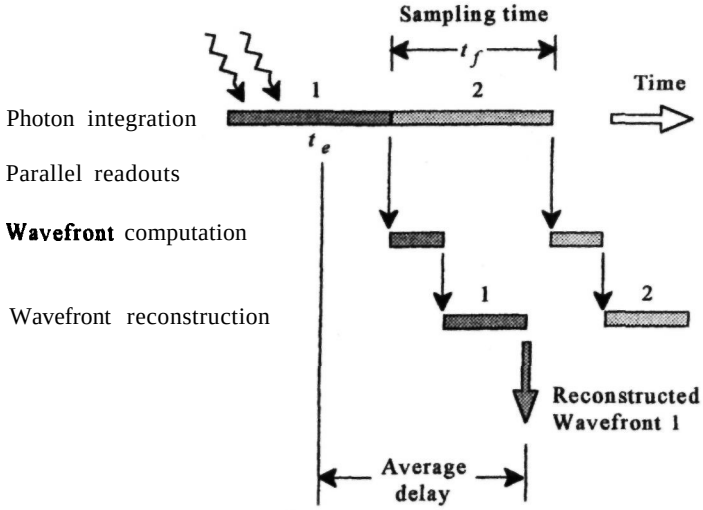


Figure 9.18 Time line for wavefront sensing with a natural star, using a parallel detector, such as an APD.

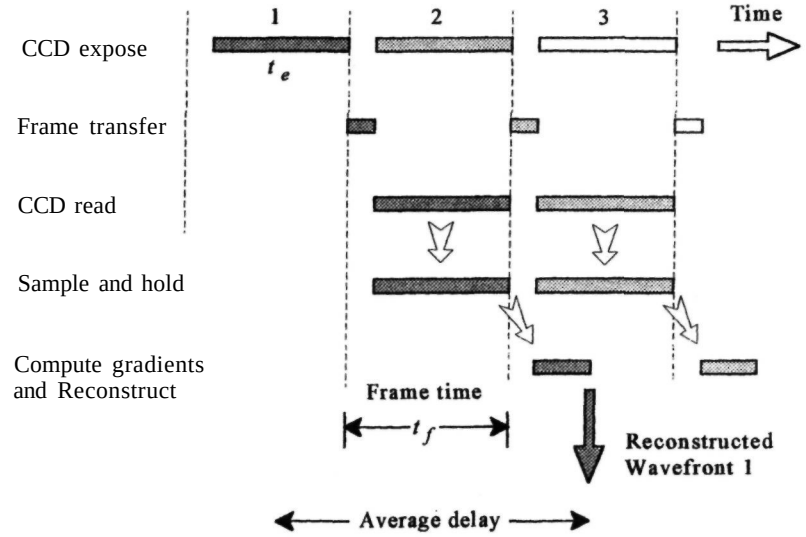


Figure 9.19 Time line for wavefront sensing with a natural star, using a frame-transfer CCD. In this example, the time delay between the average exposure and the availability of the reconstructed wavefront is about 2 frame periods.

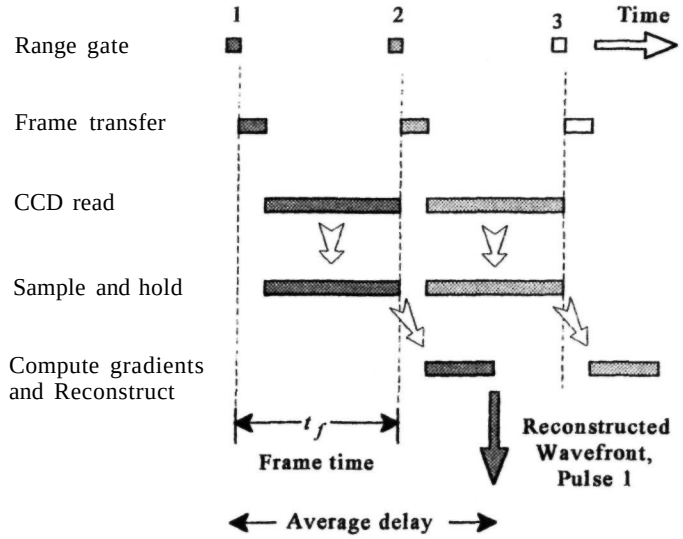


Figure 9.20 Time line for wavefront sensing with a pulsed laser beacon, using a frame-transfer CCD, which is read out for each pulse. The delay between the arrival of each pulse and the availability of the reconstructed wavefront is about 2 pulse intervals. The maximum pulse rate in this case is equal to the frame rate, f_t.

The wavefront slope computation and wavefront reconstruction are usually implemented with digital processors in which the propagation delays are in the microsecond range. With parallel detectors and data-processing architecture, the delay time between the sensing of a wavefront error and the application of the correction signal to the servo system driving the wavefront corrector can be as small as 0.1 ms, enabling a very fast correction loop (and small time delay errors). For astronomical adaptive optics, the integration time of the control system is usually in the range of about 1–10 ms, to optimize the performance with available reference sources.

Figure 9.19 illustrates the timing relationships in an adaptive optics system employing a frame-transfer CCD detector and a digital processor, when a continuous reference source is used. The key parameter in this type of system is the frame time, t_f, which determines the measurement rate of the wavefront sensor. The frame time is basically the number of pixels per port divided by the pixel readout rate per port. High-frame-rate CCD detectors have been developed for wavefront sensing using multiple output ports to minimize the frame time. In any case, the time required to read out the CCD, typically 0.5-1 ms, is the largest part of the frame time. The CCD exposure time t_e and the time required to transfer the charge packets from the exposure area to the storage area for readout must both fit within the basic frame time. The transfer time is usually on the order of 10 μs and can often be neglected.

Calibration of the detector outputs and the computation of wavefront slopes are implemented with a pipeline processor that has a small propagation delay. Wavefront reconstruction is usually performed using a parallel processor. The delay between the average exposure and the availability of the reconstructed wavefront is typically about 2 frame periods. The average time delay between sensing a wavefront error with a CCD detector and applying the correction to the servo system is therefore in the region of 1–2 ms, depending on the frame rate.

It should be recalled that when a CCD detector is used with a natural (continuous) reference source, the detector collects photons continuously (except for the small frame transfer time), independently of its readout rate. The integration time relevant to the photon count and signal-to-noise ratio is determined by the bandwidth of the filter in the control loop, as described in section 9.4.2, and not by the CCD frame rate. The frame rate (or its reciprocal, the frame time) is one component of the total time delay.

The time line for the third type of system, shown in figure 9.20, is similar to the previous one, modified to show operation with pulsed reference sources. The pulse rates used to generate laser beacons in adaptive optics vary over a wide range, from a few hertz with dye lasers up to more than 30 kHz with copper vapor lasers. To achieve continuous closed-loop compensation a pulse rate of 1 kHz or more is required. Another requirement is the capability to separate successive pulses by range gating, especially for sodium beacons, where the return must be separated from Rayleigh scatter. To achieve this requirement, lower pulse rates (less than a few kilohertz) are preferred. A good compromise for these conflicting requirements is a pulse rate of about 1 kHz, especially as this happens to be an achievable frame rate in CCD sensors. Higher pulse rates can be handled in the same way as continuous reference sources, as far as the detector is concerned.

In figure 9.20, the pulse rate is equal to the CCD frame rate, with the range gate occurring just before the frame transfer, although, in fact, it can occur anywhere in the frame time. The detector frame rate is assumed to satisfy the requirement for continuous closed-loop operation. A higher pulse rate can be accommodated by integrating more pulses on the CCD, maintaining the same frame rate. The accumulated pulses are read out once per frame. From this point on, operation is similar to that shown in the previous figure. The average time delay in the wavefront sensor using pulsed operation is seen to be very similar to that using continuous operation, nominally around 2 frame times of the CCD camera, amounting to 1 or 2 ms. Again, the photon integration time is determined by the characteristics of the low-pass filter in the control loop. The number of photoelectrons collected per measurement is the average number of events per pulse multiplied by the number of pulses per integration time.

9.4.4 Wavefront Reconstruction Errors

Error Propagation

Wavefront sensors making zonal measurements require a wavefront reconstruction process to convert the individual measurements into a spatially continuous representation of the wavefront. The error propagation of this process is defined as the ratio of the mean-square error of the reconstructed array to that of the mean-square error of the measured slope vectors. When wavefront reconstruction is modeled simply as determining the values of a two-dimensional array of points connected by linear slope vectors, then the value of the error propagator is found to depend on the exact geometrical relation between the slope vectors and the points. As discussed in section 8.2, a reconstructor using displaced x and y slope vectors (network 1) appears to have significantly smaller error propagation than one using superimposed vectors (network 2). This difference was resolved in section 8.6, where it was shown that when the reconstruction process is correctly modeled, by estimating the slopes and wavefront over the full area of the aperture, the dependence on slope vector geometry vanishes. In other words, the differences in error propagation are in the simplified reconstruction

models, rather than in the systems they are supposed to represent.

The error propagator for the reconstruction of wavefront slope measurements has a value on the order of unity and increases as the logarithm of the number of subapertures in the array. As shown in section 8.2.4, it can be expressed in the form

$$E \equiv \frac{\sigma_w^2}{\sigma_p^2} \approx \kappa_1 + \kappa_2 \ln(N^2) \tag{9.58}$$

where

σ_w = mean-square error of reconstructed wavefront
ffp = mean-square error of measured subaperture tilts in one dimension
N^2 = number of measurement subapertures
κ_1, κ_2 = constants

In the case of curvature sensors, the error propagator is proportional to the square of the number of elements, producing significant errors in large arrays.

Checkerboard Error

A different type of error occurs in wavefront sensors in which *x*- and *y*-axis slope measurements are made in congruent subapertures (network 2). Shack-Hartmann sensors usually (but not necessarily) employ this geometry. In the reconstruction process, this scheme results in two separate networks of interconnected points, which may have different piston components [Herrmann 1980]. The problem is that turbulence-degraded wavefronts really do contain a checkerboard component, which, in this case, is not sensed and therefore cannot be extracted from the slope data. In principle, the checkerboard component could be determined from a knowledge of the wavefront statistics, as suggested by Wallner [1983].

Differences in the relative phase of the two networks show up as a "waffle" pattern in the reconstructed wavefront [Wild, Kibblewhite. Fang Shi, et al. 1994; Wild, Kibblewhite, and Scor 1994]. The problem may be avoided completely by making the *x* and *y* slope measurements in displaced subapertures (network 1), resulting in a single network of connected points. If this is not possible, the best approach is simply to set the average values of the two networks to zero. The error introduced by doing this is usually negligible. Wallner [1985] has shown that for wavefronts with a Kolmogorov spatial spectrum, the mean-square error due to the checkerboard component has the same $(d/r_0)^{5/3}$ dependence as the fitting error, but with a much smaller coefficient. If the average wavefront values of the two networks are set to zero, the resulting mean-square error is given by

$$\sigma_C^2 = a_C \left(\frac{d}{r_0}\right)^{5/3} \tag{9.59}$$

The coefficient a_C depends on the overall size of the reconstruction array. For a 3 x 3 array, it has a value of 0.3 rad^2. For arrays larger than 4×4, when edge effects are less important, a_C varies inversely with the total number of subapertures and is closely approximated by

$$a_C = 1.6(n-1)^{-2} \tag{9.60}$$

where n is the number of reconstructed points across the array. In the case of a 10 x 10 array, the value of a_C is 0.020 rad^2. In comparison, the wavefront fitting error coefficient for most deformable mirrors has a value of about a_F = 0.3 rad^2. Therefore, with normal turbulence, the checkerboard problem can be solved, for all practical purposes, by setting the piston values of the two networks to zero.

9.4.5 Wavefront Fitting Error

The wavefront fitting error has a special role in adaptive optics, as it sets a fundamental limit to system performance. No matter how bright the reference source, how efficient the wavefront sensor, or how small the time delays, the image quality is determined ultimately by the ability of the correction device to flatten the wavefront. The fitting error is inversely proportional to the number of degrees of freedom in the wavefront compensator; in most cases, the number of correction zones or modes is built into the hardware design and is not easily changed. The complexity of an adaptive optics system is roughly proportional to the number of actuators or degrees of freedom, given by $N_A = (D/d)$. As a consequence, the choice of N_A, or of subaperture size d, is one of the most critical decisions to be made in the design of an adaptive optics system, a choice that governs both its performance and its total cost.

Fitting errors for several types of zonal wavefront correctors were derived by Hudgin [1977a]. For correction of Kolmogorov turbulence, the general form of the mean-square fitting error is

$$\sigma_F^2 = a_F \left(\frac{d}{r_0}\right)^{5/3} \tag{9.61}$$

where

d = subaperture size
r_0 = turbulence coherence length
a_F = fitting error coefficient, dependent on the influence function of the corrector

It is noted that r_0 increases as the 6/5 power of wavelength, so that the fitting error for a fixed subaperture size decreases at longer wavelengths.

Values of a_F for various types of wavefront corrector are shown in table 9.3. The piston-only function is the simplest type of correction and, not surprisingly, has the largest value of a_F, showing that it is very inefficient on a per-segment basis. With correction devices such as liquid crystals, this

Table 9.3 Wavefront Fitting Error Coefficients

Subaperture Influence Function	Actuators per Subaperture	Constant a_F (rad^2)
Piston only	1	1.26
Piston + tilt	3	**0.14**
Continuous-plate (super-Gaussian)	1	0.28-0.34
Pyramid	1	0.28
Gaussian	1	0.24

inefficiency is compensated by the fact that it is economically feasible to use a large number of segments. To equalize the fitting errors for different types of actuator, equation (9.61) shows that the number of zones required must be in the ratio $(d_2/d_1)^2 = (a_{F1}/a_{F2})^{6/5}$. To obtain the same error as that of a device with a Gaussian influence function, a piston-only corrector therefore requires $(1.26/0.24)^{6/5}$ or 7.3 times the number of zones.

Segmented mirrors usually employ piston + tilt functions, having 3 degrees of freedom (three actuators) per segment. The fitting error coefficient, based on segment size, is much smaller than that of a continuous-plate mirror. However, it turns out that a piston + tilt segmented mirror requires a larger total number of actuators to obtain the same residual wavefront error. The ratio of the number of subapertures required is $(a_{F1}/a_{F2})^{6/5} = (0.14/0.28)^{6/5} = 0.435$. Multiplying this by three actuators per subaperture, a factor of 1.3 is obtained. A segmented mirror therefore requires 30% more actuators than a continuous-plate mirror to obtain the same fitting error.

The influence functions of continuous-plate mirrors can be controlled by changing the relative stiffness of the faceplate and the actuators, as explained in section 6.4. When the actuator stiffness is higher than that of the faceplate, the influence function tends to be "super-Gaussian" (that is, having an exponent greater than 2), leading to a compact shape in which the response is near zero at the adjacent actuator. This shape minimizes the cross-coupling between adjacent zones, simplifying the control algorithm. However, as can be seen from table 9.3, the fitting error coefficient tends to be greater than that for a strictly Gaussian function.

When the actuator stiffness is less than that of the faceplate, the influence function gets broader and assumes a Gaussian (or even a sub-Gaussian) shape, having a somewhat lower fitting error coefficient. While this can lead to a more efficient deformable mirror in terms of correcting wavefront error, it requires a full actuator control matrix in which all overlapping influence functions are taken into account.

Fitting errors have been evaluated earlier in terms of the mean-square wavefront error. The effect of small errors (less than A/10) is to reduce the normalized peak intensity (Strehl ratio) of the image. The effects of larger wavefront errors, such as would be encountered in an adaptive optics system using partial compensation, were investigated in section 4.4. The parameter varied was the ratio of the subaperture size, d, to the turbulence parameter r_0. To summarize these results for images with partial zonal compensation, it was shown that:

- The image structure and Strehl ratio are determined entirely by the d/r_0 ratio.
- The radius of the image core stays close to its diffraction-limited size, even when the Strehl ratio is reduced to 0.1, corresponding to $d/r_0 = 3.4$. At this level of error, there is still a large ratio between the peak of the image core and the surrounding halo.
- When d/r_0 reaches 3.8, the peak contrast drops rapidly, and for larger values the core is buried in the halo, the diameter of which is determined by λ/r_0.

9.4.6 Tracking Errors

When laser beacons are used as wavefront reference sources, a fixed guide star is necessary to stabilize the angular position of the image. The stabilization is imperfect because of random fluctuations due to the angular offset between the science object and the fixed guide star. The tilt component of angular anisoplanatism has been analyzed in section 7.4, in which it was shown that the position error from this source varies monotonically with isoplanatic angle θ and is independent of wavelength, although it depends on the atmospheric turbulence structure and telescope diameter in a complicated way. Image stabilization is necessary for all astronomical observations, except for those requiring very short exposures of less than about 1/100 s. The need for fixed guide stars is the main factor determining the sky coverage of adaptive optics systems.

The basic expressions for anisoplanatic tilt error developed by Sasiela [1988] using the transverse filtering approach are given in equations (7.53) and (7.54). These equations involve a series of terms in ascending powers of θ and D. It is shown in section 7.4 that at least five terms are needed to obtain reliable predic-

tion of tilt errors over the range of parameters used in astronomical adaptive optics, especially for observations at IR wavelengths.

Although the anisoplanatic tilt error produced by a given atmospheric turbulence profile is independent of wavelength, its effect on the observed image must be assessed in relation to the Airy disk, the radius of which is directly proportional to wavelength. It was shown in section 4.5.3 (equation 4.66) that there is a direct relation between the tilt error normalized to the radius of the Airy disk and the resulting reduction in peak intensity. For example, an rms angular tilt error of 0.18 Airy radius reduces the peak intensity by a factor of 0.8. As the wavelength increases, the allowable tilt error and, consequently, the isoplanatic angle θ also increase. Each guide star therefore serves a larger area of sky at longer observing wavelengths (even if the star position sensor operates in the visible spectrum), resulting in better sky coverage for IR observations. As a separate matter, it should be noted that *tracking* the guide star at IR wavelengths is also advantageous, because of the high radiation of many IR stars and also because the star may be within the large isoplanatic patch of the laser beacon at IR wavelengths.

The overall tilt component of turbulence is usually corrected separately from the higher order wavefront errors, using a two-axis tracking mirror to eliminate image motion. Errors in tracking limit the angular resolution of long-exposure images. The exposure times of astronomical images are typically much longer than the characteristic time scale of atmospheric turbulence, so precise image stabilization is of vital importance.

When self-referencing or offset referencing is employed, in which the science object itself or a nearby star functions as the reference, the overall tilt information to drive the tracking mirror is normally extracted from the wavefront sensor (or from the reconstructor), so a separate tracking sensor is not necessary. A reference source that is bright enough to allow measurement of zonal gradients provides adequate signal-to-noise ratio for whole-aperture tracking, and the isoplanatic angle for overall tilt is always larger than that for higher order wavefront errors, so for self- and offset-referencing a separate tracking sensor is not required.

In the case of laser beacon systems, a separate tracking sensor using a natural star is essential, because a beacon launched from the ground cannot function as an absolute position reference. Although laser beacons can be placed in almost any position in the sky, natural stars of sufficient brightness have a relatively thin distribution, and so the sky coverage of compensated telescopes using laser beacons is determined solely by the availability of stars suitable for image stabilization.

The basic errors involved in compensating the angular position of an object by tracking an offset guide star are:

1. the measurement error of the tracking sensor itself, dependent on the signal-to-noise ratio of the detector output, the nature of the reference source, and the turbulence strength;
2. the temporal error due to the response time or bandwidth of the tilt compensation servo system;
3. tilt anisoplanatism, which depends on the angle between the science object and the reference star, the atmospheric turbulence structure, and the observation wavelength.

Tilt Measurement Errors

Tracking sensors normally employ a quadrant detector to sense the position of the reference object, similar to one channel of a Shack-Hartmann sensor. For a point reference source, the single-axis mean-square angular measurement error in radians squared, from equation (5.13), is

$$\sigma_{TM}^2 = \left[\frac{3\pi K_q \lambda_S}{16(SNR)D}\right]^2 \qquad r_0 > D$$
$$= \left[\frac{3\pi K_q \lambda_S}{16(SNR)r_0}\right]^2 \qquad D > r_0 \tag{9.62}$$

where

K_q = loss factor due to gap between quadrant detector elements (= 1.3-1.5)
SNR = signal-to-noise ratio, dependent on photon count and detector noise
λ_S = reference source wavelength, m,
D = Aperture of tilt sensor, m,
r_0 = effective turbulence parameter at reference wavelength, m (can be large if reference source is within isoplanatic angle)

The tilt error depends greatly on the diffraction-limited aperture size, for which two conditions are stated. When $r_0 > D$, the turbulence coherence length is larger than the aperture, so the full aperture is near diffraction-limited and D appears in the denominator. This is the most favorable condition and obtains when the reference star is compensated by the adaptive optics. This occurs either when dual adaptive optics is used, or when an IR guide star is used, in which case it may also be compensated by the primary adaptive optics because of the large isoplanatic angle at IR wavelengths. When $r_0 < D$, which is usually the case when the tilt reference source is not compensated, the diffraction-limited aperture is limited in size to approximately r_0, which then appears in the denominator, resulting in a larger tracking error.

Temporal Error

The temporal error of the tilt servo depends on the bandwidth of the atmospheric tilt fluctuations in relation to that of the tilt correction system. Tyler [1994a] has defined a fundamental tracking frequency f_T, (analogous to the Greenwood frequency f_G described in section 9.4.3), which characterizes the dynamics of wavefront tilt. When the tilt is corrected with a first-order servo having a 3-dB bandwidth of f_{3dB}, the single-axis mean-square angular error, in radians squared, can be expressed as

$$\sigma_{TT}^2 = \left(\frac{f_T}{f_{3dB}}\right)^2 \left(\frac{\lambda}{D}\right)^2 \tag{9.63}$$

where

$$f_T = K_T D^{-1/6} \lambda \sec^{1/2}(\zeta) v_2^{1/2} \tag{9.64}$$

K_T = tracking frequency constant
= 0.331 for average gradient tilt
D = tracking aperture, m,
λ = tracking reference wavelength, m
ζ = zenith angle
v_2 = second-order wind velocity moment

$$= \mu_0 v^2 = \int dh\, C_N^2(h) v^2(h) \tag{9.65}$$

v = turbulence weighted wind velocity

The tracking frequency f_T is normalized so that when $f_T = f_{3dB}$, the one-axis rms jitter is equal to λ/D.

Anisoplanatic Error

The tilt anisoplanatic error σ_{TA}^2 has been analyzed in section 7.4, where it was shown that for field angles up to about $D/40{,}000$, where D is the telescope aperture in meters, the single-axis error variance (in radians squared) due to angular anisoplanatism is

$$\sigma_{TA}^2 = 5.34 \sec^3(\zeta) \mu_2 \theta^2 D^{-7/3} \tag{9.66}$$

where

μ_2 - second-order turbulence moment
θ — angular separation

For a small separation angle θ, the mean-square wavefront phase error may be approximated as

$$\sigma_{TA}^2 = \left(\frac{\theta}{\theta_{TA}}\right)^2 \tag{9.67}$$

where

$$\theta_{TA} = \left[0.668 k^2 \sec^3(\zeta) \mu_2 D^{-1/3}\right]^{-1/2} \tag{9.68}$$

Total Tracking Error

The total mean-square tracking error is given by the sum of the individual variances

$$\sigma_\alpha^2 = \sigma_{TM}^2 + \sigma_{TT}^2 + \sigma_{TA}^2 \tag{9.69}$$

Tracking error may be expressed either as an angular tilt error σ_α (radians of angle), or as the equivalent wavefront error σ_w over an aperture D (radians of phase). The relationship is

$$\sigma_w = \frac{k D}{4} \sigma_\alpha \tag{9.70}$$

where $k = 2\pi/\lambda$.

9.5 Performance of Adaptive Optics Systems Using Natural Stars

9.5.1 *Error Model*

The mean-square wavefront (phase) error of an adaptive optics system may be expressed as the sum of individual error variances in the form

$$\sigma_S = \text{ffM} + \sigma_T + \sigma_F^2 + \sigma_A^2 + \sigma_D^2 + \sigma_{TR}^2 \tag{9.71}$$

where

CTM = wavefront sensor measurement error, equation (9.43) or (9.44)
σ_T^2 = temporal error, equations (9.53) and (9.57)
σ_F = wavefront fitting error, equation (9.61)
σ_A^2 = angular anisoplanatic error, equation (9.44)
σ_D = atmospheric dispersion (multispectral) error, equation (9.41)
^{CT}TR = phase error due to residual overall tilt.

This formulation assumes that the error components are completely independent, which may not be strictly true. Straightforward addition of the error variances tends to overestimate the total wavefront error, which is justified in the design of adaptive optics systems because it tends to compensate for small errors that are neglected. The wavefront reconstruction errors, equations (9.58) and (9.59), have been omitted deliberately as they are negligible with wavefront slope reconstructors of practical size.

In adaptive optics systems, wavefront errors are compensated by applying negative feedback to null the residual errors. Implicit in equation (9.71) is the assumption that the feedback loop performs perfectly, leaving only these residual errors. In the first-order feedback loops used in most adaptive optics systems, each loop has a large, but not infinite, gain at zero frequency, so steady-state errors are not reduced to zero. This has no importance for compensating turbulence, which has zero mean value, but it implies that optical alignment errors such as defocus are not completely eliminated by the feedback.

In the case of adaptive optics systems using natural stars, the same reference star is normally used to measure both the wavefront and the overall tilt. The photons used to measure the x and y tilt errors are collected using the whole telescope aperture, which is usually several meters in diameter. The signal-to-noise ratio of the tilt sensor is consequently much

greater than that of the wavefront sensor, which collects photons over much smaller subapertures that typically span less than half a meter. In this case, the overall tilt errors are negligible compared with the wavefront errors.

It is shown in section 9.5.4 that to maximize performance with faint reference stars, the subaperture size should be at least 0.15 m, and that good performance is obtained over the wide range of operating conditions typical of astronomical adaptive optics, using subapertures between 0.2 and 0.4 m. In the following performance calculations, a subaperture size of 0.2 m is employed. The importance of optimizing the photon integration time is discussed in section 9.5.3. The optimal integration time is the value that minimizes the total time-sensitive error $(\sigma_{BM}^2 + \sigma_{BT}^2)$ for each star magnitude.

9.5.2 Signal-to-Noise Ratio Using Natural Stars

The signal-to-noise ratio of the wavefront sensor depends primarily on the number of photons from the reference source that are detected and counted. Noise sources include the shot noise of the signal and background light, together with noise electrons added in the detection process.

The photon flux, N_S, in photons per second, received from a star in each subaperture at the input of the wavefront sensor is given by

$$N_S = (H_\lambda \lambda_S / hc) T_A \Delta\lambda \, d^2 T_O \qquad (9.72)$$

where

H_λ = spectral irradiance from star outside the atmosphere at wavelength λ_S, W m^{-2} μm^{-1}
λ_S = mean sensor wavelength, m
$\Delta\lambda$ = spectral bandwidth of wavefront sensor, μm
h = Planck's constant, 6.626 x 10^{-34} J s
c = velocity of light, 3 x 10^8 m s^{-1}
T_A = atmospheric transmission at mean sensor wavelength
d = subaperture dimension at the entrance pupil, m
T_O = optical transmittance from primary mirror to wavefront sensor input

The spectral irradiance from a G-type star of visual magnitude zero outside Earth's atmosphere at $\lambda = 0.55$ μm is $H_\lambda = 3.9 \times 10^{-8}$ W m^2 per μm [Wolfe and Zissis 1985]. The photon flux in each subaperture at a mean wavelength of 0.55 μm may be expressed in terms of visual magnitude m_V as

$$N_S = 1.08 \times 10^{11} (10^{-0.4 m_V}) T_A \Delta\lambda \, d^2 T_O \qquad (9.73)$$

One channel of a wavefront sensor detector using a frame-transfer CCD with a continuous reference source is modeled in figure 9.21. The photon flux reaching each subaperture is N_S photons per second and the optical transmission within the wavefront sensor to each detector is T_W. This optical transmission depends on the exact configuration of the sensor, and includes the attenuation of the optical components and shutter, together with the beam-division factor if multiple detector arrays are used. The advantages of using multiple detector arrays have been discussed in section 9.4.2.

The photoelectron count from the reference source is given by

$$n_c = N_S t_i T_W \eta \qquad (9.74)$$

where

n_c = number of events counted
t_i = effective integration time of the system
η = quantum efficiency of the photon detector

The CCD detector array is read out at a frame rate of $f_f = 1/t_f$, which is the basic sampling frequency in the system. To ensure stability of the feedback loop, f_f must be six to 10 times higher than the servo bandwidth, f_s, as discussed in section 8.5. The time and frequency relationships in the detection process are depicted in figure 9.22. The frame time t_f includes the transfer of the charges accumulated on detector pixels during the exposure time t_e into the storage

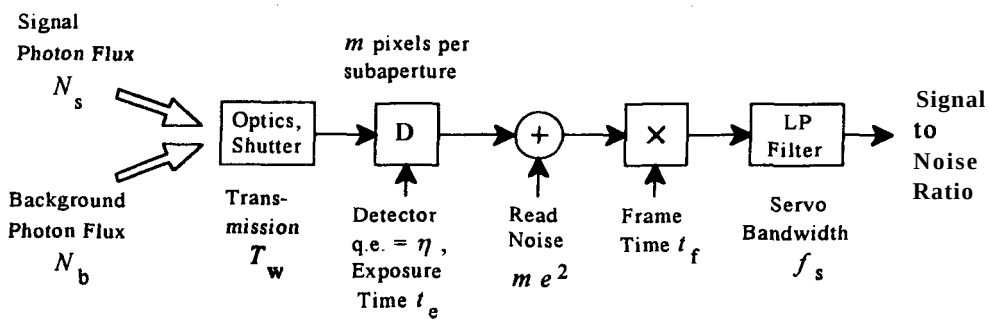


Figure 9.21 Model for photon detection in one subaperture using a continuous reference source such as a natural star, showing the factors that determine the signal to noise ratio of the wavefront sensor. q.e., Quantum efficiency (of detector); LP, Low pass.

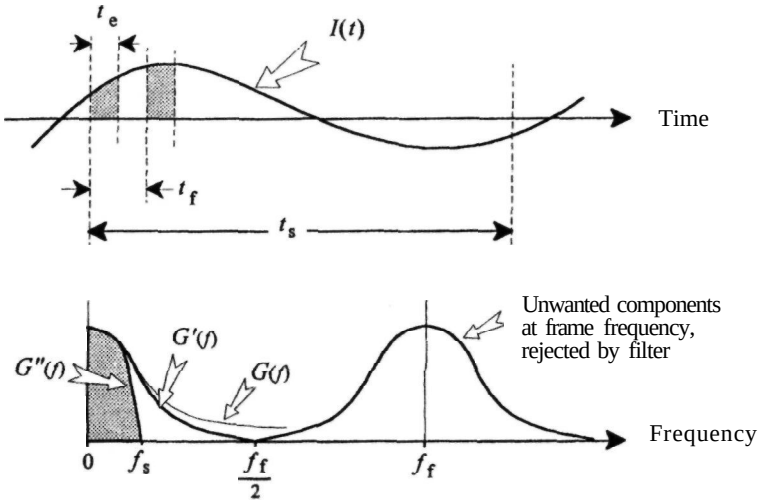


Figure 9.22 Time and frequency relationships in a wavefront sensor using a CCD detector. The detector frame rate $f_f = 1/t_f$ is the basic sampling frequency in the system and is usually six to 10 times higher than the servo bandwidth, $f_s = 1/t_s$, to ensure stability of the control loop. The effective photon integration time is determined by t_s. The spectrum of the raw turbulence (gradients) is $G(f)$ which is band-limited by the detector exposure time to approximately half of the frame frequency.

register on the CCD chip, followed by the readout itself. While one frame is being read out, the next frame is being exposed. The maximum exposure time is always somewhat less than the frame time because of the need for charge transfer.

The frequency spectrum of the wavefront gradients in each subaperture, $G(f)$, is band-limited by spatial averaging over the subapertures. It is further limited by temporal averaging over the exposure time of the CCD array, and by the frame (sampling) rate to give the spectrum $G'(f)$,corresponding to a bandwidth of about $f_f/2$. The time-sampled CCD output consists of frequency components at multiples of the frame rate f_f, each modulated with sidebands extending over $\pm G'(f)$.The low-pass filtering is beneficial in that it reduces or eliminates aliasing due to the overlap of the first-order sideband centered at f_f with the baseband spectrum. The low-pass filter in the servo loop cuts off at frequency f_s, which is normally about 1/10 of f_f, further reducing the disturbance spectrum to $G''(f)$.

One result of limiting the servo bandwidth is that turbulence components at frequencies higher than $G''(f)$ are not corrected by the adaptive optics. However, the reduced bandwidth also has the effect of increasing the integration time t_i, during which the photon counts are accumulated for each wavefront measurement, thereby increasing the signal-to-noise ratio of the wavefront sensor. The integration time is an important parameter in adaptive optics systems because it controls two major error components: the signal-to-noise ratio (and therefore the wavefront measurement error) and the temporal error. The wavefront measurement error varies approximately as $(1/t_i)^{1/2}$, while the temporal error varies as $(t_i)^{5/6}$. Consequently, there is an optimum value of t_i that minimizes the sum of these errors. Optimization of the integration time is covered in section 9.5.3.

The effective integration time over which reference source photons are accumulated depends on the characteristics of the low-pass filter in the servo loop; it is calculated as follows. The required integration time extends from $-t_i/2$ to $t_i/2$, and the temporal response, normalized to give unit area, is $h(t) = 1/t_i$. The equivalent filter response is the Fourier transform of $h(t)$:

$$F(\omega) = \int_{-\infty}^{\infty} dt\, h(t) e^{-i\omega t} = \int_{-t_i/2}^{t_i/2} dt \frac{1}{t_i} e^{-i\omega t} = \operatorname{sinc}\left(\frac{\omega T}{2}\right) \tag{9.75}$$

The photon pulses to be counted are modeled as impulses with a white-noise spectrum. If N_0 is the power spectral density in units squared per hertz, then the noise power passed by a filter with response $F(\omega)$ is

$$N = N_0 \int_0^{\infty} df |F(f)|^2 = \frac{N_0}{2\pi} \int_0^{\infty} d\omega |F(\omega)|^2 \tag{9.76}$$

For uniform integration over time t_i, the value of $F(\omega)$ from equation (9.75) is substituted to obtain

$$N = \frac{N_0}{2\pi} \int_0^{\infty} d\omega \operatorname{sinc}^2\left(\frac{\omega t_i}{2}\right) = \frac{N_0}{2t_i} \tag{9.77}$$

If the noise bandwidth in hertz is defined as $f_N = N/N_0$, then the basic relation between the noise bandwidth and the integration time is

$$t_i = \frac{1}{2f_N} \tag{9.78}$$

If the band-limiting filter is a first-order filter with time constant r, the filter response is

$$F(\omega) = \frac{1}{1 + i\omega\tau} \tag{9.79}$$

The noise bandwidth for this filter is

$$f_N = \frac{1}{2\pi}\int_0^\infty d\omega \frac{1}{1+(\omega\tau)^2} = \frac{1}{4\tau} \tag{9.80}$$

The effective integration time for a first-order filter is therefore $1/(2f_N) = 2T$ (two time-constants). If the half-power (3-dB) frequency of the filter is $f_{3dB} = 1/(2\pi\tau)$, then the effective integration time is expressed conveniently as

$$t_i = 2\tau = \frac{1}{\pi f_{3dB}} \tag{9.81}$$

This is the photon integration time assuming that the detector is continuously exposed. It must be reduced by the ratio t_e/t_f to give the final integration time. The time required to shift the charges to the storage register on a CCD is on the order of 10 μs, whereas frame times are typically 500μs or more, so, in practice, the maximum value of t_e/t_f is close to unity. Even though the transfer time is short, some smearing of the CCD image occurs if it is exposed to light continuously, so a shutter is often used to control the actual exposure t_e.

Detector Read Noise

During the readout of each pixel, noise is added by the charge-to-voltage conversion amplifier on the CCD chip. The root-mean-square value of this noise is specified as e electrons per pixel at a given pixel- or frame-rate. If m is the number of pixels per measurement (four in the case of a Shack-Hartmann sensor using quadrant detectors), then the mean-square read noise per measurement is me^2.

Background Noise

The main source of background noise at visible and near-IR wavelengths is radiation from the sky adjacent to the reference source. Sky background noise is minimized by using the smallest practical field of view in the wavefront sensor, which may be implemented by a variable field stop in the reference detector path. To allow for initial acquisition of the reference source, the field of view must encompass the source size plus the uncorrected seeing angle. Obviously, the signal-to-noise ratio under these conditions must be sufficient to allow the adaptive optics loop to converge. The background noise therefore determines the minimum reference source brightness necessary to initiate closed-loop operation. Once the system has locked on to the reference and the wavefront excursions are diminished, the field of view may be reduced to increase the signal-to-noise ratio, improving the precision of the compensation.

The importance of sky radiation depends on the function of the wavefront sensor, especially the wavelength and type of reference source. With adaptive optics systems using natural stars, both the wavefront and the overall tilt correction are obtained from the same reference source, and relatively bright stars ($m_V \sim 10$) must be used. At visible wavelengths, the sky radiation is relatively low and the exposure times required for wavefront measurement are so short (milliseconds) that background radiation may usually be neglected. To take an extreme example, daylight operation of adaptive optics has been achieved with bright stars, using a shearing interferometer sensor with a 5 arc-second field stop [Shelton et al. 1993]. The opposite extreme is represented by overall tilt sensing using stars as dim as 20th magnitude, for which sky radiation becomes the limiting factor.

Taking into account all noise sources, the expression for wavefront sensor signal-to-noise ratio, from equation (5.17), is

$$SNR = \frac{n_c}{\left\{n_c + m\left[n_b^2 + (e/G)^2\right]\right\}^{1/2}} \tag{9.82}$$

where

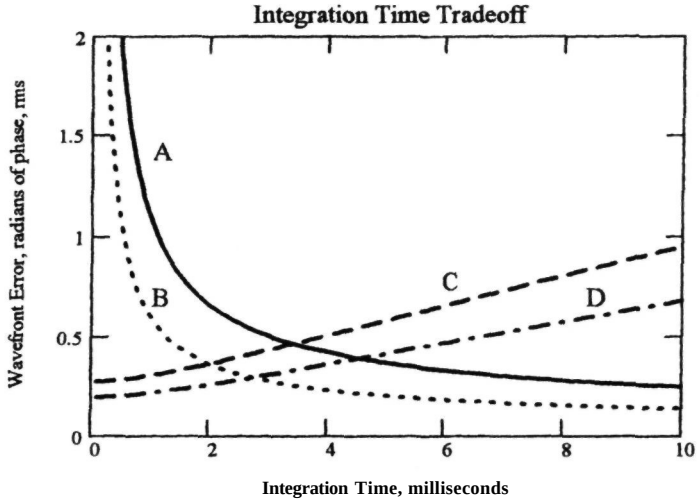
- n_c = number of photons from the reference source counted in the sensor integration time
- e = read noise per pixel in electrons rms
- n_b = number of detected background electrons per pixel
- m = number of pixels per subaperture
- G = gain of the intensifier.

For an unintensified detector, such as a bare CCD, the gain ($G = 1$). For night-time observations with natural star systems, n_b is usually negligible and the signal-to-noise ratio using a CCD detector reduces to

$$SNR = \frac{n_c}{\left\{n_c + me^2\right\}^{1/2}} \tag{9.83}$$

9.5.3 Optimization of Integration Time

The integration time is one of the few parameters in an adaptive optics system that are really "free" in the sense that there are no physical constraints. The integration time has a large effect on performance with little, if any, cost impact (other than to make it variable), so it makes sense to optimize it for each set of operating conditions. The measurement error produced by a Shack-Hartmann wavefront sensor [equation (5.16)], and the temporal error [equations (9.53) and (9.57)] are plotted as a function of integration time in figure 9.23. The photon count and signal-to-noise ratio improve as the integration time increases resulting in a smaller measurement error, which is shown for two values of r_0. On the other hand, a



A sensor error, $r_0 = 0.1$ m at $\lambda = 0.5\ \mu$m

B sensor error, $r_0 = 0.2$ m

C temporal error, Greenwood frequency $f_g = 39$ Hz

D temporal error, Greenwood frequency $f_g = 26$ Hz

Conditions: sensor wavelength = 0.7 μm, zenith angle = 30 degrees, atmospheric transmission = 0.8, reference source visual magnitude = 10, overall quantum efficiency = 0.4, sub aperture size = 0.25 m, pixels per measurement = 4, read noise per pixel = 5 electrons.

Figure 9.23 Integration time tradeoff. As the integration time is increased, the wavefront sensor error falls because more photons are collected and the signal-to-noise ratio improves, but the temporal error rises because of the longer lime delay. The optimal integration time also depends on the reference source size and brightness, and on the subaperture size in relation to the turbulence parameter r_0.

longer integration time increases the temporal error. So for each set of operating conditions, there is an optimum integration time t_{opt} at which the sum of the time-dependent errors is minimized.

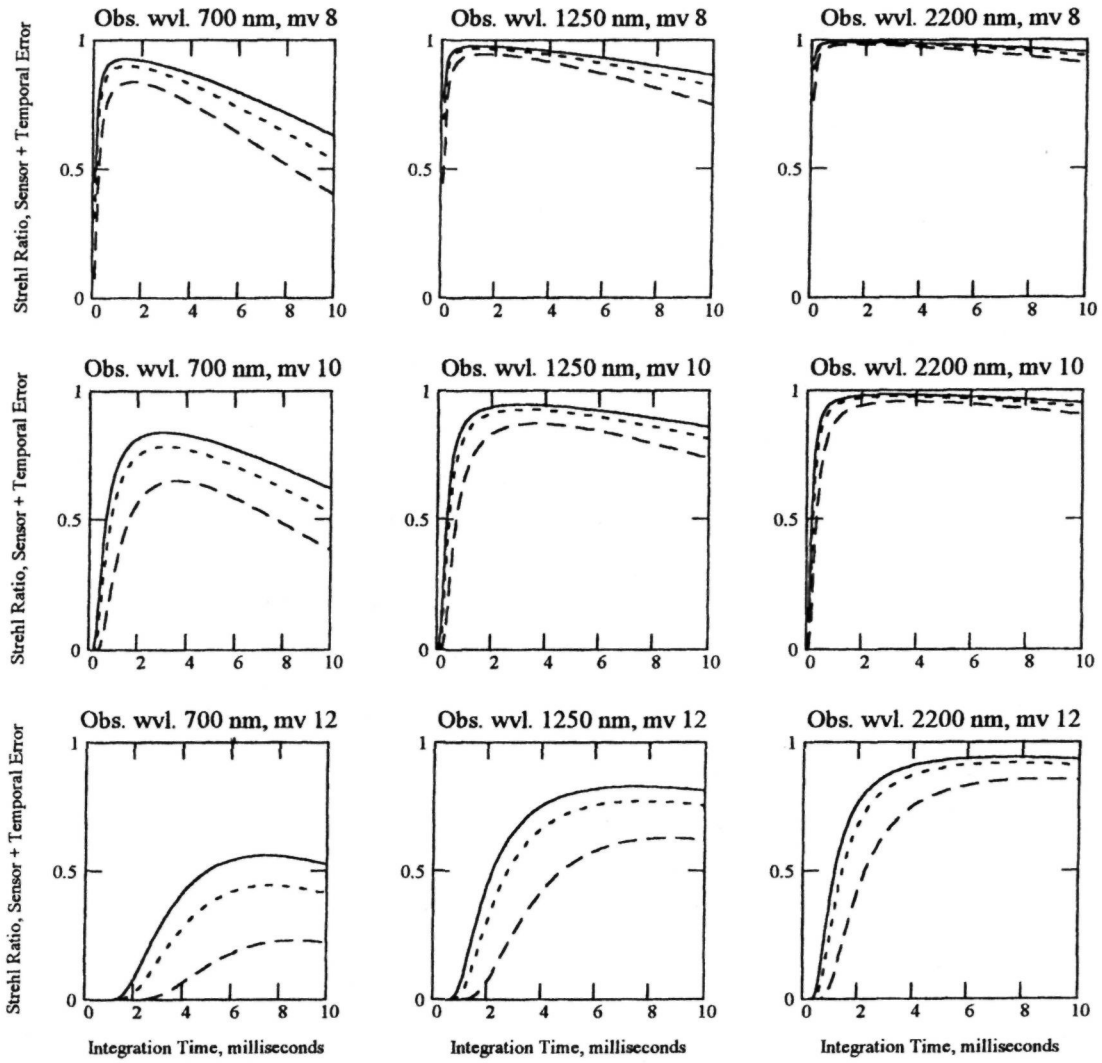
The Strehl ratio produced by the time-dependent errors is shown for a range of conditions in figure 9.24. Charts are plotted for observation wavelengths of 0.7, 1.25, and 2.2 μm, with reference source visual magnitudes of 8, 10, and 12. In each case, the Strehl ratio is shown for r_0 values of 0.1, 0.15, and 0.2m (at A = 0.5 μm), with the integration time varied from 0 to 10 ms. The optimum integration time is independent of the observation wavelength, which merely scales the wavefront error (for a fixed subaperture size). As expected, the reference source magnitude has the dominant effect, with t_{opt} being roughly proportional to $n_c^{-1/2}$, where n_c is the number of photons counted. The optimum integration time is weakly dependent on r_0, with an increase in turbulence strength requiring a somewhat longer integration time. The curves drop sharply for integration times less than the optimum, indicating that it is better to err on the side of generosity if in doubt. These results suggest that the integration time may be optimized, in practice, by adjusting it to maintain a predetermined photon count, which is relatively easy to implement with a real-time control loop. This subject is discussed in section 9.7.

9.5.4 Optimization of Number of Actuators

If the number of actuators or degrees of freedom in an adaptive optics system is increased, the fitting error improves, but (for a constant reference source magnitude) the number of photons collected per degree of freedom decreases so the measurement precision falls. Consequently, for each set of operating conditions, there is an optimum number of degrees of freedom that minimizes the total error.

The basic expression for the angular error of a bicell tilt sensor using a point source, from equation (5.13) is

$$\sigma_\alpha = \frac{3\pi}{16}\frac{\lambda}{d}\frac{1}{SNR} \quad \text{rad}$$



Conditions:

Sensor wavelength, μm	**λs =0.7**	**Overall quantum efficiency**	**n =0.4**
Zenith angle, degrees	**z = 30**	**Subaperture size, meters**	**d =** 0.25
Atmos. transmission	**Ta = 0.8**	**Read noise per pixel**	**e = 5**
Sensor spectral band, μm	**8 = 0.3**	**Pixels per measurement**	**N =4**

Figure 9.24 Optimization of integration time. The chart plots the combined Strehl ratio of sensor and temporal errors against *t,* for a range of operating conditions.

where

λ = mean operating wavelength
d = subaperture size
SNR = voltage signal-to-noise ratio of the detected signal.

The angular error is proportional to the diffraction-limited spot diameter X/d, divided by the signal-to-noise ratio SNR. To find the wavefront phase error per subaperture, the angle must be multiplied by $\pi d/2\lambda$. The signal-to-noise ratio is proportional to the square root of the number of photons counted, which for a given source brightness may be stated as $SNR \propto \sqrt{d^2} \propto d$ For the usual case in which $d \approx r_0$, the dependency of the wavefront measurement error on subaperture size is therefore

$$\sigma_M \propto \frac{1}{d} \quad \text{rad of phase per subaperture}$$

For a given telescope aperture, as the number of actuators increases, the subaperture size d decreases, resulting in a larger wavefront measurement error. If d is larger than the turbulence coherence length r_0, the spot diameter is determined by the ratio λ/r_0, and the phase error per subaperture is proportional to $1/r_0$; that is, independent of d.

The subaperture size also controls the fitting error, which, from equation (9.61), is

$$\sigma_F \propto \left(\frac{d}{r_0}\right)^{5/6}$$

These errors are shown as a function of subaperture size d in figure 9.25, for two values of r_0. The change in the slope of the sensor error when $d > r_0$ is clearly seen. When the mean-square values of these errors are added in quadrature and expressed as a Strehl ratio, the optimal subaperture size is defined by the peak of the curve. The function is highly asymmetric, dropping very steeply for subaperture sizes below the optimum.

The optimization of subaperture size is explored over a wider range of parameters in figures 9.26 and 9.27. These charts give a comprehensive view of the influence of subaperture size on overall performance, which is a more informed approach than seeking the peak of a single analytical function. The figures show, surprisingly, that the optimum value of d does not depend on either the observation wavelength or the sensor wavelength, although the peak tends to be sharper at short observation wavelengths. The main factor determining the optimum value of d is the photon flux from the reference source.

Wavefront sensor errors are dominant over the fitting errors so far as the optimization of subaperture size is concerned. When $d > r_0$, the wavefront sensor error depends on the value of r_0 at the sensor wavelength, as shown in figure 9.25, and not on d. This is the reason that the optimum subaperture size does not increase at longer observation wavelengths, as might be expected from consideration of the fitting error only. The read noise of the wavefront sensor (CCD) detector also has an influence on the optimum subaperture size, because the addition of noise electrons requires a larger signal flux to achieve the same signal-to-noise ratio. At low photon fluxes, the effect of large read noise is to increase the optimum subaperture size.

The net result of all of these factors is that the optimal value of d occurs between 0.15 and 0.25 m for a wide range of operating conditions. Furthermore, the penalty for using subapertures as large as 0.4 m tends to be less than that for using smaller sizes, suggesting that larger subapertures are a more robust solution. Reducing the number of actuators has a favorable effect on the hardware complexity and cost. Small subapertures are not a good choice for astronomical adaptive optics because, although they may give good performance with bright stars, the steep drop at lower photon rates kills the performance and limits the general utility of the instrument.

9.5.5 Performance and Sky Coverage

The overall performance of adaptive optics systems using natural stars can now be calculated using the error models summarized in the previous sections. The performance criterion of most general interest is the Strehl ratio of the compensated image, and this is determined as a function of the reference star magnitude and the atmospheric conditions, specified by the turbulence models described in section 9.2. For each condition, the integration time is optimized. As shown in section 9.5.4, the subaperture size is not critical and a value of d = 0.2 m is used in the following examples. The performance chart shown in figure 9.28 contains four plots:

1. on-axis Strehl ratio versus reference star magnitude at observation wavelengths of 0.7, 1.25, 1.62, and 2.2 μm;
2. optimized integration time for each star magnitude;
3. limiting magnitude of the reference star as a function of Strehl ratio at each observation wavelength;
4. field angle within which a given Strehl ratio is obtained, at each observation wavelength.

Natural-star adaptive optics systems perform best at IR wavelengths, particularly in the H- and K-bands, where operation is possible with stars of m_V = 14 over small fields of view. In the V- and R-bands, stars brighter than m_V = 10 are required to obtain a useful Strehl ratio and the compensated field is only a few arc seconds, as shown in figure 9.29.

The sky coverage of adaptive optics systems using natural stars may be obtained in the following way. Data on star distribution versus visual magnitude are found in *The Infrared Handbook* [Wolfe and Zissis 1985]. For use in performance calculations, it is convenient to model the star densities analytically, as described by Parenti and Sasiela [1994]. The models are

Galactic equator

$$\mathcal{D}_E \text{ as } 3.97 \times \exp(m_V) \quad \text{stars per radian squared}$$

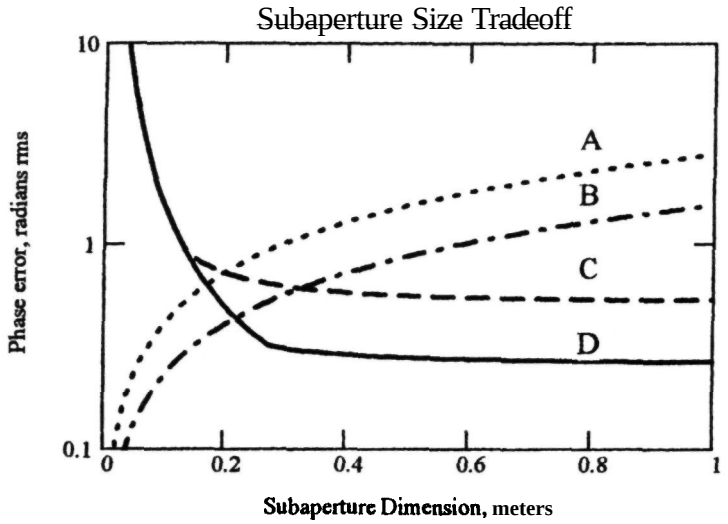
Galactic pole

$$\mathcal{D}_P = 1.27 \times 10^{-4}(m_V)^{8.2}$$

The average density of stars brighter than a given magnitude m_V may be modeled as

$$\mathcal{D}_A = 3 \times \exp(0.9 m_V) \quad \text{stars per radian squared}$$

These star densities are plotted in figure 9.30.



A fitting error, $r_0 = 0.1$ m C sensor error, $r_0 = 0.1$ m
B fitting error, $r_0 = 0.2$ m D sensor error, $r_0 = 0.2$ m

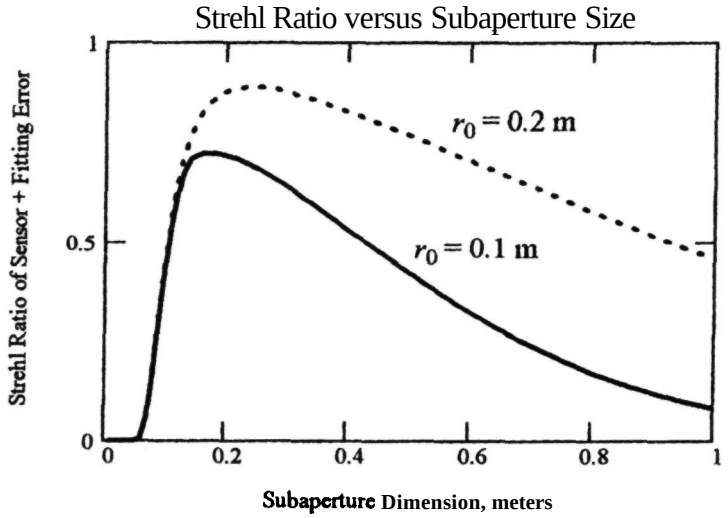
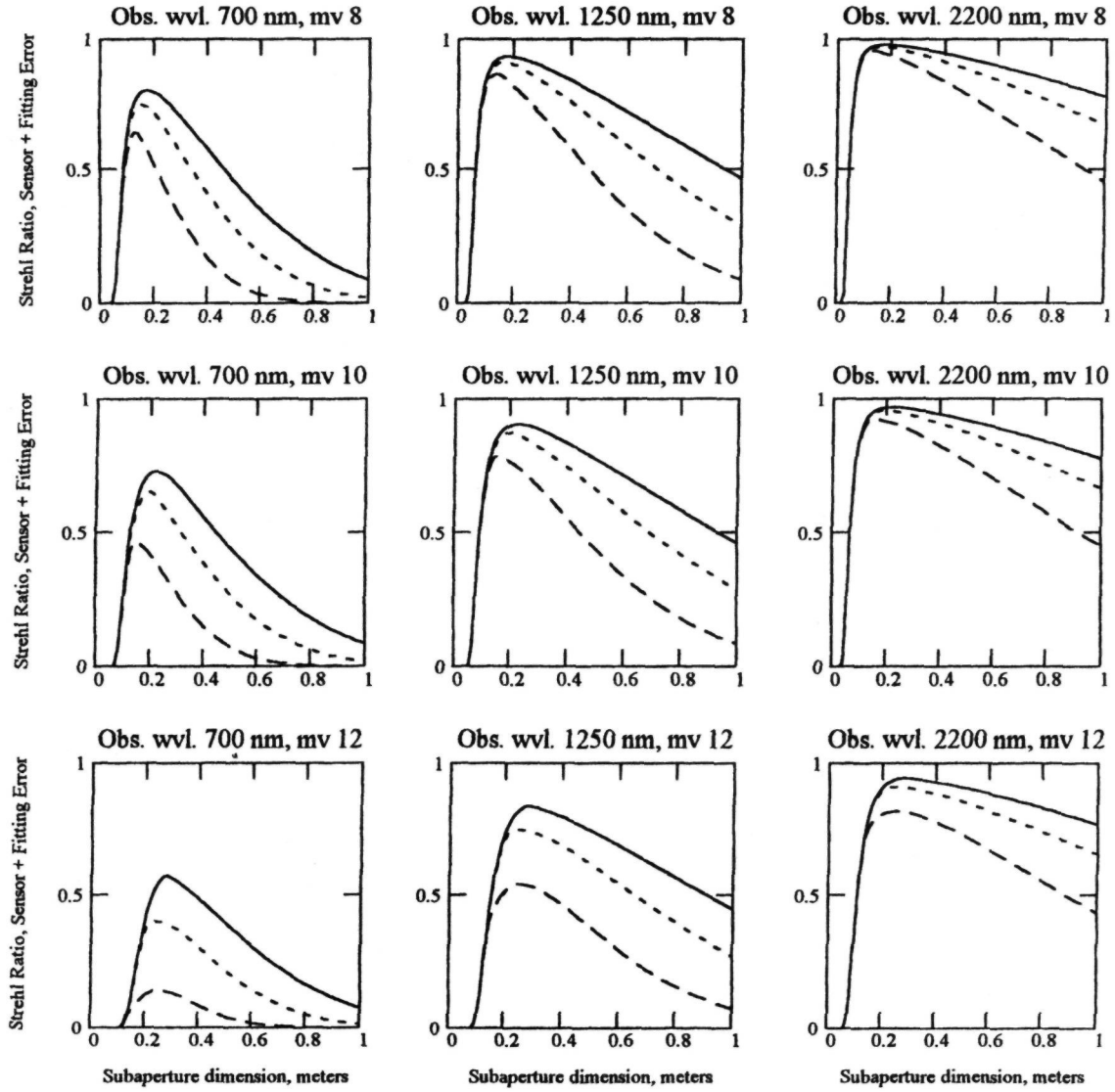


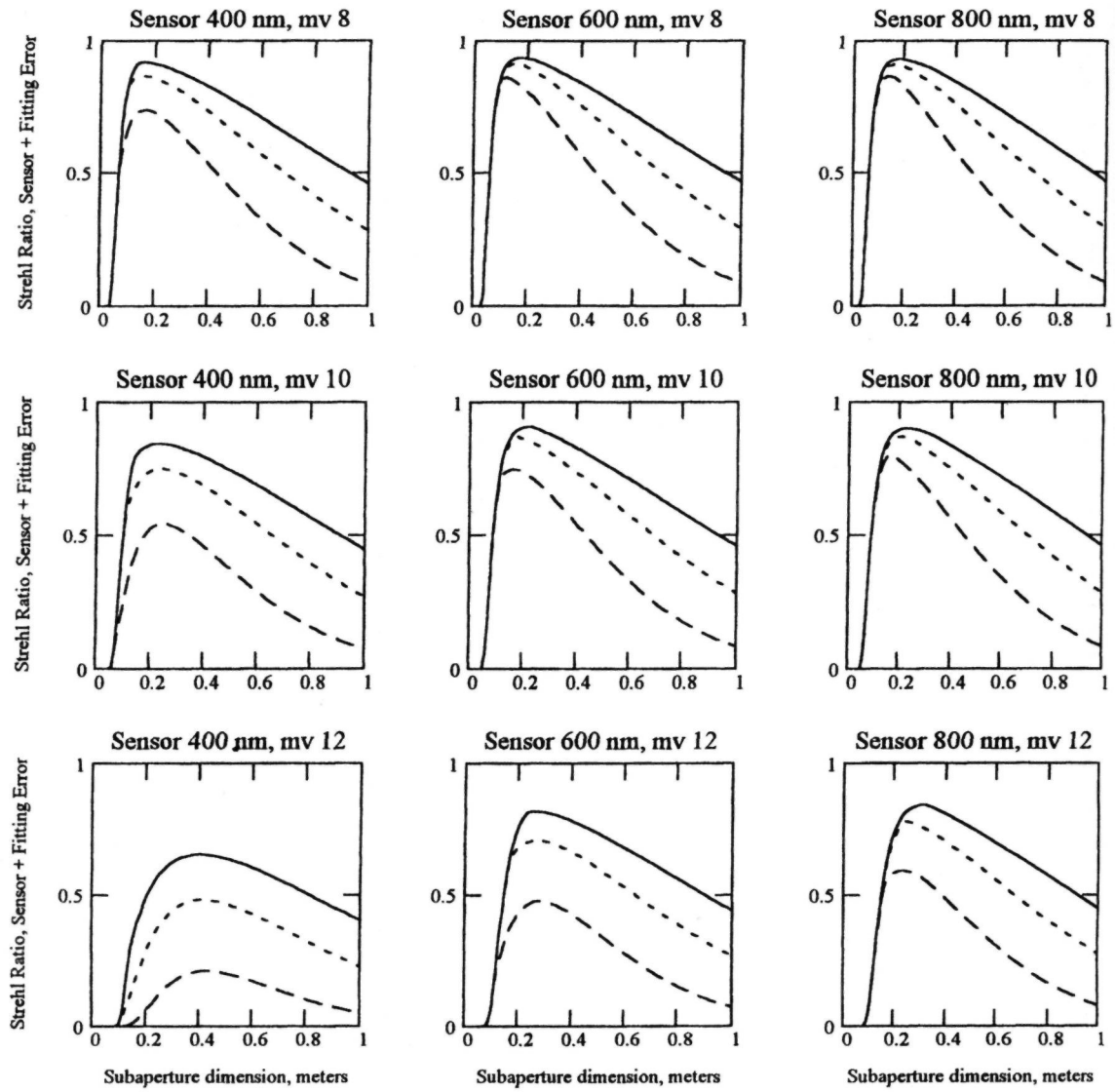
Figure 9.25 Subaperture size tradeoff. As the subaperture size is increased, the wavefront sensor error falls because more photons are collected, improving the signal-to-noise ratio, but the fitting error increases. Note that the wavefront sensor error curve flattens out when the subaperture size exceeds r_0, because the angular resolution of the sensor is then determined by the turbulence scale and not by the subaperture diameter. The combined sensor and fitting errors produce a clearly defined maximum in the Strehl ratio.



Conditions:

Sensor wavelength, μm	λs = 0.7	Overall quantum efficiency	η = 0.4
Zenith angle, degrees	z = 30	Exposure time: optimized	
Atmos. transmission	Ta = 0.8	Read noise per pixel	e = 5
Sensor spectral band, μm	5 = 0.3	Pixels per measurement	N = 4

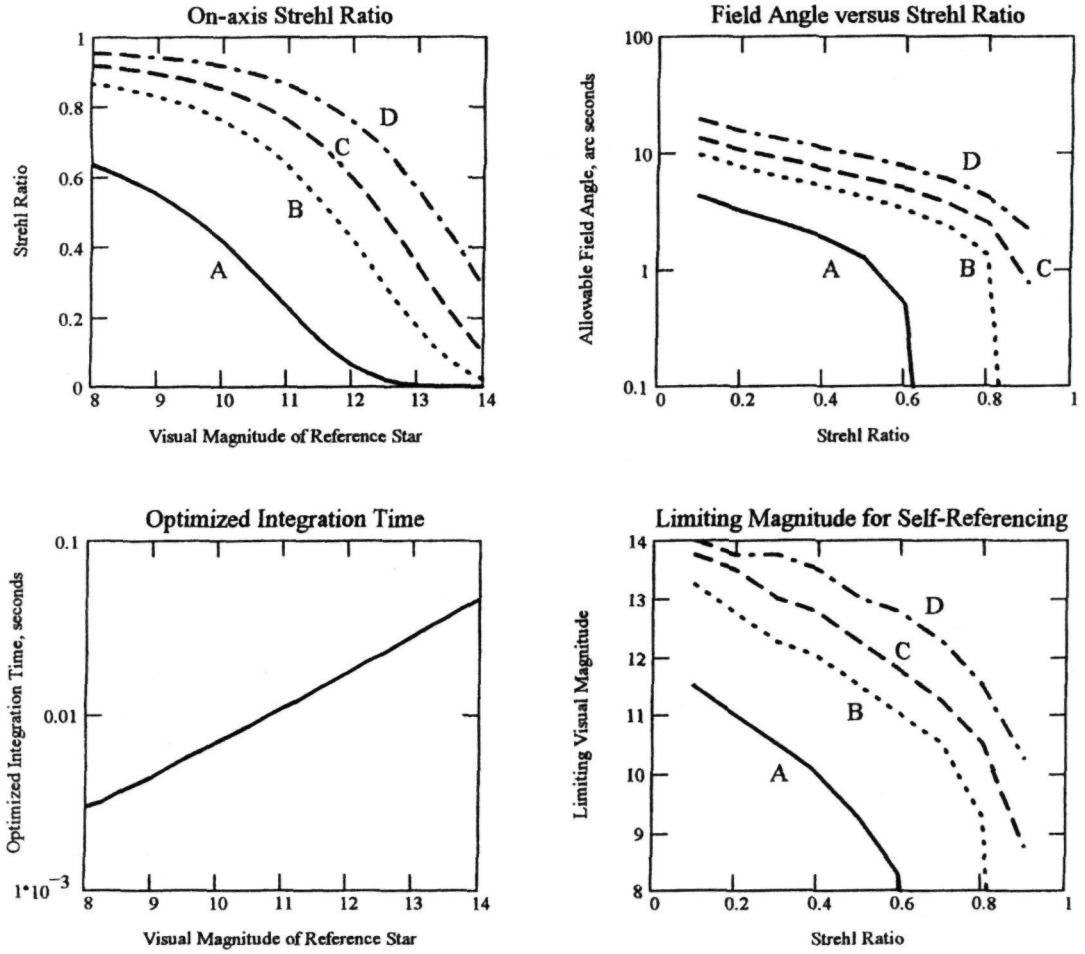
Figure 9.26 Optimization of subaperture size. The chart plots the combined Strehl ratio of sensor and fitting errors against d, for a range of operating conditions. The subaperture size is controlled mainly by the brightness of the reference source.



Conditions:

Observation wvlngth, μm	λo = 1.25	Overall quantum efficiency	n = 0.4
Zenith angle, degrees	z = 30	Exposure time: optimized	
Atmos. trans. 0.6 - 0.8		Read noise per pixel	e = 5
Sensor spectral band, μm	5 = 0.3	Pixels per measurement	N =4

Figure 9.27 Effect of sensor wavelength on optimum subaperture size. The Strehl ratio due to sensor and fitting errors is shown for sensor wavelengths of 0.4, 0.6, and 0.8 μm. The exposure time is optimized for each reference source magnitude.



Conditions: **subaperture** size d = **0.2 m,**
turbulence parameter r_0 = **0.15 m**
zenith angle ζ = **30 degrees**
isoplanatic angle θ_0 = **12** μ **radians**
Greenwood frequency^ = 31 Hz
CCD read noise = 5 electrons per pixel

A observation wavelength 0.70 μ**m**
B observation wavelength 1.25 μ**m**
C observation wavelength 1.62 μ**m**
D observation wavelength 2.20 μ**m**

Figure 9.28 Performance of an adaptive optics system using a natural star as the reference source for both wavefront and tilt compensation. The wavefront sensor operates in the visible band centered at 0.55 μm, and data are plotted at observing wavelengths of 0.7, 1.25, 1.62, and 2.2 μm. The servo loop integration time is optimized for each reference star magnitude.

The sky coverage for a given overall Strehl ratio is found as follows. First, the total wavefront error corresponding to the given Strehl ratio is determined:

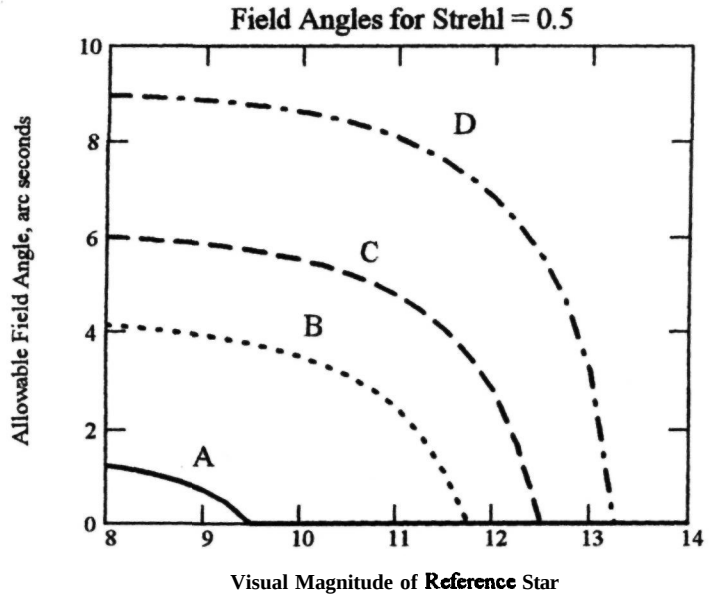
$$\sigma_S^2 = -\ln(S) \tag{9.84}$$

The total error has three constituents: the errors proportional to star magnitude, $(\sigma_M^2 + \sigma_T^2)$; the isoplanatic error σ_A^2; and other fixed errors σ_R. The allowable anisoplanatic error may then be determined as a function of star magnitude:

$$\sigma_A^2 = -\ln(S) - \sigma_M^2 - \sigma_T^2 - \sigma_R^2 = \left(\frac{\theta_A}{\theta_0}\right)^{5/3} \tag{9.85}$$

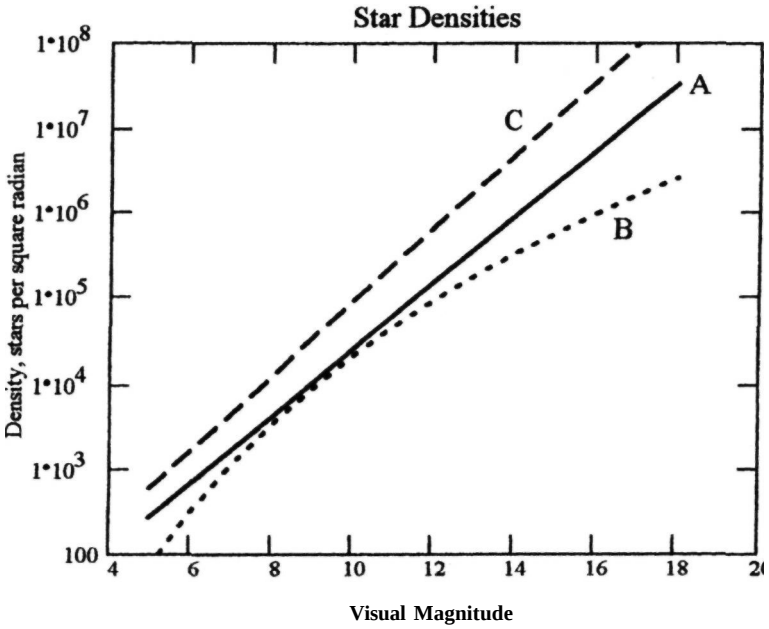
The angular radius θ_A (around the reference star) at which the anisoplanatic error equals the allowed threshold for each star magnitude is then calculated, using

$$\theta_A = \theta_0(\sigma_A^2)^{3/5} \tag{9.86}$$



A observation wavelength 0.70 μm
B observation wavelength 1.25 μm
C observation wavelength 1.62 μm
D observation wavelength 2.20 μm

Figure 9.29 Compensated field angles achievable at various observation wavelengths for a Strehl ratio of 0.5, using a natural star as the reference source. The wavefront sensor operates at 0.55 μm and the integration time is optimized at each magnitude.



A average star density model
8 galactic pole model
C galactic equator model

Figure 9.30 Density of stars brighter than a given visual magnitude. Models for the galactic pole and equator are based on data from *The Infrared Handbook* (Wolfe and Zissis 1985).

Using the average star density $\mathscr{D}_A$ as a function of magnitude, the fractional sky coverage is given by

$$\mathscr{S} = \pi(\theta_A)^2 \mathscr{D}_A \tag{9.87}$$

The fractional sky coverage for adaptive optics using natural stars is shown in figure 9.31 for Strehl ratios of 0.1–0.5, at observation wavelengths of 0.7, 1.25, 1.62, and 2.2 μm. Other conditions are specified in the figure. Even in the K-band at 2.2 μm, only about 1/1000 of the sky is accessible using this mode of operation.

The utility of natural-star operation should not be judged solely on sky coverage. Rather, it should be regarded as an economical and efficient method of implementing adaptive optics in those special cases in which a suitable field star is proximate to the science object.

9.6 Performance of Laser Beacon Systems

9.6.7 *Error Model*

Adaptive optics systems using laser beacons employ separate compensators for wavefront correction and overall tilt correction. The total error is composed of two parts:

1. the phase error of the wavefront compensator, using one or more laser beacons as the reference source;
2. the angular error of the overall tilt compensator, using a natural guide star.

These errors will be determined and analyzed separately before being combined into an overall performance model.

The mean-square wavefront (phase) error of a laser beacon system may be expressed as the sum of the individual error variances in the form

$$\sigma_W^2 = \sigma_{BM}^2 + \sigma_{BT}^2 + \sigma_F^2 + \sigma_{FA}^2 + \sigma_{MB}^2 + \sigma_D^2 \tag{9.88}$$

where

- σ_{BM} = beacon measurement error, equations (9.43) or (9.44)
- CT|T = temporal error, equations (9.53) and (9.57)
- CT_F = wavefront fitting error, equation (9.61)
- CT_{FA} = cone or focal anisoplanatic error, equation (7.35)
- σ_{MB} = error due to stitching multiple beacons, equation (9.30)
- CTQ = atmospheric dispersion (multispectral) error, equation (9.41)

This formulation assumes that all the error components are uncorrelated, which may not be strictly true, leading to a small overestimation of the error. The wavefront reconstruction errors, equations (9.58) and (9.59), have been omitted deliberately as they are negligible with wavefront slope reconstructors of practical size.

For the operating conditions typical of astronomical adaptive optics, it was shown in section 9.5.4 that the subaperture size is not a critical factor and that good performance is obtained with subapertures between 0.15 and 0.4 m. In the following performance calculations, a subaperture size of 0.2 m is used. The photon integration time is optimized by calculating the total time-sensitive error $(\sigma_{BM}^2 + \sigma_{BT}^2)$ for each value of laser power and determining the integration time that minimizes it.

The mean-square angular tilt error produced by tracking a fixed guide star may be expressed as

$$\sigma_\alpha^2 = \sigma_{TM}^2 + \sigma_{TT}^2 + \sigma_{TA}^2 \tag{9.89}$$

where

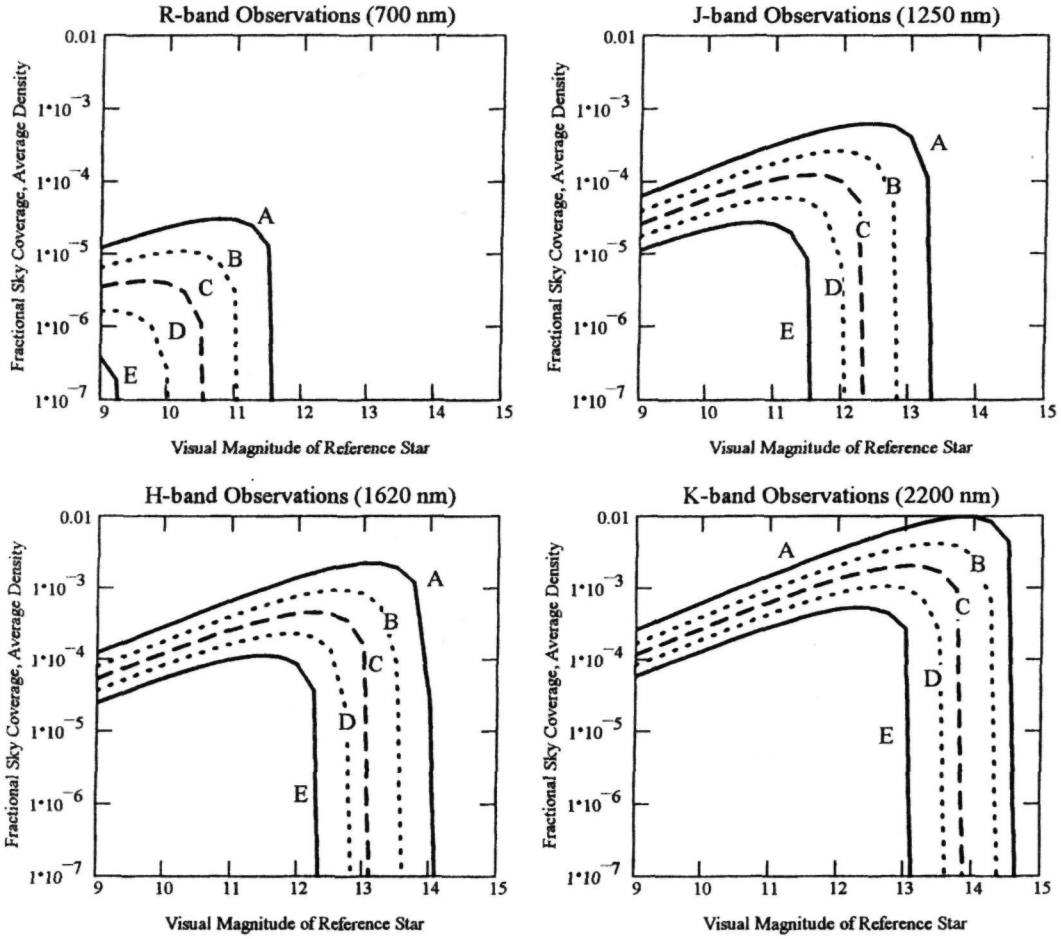
- σ_{TM} = tilt measurement error, using a quadrant detector with a natural star as the reference source, equation (9.62)
- σ_{TT} = temporal error due to the bandwidth of the tilt correction loop (pure time delays should be negligible in this case), equation (9.63)
- ^{CT}TA = error due to tilt anisoplanatism, dependent on the angular separation of the natural star from the science object, equation (9.66)

The tilt sensor is assumed to use the full aperture of the telescope and, again, the optimum tilt integration time is computed by minimizing the value of $(\sigma_{TM}^2 + \sigma_{TT}^2)$. To determine overall performance, the angular error σ_α^2 produced by tracking the fixed guide star must be combined with the wavefront phase error σ_W^2 due to the beacon wavefront compensator. Image motion due to angular tilt is modeled by using Gaussian image profiles as discussed in sections 4.5 and 4.6. The structure of the image is modified in two ways:

- Random angular tilt errors reduce the intensity of long-exposure images. The intensity reduction factor F_C due to an angular tilt variance of σ_α^2 for an aperture of D and observation wavelength λ, from equation (4.61), is

$$F_C = \left[1 + 5.17\left(\frac{D}{\lambda}\right)^2 \sigma_\alpha^2\right]^{-1}$$

When the laser beacon system produces a well-defined image core (as it should), the peak intensity of the image (Strehl ratio) is therefore reduced by this factor.



Conditions: subaperture size d - 0.2 m,
r_0 = 0.15 m at 0.5 μm,
θ_0 = 12 μradians (2.5 arc seconds)
actual zenith angle ζ — 30 degrees
Greenwood frequency f_G = 31 Hz
CCD read noise = 5 electrons per pixel.

A sky coverage for Strehl ratio = 0.1
B sky coverage for Strehl ratio = 0.2
C sky coverage for Strehl ratio = 0.3
D sky coverage for Strehl ratio = 0.4
E sky coverage for Strehl ratio = 0.5
(Strehl ratios at observation wavelengths)

Figure 9.31 Sky coverage of adaptive optics using natural stars. The wavefront sensor operates in the visible band, centered at 0.55 μm. The integration time is optimized for each star magnitude. Sky coverage is based on the global average distribution of stars.

- Angular tilt errors increase the radius of long-exposure images. The factor by which the image core radius is increased is $F_C^{-1/2}$ as given in equation (4.63).

It was shown in section 4.5.3 that if the angular tilt error is expressed in Airy disk units of $1.22\lambda/D$, then the relation between the wavefront error σ_ϕ^2, the tilt error σ_α^2, and the resulting Strehl ratio S_C is independent of wavelength and aperture size. The relation is given by

$$\sigma_A = 0.36\left[\frac{\exp(-\sigma_\phi^2)}{S_C} - 1\right]^{1/2} \tag{9.90}$$

where CT_A is the tilt error in Airy disk units. This expression gives the tradeoff between σ_A and σ_ϕ for given values of S_C and is shown in figure 9.32. For example, to achieve an overall Strehl ratio of 0.6 with a wavefront phase error of 0.5 rad rms, the allowable tilt error is 0.2 of the Airy radius $(0.244\lambda/D)$.

The sky coverage of a laser beacon system is determined entirely by the availability of natural

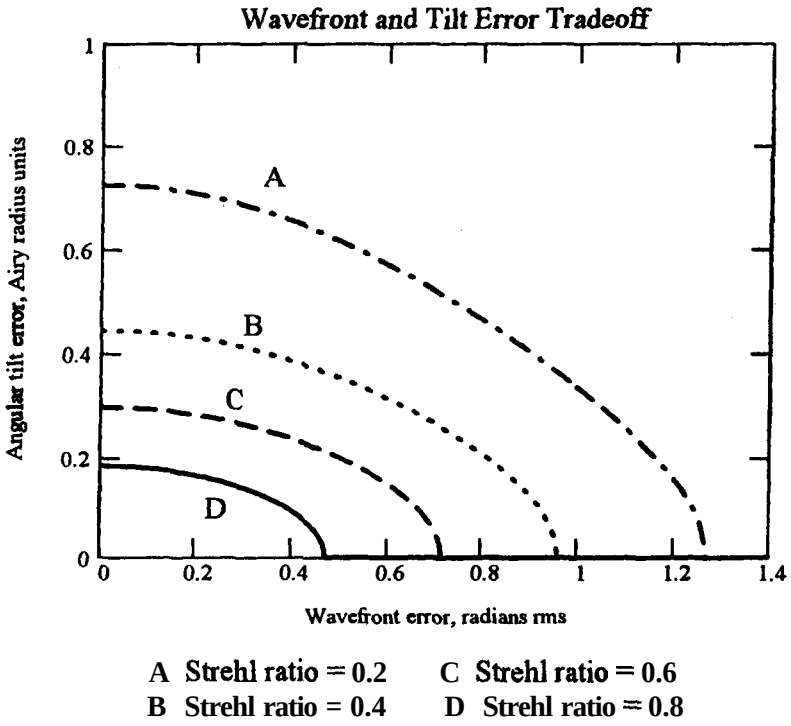


Figure 9.32 Strehl ratios due to combined wavefront and tilt errors. The units are chosen to make this chart independent of wavelength and telescope aperture.

guide stars for the tilt sensor. Because of the much fainter stars that can be used for tilt referencing, compared with those required for wavefront referencing, sky background radiation plays an important role in determining the sky coverage, especially when IR stars are used. This becomes an important consideration when the natural star used as the tilt reference shares the wavefront compensation provided by the beacon.

The first step in calculating the performance of adaptive optics systems using laser beacons is to compute the signal-to-noise ratios produced by the beacon and the natural guide star.

9.6.2 Laser Beacon Signal-to-Noise Ratio

The number of photons received from a laser beacon, within each subaperture at the input to the wavefront sensor, from equation (7.1), is

$$n_p = \left(\frac{\lambda_b}{hc}\right) T_A^2 \frac{\sigma_B N(z)}{4\pi} \left(\frac{\Delta z}{z^2}\right) (E\, T_L d^2 T_O) \qquad (9.91)$$

where

n_p = expected number of photons per pulse
E = laser pulse energy, J
λ_b = beacon wavelength, m
h = Planck's constant, 6.626 x 10^{-34} J s
c = velocity of light, 3 x 10^8 m s^{-1}
T_A = one-way transmission of atmosphere between telescope and beacon
σ_B = effective backscatter cross-section, m^2
$N(z)$ = number density of scatterers at range z, m^{-3}
Az = receiver range gate length, m
z = range to center of range gate, m
d = subaperture size at telescope entrance pupil, m
T_L = transmission of laser path to projection aperture
T_O = transmission of optical path from primary mirror to wavefront sensor

The values of most of these parameters depend on the type of laser beacon; that is, whether it employs Rayleigh scattering or sodium resonance fluorescence. For Rayleigh systems, the backscatter parameter $\sigma_B N(z)$ may be calculated using equation (7.4). Typical values are given in table 9.4. The values for $0.589\,\mu m$ are included to allow the Rayleigh

Table 9.4 Calculated Values of Rayleigh Backscatter Parameter

Beacon Height (km)	360 nm $\sigma_B N(z)$	400 nm $\sigma_B N(z)$	450 nm $\sigma_B N(z)$	500 nm $\sigma_B N(z)$	550 nm $\sigma_B N(z)$	589 nm $\sigma_B N(z)$	600 nm $\sigma_B N(z)$	700 nm $\sigma_B N(z)$
8	34.3	22.5	14.0	9.2	6.3	4.8	4.4	2.4
10	27.2	17.8	11.1	7.3	5.0	3.8	3.5	1.9
12	21.6	14.2	8.8	5.8	3.9	3.0	2.8	1.5
15	14.7	9.6	6.0	3.9	2.7	2.0	1.9	1.0
20	7.1	4.7	2.9	1.9	1.3	0.99	0.92	0.49
25	3.4	2.3	1.4	0.92	0.63	0.48	0.44	0.24
30	1.7	1.1	0.68	0.45	0.31	0.23	0.22	0.12
50	0.092	0.060	0.038	0.025	0.017	0.013	0.012	0.006

Units of $\sigma_B N(z)$ are photons per million meters.

scatter generated by sodium beacon lasers to be evaluated. The length of the range gate *Az* for Rayleigh scattering is typically 1 km.

Values for one-way atmospheric transmission, derived from data in chapter 7 of the *Handbook of Geophysics and Space Environments* [Valley 1965], are given in table 9.5.

The lidar equation can be expressed in the following form for use in system optimization:

$$n_p = \left[\frac{\lambda_b}{hc}\frac{\sigma_B N(z)\Delta z}{4\pi}\frac{T_A^2}{z^2}\right](E T_L T_O d^2) = F_B(E T_L T_o d^2) \quad (9.92)$$

where F_B is the beacon efficiency factor linking the number of photons received per laser pulse, n_p, to the discretionary parameters that will be optimized in the system design: the transmitted pulse energy (ET_L), the optical transmission 7b, and the area of the receiving subaperture d^2. The units of F_B are photons per joule per meter squared.

In the case of Rayleigh beacons, the wavelength and beacon height are not fixed, so the value F_B will depend on these parameters. It also depends on the zenith angle. Values of F_B for Rayleigh beacons at wavelengths between 0.360 and 0.700 μm, based on the data in tables 9.4 and 9.5, are given in table 9.6.

For sodium beacons, the mean height and wavelength are fixed by the characteristics of the sodium layer. The mean height is 92 km and its depth at half maximum is about 10 km. Because the depth of the sodium layer is limited naturally, its atom density is usually quoted as a column density, equivalent to $N(z)\Delta z$, which varies from about 3 x 10^{13} m^{-2} in the summer to about 1 x 10^{14} m^{-2} in the winter (see figure 7.8). It is reasonable to use an average value of 5 x 10^{13} m^{-2} for design purposes. The emission cross-section σ_B for sodium depends on the degree of saturation, as described in section 7.2.5. When the saturation time τ_s is longer than the natural decay time τ_n, saturation is negligible and the emission cross-section is $\sigma_B \approx 4 \times 10^{-16}$ m^2.

Values of F_B for unsaturated sodium beacons at zenith angles of 0°, 30°, and 45° are given in table 9.7.

The detection model for pulsed laser beacons is shown in figure 9.33. The main difference between this mode and continuous operation is that the photons are received in brief packets, the timing and duration of which are determined by the laser and range gate. For Rayleigh beacons, the range interval *Az* is usually about 1 km, so that with a short laser pulse (< 1 μs) the duration of each photon packet is about $2\Delta z/c \approx 7\,\mu$s. In the case of sodium beacons, the average depth of the sodium layer is about 10 km, so the full duration of the return is about 66 μs. In both cases, the photon packets are much shorter than the (millisecond) readout times of typical wavefront sensor CCDs. The number of photons received from the beacon per pulse is computed using equation (9.92).

To achieve continuous compensation of atmospheric turbulence at visible wavelengths, the laser pulse rate required is at least 1 kHz, which is compatible with the frame rate of CCD cameras. Detector readouts are synchronized to the laser pulses, with an appropriate time delay, depending on the range of the beacon. Operation is then very similar to that with a continuous reference source, using a servo bandwidth of about 1/10 of the pulse rate. The number of photons per measurement for a pulsed source is

$$n_c = n_p M_p T_W \eta \quad (9.93)$$

where

"P = number of photons received from the beacon per subaperture per pulse

M_p = number of laser pulses received within the integration time t_i

Table 9.5 One-Way Atmospheric Transmission for Clear Standard Atmosphere

Beacon Height (km)	Zenith Angle (degrees)	360 nm T_A	400 nm T_A	450 nm T_A	500 nm T_A	550 nm T_A	589 nm T_A	600 nm T_A	700 nm T_A
8	0	0.51	0.61	0.69	0.74	0.77	0.78	0.79	0.82
	30	0.46	0.57	0.65	0.70	0.74	0.75	0.76	0.80
	45	0.39	0.50	0.59	0.65	0.69	0.71	0.71	0.76
10	0	0.49	0.59	0.67	0.73	0.76	0.78	0.78	0.82
	30	0.43	0.55	0.63	0.69	0.73	0.75	0.75	0.80
	45	0.36	0.48	0.57	0.64	0.68	0.70	0.71	0.76
12	0	0.47	0.58	0.66	0.72	0.75	0.77	0.78	0.82
	30	0.42	0.53	0.62	0.68	0.72	0.74	0.75	0.79
	45	0.34	0.46	0.56	0.63	0.67	0.69	0.70	0.75
15	0	0.45	0.56	0.65	0.71	0.75	0.77	0.77	0.81
	30	0.40	0.52	0.61	0.67	0.71	0.74	0.74	0.79
	45	0.32	0.44	0.55	0.62	0.66	0.69	0.69	0.75
20	0	0.43	0.55	0.64	0.70	0.74	0.76	0.76	0.81
	30	0.38	0.50	0.60	0.66	0.70	0.73	0.73	0.79
	45	0.30	0.43	0.54	0.61	0.65	0.67	0.68	0.74
25	0	0.42	0.54	0.64	0.70	0.73	0.75	0.75	0.81
	30	0.37	0.49	0.60	0.66	0.69	0.71	0.72	0.78
	45	0.30	0.42	0.53	0.60	0.64	0.66	0.67	0.74
30	0	0.42	0.54	0.64	0.69	0.72	0.74	0.74	0.81
	30	0.37	0.49	0.59	0.66	0.69	0.70	0.71	0.78
	45	0.29	0.42	0.53	0.60	0.63	0.65	0.66	0.74
50	0	0.42	0.54	0.64	0.69	0.72	0.73	0.74	0.81
	30	0.37	0.49	0.59	0.65	0.68	0.70	0.70	0.78
	45	0.29	0.42	0.53	0.59	0.63	0.64	0.65	0.74

Detector Read Noise

During the readout of each pixel, noise is added by the charge-to-voltage conversion amplifier on the CCD chip. The root-mean-square value of this noise is specified as e electrons per pixel at a given pixel- or frame-rate. If m is the number of pixels per measurement (four in the case of a Shack-Hartmann sensor using a quadrant detector), then the mean-square read noise per measurement is me^2.

Background Noise

Background noise is added by radiation from the sky adjacent to the beacon. Laser beacons typically subtend an angle of about 2 arc seconds, requiring a relatively large field stop; but, at the UV and visible wavelengths used for Rayleigh and sodium beacons, the sky radiation is relatively low and the range gates are so short (milliseconds) that background noise is usually negligible. The effect of background noise on overall tilt sensing with field stars is much more serious, as discussed in section 9.6.3.

Taking into account all noise sources, the signal-to-noise ratio of the beacon sensor is

$$SNR = \frac{n_c}{\left\{n_c + m\left[n_b^2 + (e/G)^2\right]\right\}^{1/2}}$$

where

- n_c — number of photons per measurement, from equation (9.93)
- e = read noise per pixel in electrons rms
- n_b = number of detected background electrons per pixel
- m = number of pixels per subaperture
- G = the gain of the intensifier

For an unintensified detector, such as a bare CCD, the gain $G = 1$. When n_b is negligible, the signal-to-noise ratio using a CCD detector reduces to

$$SNR = \frac{n_c}{(n_c + me^2)^{1/2}}$$

Table 9.6 Efficiency Factors for Rayleigh Beacons

Beacon height, km	Zenith angle degrees	360 nm F_B	400 nm F_B	450 nm F_B	500 nm F_B	550 nm F_B	589 nm F_B	600 nm F_B	700 nm F_B
8	0	20.2	21.1	18.7	15.6	12.7	10.8	10.3	7.1
	30	12.3	13.6	12.5	10.6	8.8	7.5	7.2	5.0
	45	5.8	7.0	6.9	6.0	5.1	4.4	4.2	3.0
10	0	9.3	10.0	9.1	7.7	6.3	5.4	5.2	3.6
	30	5.6	6.4	6.0	5.2	4.3	37	3.6	2.5
	45	2.5	3.3	3.3	2.9	2.5	2.2	2.1	1.5
12	0	4.7	5.3	4.9	4.2	3.4	2.9	2.8	2.0
	30	2.8	3.3	3.2	2.8	2.4	2.0	2.0	1.4
	45	1.3	1.7	1.7	1.6	1.4	1.2	1.1	0.82
15	0	1.9	2.2	2.1	1.8	1.5	1.3	1.2	0.84
	30	1.1	1.4	1.3	1.2	1.0	0.87	0.83	0.59
	45	0.49	0.67	0.72	0.66	0.57	0.50	0.48	0.35
20	0	0.48	0.69	0.54	0.47	0.39	0.33	0.32	0.23
	30	0.28	0.44	0.35	0.32	0.27	0.23	0.22	0.16
	45	0.12	0.23	0.19	0.18	0.15	0.13	0.13	0.096
25	0	0.14	0.17	0.17	0.14	0.12	0.10	0.096	0.070
	30	0.082	0.11	0.11	0.096	0.080	0.068	0.066	0.049
	45	0.035	0.052	0.057	0.053	0.045	0.039	0.038	0.029

Units of F_B arc millions of photons per joule and meter squared.

9.6.3 Adaptive Optics Performance with a Sodium Beacon

Using the error model equations outlined in sections 9.6.1 and 9.6.2, the performance of an adaptive optics system using a single sodium beacon for wavefront compensation has been calculated as a function of laser power. The results are summarized in figure 9.34. The parameters of this system, listed fully in table 9.8, are believed to be typical of current technology, so the performance is representative of achievable systems. The atmospheric turbulence model used is the HV 15-12, as defined in table 3.1.

The top chart in figure 9.34 shows the wavefront compensation Strehl ratio as a function of the average laser power. It is assumed that the pulse rate remains constant at 1 kHz and that the pulse energy varies between 2 and 50 mJ. The advantage of observing at IR wavelengths is clearly apparent. At these wavelengths, reasonable performance is obtained with 10W of laser power and little advantage is gained by using more than 25 W. At 0.7 μm it is a different story, the maximum Strehl ratio being limited to about 0.2, even at high laser powers. This limitation is not caused by any single error source.

The error components at a laser power of 10 W are shown in table 9.9. No single error source is dominant, indicating that the system is well optimized. The largest error contribution at both 0.7 and 1.62 μm is the beacon measurement.

The optimum integration time and photon count per subaperture are shown in the lower charts in figure 9.34. The integration time also varies with atmospheric conditions, so that, although the laser power is fixed, it is still advantageous to optimize the integration time adaptively during operation of the system. The optimum photon count varies between about 100 and 300 per subaperture per measurement, with a value of 150 for a 10 W laser under the conditions shown.

Table 9.7 Efficiency Factors for Unsaturated Sodium Beacons

Zenith angle (degrees)	Atmospheric Transmission	F_B (photons J^{-1} m^{-1})
0	0.73	2.97 x 10^5
30	0.70	2.36 x 10^5
45	0.63	1.56 x 10^5

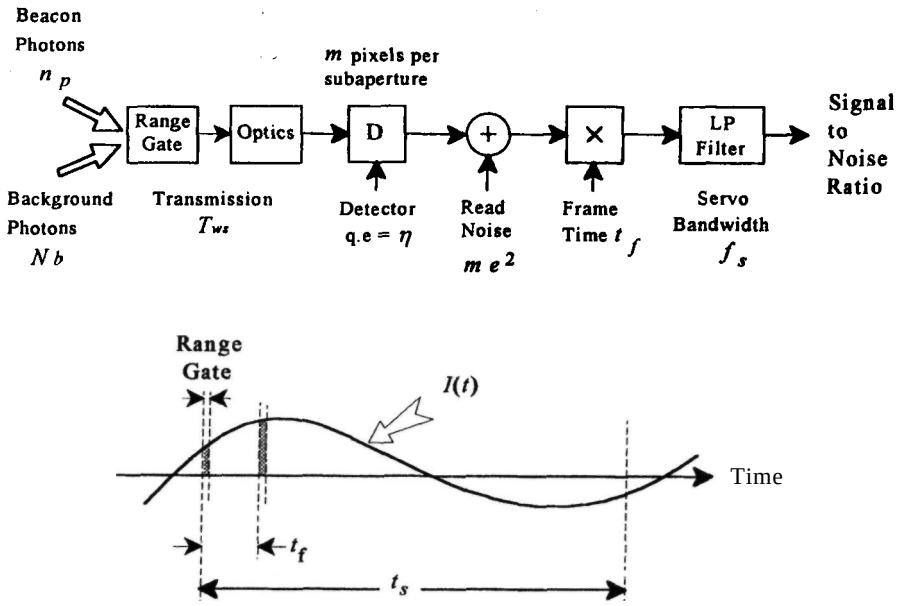


Figure 9.33 Photon detection model for a pulsed laser beacon. The detector frame rate/r = $1/t_f$ is the basic sampling frequency in the system and is usually equal to the pulse rate. The number of pulses integrated for each wavefront correction is determined by the servo bandwidth.

9.6.4 Signal-to-Noise Ratio of Tilt Sensor Using a Natural Star

The photon flux, N_s, in photons per second, received from a star at the input of the tilt sensor is given by

$$N_s = (H_\lambda \lambda_s / hc) T_A \, \Delta\lambda \, D^2 T_{ot} \qquad (9.94)$$

where

= spectral irradiance from star outside the atmosphere at wavelength λ_s, $W m^{-2} \mu m^{-1}$

λ_s = mean sensor wavelength, m

$\Delta\lambda$ = spectral bandwidth of tilt sensor, μm

h = Planck's constant, 6.626 x 10^{-34} J s

c = velocity of light, 3 x 10 ras"

T_A = atmospheric transmission at mean tilt sensor wavelength

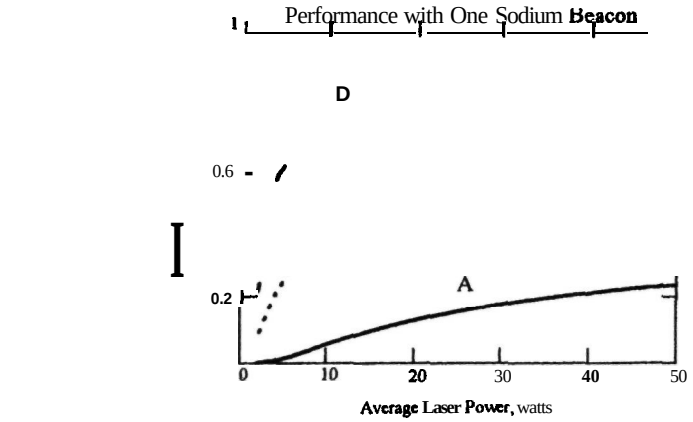
D = telescope aperture diameter at the entrance pupil, m

T_{ot} = optical transmittance from primary mirror to tilt sensor input

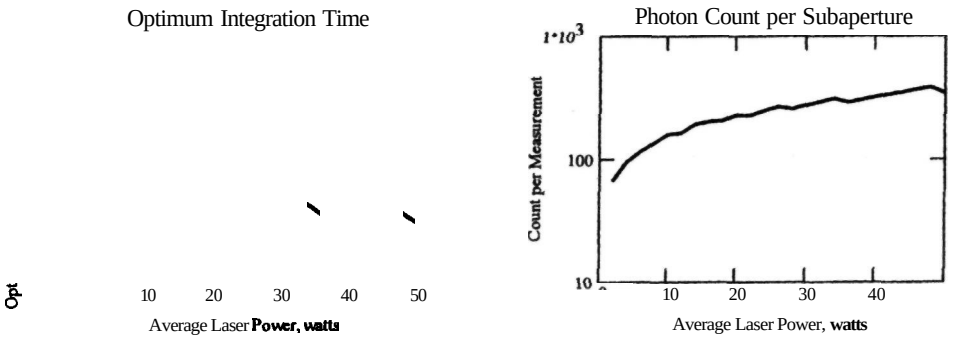
Using equation (9.73), the photon flux from a G-type star may be expressed in terms of visual magnitude m_V as

Table 9.8 Parameters of Adaptive Optics System Using a Sodium Beacon

Telescope aperture, m	4	Atmospheric transmission (one way)	0.7
Subaperture size, m	0.2	Telescope optical transmission	0.5
Turbulence parameter, m	0.15	Wavefront sensor transmission	0.8
Greenwood frequency, Hz	31.4	Laser power, W	2-50
Zenith angle, degrees	30	Pulse rate, Hz	1000
Height of sodium layer, km	92	Laser path transmission	0.5
Wavelength, μm	0.589	Detector quantum efficiency	0.8
Backscatter cross-section, m^2	4×10^{-16}	Read noise per pixel, rms	5
Sodium column density, m^{-2}	5 x 10"	Pixels per measurement	4
Beacon angular size, μrad	10	Focal anisoplanatism, d_0,m	5.02



A observation wavelength 0.70 μm
B observation wavelength 1.25 μm
C observation wavelength 1.62 μm
D observation wavelength 2 20 μm



Conditions: telescope aperture D = 4 m, subaperture size d = 0.2 m, turbulence parameter r_0 - 0.15 m at λ = 0.5 μm, zenith angle ζ = 30 degrees, Greenwood frequency f_G = 31.4 Hz, laser pulse rate = 1000 Hz, beacon angular size = 10 μrad., CCD read noise = 5 electrons per pixel.

Figure 9.34 Performance of adaptive optics using a single sodium beacon, as a function of the average laser power. The integration time (number of laser pulses) per measurement is optimized at each power level.

Table 9.9 Wavefront Error Components of an Adaptive Optics System Using a Single Sodium Beacon

Error source	Obs. Wvl. 0.7 μm		Obs. Wvl. 1.62 μm	
	Error	Strehl	Error	Strehl
Beacon measurement error σ_{BM}	1.10	0.33	0.21	0.81
Temporal error σ_{BT}	1.06	0.35	0.20	0.82
Wavefront fitting error σ_{FIT}	0.27	0.77	0.05	0.95
Cone error σ_{CONE}	0.48	0.62	0.09	0.91
Multiple beacon error σ_{MB}	0	1.00	0	1.00
Dispersion error σ_{DIS}	0	0	0.01	0.99
Totals	2.91	0.054	0.56	0.57

Obs. Wvl., observing wavelength.

$$N_s = 1.08 \times 10^{11}(10^{-0.4m_v})T_A\,\Delta\lambda\,D^2\,T_{ot} \quad (9.95)$$

The photoelectron count n_c from a continuous source is given by

$$n_c = N_s t_t T_{ST} \eta \quad (9.96)$$

where

t_t = effective integration time of the tilt servo
T_{ST} = optical transmission in the sensor to the quadrant detector
η = quantum efficiency of the photon detector

When a first-order filter with half-power (3-dB) bandwidth of $f_{3\,dB}$ is used in the tilt servo loop, it was shown in section 9.5.2 that the effective integration time is

$$t_t = 2\tau = \frac{1}{\pi f_{3\,dB}} \quad (9.97)$$

When a quadrant detector is used for sensing tilt, the readout time is negligible.

Background (Sky) Noise

Background noise can be a significant factor in tilt sensing because of the longer integration times and dim stars that are employed. To obtain good sky coverage, stars of 16th to 20th magnitude must be used, making it essential to limit the background noise to the absolute minimum. When using IR stars in the H-band (1.6 μm), sky radiation becomes the major limitation to the signal-to-noise ratio.

In a conventional tracking sensor using a quadrant detector in the image plane, the four pixels forming the quadrant define the field of view, with each pixel receiving radiation from the reference source and the sky. The overall size of the detector array must be sufficient to acquire the star with uncorrected image motion, which is considerably larger than the (short-exposure) spot size. In section 5.3.1, it was shown that for a quadrant (2 x 2) detector, the minimum pixel size is approximately equal to the spot size, but with a 4 x 4 array of the same overall size, the pixels are proportionally smaller. This suggests the possibility of using a 4 x 4 array for initial acquisition of the reference source, subsequently using only the central 2×2 elements for normal operation. Of course, the signal-to-noise ratio with the full array must be sufficient to allow the tracking loop initially to acquire the star.

If the sky spectral radiance is N_λ photons $s^{-1}\,m^{-2}$ arc $seconds^{-2}\,\mu m^{-1}$, then the detected background photon flux is

$$N_B = N_\lambda\,\Delta\lambda\,\tfrac{\pi}{4}D^2\beta^2 T_B \eta_B \quad \text{events per second} \quad (9.98)$$

where

$\Delta\lambda$ = spectral bandwidth of the tilt sensor, μm
D = telescope aperture, m
β = angular size of a detector pixel, arc seconds,
T_B = optical transmission from the primary mirror to the tilt detector
η_B = quantum efficiency of the tilt detector to background radiation

The value of N_λ depends on the mean wavelength; typical values are given in table 9.10. The number of background noise electrons is then $n_b = N_B t_t$, where t_t is the tilt sensor integration time.

The read noise of the tilt sensing array depends on the type of array. With an avalanche photodiode array operating in the pulse-counting mode, the noise should be negligible. A silicon quad cell used at visible wavelengths should also have low noise because of the small number of cells and the relatively long integration time. Infrared arrays generally have a higher noise level than those operating at visible wavelengths. The tilt sensor signal-to-noise ratio, taking into account all these noise sources, is found using equation (9.82).

9.6.5 Performance and Sky Coverage of Laser Beacon Systems

As discussed in section 9.1.3, there are several ways to achieve image stabilization with a natural guide star. The performance of three methods is determined in this section:

- uncompensated guide star;
- guide star compensated by a second beacon (dual adaptive optics);
- guide star compensated by the primary beacon (shared compensation).

The main issue when using natural guide stars is obtaining the maximum sky coverage; with a given

Table 9.10 Typical Values of Sky Radiance

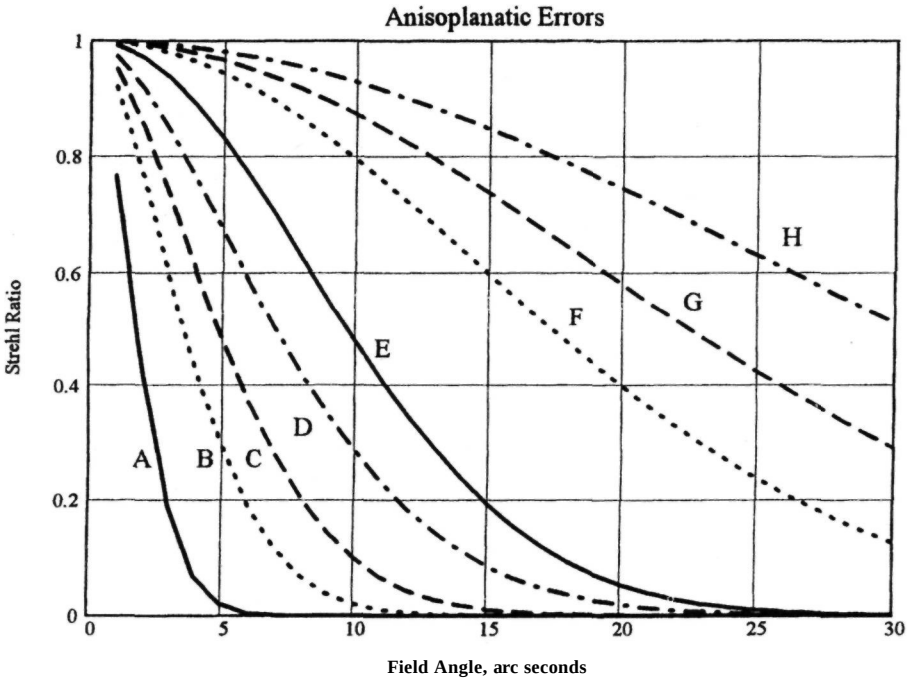
Mean Wavelength μm	Sky Radiance N_λ
0.7	360
1.25	4000
1.62	9000

population of stars, this requires maximizing the sensitivity of the tilt sensor. The angular precision of a tilt sensor such as a quad cell depends on the angular size of the reference source, which is normally enlarged by atmospheric turbulence. A considerable improvement in sky coverage is therefore obtained by compensating the image of the guide star. Natural guide stars used for tilt referencing are too faint to use in the self-referencing mode, so the most general solution is to use a separate beacon. This method, known as dual adaptive optics, further complicates what is already a complex system.

Shared compensation takes advantage of the fact that the isoplanatic angle grows as $\lambda^{6/5}$, so the compensated area surrounding the primary beacon is large at IR wavelengths. Isoplanatic errors, expressed as Strehl ratios for convenience, are shown as a function of field angle in figure 9.35. At a wavelength of $1.6\,\mu$m, the isoplanatic angle for a Strehl ratio of 0.4 is 10 arc seconds and useful compensation is obtained out to about 15 arc seconds. Within this radius from the primary beacon, much fainter (IR) stars can be used than would be possible without compensation. The sky coverage for all three of these adaptive optics operating modes is determined in this section.

Parameters of the basic tilt correction system are listed in table 9.11.

The performance of an overall tilt correction system using uncompensated natural stars is shown in figures 9.36 and 9.37 for observation wavelengths of 0.7, 1.25, and 2.2 μm. For uncompensated stars, the effective angular diameter of the reference source is determined by the ratio of the sensor wavelength to the turbulence parameter, λ_S/r_0. For the conditions in table 9.11 and for a sensor wavelength of $\lambda_S = 0.7\,\mu$m, the reference source diameter is 0.7 arc second. The method of calculating the sky cover-



A wavefront error, 0.70 μm
B wavefront error, 1.25 μm
C wavefront error, 1.62 μm
D wavefront error, 2.20 μm
E overall tilt error, 0.70 μm
F overall tilt error, 1.25 μm
G overall tilt error, 1.62 μm
H overall tilt error, 2.20 μm

Figure 9.35 Anisoplanatic errors for wavefront and tilt expressed as Strehl ratios, as a function of field angle at four observation wavelengths. The turbulence model is Mauna Kea (MK) average, for which $r_0 = 0.24$ m and θ_0 — 1.9 arc seconds at $\lambda = 0.500\,\mu$m. The zenith angle is 30°. The anisoplanatic tilt error is weakly dependent on the telescope aperture, which is 4 m in this example.

Table 9.11 Parameters of Tilt Compensation System Using Natural Stars

Turbulence model: Mauna Kea (MK) average (with turbulence peak)

Telescope aperture, m	4	Atmospheric transmission	0.7
Turbulence parameter, m @ 0.5	0.24	Telescope optical transmission	0.5
Tyler frequency, Hz @ 0.5	3.8	Wavefront sensor transmission	0.8
Zenith angle, degrees	30	Detector quantum efficiency	0.8
Sensor wavelengths, μm	0.70	Read noise per pixel, rms @ 0.7	2
	1.60	@ 1.6	5
Sensor bandwidth, μm	0.3	Pixels per measurement	4
Sky photons, $s^{-1}\mu m^{-1}$ arcsec^{-2}m^{-2} @ 0.7	360		
@ 1.6	9000		

age has been described in section 9.5.5, using an average star distribution model.

The sky coverage of this tilt compensation system depends on both the observation wavelength and the required Strehl ratio. Figure 9.36 shows that for an observation wavelength of 0.7 μm, stars of about m_V = 16 are required to achieve a tilt stabilization Strehl ratio of 0.5, limiting the coverage to about 1%, while at an observation wavelength of 2.2 μm, the same Strehl ratio can be achieved with stars of about m_V = 18, increasing the coverage to nearly 100%. The allowable field angles over which the required tilt compensation can be obtained are shown in figure 9.37. The sharp cutoff in field angle occurs at the star magnitude for which the tilt detection errors become equivalent to the specified Strehl ratio, reducing the allowable isoplanatic error to zero. It is this critical star magnitude that determines the sky coverage.

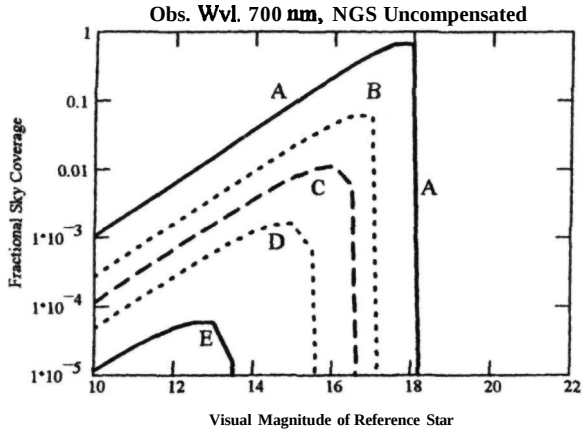
If the wavefront received from the tilt reference star is compensated (for example, by using a second laser beacon), then the effective angular size of the reference source shrinks to its diffraction-limited size of approximately λ_S/D. As explained in section 4.4, the core of the image retains this size even with a mean-square residual wavefront error as large as 2 rad^2. Therefore, it may be sufficient to compensate the tilt reference source only partially, although this needs careful evaluation as the presence of the halo in a partially compensated image reduces its contrast, and therefore decreases the signal-to-noise ratio of the wavefront sensor.

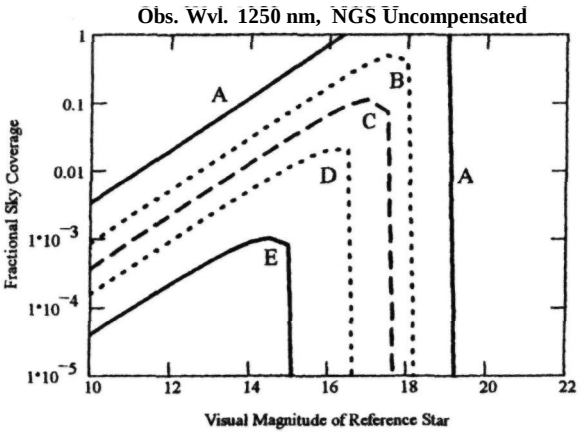
The performance of a tilt sensor operating at 0.7 μm with a compensated source is shown in figure 9.38. Wavefront compensation enables the tilt sensor to operate with much fainter sources, considerably improving the sky coverage. At an observation wavelength of 0.7 μm, the limiting magnitude for a tilt stabilization Strehl ratio of 0.5 is m_V = 19.5, giving about 30% sky coverage. For the same Strehl ratio, the sky coverage grows to 50% at 1.25 μm and is complete at wavelengths over 1.6 μm. Full compensation of the tilt reference star increases the sky coverage by a factor of about 30, compared with an uncompensated reference. The allowable field angles obtained for various (tilt) Strehl ratios are shown as a function of star magnitude in figure 9.39. The effect of compensation is to increase the limiting magnitudes for each field angle; the field angles themselves are not increased because they are determined by the anisoplanatic errors.

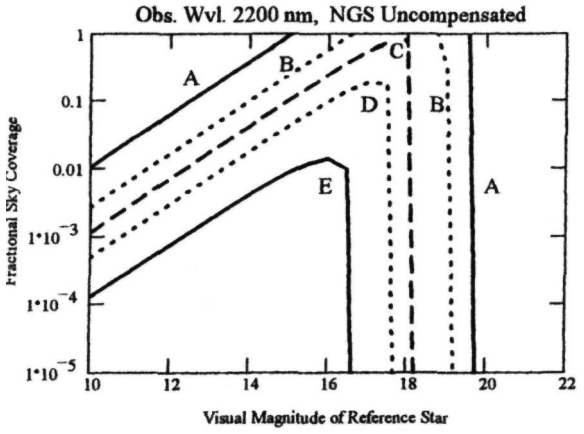
The sky coverage obtained with the third mode of tilt referencing, employing an IR star that is compensated by the primary beacon, is shown in figure 9.40. The use of an IR wavefront sensor reduces performance in two ways':

- The detector noise is usually greater than at visible wavelengths, reducing the signal-to-noise ratio.
- Sky background radiation is much higher, making it necessary to use the smallest possible pixels.

The second factor is probably the most critical. The size of the (diffraction-limited) image core in this case is about 0.01 arc second. The sky coverage chart in figure 9.40 assumes a tilt sensor pixel size of 0.1 arc second, which, on paper, gives an adequate dynamic range once the tilt loop is closed. However, in order to "capture" the tilt reference star with pixels of this size, the telescope must be pointed to within ±0.1 arc second of its true position. This operation would be facilitated by using a larger number of detector pixels for initial acquisition and discarding them once the tracker has locked on to the target, in order to reduce detector and background noise. The allowable field angles for this mode of operation are shown in figure 9.41. The performance using shared compensation appears to fall in between that of the other two, giving a useful increase in sky coverage with minimal increase in cost and complexity.





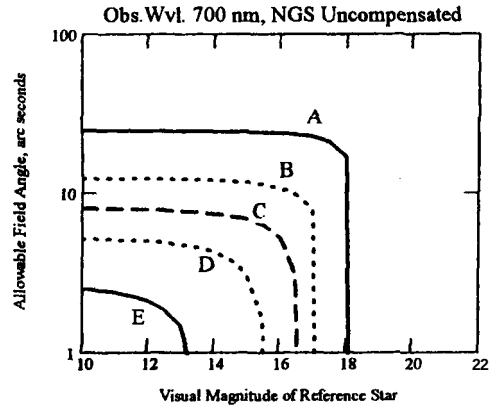


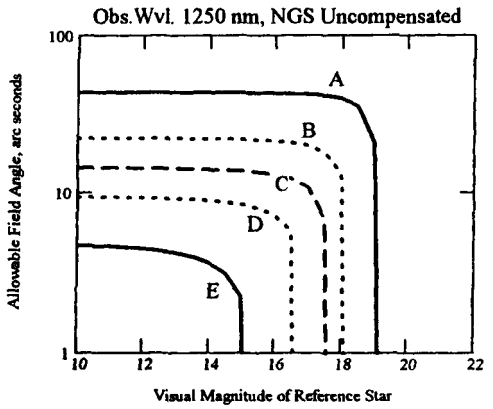
Conditions:

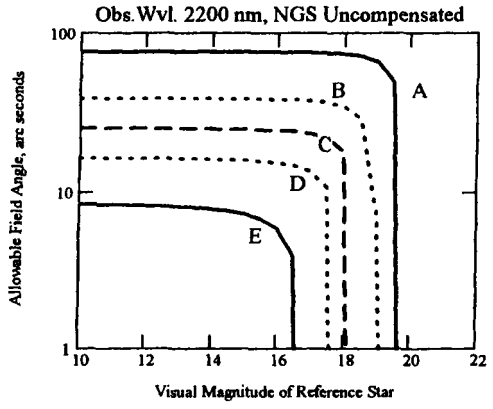
Telescope aperture = 4 m
Turbulence model = MK average
Turbulence r_0 = 0.24 m
Tyler frequency = 3.8 Hz
Zenith angle = 30 degrees
Sensor wavelength = 0.7 μm
Sensor spectral band = 0.3 μm
Sky spectral radiance
= 360 photons s^{-1} m^{-2} arcsec^{-2} μm^{-1}
Effective optical aperture = 0.33 m
Pixel angular size — 1 arcsec
Detector quantum efficiency = 0.8
Read noise per pixel = 2 electrons

A Strehl ratio = 0.1
B Strehl ratio = 0.3
C Strehl ratio = 0.5
D Strehl ratio = 0.7
E Strehl ratio = 0.9

Figure 9.36 Sky coverage for tilt referencing with uncompensated natural stars, at observation wavelengths of 0.7, 1.25, and 2.2 μm. The integration time is optimized for each star magnitude. The Strehl ratio is the contribution of tilt stabilization only.







A Strehl ratio = 0.1
B Strehl ratio = 0.3
C Strehl ratio = 0.5
D Strehl ratio = 0.7
E Strehl ratio = 0.9

Figure 9.37 Allowable field angles for the tilt reference conditions shown in figure 9.36, using an uncompensated natural star at a tilt sensor wavelength of $0.7\mu m$. The sharp cutoff in field angle occurs when the tilt detector errors become equal to the specified Strehl ratio, with the result that the allowable anisoplanatic error drops to zero.

9.7 Adaptive Parameter Control

9.7.1 The Environment of Adaptive Optics

Adaptive optics systems operate in a complex environment dominated by natural phenomena whose characteristics cannot be controlled. The basic design of these systems can be optimized for statistically average observing conditions and requirements, but the inevitable variations in these statistics have a large effect on system performance, often within short time periods. Changes in the structure and strength of atmospheric turbulence, defined by the refractive index structure parameter $C_N^2(h)$, and changes in the wind profile $v(h)$ combine to produce random variations in the normalized parameters r_0, θ_0, and f_G. Other factors that may vary during an observing session are the atmospheric transmission T_A and the brightness and apparent size of the reference source, especially if it is a laser beacon.

The effect of these short-term variations can be minimized by adjusting the parameters of an adaptive optics system in real time, without interrupting its operation. Some possible approaches are outlined in the following sections.

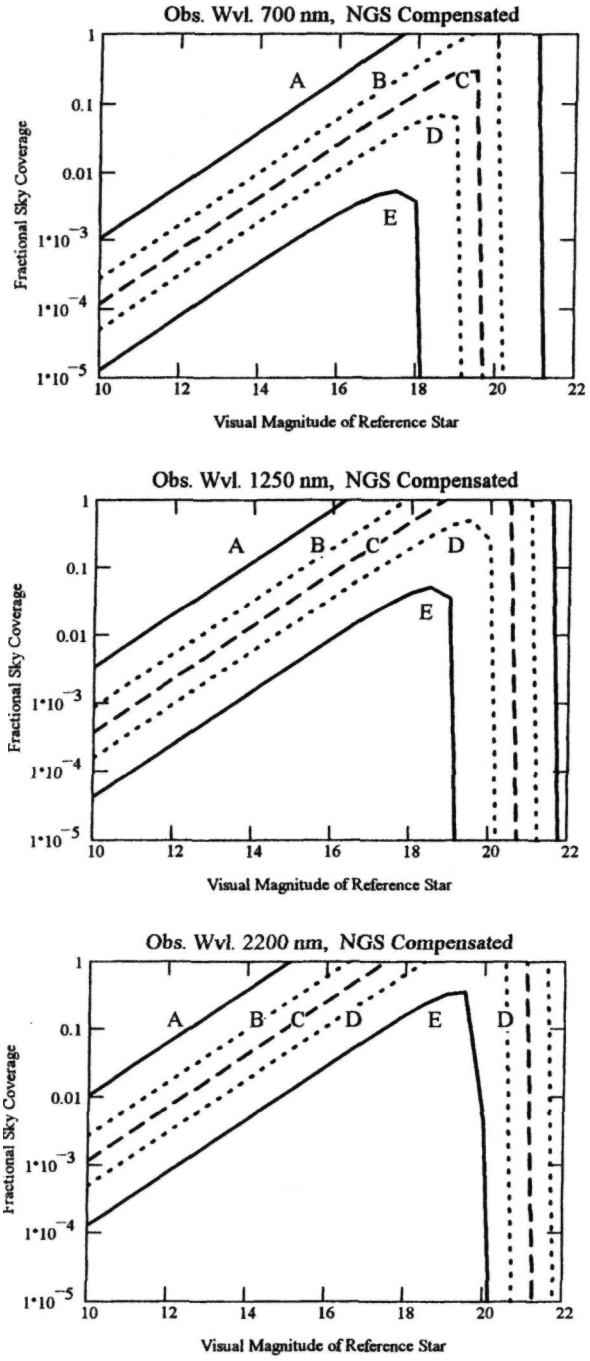
9.7.2 Parameters To Be Controlled

Adaptive optics parameters that may be optimized during operation include:

- photon integration time t, or control loop bandwidth f_S;
- wavefront sensor sensitivity or dynamic range;
- reconstruction matrix weights.

The photon integration time is the most important of these and its effect on overall performance has been discussed in section 9.5.3. It is simple enough to calculate the optimum integration time when modeling system performance, but a different approach is needed to optimize a working system in real time. As adaptive optics systems generally have separate control loops for high-order wavefront correction and overall tilt correction, the time constant of each loop must be optimized.

In practical wavefront sensors, there is a tradeoff between sensitivity and dynamic range. In the case of a slope sensor, the sensitivity or transfer function is defined in terms of volts per radian of wavefront slope; the maximum output voltage is always limited, so, as the sensitivity is increased, the dynamic range inevitably decreases. In the case of Hartmann sensors, the sensitivity is not easily changed because it is built into the hardware. With shearing interferometers, the sensitivity may be controlled over a large range by changing the shear, which can be done in real time without interrupting operation. The sensi-



Conditions:

Telescope aperture = 4 m
Turbulence model = MK average
Turbulence r_0 = 0.24 m
Tyler frequency = 3.8 Hz
Zenith angle = 30 degrees
Sensor wavelength = 0.7 μm
Sensor spectral band = 0.3 μm
Sky spectral radiance
= 360 photons s^{-1} m^{-2} arcsec^{-2} μm^{-1}
Effective optical aperture = 4 m
Pixel angular size = 1 arcsec
Detector quantum efficiency = 0.8
Read noise per pixel = 2 electrons

A Strehl ratio = 0.1
B Strehl ratio = 0.3
C Strehl ratio = 0.5
D Strehl ratio = 0.7
E Strehl ratio = 0.9

Figure 9.38 Sky coverage for tilt referencing with compensated natural stars, at observation wavelengths of 0.7, 1.25, and 2.2 μm. The integration time is optimized for each star magnitude. The Strehl ratio is the contribution of tilt stabilization only.

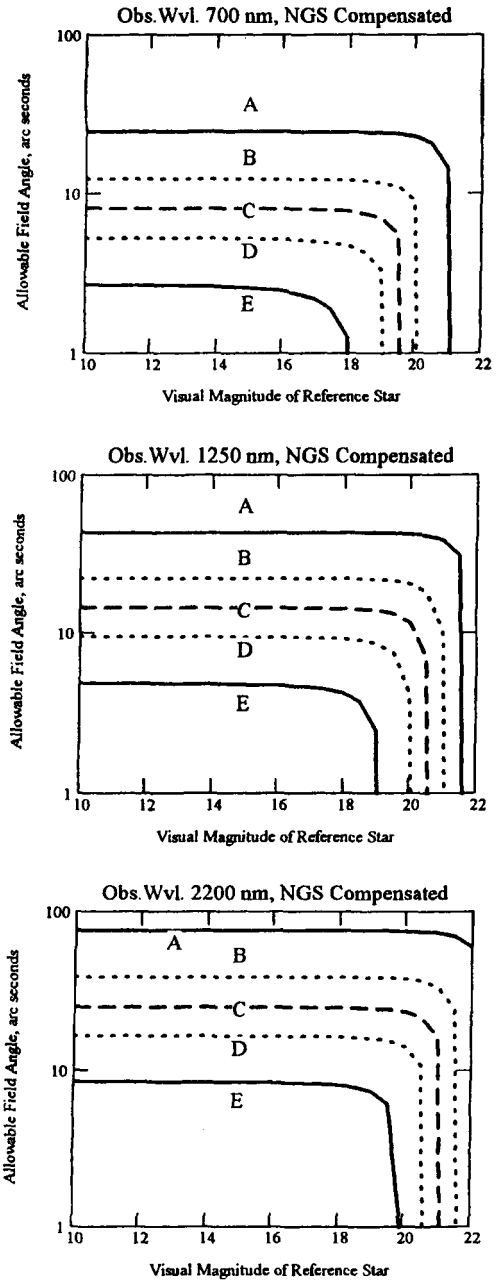


Figure 9.39 Allowable field angles for the tilt reference conditions shown in figure 9.38, using a compensated natural star at a tilt sensor wavelength of $0.7\mu m$. The sharp cutoff in field angle occurs when the tilt detector errors become equal to the specified Strehl ratio, with the result that the allowable anisoplanatic error drops to zero.

tivity/dynamic range tradeoff becomes an issue in two operational situations:

1. During initial acquisition, when the residual wavefront error is large. Convergence of the control loop is facilitated when the initially large error is within the linear range of the wavefront sensor. Once the system has converged and the residual error is small, the sensitivity of the wavefront sensor may be increased to improve the precision of the correction.
2. When a dropout occurs because of a sudden increase in turbulence, causing the system to lose track of the reference source. To regain control, it is necessary to repeat the acquisition procedure, which is facilitated by increasing the dynamic range.

It should be noted that the sensitivity/dynamic range tradeoff is determined by the design of the wavefront sensor itself and is not equivalent to changing the gain of the feedback loop.

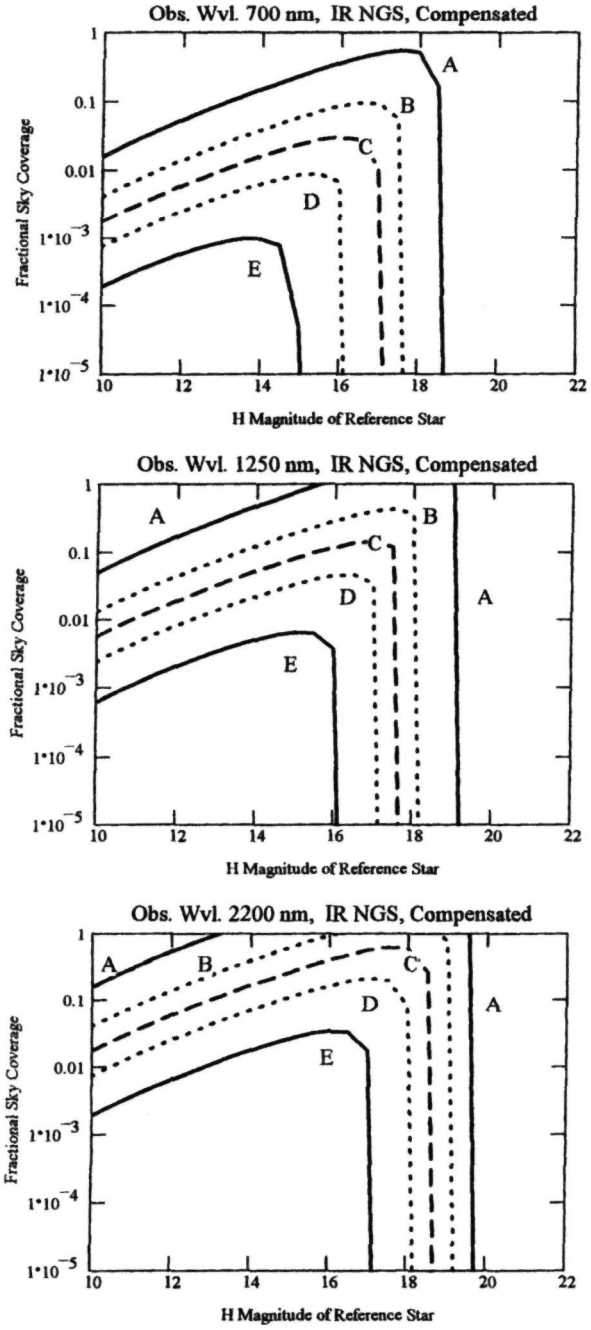
Optimization of the reconstruction matrix weights takes account of many variable factors in an adaptive optics system. In the optimal wavefront correction approach to system design, described in section 8.6, the reconstruction matrix contains essentially all of the parameters required to optimize the system; that is, it defines the adaptive feedback loop. When the reconstructor is viewed as one element of the control system, with a more limited function, it can still have a considerable effect on system performance. In particular, the contributions of individual subapertures to the overall wavefront measurement can be adjusted according to their individual signal-to-noise ratios. Possible methods of implementing these controls are discussed in the following section.

9.7.3 *Control Methods and Algorithms*

Adaptive parameter control may be considered as a high-level control loop superimposed on the basic wavefront correction system. The function of this high-level control is to compensate for changes in the statistics of the turbulence. Although a real-time response is required, the time scale of these statistical variations is two or three orders of magnitude longer than that of the wavefront correction itself. Statistical changes occur over periods of seconds, rather than milliseconds, so the functions of the loops are easily separated. Viewed in this light, adaptive parameter control may be implemented as a closed-loop control system, in which case it is necessary to find measurable parameters that are sensitive to the quality of the compensated image.

Performance measures that may be used to monitor the operation of an adaptive optics system and to maximize its performance in real time include:

residual wavefront error,



Conditions:

Telescope aperture = 4 m
Turbulence model = MK average
Turbulence r_0 = 0.24 m
Tyler frequency = 3.8 Hz
Zenith angle = 30 degrees
Sensor wavelength = 1.6 μm
Sensor spectral band = 1 μm
Sky spectral radiance
= 9000 ph. s^{-1} m^{-2} $arcsec^{-2}$ /an^{-1}
Effective optical aperture = 4 m
Pixel angular size = 0.1 arcsec
Detector quantum efficiency — 0.8
Read noise per pixel = 5 electrons

A Strehl ratio = 0.1
B Strehl ratio = 0.3
C Strehl ratio = 0.5
D Strehl ratio = 0.7
E Strehl ratio = 0.9

Figure 9.40 Sky coverage for tilt referencing with a fully compensated IR star at 1.6 μm. The compensation may be produced by laser beacon adaptive optics if the star is within its isoplanatic angle at 1.6 μm. The integration time is optimized for each star magnitude. The Strehl ratio is the contribution of tilt stabilization only.

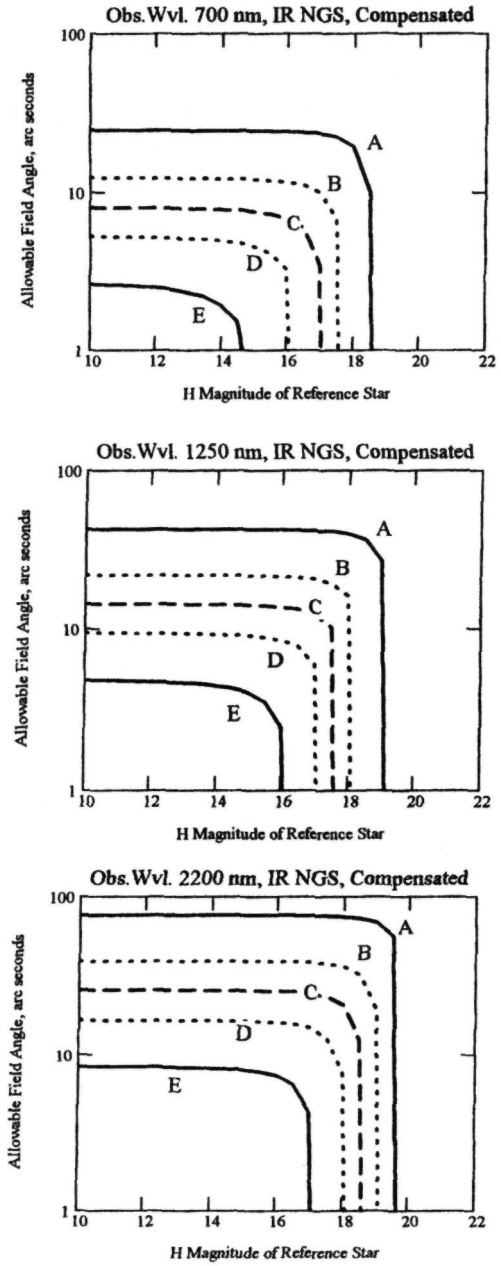


Figure 9.41 Allowable field angles for the tilt reference conditions shown in figure 9.40, using a compensated IR star at a tilt sensor wavelength of 1.6 μm. The sharp cutoff in field angle occurs when the tilt detector errors become equal to the specified Strehl ratio, with the result that the allowable anisoplanatic error drops to zero.

- signal-to-noise ratio in feedback control loops,
- frequency spectrum of wavefront sensor output,
- Strehl ratio of compensated image,
- wavefront corrector (actuator) drive voltages.

Closed-loop implementation involves monitoring one or more of these parameters and computing the required correction to be applied to the integration time, wavefront sensor sensitivity, or reconstruction weights. A variation of this approach is to use trial and error, systematically changing the controlled quantity by a small amount and observing its effect on the residual error. If the performance improves, the change is kept; if not, a change of the opposite polarity is applied. Obviously, it is preferable to look for the cause rather than the effect of performance changes, by measuring the most fundamental parameters. In the case of photon integration time, this is straightforward.

The optimum integration time is determined by minimizing the sum of the wavefront measurement error σ_M and the temporal error σ_T^2. By using simplified expressions for these errors, it is possible to find the explicit dependence of the integration time t on the photon flux and atmospheric parameters. If other noise sources are ignored, the signal-to-noise ratio of the wavefront sensor is equal to the square root of the number of photons counted, or $SNR \sim \sqrt{(N_S t)}$, where N_S is the photon counting rate per subaperture (events per second) and t is the integration time. In astronomical adaptive optics, the subaperture size d is usually greater than the turbulence parameter r_0; using equation (9.43), the mean-square wavefront measurement error for an unresolved reference source may be expressed as

$$\sigma_M^2 = \frac{K}{SNR^2 r_0^2} = \frac{K}{N_S t\, r_0^2} \tag{9.99}$$

The mean-square temporal error [using equation (9.53)] is given by

$$\sigma_T^2 = \left(\frac{f_G}{f_S}\right)^{5/3} = (\pi\, t f_G)^{5/3} \tag{9.100}$$

where f_G is the Greenwood frequency of the turbulence and $f_S = 1/nt$ is the servo bandwidth. The sum of these errors is

$$\varepsilon = \sigma_M^2 + \sigma_T^2 = \frac{K}{N_S r_0^2} t^{-1} + (\pi f_G)^{5/3} t^{5/3} \tag{9.101}$$

The value of / that minimizes the error is found by differentiating e with respect to t and setting the result to zero:

$$\frac{d\varepsilon}{dt} = -\frac{K}{N_S r_0^2} t^{-2} + \frac{5}{3}(\pi f_G)^{5/3} t^{2/3} = 0 \tag{9.102}$$

The optimum integration time is then

$$t_{opt} = \left[\frac{K}{N_S r_0^2 f_G^{5/3}}\right]^{3/8} \tag{9.103}$$

where K is a system constant. The photon counting rate N_S is easily obtained from the detector outputs, and the value of the turbulence bandwidth f_G may be deduced from the temporal frequency spectrum of the wavefront tilt measurements produced by the wavefront sensor. The turbulence parameter r_0 may be found by monitoring the mean-square value of the drive signals to the wavefront corrector. In this way, the integration time can be updated in real time to optimize the system performance.

A similar process may be used to optimize the wavefront sensor sensitivity. Using equation (9.44), the wavefront error of a shearing interferometer may be expressed as

$$\sigma_{MS}^2 = \frac{K_S}{(SNR)^2 \gamma^2 s^2} \tag{9.104}$$

where

γ - fringe contrast (modulation)
s = shear distance

The sensitivity is controlled by changing the shear distance, which is normally a fraction of the subaperture size d. However, changing the shear distance also changes γ, so to minimize the measurement error, the product γs must be maximized. The relation between γ and s is given in equation (9.49), which for $d \gg s$ may be approximated as

$$\gamma^2 \approx 1 - \left(\frac{Ks^{4/3}}{r_0^{5/3}}\right) \tag{9.105}$$

The maximum value of $\gamma^2 s^2$ is obtained when the shear $s \approx Kr_0^{3/}$. In other words, the optimal shear distance of the wavefront sensor depends on r_0, as expected. The control algorithm to optimize shear is therefore to monitor r_0 by measuring the mean-square value of the drive signals to the wavefront corrector and to make adjustments proportional to "o •

9.8 Multiconjugate Compensation

9.8.7 *Introduction*

A wavefront corrector such as a deformable mirror provides optical pathlength correction at a single plane, whereas atmospheric turbulence is a three-dimensional phenomenon involving a continuous distribution on which several thin layers of high turbulence may be superimposed. When a single corrector is employed, the total turbulence in the direct path between the reference source and the telescope is compensated perfectly, but the field angle over which this correction is effective is limited by anisoplanatism and is consequently extremely small, typically a few arc seconds, especially when strong turbulence is present at high altitudes. The corrected field of view may be extended by using multiconjugate compensation, in which multiple wavefront correctors are optically conjugated to turbulent layers in the atmosphere. It is of considerable interest to find the relation between the number of correctors and the compensated field angle; this relation depends on the strength and distribution of the turbulence, as well as the observing wavelength.

Two separate problems are encountered when implementing multiconjugate compensation:

- measurement of the turbulence in three dimensions;
- optimal placement and control of a limited number of wavefront correctors.

The first problem can be solved, in principle, by using multiple reference sources to probe the turbulence distribution over the required field of view. A proposed method of measuring multiple turbulent layers using an array of laser beacons is described in section 9.8.2. Even when the turbulence distribution is known, there remains the problem of optimally locating the correctors in the optical path, as well as computing the drive signals for each actuator. An approach to determining the optimal corrector locations is summarized in section 9.8.3.

Wide-field compensation remains one of the unsolved problems in adaptive optics. A practical goal would be to achieve good compensation of low-contrast detail at visible wavelengths over a field of 1 minute of arc, which would encompass the entire disks of the planets Jupiter or Saturn.

9.8.2 *Multiconjugate Wavefront Measurement*

The measurement of distributed turbulence and methods of controlling multiple correctors are difficult problems that have received little attention in the literature. Dicke [1975] suggested the idea of separating high- and low-level turbulence on the basis of frequency spectra due to wind speed. More recently, Johnston and Welsh [1994] have described a method of measuring two widely separated turbulent layers using an array of four laser beacons and a natural guide star. (Their analysis ignores the conjugate image sequencing problem noted in section 2.5.) Johnston and Welsh find that with a model atmosphere in which 90% of the turbulence is at 1 % of the laser beacon altitude, and the remaining 10% of the turbulence is at 10% of the beacon altitude, a two-level correction system enlarges the corrected field angle by a factor of about 3 compared with a single-corrector system, for the same isoplanatic error. They also find that with the assumed turbulence distribution, in which there is a large separation between the two layers,

the anisoplanatic angle is not sensitive to the location of a single corrector. The only case in which the position of a single corrector can increase the isoplanatic angle materially is when all turbulence is concentrated into a single layer.

These results are consistent with the analysis of Wallner [1994], described in section 9.8.3, in which little improvement in isoplanatic angle was obtained with a typical distributed turbulence model by moving the corrector from the telescope pupil up to the mean turbulence height. To obtain a significant gain in isoplanatic angle, at least two correctors are necessary; this requires multiple reference sources to provide the wavefront measurements.

9.8.3 Optimal Corrector Locations

The multiconjugate compensation system is modeled as a number of discrete wavefront correctors, each optically conjugate to a specific height h_i, as shown in figure 9.42(a). Each corrector compensates the turbulence within a discrete interval of the optical path, between heights H_i and H_{i-1} . The problem to be solved may be stated as follows: Given an atmospheric turbulence distribution $C_N^2(h)$ and a total number of correctors m, where should the correctors be placed and what is the resulting isoplanatic error?

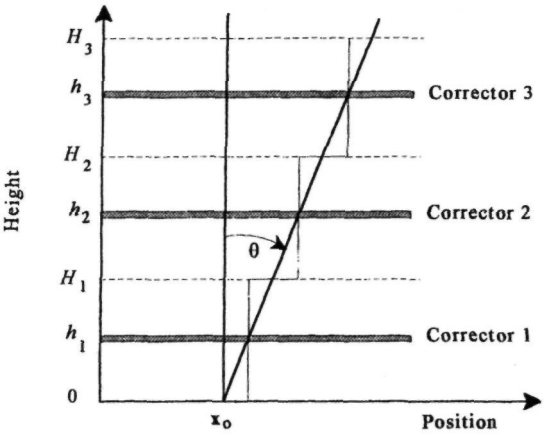
It will be assumed that each corrector is capable of providing perfect compensation for its assigned layer and that the beam path is near vertical.

The model for a single corrector is depicted in figure 9.42(b). The optical path difference measured along vertical ray paths such as AB, between layer boundaries H_i H_{i-1} is complemented by an equal and opposite path difference in the corrector located at height h_i. A ray at an angle *0* to the vertical is displaced at the corrector location by a distance *Oh-,*. The anisoplanatic error for this section is the difference in optical path lengths between the vertical ray paths AB and CD, where the latter is displaced by *Oh,*. The "vertical" path AB, in this case, is simply the path measured by the multiconjugate wavefront sensor, assuming perfect measurement and implementation of the turbulence in each layer, while the path CD is equivalent to that followed by the ray at angle *0*.

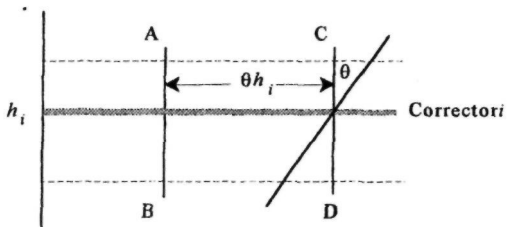
Wallner [1994] has shown that, for ι layers, the optical path error for a ray at position *x* in the aperture may be expressed as

$$\varepsilon = \sum_i \int_{H_{i-1}}^{H_i} [n(\mathbf{x} + \theta h_i, h) - n(\mathbf{x} + \theta h, h)]dh \quad (9.106)$$

where n(x, h) is the index of refraction at position x, height *h*. Assuming Kolmogorov turbulence, the



(a) Multilayer Mode)



(b) One Layer

Figure 9.42 Model for multiconjugate compensation.

mean-square anisoplanatic error is then

$$\sigma_\varepsilon^2 = 2.914 \sum_i \int_{H_{i-1}}^{H_i} C_N^2(h)|\theta h_i - \theta h|^{5/3} dh$$

$$= 2.914|\theta|^{5/3} \sum_i \left[\int_{H_{i-1}}^{h_i} C_N^2(h)|h_i - h|^{5/3} dh + \int_{h_i}^{H_i} C_N^2(h)|h - h_i|^{5/3} dh \right] \quad (9.107)$$

The conditions that minimize the mean-square error are found by setting the appropriate derivatives to zero. The condition for optimum corrector heights h_i is

$$\int_{H_{i-1}}^{h_i} C_N^2(h)|h_i - h|^{2/3} dh = \int_{h_i}^{H_i} C_N^2(h)|h - h_i|^{2/3} dh \quad (9.108)$$

The solution to this equation depends on the turbu-

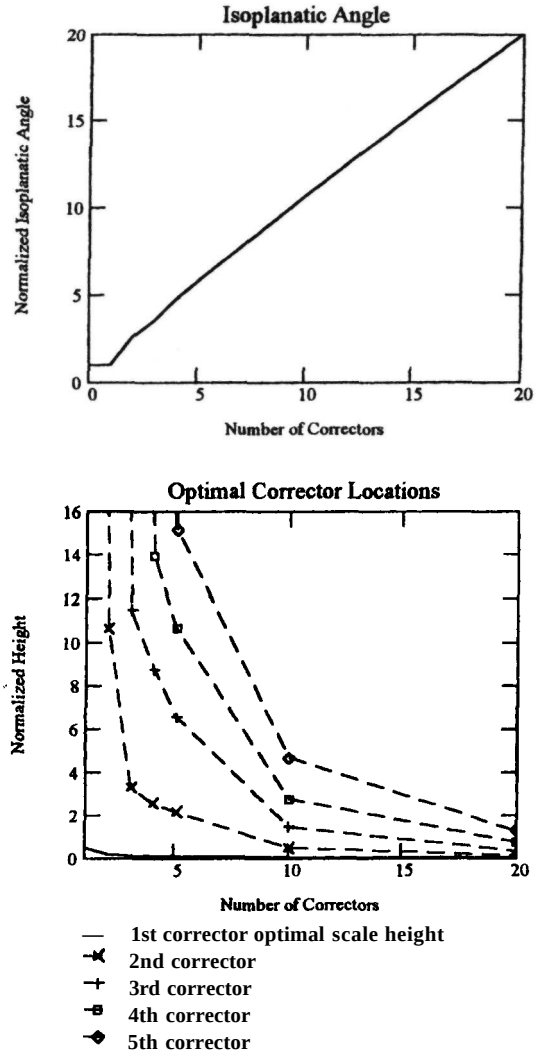


Figure 9.43 Multiconjugate compensation: improvement in isoplanatic angle and optimal corrector scale heights for the HV 5-7 turbulence model, as a function of the number of correctors employed [Wallner 1994].

lence distribution C_N^2 . The condition for optimum section heights H_i is

$$H_i = \frac{h_i + h_{i+1}}{2} \tag{9.109}$$

The section boundaries are therefore midway between the corrector locations.

For uniform turbulence from zero to distance H_m, such as that encountered in a horizontal path, the optimal corrector locations are

$$h_i = \frac{H_i + H_{i+1}}{2} \tag{9.110}$$

which specifies uniform spacing of correctors, the expected result. For the Hufnagel-Valley turbulence profile, defined in equation (3.18), the optimal corrector heights h_i and the resulting isoplanatic angles θ_0 are shown in figure 9.43 as a function of the number of correctors. In this chart, $m = 0$ represents a single corrector at the telescope aperture, while $m = 1$ represents a single corrector at the optimal height. The HV 5-7 turbulence model gives $r_0 = 0.05$ m and $\theta_0 = 1\,\mu$rad; it contains both low-level and high-level turbulence. With this profile, placing a single corrector at the optimum height makes little difference, but two correctors improve the isoplanatic angle by a factor of 2.58. Beyond this point, the isoplanatic angle improves in direct proportion to the number of correctors, up to at least 20. The area of sky over which compensation is effective therefore increases approximately as the square of the number of correctors. Even two correctors, optimally located, produce a worthwhile improvement of over six times in the compensated area.

Some of the practical issues involved in the use of multiple compensators have been discussed in section 2.5.5, where it was pointed out that the conjugate images produced by an optical relay do not naturally occur in the correct sequence for wavefront compensation. Optical images are formed in the same sequence as the objects, so that the conjugate image of the nearest layer is furthest from the telescope objective. For multiconjugate compensation, it is necessary to correct the layers in reverse order, starting with the layer nearest to the telescope, thereby requiring that its corrector also be nearest to the telescope. An optical relay system is required to rearrange the conjugate images of the layers in the correct sequence. Another problem is that the planes conjugate to high-altitude layers are close to the reference image plane, where the beam diameter is much smaller than that of conventional deformable mirrors. Optical configurations for three- and two-level correction that solve these problems have been shown in figures 2.17 and 2.18, respectively. To extend the compensated field of view therefore requires not only multiple correctors and their drive systems, but also involves the addition of several optical elements in the telescope imaging path to control the positions of the conjugate images.

10 Astronomical Adaptive Optics Programs

10.1 Introduction

This chapter contains a capsule summary of the organization, goals, and achievements of the major research programs in astronomical adaptive optics around the world, as they existed at the end of 1997. Most of these efforts are associated with specific telescopes. The technology and applications of adaptive optics are rapidly changing, so the summaries presented here should be regarded as a starting point for further investigation. Most organizations have sites on the World Wide Web, from which the latest information may be obtained. In the final sections of this chapter, current trends in adaptive optics are summarized and prospects for the future are discussed.

Some perspective on the current activity in adaptive optics can be obtained by considering its relation to the present boom in telescope building, which is certainly the biggest in history. This activity comes after a long period of stagnation, from about 1950 to 1980, in which telescope technology reached a plateau, with maximum apertures of about 5 m. This is the largest practical diameter for conventional "stiff" mirrors made of glass, which have a diameter-to-thickness ratio of between 6:1 and 8:1. Above this size, the weight increases so rapidly that the sag due to gravity becomes unacceptable. During this period, the main advances came from better detectors: notably solid-state detectors, such as charge-coupled devices (CCDs), with high efficiency and broad spectral range.

The quantum efficiency of detectors at visible wavelengths is now approaching 100% and that of infrared (IR) detectors, although lower, is rapidly improving. Furthermore, the development of adaptive optics to compensate atmospheric turbulence has enabled diffraction-limited performance to be obtained from large ground-based telescopes at near-IR wavelengths; it is expected that this capability will gradually be extended down to visible wavelengths.

With these advances in detector technology, the only way of further improving the information-gathering power of astronomical telescopes is to increase the collecting area and/or the overall size of the aperture, using large telescopes or arrays of telescopes. To meet this need, three new approaches to the fabrication of primary mirrors have been developed. These are all represented in the current group of large telescopes for which adaptive optics is planned.

The new methods of primary-mirror fabrication are:

1. Segmented mirrors, in which a number of separately fabricated panels, usually hexagonal, are assembled into a phase-continuous surface. The Keck I and II primary mirrors are examples of this method, each being composed of 36 independently supported panels, spanning a diameter of 10 ms.
2. Meniscus mirrors, which are thin slabs of quartz or ceramic, with a diameter-to-thickness ratio of about 40:1. They are supported by several hundred actuators using active control to maintain the optical figure. The four 8.2-m European Southern Observatory/Very large

Telescope (ESO VLT) mirrors are of this type, as are the two Gemini 8-m mirrors and the 8.2-m primary mirror for the Subaru telescope of the National Astonomical Observatory of Japan (NAOJ).

3. Honeycomb sandwich mirrors, which are cast using borosilicate glass, a material similar to Pyrex, with a low melting point. This process was developed at the Steward Observatory of the University of Arizona [Angel, Davison et al. 1990]. These mirrors have approximately the same stiffness as a solid piece of glass but are only one-quarter of the weight. Because of their thin structure, they quickly come into thermal equilibrium with the ambient air. Honeycomb mirrors are used in the 6.5-m Multi Mirror Telescope (MMT) conversion at Mt Hopkins and the 6.5-m Magellan Telescope at Las Campagnas, Chile. Two 8.4 m mirrors of this type are used in the Large Binocular Telescope (LBT).

Two factors determine the information-gathering capability of an astronomical telescope system: its photon sensitivity and its angular resolution. For a given detector quantum efficiency, the photon sensitivity is proportional to the area of the aperture. Given diffraction-limited operation, the angular resolution is determined by the maximum dimension of the aperture and the wavelength. In a conventional telescope with a circular aperture, the diameter and area have a fixed relationship, so the sensitivity and resolution are not independent.

However, by combining the beams from two or more separate but otherwise identical telescopes, it is possible to improve the angular resolution without increasing the basic telescope size. Interferometric imaging appears to be the most economical way of increasing the information-gathering power of large telescopes without incurring the enormous cost of making even bigger primary mirrors. This approach has, in fact, been taken with three of the largest instruments: the ESO VLT (consisting of four 8.2-m units), the Keck telescopes (two 10-m units), and the LBT (two 8.4-m mirrors on a common mount). It is planned to use all of these telescopes in the interferometric imaging mode, for which adaptive optics is essential.

10.2 Adaptive Optics Programs

In reviewing the development of adaptive optics, it is important to keep in mind the vast disparity in motivation and goals between military and scientific programs; this disparity shows up as a large difference in the scale and complexity of the equipment developed. Equipment for military applications must work reliably under the worst conditions, and must produce a specified level of performance. For example, Military adaptive optics systems are required to work at short wavelengths in locations where turbulence levels may be high. Satisfying such requirements usually requires advancing the state of the art, an expensive proposition. Astronomers, on the other hand, usually work in good seeing conditions and are able to exploit small improvements in technology that allow more information to be extracted from their observations. For example, the use of partial adaptive optics at near-IR wavelengths, producing a modest improvement in angular resolution, has led to significant scientific results [Lena 1994, Roddier et al. 1994]. There is also a considerable difference in the instruments used in the military and astronomical communities: in the former, imaging is of prime importance, with high angular resolving power being the major goal. Multispectral imaging bands used in military and earth resources analysis are relatively broad and can be separated with interference filters. In contrast, much of the information obtained on the universe is contained in the spectra of faint objects, so the spectrographs used in astronomy must measure every photon with exquisite spectral resolution.

The difference between the objectives of the national defense and astronomical communities is one reason for the long delay in applying adaptive optics technology to astronomical telescopes. Although the idea of adaptive optics was first suggested by an astronomer, the basic technology and first operating systems were developed under government-sponsored defense programs. The defense community must continually push the limits of technology to keep ahead of assumed adversaries; it usually takes some time for the value of new technology to be appreciated and to be applied to scientific work.

The first adaptive optics system to compensate a large telescope was installed in 1982 on the 1.6-m satellite-tracking telescope at the Air force MAUI Optical Station (AMOS) on Mt Haleakala, Maui, Hawaii (now operated by the U.S. Air Force Phillips Laboratory). This system was designed for the specific purpose of imaging low-Earth-orbit satellites at visible wavelengths. Since then, many experimental adaptive optics systems have been tested on astronomical telescopes throughout the world. These pioneering systems are of many different designs and have a variety of purposes, ranging from low-cost instruments with as few as 6 degrees of freedom, intended for operation in the IR spectrum, up to a 940-actuator system capable of compensating a 3.6-m telescope at visible wavelengths. The purpose of these systems has been to evaluate the utility of adaptive optics for space surveillance and astronomy, and to obtain practical experience in their design and operation. With a few notable exceptions, the early systems employed natural stars as reference sources.

Adaptive optics, as developed for military applications, acquired the reputation of being complex and expensive. In fact, much of the complexity and expense can be avoided by observing at longer wave-

lengths that are favorable to both astronomy and adaptive optics. Credit must be given to the astronomers who realized that even minimal adaptive optics systems, using fewer than 20 subapertures, could significantly improve imaging in the near-IR with large telescopes, producing useful astronomical results. This approach was pioneered by two groups: the ESO/France collaboration and The Institute for Astronomy led by F. Roddier at the University of Hawaii. This trend continues in adaptive optics systems now being designed for the new generation of 8-m telescopes, which are optimized for IR observations.

The need for bright reference sources has impeded the acceptance of adaptive optics in astronomy. The development of laser beacons in the 1980s was a major advance in adaptive optics technology and it has vastly improved the potential of adaptive optics for ground-based astronomy. As a result, there has been an explosion of interest in adaptive optics in the astronomical community and most of the new generation of 6- to 10-m astronomical telescopes have included it in their basic design. These systems are usually designed to operate with both natural stars and sodium layer beacons.

The major adaptive optics development programs in progress in 1997 are identified in Table 10.1. Capsule summaries of the activity of each development group or program are given in section 10.3. These should be considered as snapshots of a rapidly changing field of endeavor.

10.3 Program Descriptions

10.3.1 United States Air Force Phillips Laboratory, Kirtland Air Force Base, New Mexico

The USAF Phillips Laboratory operates the Starfire Optical Range (SOR), located near Albuquerque, New Mexico. It was here that the feasibility of using laser-generated beacons to measure the wavefront distortion produced by atmospheric turbulence was demonstrated in 1983 [Fugate et al. 1991]. Since then, many adaptive optics experiments using both natural stars and laser beacons have been conducted using the 1.5-m aperture telescope at SOR [Fugate et al. 1994]. This telescope has an f/1.5 parabolic primary mirror and a hyperbolic secondary mirror, producing a 10-cm diameter output beam that feeds the adaptive optics through a coude path. The alt-azimuth telescope is mounted on a hollow concrete pier with its elevation axis about 12 m above the ground. The adaptive optics components are mounted on an optical bench in a temperature-controlled room below the telescope. All the electronics equipment is located in a separate room above the optics area to remove heat sources from the optical path.

The laser is located in a separate building and is of the copper-vapor type, with outputs at 0.5106 and 0.5782 μm. The pulse rate is 5000 per second, with a pulse width of 50 ns. The laser produces Rayleigh scattering and is focused at an altitude of 10 km, with a range gate of 2.4 km. The imaging performance of the system is measured with a high-resolution CCD camera operating at 0.88 $\pm$0.05 μm. A comprehensive diagnostics capability was built into the system to permit real-time sampling, display, and recording of data.

The Generation I adaptive optics system, first operated in 1989 [Fugate et al. 1994], employed a continuous-facesheet deformable mirror with 149 independent actuators, plus 52 slaves around the periphery for edge control. A Shack-Hartmann wavefront sensor was used with an intensified, range-gated Reticon detector array. The reconstructor employed digital matrix-multiplication and the maximum closed-loop bandwidth was 65 Hz. The laser power was 75 W average. In tests conducted in November 1989, with r_0 between 0.14 and 0.19 m and a Greenwood frequency of 30 to 35 Hz at X — 0.88 μm, the uncompensated image diameter was between 1.5 and 2.5 arc seconds. The adaptive optics improved the Full-width half-maximum (FWHM) to 0.16 arc second with an average Strehl ratio of 0.2 when using a natural guide star (self-referencing), while the FWHM was 0.18 arc sec and the average Strehl ratio was 0.13 when using the Rayleigh beacon. The theoretical diffraction-limited FWHM was 0.12 arc second.

The upgraded Generation II system (1992) employed a 180-W copper vapor laser, an improved wavefront sensor using an intensified CCD array, and a deformable mirror with 241 independent actuators plus 64 slaves. The closed-loop bandwidth of the control system was increased to 143 Hz. Range gating was performed by pulsing the image intensifiers. The average Strehl ratio achieved was about 0.51 when using a bright natural star (m_V = 0) as the reference source, and 0.32 when using the Rayleigh laser beacon. The reduction in the Strehl ratio with the laser beacon was caused by focal anisoplanatism, which was due to the relatively low altitude (10 km) of the Rayleigh beacon.

The adaptive optics system was further upgraded in 1993 with an improved Shack-Hartmann sensor that used a 64 x 64 low-noise (nonintensified) CCD array, together with a 4 x 4 array of avalanche photodiodes for overall tilt tracking with a natural guide star. Range gating was accomplished with a mechanical chopper (Fugate et al. 1993, Pennington et al. 1995). The overall tracking performance was reported by Fugate [1994]. The system should have a tracking error of about 20 milli-arc-second rms for a guide star of m_V = 13, but it was not possible to operate with guide stars fainter than 9th or 10th mag-

Table 10.1 Adaptive Optics Programs Summary

Key	Program Sponsor	Telescope and Location	AO System Characteristics	WFS source	Technical Reference
1	USAF Phillips Laboratory	1.6-m AMOS, Mt Haleakala, Maui, HI	Itek CIS, 168 Actuators, visible observations	LEOS, NS	Goldberg et al. 1992, Hardy 1993
		1.5-m SOR, Albuquerque, NM	USAF, 241 actuators, visible and NIR observations	NS, RB, SB,	Fugate et al. 1993, 1994
		3.5-m SOR, Albuquerque, NM		NS, RB, SB	
		3.67-m AEOS, Mt Haleakala, Maui, HI	HDOS, 940 actuators, 32 x 32 sensor, 200 Hz bandwidth, 0.7 to 5 μm observations	NS, RB, SB	HDOS Report PA4764A, 1995
2	ESO/France	3.6-m ESO, La Silla, Chile	Come-On, 19 actuators, NIR observations	NS	Rigaut et al. 1991
		3.6-m ESO, La Silla, Chile	Come-On plus, 52 actuators, NIR observations	NS	Rousset et al. 1994
		3.6-m ESO, La Silla, Chile	ADONIS, 64 actuators, NIR observations		Hubin et al. 1993, Beuzit et al. 1994
3	ESO VLT	4 x 8.2-m VLT, Cerro Parana), Chile	Nasmyth AO System, 250 actuators, visible and IR sensors, NIR observations	NS	Hubin et at. 1994
4	Steward Observatory, University of Arizona	6 x 1.8-m MMT, Mt Hopkins, AZ	Coalign six segments, NIR observations	NS	Gray et al. 1995
	Smithsonian Astrophysical Observatory	6.5-m SMT Mt Hopkins, AZ	Adaptive secondary with 300 actuators, NIR observations	SB	Sandler, Lloyd-Hart et al. 1995
5	University of Hawaii and CFHT	3.6-m CFHT, Mauna Kea, HI	Curvature sensing and correction, 13 actuators and APDs, NIR observations	NS	Roddier et al 1994, 1995
6	CFHT and Dominion Astrophysical Observatory	3.6-m CFHT, Mauna Kea, HI	PUEO AO Bonnette, Curvature sensing and correction, 19 actuators, visible and NIR observations	NS	Arsenault et al. 1994
7	NSF, Mt Wilson Observatory	60-inch (coude) Mt Wilson, CA	ACE, 69 actuators, shearing interferometer sensor, visible observations	NS	Shelton et al. 1993
		100-inch (Cassegrain), Mt Wilson, CA	241 actuators, Hartmann wavefront sensor, visible observations.	NS	Shelton et al. 1995
8	National Science Foundation and University of Illinois	100-inch (coude), Mt Wilson, CA	UnISIS, 169 segment, visible and NIR observations	RB 0.35 μm	Thompson 1994

Table 10.1 Adaptive Optics Programs Summary *(continued)*

Key	Program Sponsor	Telescope and Location	AO System Characteristics	WFS source	Technical Reference
9	University of Chicago	1.04-m, Yerkes Observatory, WI	WCE, 69 actuators, visible observations	NS	Wild, Kibblewhite, Fang Shi et al. 1994
		3.5-m ARC, Apache Pt, NM	ChAOS, 97 actuators, visible and NIR observations	NS	Kibblewhite, Wild et al. 1994
		3.5-m ARC Apache Pt, NM	ChAOS, 249 actuators, visible and NIR observations	NS, SB	
10	Lawrence Livermore National Laboratory	3-m Shane, Lick Observatory	36 active actuators, NIR observations	SB	Olivier et al. 1995
11	U.K. Adaptive Optics Programme (Particle Physics and Astronomy Research Council)	4.2-m Herschel, La Palma, Canary Islands	MARTINI, 6 segment, ELECTRA, 76 segment, NAOMI, 76 segment, all NIR observations	NS	Myers et al. 1995, Doel et al. 1996
12	Johns Hopkins University	Adaptive Optics Coronagraph	Curvature sensor, 61 actuators, membrane mirror 0.4–1 μm	NS	Clampin et al. 1991
13	Keck Observatory	2 x 10-m Keck Telescopes, Mauna Kea, HI	Keck-II Nasmyth focus 228 active actuators, 0.8-5.0 μm observations	NS, SB	Gleckler and Wizinowich 1995
14	Gemini Project (U.S.A., U.K., Canada, Chile, Argentina, and Brazil)	8-m North, Mauna Kea, HI 8-m South, Cerroo Pachon, Chile	Active primary, 120 actuators. Secondary corrects tip-tilt and focus. AO @ MK for NIR observations at 0.85-2.5 μm	NS (LB later)	Gillett and Mountain 1996
15	Max-Planck Institute for Astronomy, and Max-Planck Institute for Extra-terrestrial Physics	3.5-m MPIA, Calar Alto, Spain	ALFA, 97-actuator deformable mirror, Shack-Hartmann sensor, sodium layer beacon	SB	
16	University of California and Jet Propulsion Laboratory	5-m Hale telescope, Mt Palomar, CA	241 active actuators, visual WF sensor, 500 Hz bandwidth, NIR observations, 1.2–3.5μm	NS	
17	Australian Research Council and University of Sydney	3.9-m Anglo-Australian Telescope, Siding spring NSW	19 actuator bimorph mirror, curvature sensor, NIR observations	NS	O'Byrne 1995
18	National Astronomical Observatory of Japan	8.3-m Subaru, Mauna Kea, HI	Active meniscus primary. Cassegrain Adaptive Optics for NIR, 36-actuator bimorph mirror, curvature sensor	NS	Iye 1996

(continued)

Table 10.1 Adaptive Optics Programs Summary *(continued)*

Key	Program Sponsor	Telescope and Location	AO System Characteristics	WFA source	Technical Reference
19	University of Arizona, Arcetri Astrophysical Laboratory, Ohio State University, German Consortium and Research Corporation	2×8.4-m Large Binocular Telescope, Mt Graham, AZ			

ACE, Atmospheric Compensation Experiment; ADONIS, Adaptive Optics Near Infrared System; AEOS, Advanced Electro-Optical System; ALFA, Adaptive optics with Laser For Astronomy; AMOS, Air Force Maui Optical Station; AO, Adaptive Optics; APD, Avalanche Photo-Diode; ARC, Astronomical Research Consortium; CFHT, Canada-France-Hawaii Telescope; ChAOS, Chicago Adaptive Optics System; CIS, Compensated Imaging System; ELECTRA, Enhanced Light Efficiency Co-phasing Telescope Resolution Actuator; ESO VLT, European Southern Observatory Very Large Telescope; HDOS, Hughes Danbury Optical Systems; LB, Laser Beacon; LEOS, Low Earth-Orbit Satellite; MARTINI, Multi-Aperture Real Time Image Normalization Instrument; MK, Mauna Kea; MMT, Multi-Mirror Telescope; MPIA, Max-Planck Institute for Astronomy, Heidelberg; NAOMI, Natural guide star Adaptive Optics system for Multi-purpose Instrumentation; N1R, Near Infra-Red, usually 0.9 to 3.4 μm (I, J, H, K and L bands); NS, Natural Star; NSF, National Science Foundation; PUEO, Probing the Universe with Enhanced Optics (also the name of an Hawaiian owl); RB, Rayleigh (laser) Beacon (7-20 km); SAO. Smithonian Astrophysical Observatory; SB, Sodium (laser) Beacon (90 km); SMT, Single Mirror Telescope; SOR, Starfire Optical Range; UnlSIS, University of Illinois Seeing Improvement System; USAF, United Stales Air Force; WCE, Wavefront Control Experiment; WFS, Wave-Front Sensor.

nitude because of phosphorescence produced by the laser pulse in the common optical path. Although the wavefront sensor is shuttered mechanically during the laser pulse itself, thereby eliminating fluorescence, phosphorescence persists between the pulses.

A compensated 3.5-m telescope is now being built at the Starfire Optical Range, primarily for defense-oriented research.

A major project of the Phillips Laboratory is the Advanced Electro-Optical System (AEOS), which employs a 3.67-m telescope installed at Mt Haleakala, Maui, Hawaii [Miller et al. 1993]. This state-of-the-art instrument is intended for a diverse group of researchers and operators, and will be used for space surveillance by the Department of Defense, as well as by the astronomical and atmospheric science communities. To accommodate multiple users, the AEOS facility has seven experimental laboratories, which are arranged radially and accessed by a single turning flat in the central coude room. The adaptive optics is installed in the common path between the turning fiat and the telescope, allowing each laboratory the option of a compensated wavefront.

The adaptive optics system is being built by Hughes Danbury Optical Systems (HDOS), under contract to the Phillips Laboratory. The telescope system has a spectral bandwidth of 0.7-5.0 μm and is capable of sunup operation. The adaptive optics has 941 correction channels with a bandwidth of 200 Hz. It is capable of compensating turbulence with r_0 values of 5 cm, using natural stars or laser beacons of visual magnitude between 1 and 8. The wavefront sensor is of the Shack-Hartmann type, using 32 x 32 subapertures. The detector is a 128 x 128 full-frame-transfer back-illuminated low-noise CCD array with 16 output ports; it was made by the Massachusetts Institute of Technology (MIT) Lincoln Laboratory. The read noise goal is < 13 electrons per pixel when cooled. The sensor wavelength range is 0.5-0.7 μm and its dynamic range is ±2 waves. Temporal sampling rates are 2500 and 5000 Hz. The real-time reconstructor performs the matrix multiplication using an array of multiplier-accumulators. The latency (last input to last output) is less than 25 μs. Data are transmitted over fiber optic links.

The deformable mirror, supplied by Xinetics Inc., has 941 actuators in a square array at 9-mm spacing. The surface quality of the mirror is $\lambda/48$ and the maximum displacement of adjacent actuators is ±2 μm. The bandwidth of the deformable mirror is 1000 Hz. The fast-tracking system has a control bandwidth of 300 Hz and operates with target magnitudes up to m_V = 15. The tilt control mirror has an angular excursion of 10 mrad. The control interface is designed to allow operation by one person.

10.3.2 European Southern Observatory/ France Collaboration

The ESO/France collaboration includes the European Southern Observatory, Observatoire de Paris (Meudon), Observatoire de Grenoble, and Observatoire de Lyon. Members of this group pio-

neered the application of adaptive optics to IR observations, where the technical requirements for wavefront compensation are less stringent than at visible wavelengths. A prototype system known as Come-On was first tested on the 1.52-m telescope of Haute Provence Observatory in October 1989 [Rousset, Fontanella, Kern, Lena et al., 1990, Rousset, Fontanella, Kern, Gigan et al. 1990]. This system was based on zonal wavefront sensing and correction, with a digital data processor for computing the corrections. The deformable mirror, made by Laserdot, consisted of a thin silicon facesheet controlled by 19 piezoelectric actuators. The wavefront analyzer operated in the visible band using a bright natural star as the reference source. It was of the Shack-Hartmann type with 20 subapertures, using a 100 x 100 intensified Reticon detector. The control system had a sampling rate of 100 Hz, which allowed a closed-loop bandwidth of only 9 Hz. The IR imaging camera, designed at Meudon, employed a 32 x 32 InSb array with a 0.1- or 0.05-arc-second pixel scale on the sky, and it operated in the J, H, K, L, and M spectral bands between 1.2 and 5 μm.

The Come-On system was tested at the Cassegrain focus of the ESO 3.6-m telescope at La Silla, Chile in 1990 and 1991 [Merkle et al. 1991]. The later observations used an electron-bombarded CCD camera in the wavefront sensor, which allowed full correction at 3.8 μm using a reference star of $m_V = 11.5$. In the L-band (3.4 μm), Strehl ratios of over 0.6 were achieved, while in the K-band (2.2 μm) the Strehl ratio was about 0.2. These results were very encouraging, so the system was upgraded in 1992 to improve its spatial and temporal resolution.

The upgraded Come-On Plus system [Rousset et al. 1993, 1994] employs a 52-actuator deformable mirror (DM) that has a stroke of ±5 μm. Overall tilt is corrected by a separate two-axis mirror driven by four piezoelectric actuators, with a range of 70 arc seconds. The Shack-Hartmann wavefront sensor has 32 subapertures and is equipped with two detectors. The "high-flux" detector array is an intensified Reticon, similar to that employed in the original system. This detector is readout-noise limited and is therefore used with bright reference sources of visual magnitude between 6 and 10. The "low-flux" detector is an electron-bombarded CCD camera developed by LEP, which, because of the large gain achieved in the accelerator (~2000), is always photon-noise limited. The peak efficiency of the photocathode is 10% at 550 nm. This detector is only used with reference sources of m_V = 10-15, because its life would be shortened by the damage to the CCD caused by energetic photoelectrons at higher fluxes.

The control system consists of two sections: a hard-wired computer to calculate the wavefront slopes from the video outputs of the wavefront sensor, and a matrix multiplier to compute the tilt and deformable mirror (DM) corrections, using a digital signal processor (DSP) board running at 13 x 10^6 operations per second, giving an update rate of 400 Hz. Additional flexibility is given by using two control algorithms. With bright stars and low turbulence, the standard zonal control algorithm is used, in which each actuator is controlled independently. For more stringent conditions, a modal control algorithm is available, in which the maximum number of modes is determined by the number of degrees of freedom on the deformable mirror. For each set of operating conditions (reference signal-to-noise ratio, turbulence strength, and temporal frequency), the optimal bandwidth for each mode is calculated and implemented by adjusting the loop gain in the data processor.

The Come-On Plus adaptive optics system was set up and tested on the ESO 3.6-m telescope at La Silla, Chile in December 1992 and April 1993 [Rousset et al. 1994]; it is now used for routine astronomical observations [Lena 1994]. With reference stars brighter than $m_V = 10$, the best angular resolution of 0.1 arc second is obtained in the J-Y and H-bands (1.25 and 1.6 μm), respectively. The average Strehl ratio is about 0.13 at 1.25 μm and 0.65 at 2.2 μm. With reference stars dimmer than $m_V = 11$, the performance degrades rapidly (even with modal control), mostly due to the low quantum efficiency of the wavefront sensor detectors, and the Strehl ratios are then generally less than 0.1.

The latest version of Come-On Plus has been given the acronym ADONIS (Adaptive Optics Near-Infrared System) and is intended as a common-user instrument for the astronomical community at the ESO 3.6-m telescope at La Silla where it has been in operation since 1993 [Hubin et al. 1993, Beuzit et al. 1994]. A further objective is to develop operational procedures and to test technical concepts that can be applied to adaptive optics for the ESO VLT. The control system and operator interface have been redesigned to make the adaptive optics more user friendly. Many system parameters that have to be set according to astronomical and atmospheric conditions are now optimized using an artificial intelligence system [Demailly et al. 1994]. This involves both internal and external sensors, as well as operator inputs. ADONIS can be used by astronomers without any specialized knowledge of adaptive optics.

10.3.3 European Southern Observatory/ Very Large Telescope

The ESO VLT consists of four 8-m telescopes that may be used independently or in concert. The observatory is located at Cerro Paranal, Chile. To exploit the full capability of this system for high-resolution imaging, spectroscopy, and interferometry, adaptive optics has been included in the design from the start. The original plan to install adaptive optics in the coude path of each telescope has been postponed, and the current goal (mid-1997) is to implement a

first-generation system at the Nasmyth focus of Telescope Unit 1. The proposed implementation has been described by Hubin et al. [1994].

The ESO has sponsored a long-term development program in adaptive optics, including the prototype systems Come-On and Come-On Plus, described above. The system has been upgraded with a comprehensive and user-friendly control interface and is now known as ADONIS. The software developed for ADONIS will also be used in the VLT adaptive optics. The VLT is intended for observations over a large spectral band, from the visible up to 10 μm. Experience with the Come-On system showed that adaptive optics is both effective and economical at near-IR wavelengths. Consequently, the VLT Nasmyth Adaptive Optics System (NAOS) is designed to operate over the 1-5 μm spectral band. In the nonthermal domain (1-2.5 μm), using a visible wavefront sensor, the objectives are: (1) to achieve a minimum Strehl ratio of 0.7 at 2.2 μm for median seeing conditions, when a suitable reference star is available; (2) to obtain the maximum sky coverage compatible with an image resolution of 0.2 arc second FWHM at 2.2 μm. The second objective requires the use of off-line data processing to reconstruct the point spread function from the adaptive optics data, in order to derive deconvolved images.

For work in the thermal domain (3-5 μm), an additional near-IR wavefront sensor, operating in the 1-25 μm band, will be used to increase the sky coverage. NAOS is scheduled to be installed on Unit 1 of the VLT in the year 2000.

The ESO, with its associate the Max Planck Institut fur Extraterrestrische Physik Garching, is studying the implementation of a laser beacon system for adaptive optics on the 3.6-m ESO telescope. The use of laser beacons with the 4 x 8-m VLT is also being considered.

10.3.4 Center for Astronomical Adaptive Optics, Steward Observatory, University of Arizona, Tucson, Arizona.

The Multi Mirror Telescope (MMT) at Mt Hopkins, Arizona is a joint facility of the University of Arizona and the Smithsonian Astrophysical Observatory. It originally consisted of six 1.8-m telescopes on a common alt-azimuth mount, the beams of which were combined into a single image. [Beckers et al. 1982]. The primary mirrors were not optically phased, so the angular resolution was that of a single 1.8-m aperture. To improve the superposition of the images, a low-order adaptive optics system, the FASTTRAC II, was built and tested on the MMT. This system was the successor to the FASTTRAC 1 [Close et al. 1994], which implements a tip-tilt secondary mirror on the 1.5- and 2.3-m telescopes at Steward Observatory. The FASTTRAC II [Gray et al. 1995] corrected tip-tilt errors over each of the MMT's six primary mirrors. Phase errors between the mirrors were not corrected.

Experiments were conducted with FASTTRAC II in 1995 and 1996 using a sodium laser beacon [Lloyd-Hart et al. 1995, 1997]. A 4-W ring dye laser at 589 nm was employed, with a sodium vapor Faraday cell to lock the laser to the correct wavelength. On the MMT, the laser beam was projected axially through a 48-cm diameter launch telescope, producing a beacon of brightness comparable to a star of of magnitude $m_V = 10$. The first astronomical images sharpened by a sodium beacon were obtained in May 1996 [Lloyd-Hart et al. 1997]. Images of stars in the core of M13 showed an improvement in resolution from 0.74 to 0.53 arc second in the K-band, using a fixed guide star at 41 arc seconds from the center of the science camera field. No anisoplanatism was detected across the 24-arc-second field of view. The results suggest that global tilt across the 7-m aperture of the instrument is highly correlated over larger field angles than the standard turbulence model predicts. This is good news for astronomers concerned with the sky coverage obtainable with adaptive optics on large telescopes of the 8-m class.

The experience obtained by the Center for Astronomical Adaptive Optics with the FASTTRAC II system is being applied to the design of the adaptive optics system for the 6.5-m single-mirror conversion of the MMT [Martin et al. 1997, Lloyd-Hart et al. 1997]. The change is planned for late 1997. The adaptive optics system is expected to provide diffraction-limited images from 1.6 to 5 μm, with nearly complete sky coverage [Sandler, Lloyd-Hart et al. 1995], using a sodium laser beacon projected from behind the secondary mirror. A major feature of the new 6.5-m instrument is the adaptive secondary mirror, which eliminates additional optical elements in the imaging path, thereby minimizing the background thermal radiation, an important factor in mid-IR observations. The adaptive secondary mirror is being fabricated by the Steward Observatory Mirror Laboratory [Bruns et al. 1995, Martin and Anderson 1995, Modisett and Martin 1995]. It employs a highly aspheric hyperboloidal faceplate, 2 mm thick and 64 cm in diameter. The surface is controlled by 324 voice-coil actuators with associated capacitor position sensors. A two-tier servo system is employed, using local servo loops with 10-kHz bandwidth closed around the capacitors in order to maintain the surface in the commanded position, while the adaptive optics loop is closed around the wavefront sensor with a bandwidth of about 1 kHz. The wavefront sensor is of the Shack-Hartmann type with 150 subapertures using a 4-port CCD operating at 250 kHz per port. Overall tilt sensing with a natural star is accomplished in the 1.2-2.3 μm band, in order to get shared compensation from the sodium beacon.

70.3.5 *University of Hawaii, Institute for Astronomy*

Development of a low-order adaptive optics system for near IR observations started at the University of Hawaii (UH) in 1990 [Roddier, Graves et al. 1991]. The objective was to build the simplest possible compensation system for astronomical images, using only natural stars as reference sources. The UH system is based on curvature sensing and correction [Roddier 1988a]. Most adaptive optics systems employ wavefront slope sensing, in which x and y slopes measured within an array of zones in the telescope aperture are reconstructed to obtain a two-dimensional map of the wavefront. This error map defines the drive signals that are applied to a zonal deformable mirror. In curvature sensing, the average wavefront curvature is measured in each zone of the aperture and this (scalar) quantity is used to drive an actuator that produces the compensating curvature directly in the corresponding zone of the correction device, usually a bimorph mirror. This approach simplifies (but does not eliminate) the reconstruction function. Wavefront curvature is easily measured, requiring only one photon detector per subaperture; it allows a relatively simple implementation of adaptive optics. It is, however, only suitable for low-order correction, as reconstruction errors accumulate with a large number of zones (see section 5.5). The IR observations meet this requirement, because the turbulence coherence length is usually on the order of 1 m at a wavelength of 2.2 μm.

The UH adaptive optics system employs 13 subapertures. The wavefront sensor uses a vibrating membrane mirror that defocuses the wavefront in each subaperture in opposite directions. The resulting changes in illumination are detected by an array of 13-photon-counting avalanche photodiodes, which have peak quantum efficiency of about 40% at 0.7 μm. A perfect wavefront results in equal deviations of intensity for each half-cycle of mirror excursion; defocus errors are detected by comparing the photon counts accumulated in the odd and even half-cycles. The sensor outputs are converted into control signals for the deformable mirror by means of a digital processor. The wavefront corrector is a 13-electrode bimorph mirror made by Laserdot. The advantage of using a bimorph mirror is that the electrode configuration can be made to match the sensor subapertures for maximum compensation efficiency. The sampling frequency of the control loop is 1 kHz, allowing a closed-loop bandwidth of about 100 Hz. The loop gain and the vibration amplitude of the modulation mirror are adjusted to optimize performance with different reference star magnitudes and seeing conditions.

The original experimental UH system had to use the Canada-France-Hawaii Telescope (CFHT) coude path, which attenuated the light by two stellar magnitudes and also introduced chromatic errors. In spite of these problems, near-IR images with 0.1-arcsecond resolution of the proto-planetary nebulae "Frosty Leo" and the "Red Rectangle" were obtained [Roddier et al. 1994, 1995]. An important characteristic of the UH system is its ability to work with faint reference sources. Roddier et al. report that in the H-band (1.6 μm), at the CFHT Cassegrain focus, a Strehl ratio of about 0.4 can be obtained with a reference source of magnitude 11.5 in the R-band (0.7 μm), equivalent to visual magnitude 12.5, and that a Strehl of 0.2 is possible with m_V = 15.

The adaptive optics hardware was reconfigured in 1994 for use at the Cassegrain f/36 focus of the 3.6-m CFHT telescope. Two cameras can be used simultaneously: a 1024 x 1024 pixel CCD camera and a 1024 x 1024 pixel IR camera using a Rockwell HgCdTe array with sensitivity in the 1-2.5 μm range. The image scales are 0.024 and 0.035 arc second per pixel, giving fields of view of 25 x 25 and 36 x 36 arc seconds, respectively. The compensated images are first preprocessed (background subtraction, bad pixel elimination, intensity normalization, frame recentering and averaging) and are then deconvolved using point spread functions (PSFs) obtained from similarly processed images of nearby stars. The Lucy-Richardson algorithm was found to be most suitable for images with a large dynamic range. The UH adaptive optics system has been used recently to obtain high-resolution images revealing the structure of T Tauri stars [Roddier et al. 1996].

10.3.6 Canada-France-Hawaii Telescope Corporation

The CFHT group pioneered the use of image stabilization to improve the resolution of astronomical images, using the High-Resolution Camera (HR Cam) [McClure et al. 1989, Racine and McClure 1989]. In conjunction with the excellent seeing at the site and the thermal control of the CFHT dome, this camera has obtained outstanding images at near-IR wavelengths. The CFHT group has developed an Adaptive Optics Bonnette (AOB), that is installed at the f/8 Cassegrain focus of the 3.6-m telescope, forming an interface between the telescope and the instrumentation. [Arsenault et al. 1994]. This instrument is also known as "PUEO," after a Hawaiian owl that sees particularly well at night. Physically, the bonnette is an aluminum casting 1.1 m in diameter and 28 cm thick, containing the adaptive optics components. It is a facility instrument with backup capability to ensure reliability, providing access to either the f/19.6 compensated beam or the direct f/8 telescope beam.

The adaptive optics employs a 19-electrode bimorph deformable mirror and a 19-subaperture wavefront curvature sensor, based on the University of Hawaii system described in section 10.3.5. Image motion is corrected by a separate tip-

tilt mirror. Wavefront sensing employs either the science object or a nearby star as the reference source. The wavefront sensor has an instantaneous field of view of only 3 arc second, but is mounted on an x-y-z stage so it can be placed anywhere within the 90-arc sec field of view of the bonnette. The control system provides a 90-Hz servo bandwidth using modal control. The performance of this adaptive optics system has been determined by computer simulation [Rigaut et al. 1994].

The user interface has been carefully designed to allow nonspecialists in adaptive optics to operate the instrument in an optimal fashion. The AOB has been used for scientific observations since 1996.

The instrumentation used with the CFHT adaptive optics bonnette includes visible imaging, IR imaging, and high spatial resolution spectroscopy. An integral-field spectrograph optimized for observations with the AOB has been developed by a team from the Observatoire de Lyons, France. The visible imager is a 2000 x 2000 CCD, which gives a scale of 0.045 arc second per 15-μm pixel. The field of view is 1.5 arc minutes, but is likely to be limited by isoplanatism most of the time. The University of Montreal IR Camera (MONICA) narrow-field camera, employing a NICMOS (near infrared camera/multiobject spectrometer) 256 x 256 chip, is now in operation.

10.3.7 National Science Foundation/ Mount Wilson Observatory

The 100-inch reflecting telescope at Mount Wilson Observatory, which contributed so much to astronomy in the first half of the twentieth century, is now revitalized with adaptive optics. In many ways, it is an ideal vehicle for upgrading, because of the excellent quality of its optics and the exceptional seeing conditions at the site. In addition, its modest aperture allows a high degree of correction to be achieved at visible wavelengths with available technology.

The first experiments with adaptive optics at Mt Wilson [Shelton et al. 1993] were performed on the 60-inch telescope using the Atmospheric Compensation Experiment (ACE) hardware, which was described briefly in section 1.9. The ACE was built by MIT Lincoln Laboratory and Itek Corporation, with U.S. Department of Defense sponsorship, and it was first operated in the early 1980s. It employs a 69-actuator deformable mirror and a shearing interferometer wavefront sensor. The ACE was intended mainly to compensate laser beams propagating through the atmosphere, but it is also capable of correcting wideband images. When this equipment was installed at the coude focus of the 60-inch telescope, images of first-magnitude stars were obtained with FWHM of 0.12 arc seconds at visible wavelengths. The peak intensities were 6-10 times higher than those of uncompensated images. The limiting magnitude of this system was about $m_B = 6$.

Using this experience, Shelton et al. [1995] have built an improved adaptive optics system for the 100-inch telescope. The system works at visible wavelengths using natural reference stars and is automated for one-person operation. It employs a 241-actuator continuous-plate deformable mirror and a 16 x 16 Shack-Hartmann wavefront sensor. The primary goal is to maximize the correction efficiency with faint reference sources, which is achieved as follows:

1. The adaptive optics system is installed at the Cassegrain focus to eliminate long optical paths and additional optical surfaces.
2. Available light is split between the wavefront sensor and the science camera, according to the task. If the reference star is not of scientific interest, the field is split with a mirror near the image plane; almost 100% of the reference light passes through a hole in the mirror to the wavefront sensor, while almost 100% of the science object light is reflected to the camera. The mirror contains several holes with diameters between 0.4 and 6 arc seconds, which act as a selectable field stop for the wavefront sensor. If the reference object is part of the science field, then a beam splitter is inserted in the pupil plane to divide the light between the wavefront sensor and the science camera.
3. A modal control system is used in which the number of modes actively corrected depends on the brightness of the reference source.
4. A specially designed low-noise, high-frame-rate CCD chip is used in the Shack-Hartmann wavefront sensor [Levine et al. 1994]. Improved CCD chips, such as the ADAPT II, will be used as they become available.

Communication between the optical unit and the electronics is handled by fiber optics links. The host computer provides the user interface, as well as handling the command and control functions. The wavefront processor is based on commercial DSPs. Modal control is implemented by using the appropriate control matrix, which can be recalculated and loaded in real time. The frame rate of the system, currently 200 Hz, is limited mainly by the data processing capability, but it is hoped to improve this by speeding up critical operations and data transfer.

The science camera employs a 1024 x 1024 CCD, which can be operated at a relatively high frame rate (2 Hz) for diagnostics and alignment. The field of view is normally 22.8 arc seconds at 0.022 arc second per pixel. The 100-inch telescope with adaptive optics is being used for several imaging and spectroscopic applications. A compensated image can be sent, using single-mode fibers, to existing spectrographs at Mt Wilson. For high-resolution spectroscopy, the higher Strehl ratios achievable with adaptive optics improve

the signal-to-noise ratio per exposure time and allow fainter stars to be observed; the higher stability of the PSF also improves the Doppler precision. In the "coronagraphic" mode, in which the bright reference star is excluded from the science camera, it may be possible to achieve direct imaging of circumstellar structures or even faint companions.

10.3.8 University of Illinois Seeing Improvement System

A 5-year program, started in 1993, is being conducted by the Astronomy Department of the University of Illinois, with the aim of providing the means for testing new concepts in adaptive optics. In this project, an adaptive optics system known as UnISIS (University of Illinois Seeing Improvement System) will be installed at the coude focus of the 2.5-m telescope at Mount Wilson Observatory [Thompson 1994]. Laser beacons will be incorporated into the system from the start. Initially, an excimer laser operating at 351 nm will be used to generate Rayleigh beacons at altitudes up to 18 km. This laser has previously been used at Mount Laguna Observatory [Thompson et al. 1991, Thompson and Castle 1992].

The configuration of the adaptive optics, as originally planned, is as follows: The wavefront compensator is a 169-segment (13 x 13 square) deformable mirror made by ThermoTrex Corporation (TTC). This type of mirror has 3 degrees of freedom (tip, tilt, and piston) per segment, giving a total of about 400 actuators within the inscribed circular pupil. A second segmented mirror may be added later, to test multiconjugate adaptive optics. Wavefront measurements are made using the laser beacon and a natural guide star, combined in the most efficient way. This is different from the customary approach in which the natural star is used only to measure overall tilt. In the UnISIS system, the natural guide star sensor is a neural network camera. The neural network is trained to recognize overall tip, tilt, and other low-order Zernike modes. Only when the error measured in each mode with the natural star (including the effects of both signal-to-noise ratio and isoplanatism) is greater than that of the laser beacon sensor will the latter take precedence. The neural network sensor is based on a 64 x 64 low-noise CCD made by MIT Lincoln Laboratory, using small sections of the array (between 5×5 and 9×9) for the two out-of-focus images. The laser beacon sensor is of the Shack-Hartmann type, using 18 subapertures across the diameter, corresponding to a sampling interval of 13.9 cm at the telescope pupil. With multiple guide stars, a shearing interferometer with periodic interferometric weighting [Sandler 1992] may be employed. The spectral range of the optical system is 0.35–2.2 μm. The image sensors will include a CCD camera and a near-IR camera using a NICMOS-3 array.

The excimer laser produces a beam power of about 50 W at a pulse rate of 333 Hz, corresponding to 150 mJ per pulse. The beam is launched through the main telescope to reduce divergence. A rotating shutter using an ultraviolet (UV)-grade fused silica substrate with small reflecting spots is employed to insert the laser beam into the optical train and to block the Rayleigh return from low altitudes. The laser was installed at the telescope in 1995, and plans have been made to conduct tests with three Rayleigh beacons [Thompson and Yao-Heng 1995].

70.3.9 University of Chicago, Department of Astronomy and Astrophysics

The objective of the University of Chicago adaptive optics program is to develop an instrument that can be installed as a facility in astronomical observatories, at an affordable cost. The instrument is known as ChAOS (Chicago Adaptive Optics System). The approach is to develop, from the start, a high-order compensation system that uses a sodium laser beacon, which can be upgraded in the future with minimal modifications. This is a different approach from that of other groups which start with low-order systems using natural stars. The techniques used in such systems may not be applicable to higher order systems.

The ChAOS program started in 1989, based on a proposal for a laser beacon adaptive optics system for use at IR wavelengths on the 3.5-m telescope at Apache Point, New Mexico [Kibblewhite, Wild et al. 1994]. The performance goal is to produce a Strehl ratio of 0.6 at a wavelength of 2.3 μm, in 1-arc second seeing. Several versions of ChAOS have been built: an 87-actuator system, optimized for use with a sodium beacon, was first operated in 1995, and a 201-actuator mirror has since been built [Smutko et al. 1996].

ChAOS uses a Shack-hartmann wavefront sensor with 16×16 subapertures; the camera employs a 64 x 64 CCD detector made by MIT Lincoln Laboratory with a readout rate of 2000 frames per second and read noise of 13 electrons per pixel. The deformable mirror consists of a quartz faceplate 125 mm in diameter and 1 mm thick, supported by an array of piezoelectric actuators that are mounted on an Invar baseplate. The uncorrected surface is flat to about 1/2 wave. The actuator stroke is 4 μm for a drive voltage of 400. The high-voltage drivers are multiplexed, each one driving 16 actuators. Update time for the whole mirror is 300 μs and the power required to drive the mirror is less than 0.1 W per actuator. The optics package includes calibration and diagnostic facilities.

The data processor consists of two sections, a digital signal processor (DSP) that computes the wavefront slopes, and a reconstructor that estimates the wavefront errors over the aperture; the error signals are then sent back to the DSP to compute the drive

signals to the actuators. The data processor employs 16 DSP chips and 16 multiply-accumulator chips (MACs). It takes 2 ms to compute the control signals for a 250-actuator system. The data processor uses parallel architecture, and performance can be improved by adding more cards. Several different estimating algorithms are available, selectable by the operator to match the current turbulence conditions. The operator interface is a Macintosh computer that controls the system using an interactive software package. A major effort has been made to develop user-friendly software for the ChAOS project. Prior to closing the adaptive optics loop, the software runs checkout and calibration procoedures on the system. During operation, real-time displays of photon counts, wavefront errors, atmospheric time constants, and other parameters are provided to enable the performance to be assessed. Two controls are accessible to the operator to optimize performance: one varies the amount of spatial smsoothing across the deformable mirror, while the other varies the servo bandwidth.

The sodium beacon laser is a 10-W diode-pumped sum-frequency laser developed at MIT Lincoln Laboratory [Kibblewhite, Vuilleumier et al. 1994]. It is launched from a 0.3-m diameter mirror located behind the secondary of the 3.5-m telescope.

The science instrument package is attached to the adaptive optics unit by a large precision bearing, to accommodate field rotation. The IR camera (ChAOS CAM) employs a Rockwell 256 x 256 HgCdTe array sensitive from 0.8 to 2.5 μm. ChAOS first operated on the 3.5-m telescope using natural stars in May 1995 and the laser beacon system was installed in December 1995. Using bright stars, the system achieved an angular resolution of 0.14 arc second FWHM at 0.9 μm, improving the peak intensity by a factor of 14. Using the IR camera at 1.6 μm, the angular resolution achieved was 0.2 arc second, with a peak intensity improvement of 8 times.

10.3.10 Lawrence Livermore National Laboratory

A research and development program on laser beacons for astronomy started at Lawrence Livermore National Laboratory (LLNL) in 1990, with the goal of demonstrating the feasibility of closed-loop adaptive optics using laser beacons generated in the sodium layer [Max et al. 1994]. The laser source used for the initial experiments was a high-average-power tunable dye laser, developed for the U.S. Department of Energy's Atomic Vapor Laser Isotope Separation (AVLIS) Program. The dye laser, pumped by a copper vapor laser, produces an average power of 1.1 kW at a pulse rate of 26 kHz, and an average power of 400 W at 13 kHz. The laser line width is broadened to 3 GHz to match the D_2 absorption line of sodium. The pulse length is 32 ns, and requires stretching to avoid saturation of the sodium layer.

The initial phase of the program involved measurements of the characteristics of laser beacons and was completed in 1993 [Avicola et al. 1994]. At this point, the AVLIS laser was shut down and the focus of the program shifted to development of an adaptive optics package and sodium laser for installation at Lick Observatory. The goal was to attain a full observational capability on the Shane 120-inch telescope by the end of 1996.

The first step was to build a prototype adaptive optics package for field tests [Brase et al. 1994]. This system employed a 69-actuator deformable mirror and used a natural star for wavefront sensing, at wavelengths below 0.65 μm, while observations were made in the 0.7 to 1 μm band. The package was tested in 1994 at the Lick Observatory, using the 1-*m* telescope [Olivier et al. 1994], and also with the 3-m Shane telescope [Olivier et al. 1995]. Results were consistent with expectations, the Strehl ratio of a bright star being increased by a factor of about 12 at an observation wavelength of 1 μm.

The current adaptive optics system operates on the 3-m Shane telescope, with a sodium beacon as the reference source. It employs a 127-actuator mirror made by LLNL, using PMN actuators in a triangular pattern. Sixty-one of the actuators are actively controlled. The wavefront sensor is a Shack-Hartmann with 37 subapertures, using a CCD camera package built by Adaptive Optics Associates (AOA). The CCD itself is a 64 x 64 pixel device, made by MIT Lincoln Laboratory, with a read noise of 7 electrons per pixel at a read rate of 1200 frames per second. The tip-tilt sensor used with a natural guide star uses four photon-counting avalanche diodes arranged as a quad cell. The wavefront control computer is a 160 Mflop Mercury VME system, capable of operating at sample rates up to 500 Hz. The control bandwidth of the adaptive optics system is 30 Hz.

The design of the sodium laser has been described by Friedman et al. [1994] and is summarized in section 7.6. It employs state-of-the-art solid-state pump lasers with a tunable dye laser as the conversion element. The pump laser and dye master oscillator are located in a separate room, and are connected via quartz fibers to the final dye amplifier, which is mounted on the telescope itself. The beam is launched by a refractive telescope of 30-cm aperture, attached off-axis to the main telescope structure. The laser has an average power output of 18 W at a pulse rate of 11 kHz; the pulse duration is 100 ns. The apparent size of the beacon is 1.8 arc seconds for a 10-s exposure. The return signal from the laser beacon was measured in September and November 1996 and found to be comparable with that of a natural star of $m_V = 7.0$.

In September 1996, the Lick Observatory adaptive optics system produced the first significant image

improvement obtained with a high-order adaptive optics system, using a sodium beacon. Measurements made at 2.2 μm on a star of $m_V = 9.1$, using an 80-s exposure, showed that the FWHM decreased by a factor of 2.4 (from 0.75 to 0.31 arc second) and the peak intensity increasd by a factor of 3.3, compared with the uncompensated image.

10.3.11 *United Kingdom Adaptive Optics Programme*

The U.K. Adaptive Optics Programme is funded by the U.K. Particle Physics and Astronomy Research Council (PPARC). The participants in this program include the University of Durham, The Royal Greenwich Observatory, The Royal Observatory Edinburgh, The University of Edinburgh, the University of Oxford, Imperial College London, and the Cavendish Laboratory, Cambridge. The primary purpose of this program is to equip two major telescopes with common-user adaptive optics systems: the 4.2-m William Herschel Telescope (WHT) located at the Observatorio del Roque de las Muchachos on La Palma, Canary Islands, and the 3.8-m United Kingdom Infrared Telescope (UKIRT) on Mauna Kea, Hawaii. The program also supports developments in several related fields, such as site evaluation. IR instrumentation, adaptivie secondary mirrors, partial adaptivie optics, and laser beacons. The goal is to implement adaptive optics systems that have wide astronomical applications. The main elements of this program were reviewed in a paper by Myers et al. [1995], on which the following summary is based.

Common-User Adaptive Optics Systems

The adaptive optics for the WHT and UKIRT will initially use natural stars as reference sources and will be optimized for observations in the near-IR, around 2.2 μm. They will be capable of feeding a number of instruments that will be locally supported and available to visiting astronomers, without a special operations team. The initial concept for the WHT adaptive optics system was described by Gentles et al. [1995]. It operates at the Nasmyth focus, which allows it to be mounted on a fixed optical bench. The deformable mirror can be conjugated either at the entrance pupil or at a turbulent layer in the atmosphere. The architecture of the adaptive opticis system is designed to accommodate sodium laser beacons at a later phase of the program. The UKIRT is an equatorial telescope and the adaptive optics will be implemented somewhat differently, probably using an adaptive secondary mirror.

The common-user adaptive optics system for the WHT has been given the acronym NAOMI (Natural guide-star Adaptive Optics system for Multi-purpose Instrumentation). A technical description is contained in the document AOW/GEN/AJL/7.2/07/97, dated 30 July 1997. NAOMI is a "no frills" adaptive optics system, for observations at near-IR wavelengths. It will operate initially using natural stars as reference sources, with minimal automatic control. The instrument and its operating software are based on experience with the ELECTRA system, described later. The current design employs a 76-segment mirror with 10 segments across the diameter, the WHT pupil diameter covering 7.3 segments. The wavefront sensor is of the Shack-Hartmann type, with interchangeable lenslet arrays to adjust the spatial resolution. Larger subapertures are used when the photon counts from the reference source are low. Two low-noise CCD arrays are used in the binned mode for sensing x and y wavefront gradients. The control system uses zonal processing at high light levels, with modal control available for faint reference sources. Performance of the adaptive optics is optimized manually, using real-time visual displays of wavefront distortion. Upgrade paths are provided for fully automated calibration and system optimization, and the eventual use of laser beacons.

Infrared Imager for the William Herschel Telescope

The WH Infrared Camera (WHIRCAM) is a near-IR imager using a 256 x 256 InSb detector operating in the 1-5 μm range. It was commissioned in 1995.

Site Evaluation and Modeling

The goal of this program is to conduct a comprehensive evaluation of the turbulence environment at the sites of both U.K. telescopes. The data will be used in conjunction with theoretical modeling [Wilson and Jenkins 1996] to optimize the design of adaptive optics systems and science instrumentation.

Near-Term Adaptive Optics Systems

The University of Durham adaptive Optics group have built a new IR version of MARTINI (Multi-Aperture Real time Image Normalization Instrument). The instrument is designed to operate at the Nasmyth focus of the William Herschel Telescope [Myers et al. 1994, Doel et al. 1996]. This version, known as MARTINI-III, is a low-order adaptive optics system using a correction mirror with six segments, each of which is controlled in tip, tilt, and piston by three piezoelectric actuators. It is a multispectral system in which the wavefront disturbances are sensed at visible wavelengths using a natural star, while the science observations are made with the WHIRCAM at infra-red wavelengths. MARTINI-III was commissioned in 1995 and

enhancements are continuing. The maximum frame rate of the system is 345 Hz and the limiting reference star magnitude is m_V = 13 to 14.

The University of Durham has also developed a higher-order adaptive optics system know as ELECTRA (Enhanced Light Efficiency Co-phasing Telescope Resolution Actuator), which is also intended for use on the Nasmyth focus of the WHT. ELECTRA employs a 76-segment deformable mirror, each segment driven by three piezoelectric actuators which are corrected for hysteresis using strain gauges, giving a total of 228 degrees of freedom. Overall tilt is corrected by a separate steering mirror. The wavefront sensor is of the Shack-Hartmann type. The ELECTRA control console includes a real time three-dimensional display of the uncorrected and corrected wavefronts.

Development Programs

These efforts include a study of adaptive correction using a secondary mirror [Bigelow et al. 1993, 1994], and development of a sodium laser beacon. Future activities include the development of a higher order AO system for operation at visible wavelengths, and research on the use of AO for long baseline interferometers.

10.3.12 Johns Hopkins University, Center for Astrophysical Sciences.

The Johns Hopkins University (JHU) Adaptive Optics Coronagraph (AOC) has been developed for use with telescopes of the 1–2-m class to allow the investigation of faint objects in close proximity to bright stars [Clampin et al. 1991]. These objects include circumstellar disks, brown dwarfs, and galactic nuclei. Such objects are, in many ways, ideal for adaptive optics because they include a bright reference source suitable for wavefront sensing and their angular extent is typically a few arc seconds (within the isoplanatic angle). The use of a coronagraphic imaging system (similar to the Lyot solar coronagraph) facilitates measurement of dim objects by reducing the scattered light from the parent star.

In its original form, the AOC compensated image motion only, achieving a factor of 2 increase in image resolution, as theoretically predicted. It now employs a membrane mirror with 61 actuators within the active pupil, together with 30 actuators along the boundary. The mirror stroke is ±3 μm and the bandwidth is 1 kHz. A wavefront curvature sensor with two displaced image planes is employed; the output signal is proportional to the Laplacian of the wavefront, which is the function required to drive a membrane mirror [Roddier 1988a]. The coronagraph is located in the compensated beam, following an occulting mask positioned in the image plane, which directs all the light from the central (reference) star to the wavefront sensor. The circumstellar field is then apodized and imaged on to a CCD camera.

The AOC has been used by researchers at JHU and Caltech to search for brown dwarfs, which are notoriously difficult to find because of their low brightness and proximity to a brighter companion. Using the Palomar 1.5-m telescope in October 1994 and the Hale 5-m telescope nearly a year later, a dim object of 14th magnitude at 2.2 μm, probably a brown dwarf, has been identified near the 8th magnitude M-type dwarf Gliese 229 [Nakajima 1995].

10.3.13 W. M. Keck Observatory

The AO system for the W. M. Keck Observatory is intended to provide near diffraction-limited images over the wavelength range 1.1-2.3 μm, with some correction capability over the range 0.8-5.0 μm. The first stage of this program is to install the adaptive optics on the Nasmyth platform of the Keck II 10-m telescope. The reference sources will be natural stars up to about 15th magnitude, or a single sodium laser beacon of about 10th magnitude with a natural guide star of up to 19th magnitude. The National Aeronautical and Space Administration (NASA) is planning to fund an interferometer using the two Keck telescopes; this will require a second AO system for Keck I.

The Keck II adaptive optics facility consists of four main subsystems [Gleckler and Wizinowich 1995, Wizinowich et al. 1996]:

- user interface and supervisory control;
- optics bench and system enclosure;
- wavefront controller;
- laser and beam projection optics.

The first two items are being built at the W. M. Keck Observatory and the second two at Lawrence Livermore National Laboratory.

The user interface and supervisory control subsystem [Stomski et al. 1996] provides a graphical user interface with overall control of the adaptive optics facility, using an EPICS (Experimental Physics and Industrial Control System) environment to handle the distributed processing needs. Its functions include control of AO subsystems during all phases of observing, communication with other observatory systems and science instruments, and implementation of test procedures and system diagnostics. The entire adaptive optics system is controlled from a single workstation.

The optics bench, physically mounted on the Nasmyth platform, contains all the optical components of the AO system and provides an output beam (to the science instrumentation) that has the same f-number and pupil location as that of the telescope itself; that is, f/15 with the pupil at 19.96 m [Gregory et al. 1996]. The main components in the imaging path are the image rotator, the fast tip-tilt mirror, an off-axis paraboloid (OAP) to collimate the

beam and reimage the telescope pupil on the deformable mirror, a second OAP to produce the f/15 output beam, a dichroic beam-splitter that transmits the IR imaging wavelengths, and an atmospheric dispersion corrector for the IR band. The IR beam then enters the imaging camera. Visible wavelengths reflected from the dichroic beam-splitter pass through a dispersion corrector for the visible band and are then split between the tilt/focus sensor and the wavefront sensor. The beam can also be switched to the acquisition camera. The optical system is designed with the science instruments centered on the optical axis, which means that the wavefront and tip-tilt sensors must be steerable to different field positions without disturbing the alignment of the wavefront sensor to the DM actuators. This is achieved with two field selection mirrors located in front of the wavefront sensor. In addition, the entire wavefront sensor is moved axially to accommodate the difference in focal distance between natural stars and sodium beacons. The tilt/focus sensor, required to stabilize the image when using a laser beacon, is mounted on a precision x-y-z translation stage so that it can acquire a natural star anywhere within the 2-arc minute field of view. The optics bench also includes components for alignment, calibration, and diagnostics. The electronics associated with the AO system are mounted in cabinets at the side of the Nasmyth platform.

The wavefront controller subsystem consists of the components in the two closed loops: the tip-tilt sensor driving the fast-steering mirror, and the high-order wavefront sensor driving the deformable mirror. The tip-tilt sensor uses four EG&G Optoelectronics Canada photon-counting avalanche diodes; astigmatism is introduced into the beam to enable defocus to be measured [Cohen et al. 1984]. A Shack-Hartmann wavefront sensor with 16 subapertures across a 9-m inscribed aperture is used. The detector is a 64 x 64 pixel CCD made by MIT Lincoln Laboratory, with read noise of 11 electrons at 2-kHz frame rate. It is planned to use a 349-actuator deformable mirror made by Xinetics Inc. The temperature in the dome is maintained near 0° C, at which temperature the properties of PMN material normally used for precision deformable mirrors is less than ideal. A 37-element mirror is currently being evaluated in this environment. The wavefront processor performs the standard tasks of computing the wavefront gradients from the pixel data and then reconstructing the wavefront using matrix multiplication. The processing rate is compatible with a closed loop bandwidth of 100 Hz.

The laser is of similar design to that developed by LLNL for Lick Observatory. It is a neodymium–yttrium aluminium garnet (Nd:YAG)-pumped, three-stage dye laser, giving an output of 20 W at a pulse rate of 30 kHz. The bandwidth is broadened by phase modulation to match the absorption profile of sodium. The pump lasers and dye master oscillator are located in a room in the dome; they are connected by optical fibers to the preamplifier and power amplifier stages of the dye laser, which are mounted on the Keck telescope. Also mounted on the telescope are an alignment and diagnostics table and the 0.5-m projection telescope.

Two science instruments are planned for use with the Keck II adaptive optics system: NIRC-2, a near-IR camera, and NIRSPEC, a near-IR echelle spectrograph. The camera, built at CalTech, employs a 1024 x 1024 InSb array with a scale of 0.02 arc second per pixel. The spectrograph is being built at the University of California, Los Angeles (UCLA) and will employ slit sizes as small as 0.1 arc second. The adaptive optics system using natural guide stars is scheduled for installation on the Keck II telescope in 1998 and is expected to be in operation at the end of the year. The laser facility is planned for installation in 1999.

10.3.14 *Gemini 8-m Telescopes Project*

The Gemini Project consists of two 8-m telescopes, Gemini North at Mauna Kea, for which first light is expected in 1998, and Gemini South at Cerro Pachon, Chile, with first light expected in the year 2000. An adaptive optics module using natural guide stars is planned for Gemini North, with possible future upgrade to laser beacons. It is expected that adaptive optics using laser beacons will eventually be installed at Gemini South. [Simons and Gillett, 1996].

The Gemini telescopes are equipped with fast tip-tilt-focus secondary mirrors that can correct the lowest order atmospheric aberrations, as well as wind shake, over a wide field of view. The primary mirrors are of the meniscus type, their figures being controlled with active optics that remove most low-order aberrations. The remaining optical errors, to be corrected by the adaptive optics, are therefore of relatively high order. The adaptive optics requirements have been carefully evaluated in relation to the science objectives of the Gemini program [Morris et al. 1996]. The baseline parameters are:

- operation with natural guide stars;
- Shack-Hartmann wavefront sensor, 12 x 12 subapertures;
- wavefront sensor CCD detector read noise of 3 electrons per pixel;
- deformable mirror optically conjugated to turbulence altitude;
- Strehl ratio goal of 0.1-0.8 over the 0.85-2.5 μm spectral range;
- compensated field of 20-30 arc second diameter with a Strehl ratio of at least 0.5 in the K-band (2.2 μm).

The adaptive optics package will be mounted on the Cassegrain instrument support structure of the Gemini North telescope, where the corrected beam

can be fed to any of the instrument ports by means of a fold mirror. The initial science instrumentation planned for Phase I is as follows [Gillett and Mountain 1996]:

- Near-IR camera, spectral range 1-5 μm, using a 1024 x 1024 InSb array, with plate scales of 0.02, 0.05, and 0.12 arc second per pixel, giving fields of view between 20 and 120 arc seconds. This instrument will have a very low internal background, consistent with the low emissivity of the telescope, and will be used with and without the adaptive optics, at Mauna Kea.
- Near-IR spectrograph, 1-5 μm, also using a 1024 x 1024 InSb array, with spectral resolutions of 2000 and 8000. The plate scales will be 0.05 and 0.15 arc second per pixel, giving fields of view between 50 and 150 arc seconds. An integral field module is an option. This instrument is for use at Mauna Kea.
- Two multi-object spectrographs, operating over the spectral range 0.36–1.1 μm, each employing three 2k x 4k CCD arrays, with an image scale of 0.08 arc second per pixel and a field of view of 5 arc minutes. The spectral resolution is up to 10,000. These instruments are also equipped with integral field modules. The Mauna Kea spectrograph will have coatings optimized for the red spectrum, while the Cerro Pachon instrument will be optimized for the blue spectrum.
- Mid-I R Imager, spectral range 8-30 μm, using a 256 x 256 SiAs array with a pixel scale of <0.13 arc second per pixel. This instrument will be used initially on Gemini North, but will be available to Gemini South.
- High-resolution optical spectrograph, designed for use at the Cassegrain focus of Gemini South, will have high throughput at UV wavelengths. It will use two 2k x 4k CCD arrays and will have resolutions of 50,000 and 120,000.

Development and upgrading of science instruments will continue during Phase II of the program.

10.3.15 Max-Planck-Institute fur Astronomie

The Max-Planck-lnstitute fur Astronomie (MPIA) in Heidelberg, together with the Max-Planck-Institute fur Extraterrestriche Physik, Garching, has developed an adaptive optics system for the MPIA 3.5-m telescope at Calar Alto, Spain. The system is designed for use with a sodium laser beacon and is known as ALFA (Adaptive optics with Laser For Astronomy). It employs a Shack-Hartmann wavefront sensor using a low-noise CCD detector, together with a 97-actuator, continuous-faceplate deformable mirror. A separate steering mirror is used for image stabilization. The sodium beacon is generated by a dye laser, pumped by a continuously operating argon-ion laser. The ALFA system first operated in closed-loop mode in July 1997, using fixed guide stars of magnitude $m_V = 9.7$.

10.3.16 Palomar Adaptive Optics System

A 5-year program was started in November 1994 to provide a facility adaptive optics system for the 5-m Hale telescope at Mount Palomar Observatory. The adaptive optics is being built by the Jet Propulsion Laboratory in return for 25% of the Hale telescope observing time. The instrument is mounted at the Cassegrain focus and will initially use natural stars for wavefront sensing, although it will also be capable of operation with a sodium beacon. The wavefront corrector is a continuous facesheet deformable mirror made by Xinetics, with a total of 349 actuators, of which 241 are active within the 5-m telescope aperture. The wavefront sensor operates at visible wavelengths, employing a phase-shifting interferometer with a 64 x 64 CCD camera operating at a frame rate of 500 Hz, using "skipper" technology to minimize the read noise. The IR science camera is mounted on the optical axis, with a dichroic beamsplitter to separate the light from the off-axis reference star. The optical system is optimized for observations in the K band around 2.2 μm; it is also usable in the J and L bands (1.2-3.5 μm). The first "lock" of the adaptive optics instrument is scheduled for early 1998, with delivery of the operational system later the same year.

10.3.17 Anglo-Australia Telescope

The 3.9-m Anglo-Australian Telescope (AAT) at Siding Spring, New South Wales, has been in operation since 1974 and is one of the prime instruments in the southern hemisphere. The adaptive optics program, sponsored by the Australian Research Council, is being performed by a consortium of Australian Universities and the Anglo-Australian Observatory. The objectives are to prove the effectiveness of adaptive optics at the AAT site, and to implement a system that will provide a scientifically valuable gain in the resolution and efficiency of the telescope. The first stage of the project was a fast tip-tilt correction system, installed in the west coude room of the AAT in 1995. A low-order adaptive optics system based on a 19-element bimorph mirror and a curvature sensor is now being implemented [O'Byrne 1995]. This system will produce the largest gain in image quality at infrared wavelengths and will be used primarily with the IR camera/spectrograph.

10.3.18 National Astronomical Observatory of Japan

The National Astronomical Observatory (NAO) of Japan is the sponsor of the Subaru Telescope at Mauna Kea, Hawaii. The primary mirror of this telescope is an 8.3-m diameter meniscus, with a thickness of only 20 cm. The surface figure of the mirror is controlled by an active supporting structure employing a total of 264 actuators, which apply forces to counteract the effects of gravity, temperature variations, and wind pressure. The sensors employed in this active control system are mechanical. To compensate rapidly changing wavefront errors due to atmospheric turbulence, the Subaru Instrumentation Program [Iye 1996] includes a Cassegrain adaptive optics system, optimized for observations in the near infrared. The objective of the system is to achieve diffraction-limited resolution of 0.067 arc second for observations at 2.2 μm, with useful performance over the spectral range of 1 to 5 μm. The adaptive optics is based on curvature sensing and correction, using 36 controlled elements. The curvature sensor employs 36 avalanche photodiodes, which control a 60-mm diameter bimorph deformable mirror with 36 electrodes. The control system makes 2000 corrections per second. The field of view for full correction at 2.2 μm is 60 arc seconds, with image motion corrected over 120 arc seconds. The Subaru adaptive optics system will be used with the InfraRed Camera and Spectrograph (IRCS) at wavelengths between 1.0 and 5.4 μm, and also with the Coronagraphic Imager with Adaptive Optics (CIAO), a diffraction-limited coronagraph for observations between 0.9 and 3.8 μm. First light for the Subaru telescope is planned for the summer of 1998, with full operations expected in 2000.

10.3.19 Large Binocular Telescope Project

The Large Binocular Telescope (LBT) is a collaboration between the University of Arizona, the Arcetri Astrophysical Observatory in Florence (representing the Italian astronomical community), Ohio State University, a consortium of German astronomical research institutions, and the Research Corporation in Tucson. It will be located at Mt Graham (3200 m) in the Sonora desert of Arizona. The LBT consists of two 8.4-m mirrors. with center-to-center separation of 14.4 m, on a common elevation-over-azimuth mount. The two primary mirrors have a combined collecting area equivalent ot that of an 11.8-m circular aperture, spanning a dimension of 22.8 m. The telescope structure is very compact and stiff, due to the small f-number of the primaries, f/1.14. Because the mirrors are on a common mount, the optical pathlengths remain constant as celestial objects are tracked, eliminating the need for optical delay lines.

Two sets of secondary mirrors are provided, to implement the three focal stations:

- Cassegrain focus, f/3.8, field of view 60 arc seconds;
- Gregorian (1R) focus, f/15, field of view 4 arc minutes;
- Phased combined focus (using tertiary flats), f/15, field of view 8 x 4 arc minutes.

The facility instruments initially planned for the LBT include visible and near-IR cameras, a faint-object spectrograph, and a medium-resolution IR spectrograph. Long-term plans include an interferometric imager (taking advantage of the full baseline), together with an imager and spectrograph for the thermal IR.

Adaptive optics is included in the LBT design to enable the full benefit of interferometric imaging to be obtained. When used in the interferometric imaging mode, the binocular design of the telescope allows sampling of all spatial frequencies produced by baselines of up to 22.8 m. Interferograms can be obtained with the full baseline at any position in the sky. By observing an object at different sky rotation angles, two-dimensional images can be reconstructed with the full resolution of the telescope.

10.4 Current Trends

It is evident from reviewing the projects listed above that adaptive optics is now considered to be an essential component in the new generation of large astronomical telescopes. In most cases, adaptive optics is intended (at least initially) to provide good correction at near-IR wavelengths between 1.25 and 3.4 /urn, with partial correction at shorter wavelengths. As ground-based telescopes get larger, they depend more and more on adaptive optics to achieve their full imaging diffraction-limited resolution. An efficient way to incorporate adaptive optics into an astronomical telescope is to use an adaptive secondary mirror. This approach is used in the 6.5-m MMT conversion [Bruns et al. 1995] and is also being considered for the Gemini telescopes [Bigelow et al. 1993, 1994] and the LBT [Salinari et al. 1993]. The main disadvantage of this approach is that it precludes any significant enlargement of the compensated field of view.

Another trend is also evident: adaptive optics is being retrofitted to older telescopes with apertures of up to 4 m, to make them competitive with the new generation of astronomical giants. As shown in chapter 4, the peak image intensity delivered by a fully compensated telescope is proportional to D^4, while in an uncompensated telescope it is $D^2 r_0^2$. A compensated telescope with an aperture of about 1.5 m will therefore give the same peak image intensity as an uncompensated 8-m-aperture instrument in the best seeing conditions. A large number of telescopes in the 1.5-2.5 m class exist in observatories around the world; these instruments can benefit greatly by

upgrading with adaptive optics. A relatively simple AO system can be very effective at near-IR wavelengths and there is potentially a large market for reliable and low-cost systems of this type. For retrofitting existing telescopes, a self-contained adaptive optics package that can be attached to the Cassegrain, Nasmyth, or coude focus is a good solution. The adaptive optics bonnette designed for the CFHT [Arsenault et al. 1994] is an example of this approach. The unit provides a compensated beam with the same f-number as that of the input, and the compensating optics can be switched in or out as required.

Obtaining good sky coverage at shorter wavelengths remains the major problem with astronomical adaptive optics. Laser beacons generated in the sodium layer are clearly the preferred reference sources, because of their high efficiency and smaller measurement errors due to the high altitude. The use of laser beacons increases the cost of observing time, but may enhance its value by a larger factor. Most AO systems retain the capability of using natural stars when they are available, and it makes sense to keep this option open. Useful science can be carried out with adaptive optics using natural reference sources, as shown by Roddier et al [1994] and Lena [1994].

10.5 Future Prospects

Adaptive optics is no longer in its infancy, but it still has a long way to go before it can be considered a mature technology. Major improvements are required to expand the compensated field of view, to improve compensation at short wavelengths, and to increase the sky coverage. It is possible that these problems will be solved in an unexpected way: the opportunities for new ideas and approaches are immense.

From the present perspective, several avenues of investigation can be seen. The following topics appear to be rewarding areas for further development in adaptive optics:

- Compensation devices with high spatial and temporal resolution, possibly using liquid crystal or solid-state technology. With a high actuator density, piston-only correction of phase is adequate, providing that there is minimum dead space between elements.
- Multiconjugate compensation to enlarge the field of view. Developments are required in three areas: widefield wavefront measurement, multiple wavefront correction devices, and multiconjugate processing algorithms.
- Optimal algorithms for real-time control, using the best combination of reference sources (laser beacons and natural stars), together with adjustment of spatial and temporal bandwidths, to optimize the compensation efficiency for current turbulence conditions.
- Optimal postprocessing algorithms to improve compensated image quality by using measured data on residual wavefront errors. One approach is to use partial (real-time) compensation to raise the signal-to-noise ratio of the image data only to the extent that deconvolution algorithms are effective. Diffraction-limited images are then retrieved using the measured residual wavefront errors.
- Simplified operator interface with the adaptive optics, the goal being to make the operations mostly automatic. The functions should include routine preobservation and postobservation procedures, such as readiness tests, diagnostics, selection of reference stars, and data recording, as well as real-time optimization during the observing session, including control of the beacon laser.

The ultimate goal of adaptive optics in astronomical telescopes is to make Earth's atmospheric window transparent to the observer, who will be aware only of the superb performance of the telescope and its instrumentation.

Appendix A
Estimating the Position of an Image

A.1 Introduction

The discussion of wavefront sensing techniques (chapter 5) has shown that the process of wavefront slope measurement reduces to estimating the position of a spot in the case of a Hartmann sensor, or to estimating the position of a fringe pattern in the case of a shearing interferometer. In each case, the error in the wavefront measurement is directly related to the accuracy with which the position of the spot or fringe is determined. This is limited ultimately by the quantum nature of light and it is also limited by the spatial resolution of the image sensor employed. In this appendix, based on an analysis by Wallner [1985], the error produced by an arbitrary position estimator is derived for both continuous and discrete sensors. The optimum estimator is then found. The analysis enables the performance of any type of wavefront slope sensor to be determined and compared with the optimum. The results are particularly useful in the design of adaptive optics systems, because they enable a realistic tradeoff to be made between wavefront sensor error and hardware complexity. It is found, in general, that relatively simple position detection systems can be used in wavefront slope sensors without significant loss of performance.

A.2 General Position Estimator

The distribution of photoelectrons in an image is a random function with a mean or expected value described by the intensity distribution, which is determined by the characteristics of the source of radiation and the optical configuration of the wavefront sensor. In the case of an unresolved point source and an aberration-free imaging system, the intensity distribution is simply the diffraction-limited response of the system. It is assumed that the shape of this function is known and only its position is uncertain.

Photons arrive and are detected randomly. At each point in the image, the expected value of the photoelectron density is given by the intensity function $f(x,y)$. If the image is displaced by Ax, Ay, the expected electron density will be

$$\langle \zeta(x, y) \rangle = f(x + \Delta x,\ y + \Delta y) \qquad \text{(A.1)}$$

where

$\zeta(x,y)$	=	charge density in the image, in electrons per radian squared
(*)	=	expected value
$f(x,y)$	=	nominal image intensity
Ax, Ay	=	displacement from nominal position

The displacement of the image on each axis is estimated by weighting the electron distribution with an estimating function $g(x,y)$ and integrating over the image plane. The measure of displacement in the x axis obtained in this way is

$$m_x = \int dx \int dy\ g_x(x, y)\ \zeta(x, y) \qquad \text{(A.2)}$$

The expected value of this measure is simply the weighted integral of the expected value of the electron distribution:

$$\langle m_x \rangle = \int dx \int dy\, g_x(x,y)\langle \zeta(x,y)\rangle = \int dx \int dy\, g_x(x,y)\, f(x+\Delta x,\ y+\Delta y) \tag{A.3}$$

For small displacements, the expected value of the electron distribution may be expanded about its nominal position

$$f(x+\Delta x,\ y+\Delta y) = f(x,y) + f_x'(x,y)\Delta x + f_y'(x,y)\Delta y \tag{A.4}$$

where $f_x'(x,y) = \partial/\partial x\, f(x,y)$, and so on. Substituting in equation (A.3), we obtain

$$\langle m_x \rangle = \int dx \int dy\, g_x(x,y) \times \left[f(x,y) + f_x'(x,y)\Delta x + f_y'(x,y)\Delta y \right] \tag{A.5}$$

The value of $\langle m_x \rangle$ may be adjusted to zero when $Ax = Ay = 0$ by adding a constant term to $g_x(x,y)$. If this is done,

$$\int dx \int dy\, g_x(x,y)\, f(x,y) = 0 \tag{A.6}$$

The equations for m_y are similarly functions of the y-axis estimating function $g_y(x,y)$.

In general, the one-dimensional measures of image displacement m_x and m_y depend on both Ax and Ay, and two equations must be solved simultaneously to get the desired values. The subsequent analysis is considerably simplified by confining the displacement to a single axis. If the intensity distribution $f(x,y)$ is continuous, axes may always be chosen to eliminate the Ay term from equation (A.5). In particular, *if* $f(x,y)$ is symmetrical about the x and y axes and $g_x(x,y)$ is symmetrical about the y axis, conditions that are easily arranged in practice, then the offending term will disappear.

Under these conditions, the estimated deviation in position in the x axis is given by

$$\Delta\hat{x} = \frac{m_x}{\int dx \int dy\, g_x(x,y)\, f_x(x,y)} \tag{A.7}$$

The variance of the estimated position (that is, the mean-square deviation from its average value) is then

$$\sigma^2_{\Delta\hat{x}} = \frac{\sigma^2_{m_x}}{\left[\int dx \int dy\, g_x(x,y)\, f_x(x,y)\right]^2} \tag{A.8}$$

The variance of the measure m can be shown to be

$$\sigma^2_m = \int dx \int dy\, g^2(x,y)\, f(x,y) \tag{A.9}$$

Substituting this value in equation (A.8) gives the mean-square error in estimated position, for small excursions around the null:

$$\sigma^2_{\Delta\hat{x}} = \frac{\int dx \int dy\, g^2(x,y)\, f(x,y)}{\left[\int dx \int dy\, g(x,y)\, f_x(x,y)\right]^2} \tag{A.10}$$

This equation can be used to evaluate the performance of any estimating function $g(x,y)$. For example, substituting the weighting function $g(x,y) = x$ in equation (A.10) gives the mean-square error on one axis using the image centroid as the estimate of position.

A.3 Position Sensing with Discrete Detector Arrays

The weighting function $g(x,y)$ in the previous section was assumed to be a continuous function of position. A possible method of implementation would be to use variable attenuation optical filters that have the desired transmission function, with four separate detectors to handle the positive and negative weights in x and y. Such a system would require image intensification to ensure that quantum noise was still dominant after the attenuation in each path. An alternative approach would be to use a detector array that has very high spatial resolution. Neither of these methods is attractive from a practical viewpoint. A more useful approach is to use a detector array that has a relatively small number of elements, the outputs of which are weighted and combined. This is equivalent to a weighting function that is constant over each detector area. Using square pixels with side p, such a weighting function is defined as

$$g(x,y) = g_i \qquad x_i < x < (x_i + p),\ y_i < y < iy, + p) \tag{A.11}$$

where

g_i = weight applied to rth pixel
x_i , y_i = location of lower left corner of pixel

Substituting this relation in equation (A.10) gives the mean-square error using a quantized measure:

$$\sigma^2_{\Delta\hat{x}} = \frac{\sum_i g_i^2 a_i}{\left[\sum_i g_i b_i\right]^2} \tag{A.12}$$

where

$$a_i = \int_{x_i}^{x_i+p} dx \int_{y_i}^{y_i+p} dy\, f(x,y) \tag{A.13}$$

$$b_i = \int_{x_i}^{x_i+p} dx \int_{y_i}^{y_i+p} dy\, f_x'(x,y) = \int_{y_i}^{y_i+p} dy\, [f(x_i+p,y) - f(x_i,y)] \tag{A.14}$$

These equations enable the error in determining the position of a spot or fringe to be evaluated as a func-

tion of the number of photoelectrons counted, for any detector configuration and set of weights.

A.4 Optimum Weighting Functions

The optimum set of weights that minimizes the mean-square position error when using a discrete detector array is found by differentiating equation (A.12) with respect to g_i and equating to zero. The result is

$$\frac{d\sigma^2_{\Delta\hat{x}}}{dg_i} = \frac{\left(\sum_j g_j b_j\right)^2 2g_i a_i - \left(\sum_j g_j^2 a_j\right) 2\left(\sum_j g_j b_j\right) b_i}{\left(\sum_j g_j b_j\right)^4} = 0 \tag{A.15}$$

This reduces to

$$g_i a_i \left(\sum_j g_j b_j\right) = b_i \left(\sum_j g_j^2 a_j\right) \tag{A.16}$$

which has the solution

$$g_i^* = \frac{b_i}{a_i} \tag{A.17}$$

The expression for the optimum gain or weight for each pixel is intuitively satisfying. The factor b_i is proportional to the derivative of the intensity function, so the gain is maximized at pixel locations with large slope and is zero when the slope is zero. The factor a_i is proportional to the image intensity; pixel locations having low intensity and, consequently, low quantum noise have a higher weighting because a change in value is more likely to be due to a shift in position than to quantum fluctuations.

Substituting these weights back into equation (A.12) gives the optimized variance for the measurement error in a single axis using a discrete detector:

$$\sigma^2_{\Delta\hat{x}}{}^* = \left[\sum_j \frac{b_j^2}{a_j}\right]^{-1} \tag{A.18}$$

Note that these weights are optimized only for small excursions around the null. It will be shown later that other weighting functions may give better overall performance for larger excursions.

The optimum weighting function for the continuous case can be found as the limit of equation (A.17) as the pixel size approaches zero. For small pixels, the integrands in equations (A.13) and (A.14) will not vary significantly over the area of integration, so

$$\begin{aligned} a_i &\Rightarrow p^2 f(x_i, y_i) \\ b_i &\Rightarrow p^2 f_x'(x_i, y_i) \end{aligned} \tag{A.19}$$

The optimum gain for the continuous case is then

$$g^* = \frac{f_x'(x, y)}{f(x, y)} \tag{A.20}$$

The variance of the optimized estimate for a continuous detector is

$$\sigma^2_{\Delta\hat{x}}{}^* = \left[\int dx \int dy \frac{[f_x'(x, y)]^2}{f(x, y)}\right]^{-1} \tag{A.21}$$

A.5 Performance of Hartmann Sensors

These results may now be applied to evaluate the performance of Hartmann sensors in which the error in determining the wavefront slope is directly related to the error in estimating the position of the spot. Hartmann sensors generally employ square subapertures. The intensity function is then

$$f(x, y) = I_0 \left[\frac{\sin\left(\frac{\pi dx}{\lambda}\right)}{\left(\frac{\pi dx}{\lambda}\right)}\right]^2 \left[\frac{\sin\left(\frac{\pi dy}{\lambda}\right)}{\left(\frac{\pi dy}{\lambda}\right)}\right]^2 + I_B \tag{A.22}$$

where

$$I_0 = \frac{d^2}{\lambda^2} n_s \tag{A.23}$$

and

- $f(x,y)$ = image intensity in photoelectrons per steradian
- x,y = field angle, rad
- I_0 = image intensity on axis in photoelectrons per steradian
- d = dimension of square subaperture
- λ = mean wavelength
- n_s = number of signal photoelectrons
- I_B = background intensity in image plane, in photoelectrons per steradian

The measurement errors for optimized pixel weighting are then computed using equation (A.18) for a discrete detector and equation (A.21) for a continuous detector. In the discrete case, the error depends on the position of the image center in relation to the pixel array. The most favorable image location is at the junction of four pixels. Limiting cases of the rms position error for circular and square apertures, with and without background radiation, for small and large pixel sizes are listed in table A.1.

For a circular aperture with a quad cell detector and no background radiation, the rms error is $(3\pi/16)(\lambda/dn_s^{1/2})$, which is the same as that derived by Tyler and Fried [1982]. The present analysis enables the error to be derived for any intensity function, detector size, pixel weighting, and background. For zero background, it is seen that a continuous detector produces only about half the error of a

Table A. 1 Errors in Estimating the Position of a Diffraction-Limited Spot

Position error $\sigma_{\Delta x}$ in radians rms, using optimum pixel weights

1. Circular aperture, spot width $r_a = 1.22\lambda/d$		
Background intensity	$I_B = 0$	$I_B >> n_s d^2/\lambda^2$
Pixel size $p \ll r_a$ (continuous detector)	$\sigma_{\Delta x} = 0.261\ r_a/n_s^{1/2}$	$\sigma_{\Delta x} = 0.65\ I_B^{1/2} r_a^2/n_s$
Pixel size $p \gg r_a$ (quad cell)	$\sigma_{\Delta x} = 0.483\ r_a/n_s^{1/2}$	$\sigma_{\Delta x} = 1.11\ I_B^{1/2}\ r_a p/n_s$
2. Square aperture, spot width $r_a = k/d$		
Background intensity	$I_B = 0$	$I_B >> n_s d^2/\lambda^2$
Pixel size $p \ll r_a$ (continuous detector)	$\sigma_{\Delta x} = 0.277\ r_a/n_s^{1/2}$	$\sigma_{\Delta x} = 0.755\ I_B^{1/2}\ r_a^2/n_s$
Pixel size $p \gg$ r, (quad cell)	$\sigma_{\Delta x} = 0.500\ r_a/n_s^{1/2}$	$\sigma_{\Delta x} = 1.05\ I_B^{1/2}\ r_a p/n_s$

quad cell. The error coefficients for circular apertures in table A.1 are smaller than those for square apertures for the same photon count. However, square apertures collect more light than circular ones of the same width, and, when this is taken into account, the square apertures normally used in Shack-Hartmann sensors give a better performance.

In the design of Shack-Hartmann sensors, the design parameters of interest are:

1. The size of the detector pixels in relation to that of the optical image;
2. The number of pixels required per subaperture: for example, 2×2, 4×4, and so on;
3. The optimum weighting or gain of each pixel;
4. The dynamic range (wavefront tilt excursion) of each subaperture;
5. The effect of a gap between the pixels.

These parameters are readily determined using the above analysis, enabling the design to be optimized for any set of conditions.

The relationship between pixel size and the image intensity distribution is shown in figures A.1 and A.2. The images are produced by diffraction-limited square subapertures. Figure A.1 represents a quad cell, for which the (angular) pixel size is usually larger than the diffraction image. With a displacement of $\pm 1/2$ wave in x and y, the image is mostly within one quadrant, resulting in a small linear range. The use of a 4 x 4 detector array, with a pixel size of 1.33 x λ/d, as shown in figure A.2, greatly increases the dynamic range. The linearity of the response to image displacement can be controlled by weighting the pixels, as shown in section 5.3.

Normalized position measurement errors for small excursions around the null are shown in figure A3, as a function of the pixel size. The pixel size is specified in units of k/d, where k is the mean wavelength of the wavefront sensor and d is the diameter of a circular subaperture or the side of a square subaperture. A pixel size of 0.1 units approaches the performance of a continuous detector, while a pixel size of 10 units is equivalent to a quad cell. Curves (a) and (b) show the normalized rms position errors for an unlimited number of detector elements, using circular and square apertures, respectively, and using the optimum pixel weighting as defined in equation (A. 17). It is seen that the error grows steadily as the pixel size increases, with a steep increase between $0.5\lambda/d$ and $1.0\lambda/d$, above which it remains constant.

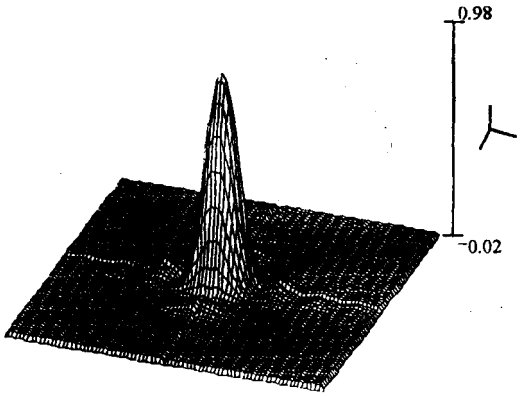
The chart shows that the error in determining the position of a spot is minimized by using small pixels, providing that the array covers the area of the spot plus the dynamic range required, which may require a large number of pixels per subaperture.The error saturates at about twice the minimum value when the pixel size reaches λ/d, remaining constant even for the large pixels used in a quad cell. Therefore, considering only the error around the null, a large increase in the number of pixels is required to improve on the performance of a simple quad cell detector.

This picture changes when the transfer function of the position sensor is considered. Figure A.4 shows the normalized error and the output voltage of the position sensor as a function of the spot displacement. The curves are plotted for a 4 x 4 array with two different weightings and also for a 2 x 2 array (quad cell). The pixel size is fixed at 1.33 x λ/d. The dotted curve shows a 4 x 4 pixel array with the optimized weights (equation A. 17) of —0.625, —1, +1, +0.625, giving minimum error in the presence of photon noise at the null position. These weights remain fixed as the spot is displaced. The solid curve shows a 4 x 4 array with weights of —3, —1, +1, +3. The dashed curve shows a quad cell with weights of —1, +1. For small spot displacements up to about 0.67 x k/d, the performance of all three systems is similar, the optimized weights giving only a very slight advantage.

For large displacements, the three systems behave quite differently. The system with optimized weights has a poor dynamic range, worse than a quad cell, with the output signal actually dropping as the displacement increases. The system shown by the solid curve has by far the best overall performance, giving maximum output and minimum error over the lar-

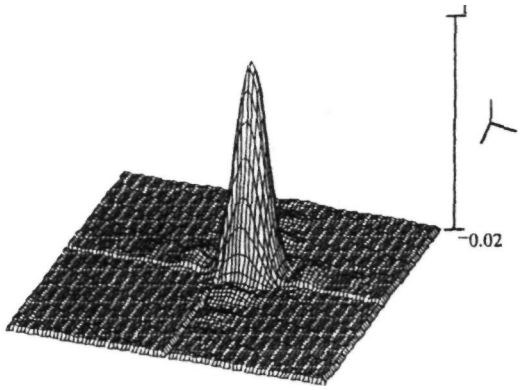
N = 2 ps = 4

Image displacement = 0



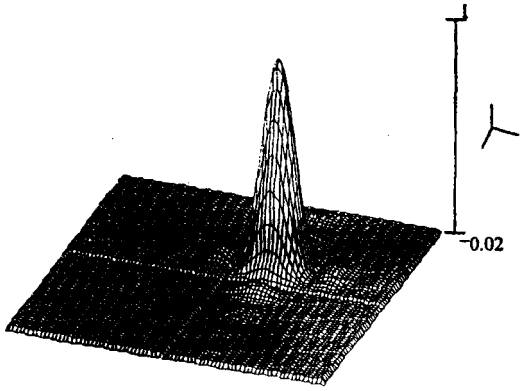
M1

Image displacement = 0.5 wave in x **and** y



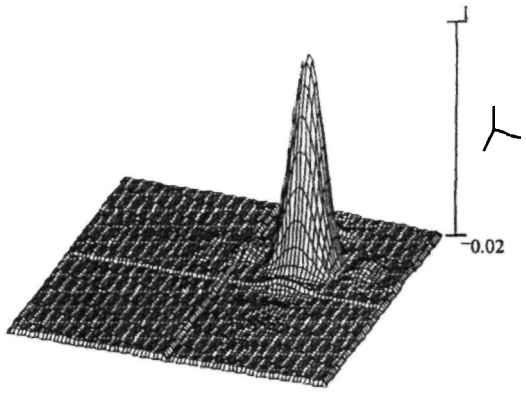
M2

Image displacement = 1.0 wave in x **and** y



M3

Image displacement = 1.5 wave in x **and** y

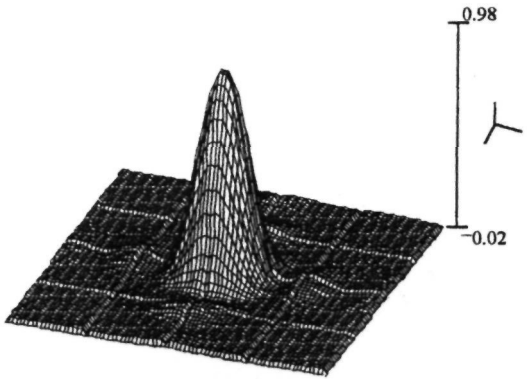


M4

Figure A.1 Intensity of a diffraction-limited spot produced by a square aperture, showing its relation to a quadrant (2 x 2) detector array for displacements of 0, 0.5, 1.0, and 1.5 waves.

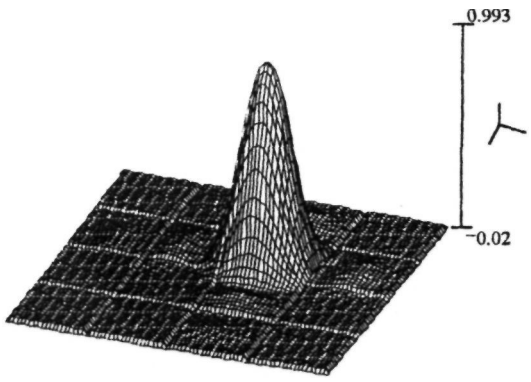
N = 4 ps = 1.33

Image displacement = 0



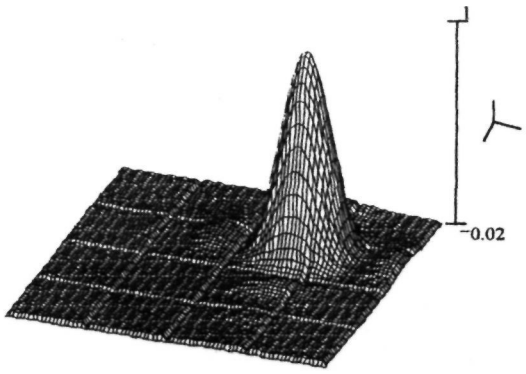
M1

Image displacement = 0.5 wave in x and y



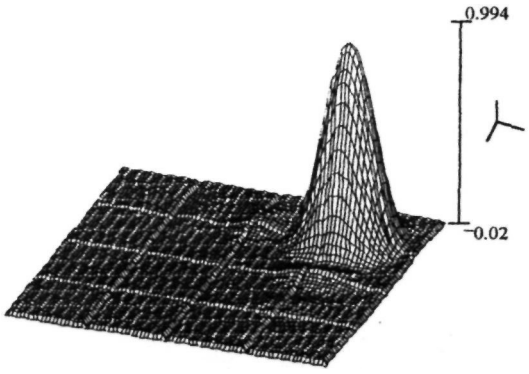
M2

Image displacement = 1.0 wave in x and y



M3

Image displacement = 1.5 wave in x **and** y



M4

Figure A.2 Intensity of a diffraction-limited spot produced by a square aperture, showing its relation to a 4×4 detector array for displacement of 0, 0.5, 1.0, and 1.5 waves.

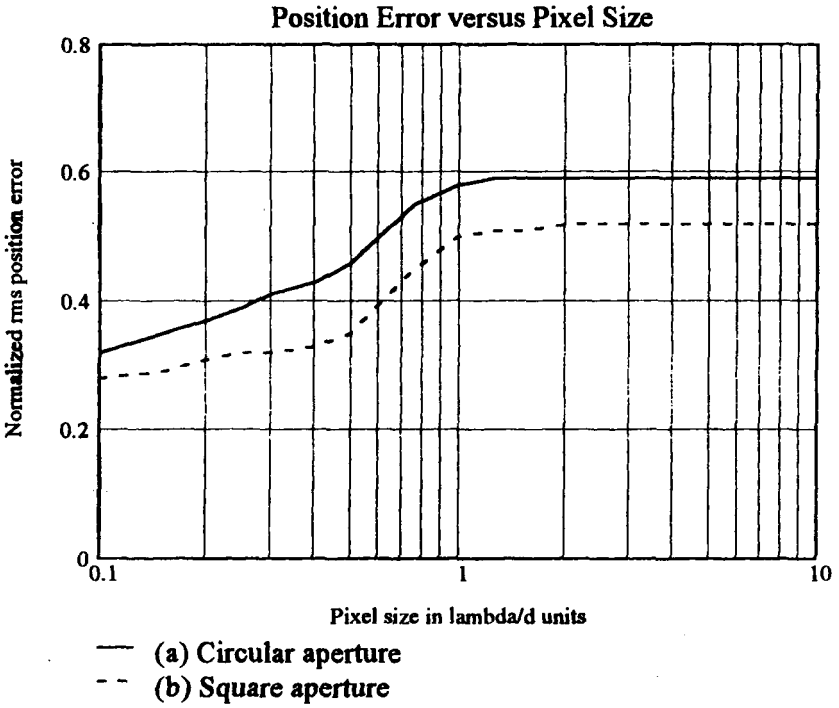


Figure A.3 Effect of the detector pixel size on the error in determining the position of a spot. The spots are produced by diffraction-limited circular and square apertures and are centered at the junction of four contiguous cells. It is assumed that the spot excursions are small compared with the spot size, which is usually the case in closed-loop tracking systems. Background light is zero. The left-hand side of the x-axis corresponds to very small pixels, approaching a continuous detector plane. The right-hand side corresponds to large pixels, equivalent to a quad cell. Changing the pixel size from very small to very large increases the error at the null by a factor of less than 2. Most of this change occurs at pixel sizes between 0.5 and 1.0 λ/d.

gest dynamic range. The system using a quad cell has significantly greater error than either of the others at moderate spot displacements, but performs well around the null. These results indicate that it is essential to explore the performance of position sensing systems over their entire dynamic range and not just around the null, where the conditions may be optimized.

The effect of a gap between pixels on the output signal and normalized error of position sensors is shown in figure A.5. Curves are shown for gaps of zero, 10%, 20%, and 30% of the pixel size. The presence of a gap always reduces the output signal and increases the error, although the latter effect is small for displacements of less than $0.5\lambda/d$. These effects become serious for gaps that exceed about 10% of the pixel size and are in addition to those produced by loss of light in the gaps.

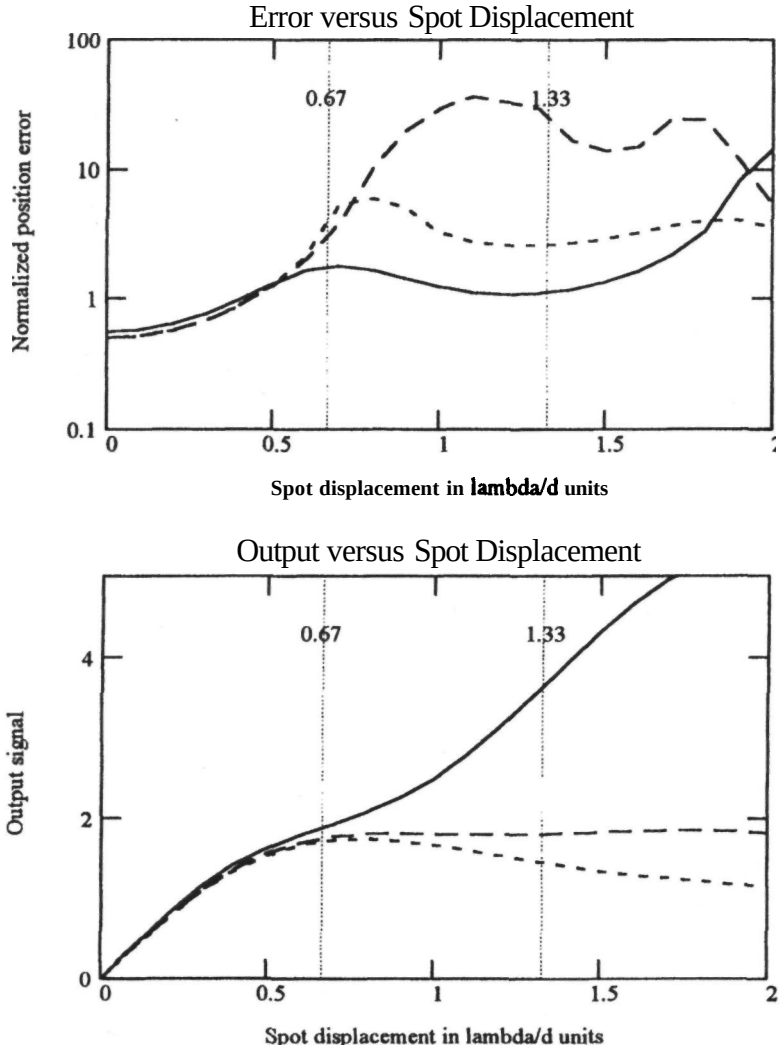


Figure A.4 Performance of position sensors as a function of the spot displacement. The spot is formed by a square subaperture of dimension *d*. The null position of the spot is at the center of the array, at the junction of four detector cells. The pixel size in the 4 x 4 array is 1.33 λ/d angular units. The optimized weights are those giving minimum error in the presence of photon noise at the null position. These weights remain fixed as the spot is moved. For displacement up to about 0.67 λ/d (half the cell size), the performance of all three systems is similar. For excursions greater than 0.67 λ/d, a 4 x 4 array with weighting of —3, −1, +1, +3 has superior performance, having a smaller error and larger dynamic range.

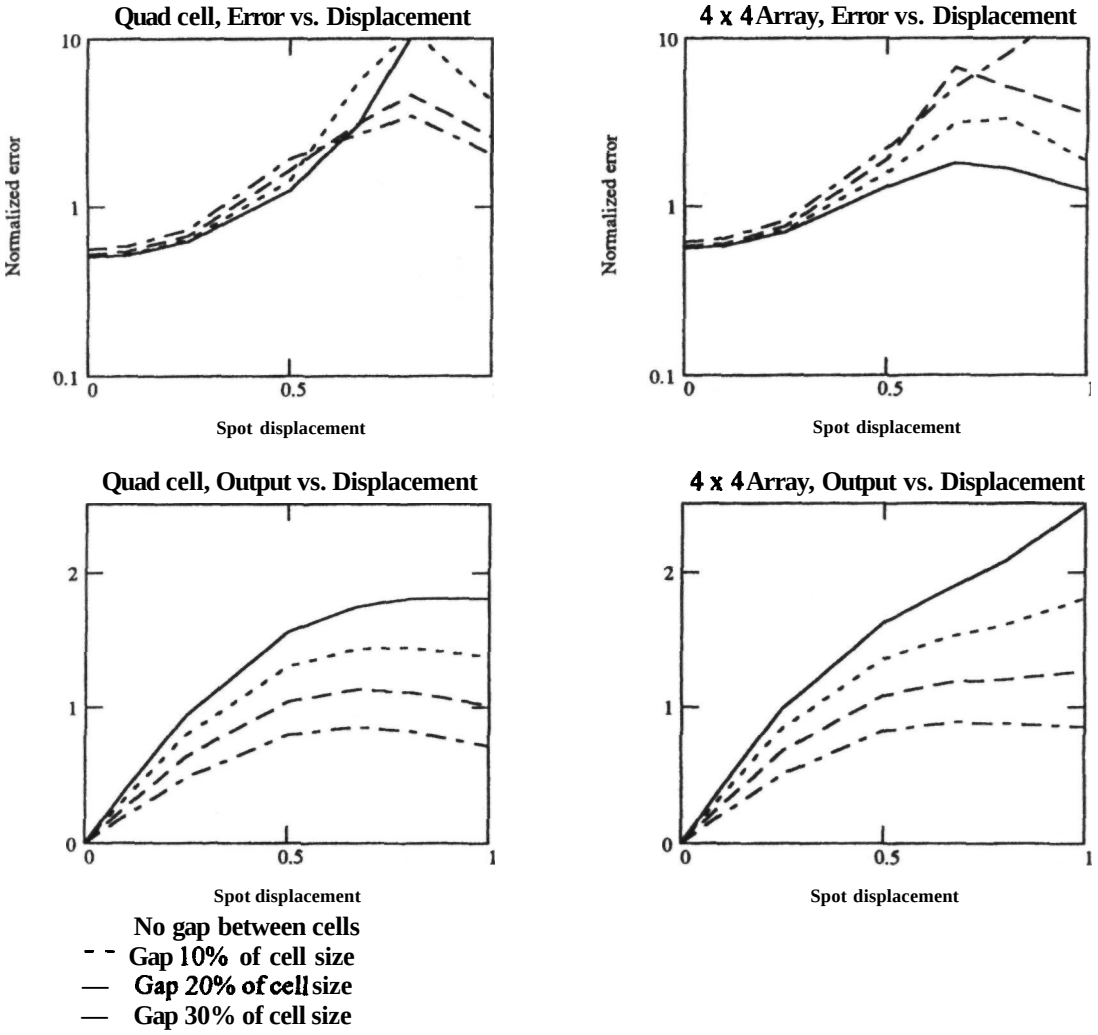


Figure A.5 Effect of a gap between detector cells on the performance of 2 x 2 (quadrant) and 4×4 position sensor arrays. A gap between adjacent detector cells always increases the error in measuring the position of a spot. The presence of a gap also reduces detector output and produces nonlinearity.

Appendix B

Active Control for Long-Baseline Interferometers

B.I Introduction

In this appendix, the effects of atmospheric turbulence on the image structure of a ground-based Michelson interferometer are analyzed and a method of improving the performance using a combination of adaptive optics and automatic fringe tracking is described. Matched filters in multiple spectral bands are used to determine the optical pathlength errors and to estimate the mutual coherence function of the source. A practical implementation of the system is described and estimates of its performance are given.

B.2 System Model

The basic interferometer model is shown in figure B.1. The two collecting apertures are separated by a variable baseline *A*. The two beams are brought to a fixed separation *D* and focused by lens *L* onto the image plane *P*, where they overlap and interfere. The effective focal length of the system is *F*.

B.2.1 Ideal System

We first analyze an ideal system with no wavefront perturbations. The spectral intensity at the focal plane for a single rectangular aperture with an unresolved circular source is

$$I_k = I_{ok}\left(\operatorname{sinc}\frac{kax}{F}\right)^2\left(\operatorname{sinc}\frac{kby}{F}\right)^2 \qquad \text{(B.1)}$$

where I_{ok} is the peak spectral intensity in photons per meter squared per micrometer, given by

$$I_{ok} = n_k \pi \alpha^2 \tau \frac{4a^2b^2k^2}{\pi^2F^2} \qquad \text{(B.2)}$$

and

n_k = spectral radiance of source, photons m sr^{-1} s^{-1} μm
a = angular radius of source, rad
r = exposure time, s
a = half-width of aperture, m
b = half-length of aperture, m
k = wave number, μm^{-} = $2\pi/\lambda$
F = focal length of aperture, m

The fringe pattern is generated in the direction of the baseline, which is defined as the x axis. If the detectors are assumed to extend over the entire pattern in the y axis, then equation (B.1) may be integrated over y, reducing it to the following function of x only

$$I_{kx} = I_{okx}\left(\operatorname{sinc}\frac{kax}{F}\right)^2 \qquad \text{(B.3)}$$

where

$$I_{okx} = n_k \pi \alpha^2 \tau \frac{4abk}{\pi F} \qquad \text{(B.4)}$$

If two such apertures, separated by distance D, are focused at the same point and there are no phase disturbances, the spectral intensity becomes

$$I'_{kx} = 2I_{okx}\left(\operatorname{sinc}\frac{kax}{F}\right)^2\left[1 + \gamma_A \cos\frac{kDx}{F}\right] \qquad \text{(B.5)}$$

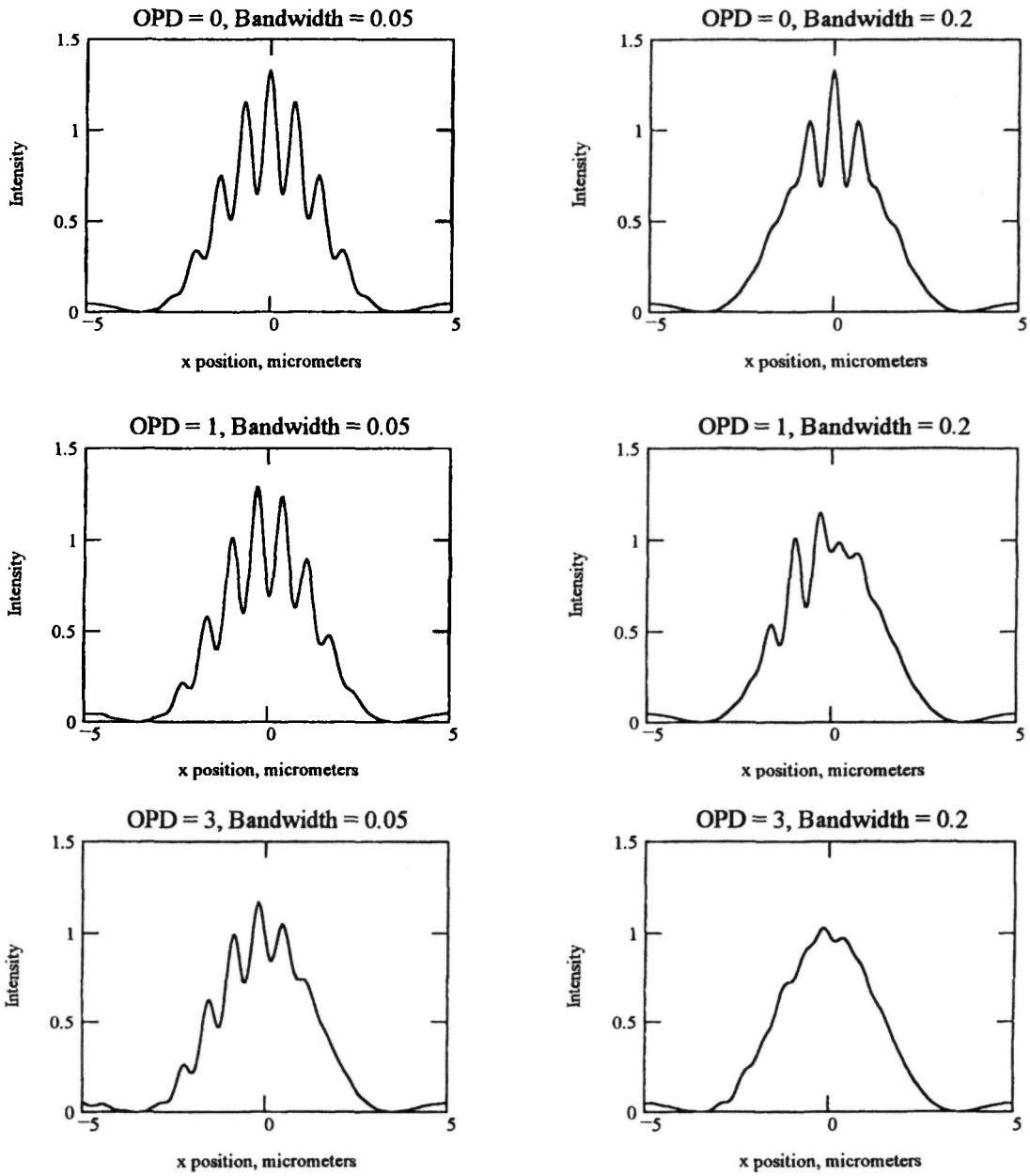


Figure B.2 Interferometer image structure, showing the effects of optical path differences (OPDs) and finite spectral bandwidths. The units for OPD and bandwidth are micrometers and the center wavelength is 0.7 μm.

phase, resulting in the unique "white-light" fringe. But measurement of the fringe phase away from the null is best achieved in narrow spectral bands, using a dispersed or "channeled" spectrum.

The condition for zero fringe contrast, which for a circular source occurs when $\gamma_A = 0$ in equation (B.6), is only satisfied at one wavelength for each value of *A*. Fringe modulation still occurs at higher and lower wavelengths, suggesting that measurement of fringe modulation is best accomplished with a spectrally separated image.

B.2.3 *Effects of Atmospheric Turbulence*

Atmospheric turbulence produces three distinct effects on the interference image in a multiaperture interferometer. These effects are:

1. differences in the absolute optical pathlength, from the object to the image, in the arms of the interferometer;
2. independent (overall) tilt of the apertures, resulting in displacement of the centroids, reducing the overlap area where the fringes are generated;
3. wavefront distortion within each of the apertures, which spreads the images over a larger area and reduces their intensity.

In the central portion of the image, these effects may be accounted for by rewriting the expression for focal plane intensity distribution [equation (B.8)], in the following form:

$$I'_{kx} = 2I_{ox}\left(\operatorname{sinc}\frac{\bar{k}ax}{F}\right)^2 \gamma_w\left[1+\gamma_A\gamma_k\gamma_P\cos\bar{k}\left(\frac{Dx}{F}+\Delta P\right)\right] \tag{B.11}$$

where

γ_k = fringe contrast factor due to spectral bandwidth and cross product of bandwidth and mean pathlength error *AP*

$$= \operatorname{sinc}\Delta k\left(\frac{dx}{F}+\Delta P\right) \tag{B.12}$$

γ_p = fringe contrast factor due to time varying pathlength error *AP*

γ_w = fringe contrast factor due to aperture wavefront error

The function of the adaptive control system is to optimize the values of γ_p and γ_w in real time.

The magnitude and temporal characteristics of the turbulence effects will now be summarized. Atmospheric turbulence may be characterized by a Kolmogorov spectrum within the inertial subrange defined by the inner and outer scales. The outer scale is limited by the height of the collecting apertures above the ground, which is usually much less than the baselines of ground-based interferometers (tens or hundreds of meters). Thus, the optical path variations in the two apertures may be considered statistically independent. To determine the temporal variation of optical path differences, the atmosphere can be treated as a series of frozen layers with varying wind velocities. In the present analysis, to obtain the basic results in the simplest way, a single layer of turbulence is used.

The appearance of spectrally dispersed fringes under various conditions is shown in figure B.3. Under ideal conditions, with exactly equal optical pathlengths and with the object unresolved by the aperture spacing, the fringes appear as shown in figure B.3(a). The fringe spacing is directly proportional to wavelength. At the center of the image, the fringes are coincident at all wavelengths, producing the unique white-light fringe. Under normal seeing conditions, the fringes are in constant motion and at one instant of time may appear as shown in figure B.3(b). Even at path equality *(R = L),* the fringes are randomly displaced in phase by atmospheric turbulence. The effect of a constant pathlength difference is to displace the central fringe toward the longer arm of the interferometer. Because of the spectral dispersion, a pathlength error makes the fringe pattern appear to rotate, as shown in the figure, with the direction of rotation revealing the polarity of the error.

The appearance of the dispersed fringes at path equality near the first null is shown in figure B.3(c). The first null ($\gamma_A = 0$) occurs when the aperture spacing (baseline) is A — $BX/6$, where λ is the observation wavelength, $B = 0.5$ if the object consists of two point sources of angular separation *6,* and $B = 1.22$ if the object is a uniform circular disk of angular diameter *6.* As the interferometer aperture spacing A is increased, the null first appears at the blue end of the spectrum and progresses through to the red end. The contrast γ_A at the edges of a band covering a 2:1 range of wavelengths spanning the null is about 40% at the red end and 15% at the blue end. It is therefore possible to maintain adaptive control over the interferometer pathlength, even in the vicinity of the null. Note that the presence of a null is made evident by the phase reversal of the fringes.

B.2.4 Fringe Detection Performance.

Successful operation of the interferometer depends on the ability to evaluate the term γ_A by measuring the fringe contrast in spite of the changes in fringe phase and the presence of the contrast degradation factors γ_P, γ_k, and γ_w, caused by atmospheric and instrumental perturbations.

Two outputs are required from the fringe detector:

1. a pathlength error signal to feed back to the pathlength compensator to minimize *AP;*
2. an estimate of the fringe contrast γ_A of the object observed.

The fringes may be detected by integrating the intensity after multiplying I'_{kx} in equation (B.11) by orthogonal "matched filter" functions. The resulting signals are

$$C = \int_{-\infty}^{\infty} dx\left(\operatorname{sinc}\frac{\bar{k}ax}{F}\right)^2 \operatorname{sinc}\frac{\Delta kDx}{F}\cos\frac{\bar{k}Dx}{F}I'_{kx}$$

$$S = \int_{-\infty}^{\infty} dx\left(\operatorname{sinc}\frac{\bar{k}ax}{F}\right)^2 \operatorname{sinc}\frac{\Delta kDx}{F}\sin\frac{\bar{k}Dx}{F}I'_{kx} \tag{B.13}$$

These spatial filters can be implemented either optically before detection, or electronically after detection. The latter more easily allows amplification of the signal before it is applied to the filters, leading to lower quantum noise.

For the narrow spectral band considered, the fringe frequency terms may be averaged over one cycle giving

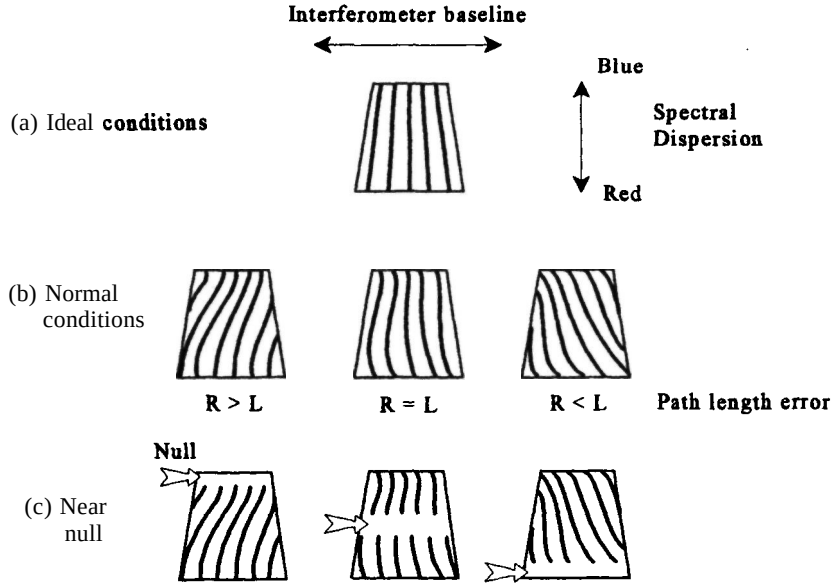


Figure B.3 Fringe patterns with spectral dispersion.

$$C = N\gamma_w\gamma_F \cos \phi(\bar{k})\, g_4(r)$$
$$S = N\gamma_w\gamma_F \sin \phi(\bar{k})\, g_4(r) \qquad \text{(B.14)}$$

where

N = total photons collected by both apertures in band Ak,

$$\gamma_F = \gamma_A \gamma_k \gamma_? \qquad \text{(B.15)}$$

$$\phi(\bar{k}) = \bar{k}\Delta P \qquad \text{(B.16)}$$

$$g_4(r) = \frac{1}{\pi}\int_0^\infty dz\, [\text{sinc}(z)]^4[\text{sinc}(zr)]^2 \qquad \text{(B.17)}$$

$$z = \frac{\bar{k}ax}{F} \qquad \text{(B.18)}$$

$$r = \frac{\Delta k}{\bar{k}}\frac{D}{a} \qquad \text{(B.19)}$$

The fringe phase can now be estimated from the two filter outputs:

$$\hat{\phi}(\bar{k}) = \tan^{-1}\left(\frac{S}{C}\right) \qquad \text{(B.20)}$$

The variance of this estimate can be expressed as

$$\sigma_\phi^2 = \frac{1}{N\gamma_w\gamma_F^2}\frac{g_6(r)}{g_4^2(r)} = \frac{1}{N\gamma_w\gamma_F^2 f(r)} \qquad \text{(B.21)}$$

where

$$g_6(r) = \frac{1}{\pi}\int_0^\infty dz[\text{sinc}(z)]^6\,[\text{sinc}(zr)]^2 \qquad \text{(B.22)}$$

The function $f(r)$ is plotted in figure B.4. The number of photons collected is proportional to Ak, which, in turn, is proportional to r, so that the signal-to-noise ratio of the phase estimate is proportional to $rf(r)$, which is also plotted on figure B.4. Increasing the bandwidth reduces the variance of the fringe phase angle estimate linearly up to $r = 1$, and little thereafter. To avoid ambiguous phase measurements, the mean-square fringe error should be limited to

$$\sigma_\phi^2 \le \left(\frac{\pi}{2}\right)^2 \qquad \text{(B.23)}$$

The pathlength error is given by

$$\Delta\hat{P} = \frac{1}{\bar{k}}\hat{\phi} \qquad \text{(B.24)}$$

To reduce the pathlength error variance, the estimates from n equal wavenumber bands may be combined, giving

$$\sigma_P^2 = \frac{1}{\bar{k}^2 n}\sigma_\phi^2 = \frac{1}{\bar{k}^2\gamma_w\gamma_F^2 Nf(r)} \qquad \text{(B.25)}$$

Here, N is the total number of photons detected in all spectral bands. To minimize the pathlength error for a fixed spectral bandwidth $2nAk$, it is necessary to minimize r. But, if r is too small, the photon error increases, violating the limit of equation (B.23). A good compromise is to choose $r \approx 1$, so that equation (B.25) becomes

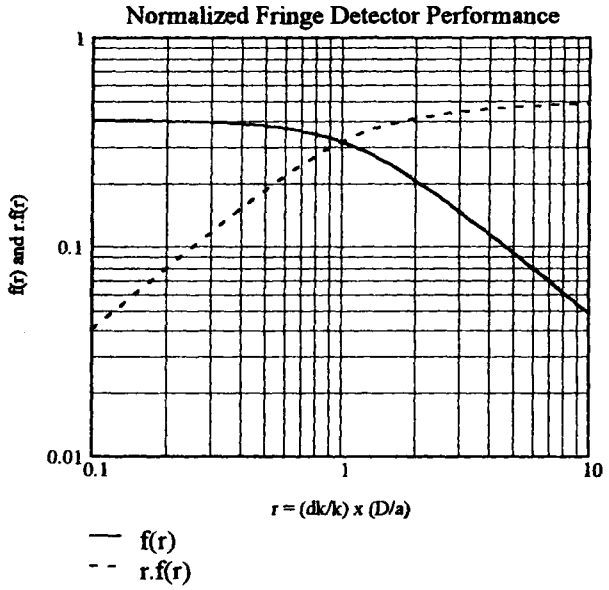


Figure B.4 Normalized performace of a fringe detection system.

$$\sigma_P^2 \approx \frac{3}{\bar{k}^2 \gamma_w \gamma_F^2 N} \tag{B.26}$$

The next step is to make an estimate of the fringe contrast, which is carried out using the cosine and sine functions defined in equation (B.14):

$$\gamma_F^2 = \frac{S^2 + C^2 - 2N\gamma_w g_6(r)}{[N\gamma_w g_4(r)]^2} \tag{B.27}$$

where N is the total number of photons collected. The required value of γ_A may then be found using equation (B.15). The actual values of γ_k, γ_w, and γ_P can then be found from the residual errors of the active compensation system, which are discussed in the next section.

The rms error in the estimate of the fringe contrast γ_A can be shown to be

$$\sigma_{\gamma_A} = [\gamma_k^2 \gamma_P^2 \gamma_w N f(r)]^{-1/2} \tag{B.28}$$

This error is inversely proportional to $\sqrt{N}$ and can therefore be minimized by integrating the photon count over an extended period of time.

B.3 Active Control System

The practical implementation of a compensated optical interferometer will now be described. The compensation system consists of three active control loops, as shown in figure B.5. These subsystems are:

- baseline control;
- wavefront compensation;
- fringe tracking.

The purpose and implementation of each of these subsystems will now be described.

B.3.1 Baseline Control

The purpose of the first control loop is to maintain approximate equality of the mean optical pathlengths in the two arms of the interferometer, while tracking the science object, and also while varying the baseline. Corrections for the observation geometry and systematic changes in optical pathlength due to Earth's rotation are calculated from a knowledge of the physical length of the baseline and the object angle in the plane of the baseline. The optical pathlength between each telescope and the image plane at which the interference fringes are obtained is then adjusted by the use of optical delay lines ("trombones"), or by moving the fringe measuring station. The mean pathlength error ΔP between the two arms of the interferometer must be minimized in order to optimize the fringe contrast factor γ_k, as shown in equation (B.12). The required bandwidth of this system is about 10 Hz, with a precision on the order of $\pm 10\,\mu$m. This baseline control system is also used to scan the optical path difference during initial acquisition of the object, in order to locate the central white-light fringe, which is detected by the fringe signal processor. Precise correction of optical path differences is performed by the fringe tracking system described in section B.3.3.

A suitable measurement technique for baseline control is two-wavelength interferometry, which enables the optical pathlengths to each of the telescopes to be measured independently. Two axial beams are transmitted through the beam splitters

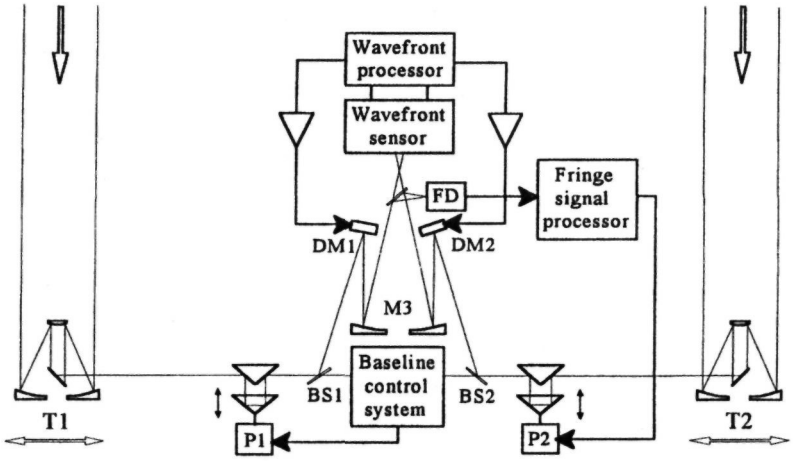


Figure B.5 Active control system for an optical interferometer.

BS1 and BS2, traversing the baseline optics to the secondaries of the two collecting apertures T1 and T2, where they are retroreflected back through the system to the fringe counter in the baseline control system. Corrections to the mean optical pathlength are made with the optical delay line P1, which has a large correction range but relatively low bandwidth ($\approx$ 10 Hz).

B.3.2 Wavefront Compensation

The second active control system is an adaptive optics seeing compensator, which provides both precise angle-tracking of the science object and correction of the wavefront distortion due to atmospheric turbulence. The purpose of this system is to maximize the fringe contrast factor γ_w, due to wavefront errors over the interfering apertures. The basic requirement is to eliminate wavefront differences between the two apertures. In general, there is little correlation between the high-order wavefront distortions in two separated apertures, so that, in practice, each aperture is compensated independently to achieve a flat wavefront. The adaptive optics systems used for this purpose have been described in the main section of this book and will only be outlined here.

The light beams collected by the two apertures of the interferometer pass through the coarse and fine pathlength correctors P1 and P2, and are then reflected by deformable mirrors DM1 and DM2, and the fixed mirror M3, to enter the wavefront sensor as shown in Figure B.5. The optical wavefront corrections for each aperture are computed in real time and fed back to the appropriate deformable mirror to compensate the distortion.

The residual mean-square wavefront error of such a system, when using a natural star as the reference source, has four main components:

$$\sigma_w^2 = \sigma_M^2 + \sigma_T^2 + \sigma_F^2 + \sigma_A^2 \qquad \text{(B.29)}$$

where

σ^2_w = total mean-square wavefront error
σ^2_M = wavefront measurement error, equation (9.43) or (9.44)
σ^2_T = temporal error, equations (9.53) and (9.56)
σ^2_F = wavefront fitting error, equation (9.61),
σ^2_A = anisoplanatic error, equation (9.28).

For values of CT^2, less than about 4 rad^2, the fringe contrast factor may be found from the relation for Strehl ratio [equation 4.40)]:

$$\gamma_w = S = \exp(-\sigma_w^2) \qquad \text{(B.30)}$$

Performance calculations for adaptive optics systems using natural stars are described in section 9.5 and performance charts are shown in figures 9.28 and 9.29. When laser beacons are used, additional error sources are involved. The performance calculations for laser beacon systems are described in section 9.6 and performance charts are given in figures 9.34 through 9.41.

B.3.3 Fringe Tracking

The fringe tracker compensates rapid fluctuations in the optical path difference between the two interferometer apertures, caused by atmospheric turbulence, thus maintaining the mean phase of the fringes within ±1/2 wavelength. This is an essential function in long-baseline interferometers because, without real-time compensation, the spectral bandwidth would have to be extremely small to obtain the required coherence length, in the presence of the large pathlength errors due to turbulence.

The essential components of the fringe tracking system are the fringe detector FD, the fringe signal processor, and the fast-acting optical delay line P2, as shown in figure B.5. The fringe detection system includes a prism that forms a spectrally dispersed image on the detector array. Spectral dispersion of the image produces a tilt in the interference fringes if the pathlengths are unequal in the two arms of the interferometer, as shown in figure B.3. The configuration of the fringe detector array is shown in figure B.6. Each row of elements corresponds to one of the n spectral bands; in the figure, $n = 5$. The detector geometry matches the fringe spacing, which is directly proportional to the wavelength. The fringes are sampled by four detector elements per cycle. The number of fringes in the baseline direction is approximately D/a, so the total number of detector elements is $4nD/a$.

A block diagram of the signal processing required to obtain the estimates of fringe phase and contrast is shown in figure B.7. The first operation is to correlate the image samples Δi_x with the matched filter functions $[W(x) \cos G(x)]$ and $[W(x) \sin G(x)]$, defined in equation (B.I4), summing them

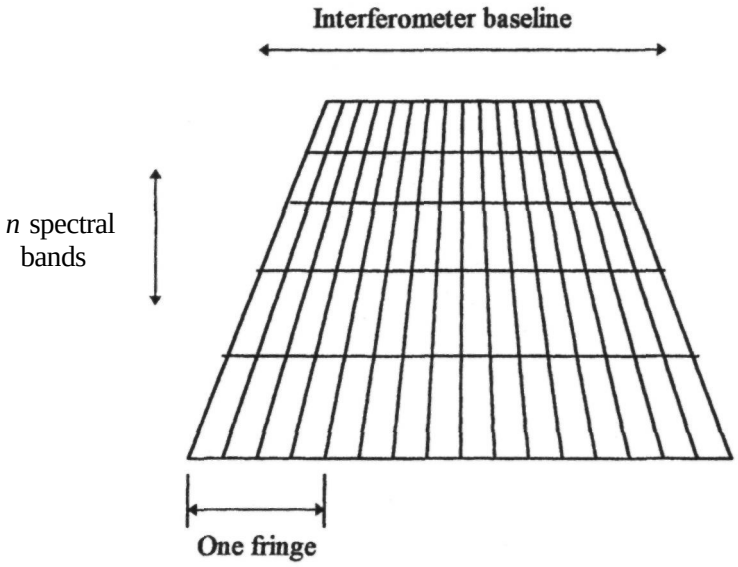


Figure B.6 Detector array for spectrally dispersed image.

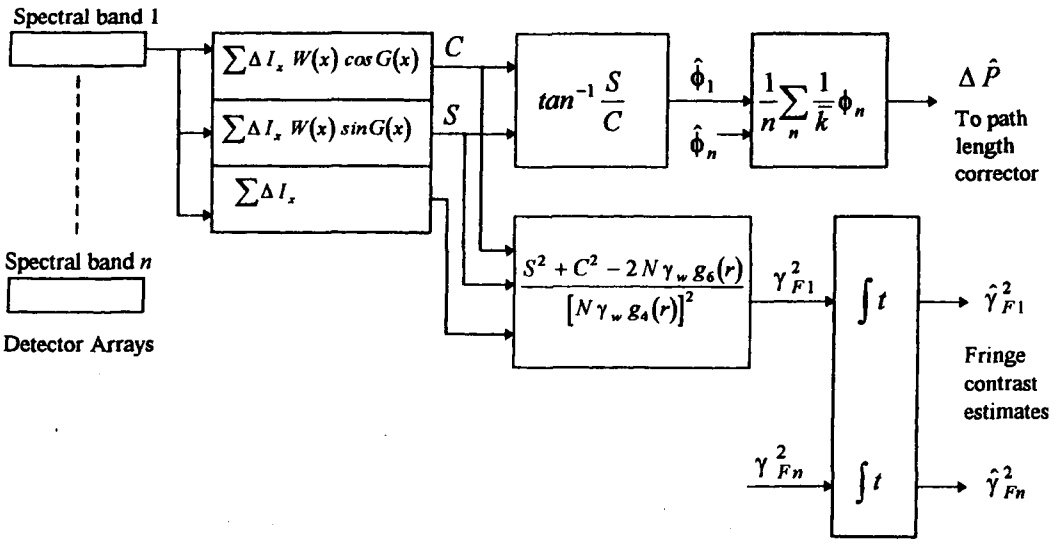


Figure B.7 Fringe detector signal processor.

over the x dimension of the image to form the outguts C and S, from which the fringe phase estimates ϕ are made for each spectral band. These estimates are then averaged and multiplied by Vk to give the path length error AP, which is fed back to the high-speed pathlength corrector P2 to close the fringe tracking loop.

B.3.4 Fringe Contrast Estimation

The overall fringe contrast γ_F is found using equation (B.27). The functions C and S have already been formed. The intensities of all detector outputs in the row are summed to obtain $N\gamma_w$, from which the functions $2N\gamma_w g_6(r)$ and $N\gamma_w g_4(r)$ are formed. The estimates of γ_F^2 are then computed for each spectral band, and time averaged. The final step is to compute the required output of the system, which is the object mutual coherence function γ_A, using equation (B.15). With active control of path length, AP will be small and γ_k will be close to unity. The values of γ_w and γ_P are obtained from the measured wavefront error σ_w and pathlength error σ_P, using equations (B.30) and (B.26).

B.4 Summary

The performance of long-baseline optical interferometers is significantly improved by the use of active control, which includes the use of fringe tracking to compensate for random pathlength errors between the apertures, together with adaptive optics to remove wavefront errors within each aperture. An efficient method of estimating the phase and the contrast of the fringes, using a channeled spectrum detector with matched filters, has been described. The use of active and adaptive control improves the performance of long-baseline interferometers at visible wavelengths by about 10 magnitudes with 1-m apertures, to as much as 15 magnitudes with 10-m apertures.

Bibliography

Acton, D. S., and R. C. Smithson, 1991. Solar astronomy with a 19-segment adaptive mirror. Proc. SPIE 1542, 159–164.

Acton, D. S., and R. C. Smithson, 1992. Solar imaging with a segmented adaptive mirror. Appl. Optics **31,** 3161-3169.

Acton, D .S., R. J. Sharbaugh, J. R. Roehrig, and D. Tiszauer, 1992. Wave-front tilt power spectral density from the image motion of solar pores, Appl. Optics **31,** 4280-4284.

Airy, G. B., 1835. Trans. Camb. Phil. Soc. 5, 283.

Aldrich, R. E., S. L. Hou, and M. L. Harvill, 1971. Electrical and optical properties of $Bi_{12}SiO_{20}$. J-Appl. Phys. **42,** 493–494.

Allen, J. G., A. Jankevics, D. Wormell, and L. Schmutz, 1987. Digital wavefront sensor for astronomical image compensation. Proc. SPIE **739,** 124-128.

Angel, J. R. P., 1994. Ground-based imaging of extrasolar planets using adaptive optics. Nature **368,** 203.

Angel, J. R. P., W. B. Davison. J. M. Hill, E. J. Mannery and H. M. Martin, 1990. Progress toward making lightweight 8 m mirrors of short focal length. Proc. SPIE **1236,** 636-640.

Angel, J. R. P., P. Wizinowich, M. Lloyd-Hart, and D. Sandler, 1990. Adaptive optics for array telescopes using neural-network techniques. Nature **348,** 221-224.

Appenzeller, I., and G. Rupprecht, 1992. FORS, The focal reducer for the VLT. ESO The Messenger, March 1992, 18.

Armstrong, J. T., 1994. Progress on the Big Optical Array (BOA). Proc. SPIE **2200,** 62-70.

Arsenault, R., D. Salmon, J. Kerr, F. Rigaut, D. Crampton, and W. Grundman, 1994. "PUEO," The Canada-France-Hawaii telescope adaptive optics Bonnette I: system description. Proc. SPIE **2201,** 833-842.

Artzner, G., 1992. Microlens arrays for Shack-Hartmann wavefront sensors. Optical Eng **31,** 1311–1322.

Avicola, K., J. M. Brase, J. R. Morris, H. D. Bissinger, J. M. Duff, H. W. Friedman, D. T. Gavel, C. E. Max, S. S. Olivier, R. W. Presta, D. A. Rapp, J. T. Salmon, and K. E. Waltjen, 1994. Sodium-layer laser guide star experimental results. J. Opt. Soc. Am. **A11,** 825-831.

Azouit, M., and J. Vernin, 1980. Remote investigation of tropospheric turbulence by two-dimensional analysis of stellar scintillation. J. Atmosph. Sci. **37,** 1550-1557.

Baba, N., T. Ogura, and N. Miura, 1993. Wave-front correction from point spread function data. Appl. Optics **32,** 4979-4983.

Babcock, H. W., 1948. A photoelectric guider for astronomical telescopes. Astrophys. J. **107,** 73-77.

Babcock, H. W., 1953. The possibility of compensating astronomical seeing. Pub. Astr. Soc. Pac. **65,** 229-236.

Babcock, H. W., 1958. Deformable optical elements with feedback. J. Opt. Soc. Am. **48,** 500.

Babcock, H. W., 1963. Astronomical seeing monitor. Pub. Astr. Soc. Pac. **75,** 1.

Babcock, H. W., 1990. Adaptive optics revisited. Science, **249,** 253-257.

Babcock, H. W., 1992. Astronomical background for adaptive optics. *Laser Guide Star Adaptive Optics Workshop: Proceedings,* ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB, NM, March 1992, 1-18.

Babcock, H. W., B. H. Rule, and J. B. Fassero, 1956. An improved automatic guider. Pub. Astr. Soc. Pac. **68,** 256.

Bahcall, J. N. (Chair), 1991. *The Decade of Discovery in Astronomy and Astrophysics.* National Academy Press, Washington, DC.

Baldwin, J. E., M. G. Beckett, R. C. Boysen, D. Burns, D. F. Buscher, G. C. Cox, C. A. Haniff, C. D. Mackay, N. S. Nightingale, J. Rogers, P. A. G. Scheuer, T. R. Scott, P. G. Tuthill, P. J. Warner, D. M. A. Wilson, and R. W. Wilson, 1996. The first images from an optical aperture synthesis array: mapping of Capella with COAST at two epochs. Astron. Astrophys. 306, L13-L16.

Barclay, H. T., P. H. Malyak, W. H. McGonagle, R. K. Reich, G. S. Rowe, and J. C. Twichell, 1992. The SWAT wavefront sensor. The Lincoln Laboratory Journal 5, 115-128.

Barletti, R., G. Cepatelli, L. Paterno, A. Righini, and N. Speroni, 1976. Mean vertical profile of atmospheric turbulence relevant for astronomical seeing. J. Opt. Soc. Am. **66,** 1380-1383.

Barrett, T. K., D. G. Sandler, R. Q. Fugate, M. T. Donovan, B. A. Ellerbroek, C. H. Higgins, and J. M. Spinhirne, 1992. Neural network wavefront sensor for natural and artificial guide stars. *Laser Guide Star Adaptive Optics Workshop: Proceedings,* ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB, NM, March 1992, 255-275.

Beatty, T. J., R. E. Bills, K. H. Kwon, and C. S. Gardner, 1988. CEDAR lidar observations of sporadic Na layers at Urbana, Illinois. Geophys. Res. Lett. **15,** 1137–1140.

Beckers, J. M., 1988. Increasing the size of the isoplanatic patch with multiconjugate adaptive optics, in *Proc. of the ESO Conference on Very Large Telescopes and Their Instrumentation.* Garching, Germany, 693-703.

Beckers, J. M., 1991. The use of differential adaptive optics for astronomical interferometry. Appl. Optics **30,** 5010–5011.

Beckers, J. M., 1992. Optimization of partial adaptive optics. Appl. Optics **31,** 424-425.

Beckers, J. M., 1993. Adaptive optics for Astronomy: principles, performance and applications. Ann. Rev. Astron. Astrophys. **31,** 13-62.

Beckers, J. M., 1994. Interpretation of out-of-focus star images in terms of wave-front curvature. J. Opt. Soc. Am. **All,** 425-427.

Beckers, J. M., C. Poland, B. L. Ulich, J. R. P. Angel, F. Low and W. Wyatt, 1982. Performance of the Multiple Mirror Telescope VI: MMT telescope coalignment system. Proc SPIE **332,** 42-49.

Beckers, J. M., F. J. Roddier, P. R. Eisenhardt, L. E. Goad, and K. L. Shu, 1986. National Optical Astronomy Observatories (NOAO) Infrared Adaptive Optics Program, I: general description. Proc. SPIE. **628,** 290-297.

Beletic, J. W., C. A. Bleau, C. M. Adkins, T. R. Collins, R. H. Platt, and R. C. Du Varney, 1993. A new approach to high speed, low noise camera design. *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics* (ESO Conference and Workshop Proceedings No. 48) Garching, Germany, 2-5 Aug., 1993, 349-354.

Belsher, J. F., and D. L. Fried, 1984. Expected antenna gain when correcting tilt-free wavefronts. The Optical Sciences Company, Placentia, CA, Report TR-576.

Benedict, R. Jr., J. B. Breckenridge, and D. Fried, (feature eds.), 1994. Atmospheric-compensation technology: Introduction. J. Opt. Soc. Am. **All,** 1-2.

Bester, M., W. C. Danchi, and C. H. Townes, 1990. Long baseline interferometer for the mid-infrared. Proc. SPIE **1237,** 40.

Beuzit, J. L., N. Hubin, E. Gendron, L. Demailly, P. Gigan, F. Lacombe, F. Chazallet, D. Rabaud, and G. Rousset, 1994. ADONIS: a user-friendly adaptive optics system for the ESO 3.6 m telescope. Proc. SPIE **2201,** 955-961.

Bigelow, B. C, D. D. Walker, R. G. Bingham, and P. D'Arrigo, 1993. A deformable secondary mirror for adaptive optics. *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics* (ESO Conference and Workshop Proceedings No. 48), Garching, Germany, 2-5 Aug., 1993, 261-266.

Bigelow, B. C, R. G. Bingham, and D. D. Walker, 1994. Optimizing a deformable secondary mirror for adaptive optics. Proc. SPIE **2201,** 800-810.

Blackwood, G. H., P. A. Davis, and M. A. Ealey, 1991. Characterization of PMN: electro-ceramic plates and select actuators at low temperatures. Proc. SPIE **1543,** 422–429.

Blazit, A., D. Bonneau, M. Josse, L. Koechlin, A. Labeyrie and J. L. Oneto, 1977. The angular diameters of Capella A and B from Two-Telescope Interferometry. Astrophys. J. **217,** L55-L57.

Bliss, E. S., J. R. Smith, and R. L. Miller, 1988. New applications and designs for deformable mirrors. Proc. SPIE **874,** 223-227.

Bonaccini, D., G. Brusa, S. Esposito, P. Salinari, and P. Stefanini, 1990. Adaptive optics wavefront corrector using addressable liquid crystal retarders. Proc. SPIE **1334,** 89-97.

Bonaccini, D., G. Brusa, S. Esposito, P. Salinari, P. Stefanini, and V. Bilotti, 1991. Adaptive optics wavefront corrector using addressable liquid crystal retarders: II. Proc. SPIE **1543,** 133-143.

Bonaccini, D., S. Esposito, and G. Brusa, 1994. Adaptive optics with liquid crystal phase screens. Proc. SPIE **2201**, 1155–1158.

Borgnino, J., and J. Vernin, 1978. Experimental verification of the inertial model of atmospheric turbulence from solar limb motion. J. Opt. Soc. Am. 68, 1056–1062.

Born, M., and E. Wolf, 1975. *Principles of Optics,* Fifth Edition. Pergamon Press, New York.

Bowen, I. S., 1950. Final adjustments and tests of the Hale Telescope. Pub. Astr. Soc. Pac. 62, 91.

Bowen, I. S., 1964. Telescopes. Astron. J. 69, 816-825.

Bowman, M. R., A. J. Gibson, and M. C. W. Sandford, 1969. Atmospheric sodium measured by a tuned laser radar. Nature **221,** 456.

Boyer, C, V. Michau, and G. Rousset, 1990. Adaptive optics: interaction matrix measurements and real time control algorithms for the COME-ON project. Proc. SPIE **1271,** 63-81.

Bradley, L. C, 1992. Pulse-train excitation of sodium for use as a synthetic beacon. J. Opt. Soc. Am. **B10**, 1931-1944.

Brase, J. M., J. An, K. Avicola, H. D. Bissinger, H. W. Friedman, D. T. Gavel, B. Johnston, C. E. Max, S. S. Olivier, R. Presta, D. A. Rapp, J. T. Salmon, and K. E. Waltjen, 1994. Adaptive optics at Lick Observatory: system architecture and operations. Proc. SPIE **2201,** 474-488.

Bridges, W. B., P. T. Brunner, S. P. Lazzara, T. A. Nussmeier, T. R. O'Meara, J. A. Sanguinet, and W. P. Brown Jr., 1974. Coherent optical adaptive techniques. Appl. Optics **13,** 291-300.

Brown R. G. W., K. D. Ridley, and J. G. Rarity, 1986. Characterization of silicon avalanche photodiodes for photon correlation measurements.!: Passive quenching. Appl. Optics **25,** 4122-4126.

Bruns, D. G, D. G. Sandler, B. Martin, and G. Brusa, 1995. Design and prototype tests of an adaptive secondary mirror for the new 6.5 m single-mirror MMT. Proc. SPIE **2534,** 130-133.

Bryngdal, O. 1965. *Applications of Shearing Interferometry. Progress in Optics IV.* North-Holland, New York, 39.

Buffington, A., F. S. Crawford, R. A. Muller, A. J. Schwemin, and R. G. Smits, 1976. Active image restoration with a flexible mirror, Proc. SPIE **75,** 90-96.

Buffington, A., F. S. Crawford, S. M. Pollaine, C D. Orth, and R. A. Muller. 1978. Sharpening stellar images. Science **200,** 489-494.

Bufton, J. L., 1973. Comparison of vertical profile turbulence structure with stellar observations. Appl. Optics **12,** 1785.

Bufton, J. L., P. O. Minott, M. W. Fitzmaurice, and P. J. Titterton, 1972. Measurements of turbulence profiles in the troposphere. J. Opt. Soc. Am. 62, 1068-1070.

Burgess, A., and H. Ghanderharian, 1984. Visual signal detection: I. Ability to use phase information. J. Opt. Soc. Am. **A1**, 900–905.

Burke, B. F. (Ed.), 1992. TOPS: Toward Other Planetary Systems, Report by the Solar System Exploration Division, NASA.

Burke, J. J., and J. B. Breckenridge, 1978. Passive imaging through the turbulent atmosphere: fundamental limits on the spatial frequency resolution of a rotational shearing interferometer. J. Opt. Soc. Am. **68,** 67-77.

Buscher, D. F., A. P. Doel, C. A. Haniff, and R. W. Wilson, 1995. Visible-wavelength diffraction-limited imaging using low-order adaptive optics. Proc. SPIE **2534,** 53-61.

Carleton, N. P., 1988. The IOTA Project, in *High Resolution Imaging by Interferometry,* ed. F. Merkle. (ESO Conference Proceedings No. 29), 939.

Carleton, N. P., W. A. Traub, M. G. Lacasse, P. Nisenson, M. R. Pearlman, R. D. Reasenberg, X. Xu, C. M. Coldwell, A. Panasyuk, J. A. Benson, C. Papaliolios, R. Predmore, F. P. Schloerb, H. M. Dyck, and D. M. Gibson, 1994. Current status of the IOTA interferometer. Proc. SPIE **2200,** 152–165.

Cathey, W. T., C. L. Hayes, W. C. Davis, and V. F. Pizzuro, 1970. Compensation for atmospheric phase effects at 10.6μm., Appl. Optics 9, 701-707.

Centamore, R. M., and A. Wirth, 1991. High bias membrane mirror. Proc. SPIE **1543,** 128-132.

Chanan, G., G. Djorgovski, A. Gleckler, S. Kulkarni, T. Mast, C. Max, J. Nelson, and P. Wizinowich, 1996. Adaptive optics for Keck Observatory. Keck Observatory Report No. 208, California Association for Research in Astronomy, W. M. Keck Observatory, Kamuela, HI.

Chen, C. T., 1984. *Linear System Theory and Design.* Holt, Rinehart and Winston, Orlando, FL.

Chen, H., and L. Shentu, 1992. Achromatic interferometer for imaging through turbulence. Appl. Optics 9, 1274-1278.

Chernov, L. A., 1960. *Wave Propagation in a Random Medium.* McGraw-Hill Book Company, New York.

Cheselka, M., and R. Angel, 1996. Automated optical aircraft spotter. Paper presented at the Laser Safety Meeting, Kirtland Air Force Base, New Mexico, 10–11 June, 1996.

Christian, C, and R. Racine, 1985. Dependence of seeing correlation on image separation at the CFH telescope on Mauna Kea. Pub. Astr. Soc. Pac. 97, 1215-1226.

Churnside, J. H. 1989. Angle of arrival fluctuations of retro-reflected light in the turbulent atmosphere. J. Opt. Soc. Am. A6, 275-279.

Churnside, J. H., and R. J. Lataitis, 1987. Angle of arrival fluctuations of a reflected beam in atmospheric turbulence. J. Opt. Soc. Am. A4, 1264-1272.

Clampin, M., 1989. Optimization of MCP intensifier tubes in astronomical adaptive optics sensors. Proc. SPIE **1114,** 152-159.

Clampin, M., S. T. Durrance, D. A. Golimowski, and R. H. Barkhouser, 1991. The Johns Hopkins adaptive optics coronagraph. Proc. SPIE **1542,** 165-174.

Clark, A. E., 1991. Giant magnetostriction materials from cryogenic temperatures to 250 °C. Proc. SPIE **1543,** 374-381.

Clark, A. E., and H. S. Belson, 1972. Giant room temperature magnetostrictions in Tb Fe_2 and Dy Fe_2. Phys. Rev. B5, 3642.

Clark, A. E., J. P. Teter, and O. D. McMasters, 1988. Magnetostriction "jumps" in twinned $Tb_{0.3}$ $Dy_{0.7}$ $Fe_{1.9}$. J. Appl. Phys. **63,** 3910.

Clark, B. G., 1980. An efficient implementation of the algorithm CLEAN. Astron. Astrophys. **89,** 377-378.

Close, L. M., D. Wittman, and D. W. McCarthy Jr, 1994. High resolution infrared imaging utilizing a tip-tilt secondary mirror. Proc. SPIE **2201,** 447-457.

Cohen, D. K., et al. 1984. Automatic focus control: the astigmatic lens approach. Appl. Optics **23,** 4.

Cohen, M., and I. Drabkin, 1948. *A Sourcebook in Greek Science.* McGraw-Hill, New York.

Colavita, M. M., M. Shao, and D. H. Staelin, 1987. Atmospheric phase measurements with the Mark III stellar interferometer. Appl. Optics **26,** 4106-4112.

Colavita, M. M., M. Shao, B. E. Hines, J. K. Wallace, Y. Gursel, F. Malbet, J. W. Yu, H. Singh, C. A. Beichman, X. Pan, T. Nakajima, and S. R. Kulkarni, 1994. ASEPS-0 testbed interferometer. Proc. SPIE **2200,** 89-97.

Coltman, J. W., and A. E. Anderson, 1960. Noise limitations to resolving power in electronic imaging. Proc. IRE **48,** 858-865.

Coulman, C. E., J. Vernin, Y. Coqueugniot, and J. L. Caccia, 1988. Outer scale of turbulence appropriate to modeling refractive index structure profiles. Appl. Optics **27,** 155-160.

Cubalchini, R., 1979. Modal wavefront estimation from phase derivative measurements. J. Opt. Soc. Am. **69,** 972-977.

Dautet, H., P. Deschamps, B. Dion, A. D. MacGregor, D. MacSween, R. J. McIntyre, C. Trottier, and P. P. Webb, 1993. Photon counting techniques with silicon avalanche photodiodes. Appl. Optics **21, 3894–3900.**

Davis, J., 1993. in *Very High Angular Resolution Imaging.* 1AU Symp. **158,** Astron. Soc. Pac, 135.

Davis, J., and W. J. Tango, 1985. A new very high angular resolution stellar interferometer. Proc. Astron. Soc. Aust. 6, 38-43.

Demailly, I., E. Gendron, J. L. Beuzit, F. Lacombe, and N. Hubin, 1994. Artificial intelligence system and optimized modal control for the ADONIS adaptive optics instrument. Proc. SPIE **2201,** 867-878.

Department of Defense, 1970. Piezoelectric ceramic for sonar transducers. Military Standard FSC5845, MIL-STD-1376 (SHIPS), 21 Dec, 1970.

DeWitt, J. H., R. H. Hardie, and C. K. Seyfert, 1957. A seeing compensator employing television techniques. Sky and Telescope. **XVII,** 8-9.

De Wolf, D. A. 1974. Waves in turbulent air: a phenomenological model. Proc. IEEE **62,** 1523-1529.

Dicke, R. H., 1975. Phase-contrast detection of telescope seeing errors and their correction. Astrophys. J. **198,** 605-615.

Downie, J. D., and J. W. Goodman, 1989. Optimal wavefront control for adaptive segmented mirrors. Appl. Opt. **28,** 5326-5332.

Driggers, R. G., C. E. Halford, and G. D. Boorman, 1990. Use of spatial light modulators in frequency modulation reticle trackers. Opt. Eng. **29,** 1398—1403.

Durrance, S. T., and M. Clampin, 1989. The Johns Hopkins University Adaptive Optics Program. Proc. SPIE **1114,** 97-104.

Dyck, H. M., J. A. Benson, and F. P. Schloerb, 1995. Imaging a binary star with a two-telescope Michelson stellar interferometer. Astron. J. **110,** 1433-1439.

Dyson, F. J., 1975. Photon noise and atmospheric noise in active optical systems. J. Opt. Soc. Am. **65,** 551-558.

Ealey, M. A., and P. A. Davis, 1990. Standard SELECT electrostrictive lead magnesium niobate actuators for active and adaptive optical components. Opt. Eng. **29,** 1373-1382.

Ealey, M. A., and J. F. Washeba, 1990. Continuous face sheet low voltage deformable mirrors. Opt. Eng. **29, 1191–1198.**

Ealey, M. A., and J. Wellman, 1989. Fundamentals of deformable mirror design and analysis. Proc. SPIE **1167,** 66-84.

Ealey, M. A., and J. A. Wellman, 1994. Xinetics low cost deformable mirrors with actuator replacement cartridges. Proc. SPIE **2201, 680–687.**

Ealey, M. A., and C. E. Wheeler, 1990. Integrated wavefront corrector. Proc. SPIE **1271,** 254-264.

Edlen, B., 1966. Metrologia 2, 71.

Ellerbroek, B. L., 1994. First-order performance evaluation of adaptive optics systems for atmospheric turbulence compensation in extended-field-of-view astronomical telescopes. J. Opt. Soc. Am. **All,** 783-805.

Ellerbroek B. L., C. Van Loan, N. P. Pitsianis, and R. J. Plemmons, 1994. Optimizing closed loop adaptive optics performance using multiple control band widths. Proc. SPIE **2201,** 935-948.

Elliot, J. L., and I. S. Glass, 1970. A quantitative fringe detector for stellar interferometry. Astron. J. **75,** 1123-1132.

Esposito, S., G. Bursa, and D. Bonaccini, 1993. Liquid crystal wavefront correctors computer

simulation results. *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics* (ESO Conference and Workshop Proceedings No. 48) Garching, Germany, 2-5 Aug, 1993, 289-293.

Fante, R. L., 1975. Electromagnetic beam propagation in turbulent media. Proc. IEEE **63,** 1669–1692.

Fante, R. L., 1976. Some results on the effect of turbulence on phase-compensated systems. J. Opt. Soc. Am. 66, 730-735.

Feinlieb, J., S. G. Lipson, and P. F. Cone, 1974. Monolithic piezoelectric mirror for wavefront correction. Appl. Phys. Lett. **25,** 311–313.

Fellgett, P. B. 1953. Concerning photographic grain, signal-to-noise ratio, and information. J. Opt. Soc. Am. 43, 271-282.

Fellgett, P. B. 1969. Design of astronomical telescope systems. *Proc. of Conference on Optical Instruments and Techniques,* ed. J. Home Dickson, Univ. of Reading, July 1969, Oriel Press, London.

Fizeau, A. H. L., 1868. Rapport sur le prix Bordin. C. R. Acad. Sci. (Paris), 66, 932.

Forbes, F., 1989. Bimorph PZT active mirror. Proc. SPIE **1114,** 146-151.

Forbes, F. F., and N. Roddier, 1991. Adaptive optics using curvature sensing. Proc. SPIE **1542,** 140–147.

Fowler, A. M., R. G. Probst, J. P. Britt, R. R. Joyce, and F. C. Gillett, 1987. Evaluation of an indium antimonide hybrid focal plane array for ground-based infrared astronomy. Opt. Eng. 26, 232-240.

Foy, R., and A. Labeyrie, 1985. Feasibility of adaptive telescope with laser probe. Astron. Astrophys. **152,** L29-L31.

Freischlad, K., and C. L. Koliopoulos, 1985. Wavefront reconstruction from noisy slope or difference data using the discrete fourier transform. Proc. SPIE **551,** 74-80.

Fried, D. L., 1965. Statistics of a geometric representation of wavefront distortion. J. Opt. Soc. Am. **55,** 1427-1435.

Fried, D. L., 1966a. Optical resolution through a randomly inhomogeneous medium for very long and very short exposures. J. Opt. Soc. Am. 56, 1372-1379.

Fried, D. L., 1966b. Limiting resolution looking down through the atmosphere. J. Opt. Soc. Am. 56, 1380-1384.

Fried, D. L., 1967. Optical heterodyne detection of an atmospherically distorted signal wave front. Proc. IEEE **55,** 57-67.

Fried, D. L., 1976. Imaging through the atmosphere. Proc. SPIE **75,** 20.

Fried, D. L., 1977. Adaptive optics for imaging within the atmosphere: reference generation and field-of-view widening. Technical Report TR-273, Optical Sciences Company, Placentia, CA.

Fried, D. L., 1978. Probability of getting a lucky short-exposure image through turbulence. J. Opt. Soc. Am. 68, 1651–1657.

Fried. D. L., 1979. Resolution, signal-to-noise ratio and measurement precision. J. Opt. Soc. Am. 69, 399-406.

Fried, D. L., 1982. Anisoplanatism in adaptive optics. J. Opt. Soc. Am. 72, 52-61.

Fried, D. L., 1990. Time-delay-induced mean-square error in adaptive optics. J. Opt. Soc. Am. A7, 1224-1225.

Fried, D. L., and J. F. Belsher, 1994. Analysis of fundamental limits to artificial guide star adaptive optics system performances for astronomical imaging. J. Opt. Soc. Am. **All,** 277-287.

Friedman, H., G. Erbert, T. Kuklo, T. Salmon, D. Smauley G. Thompson, and N. Wong, 1994. Design of a fieldable laser system for a sodiun guide star. Proc. SPIE **2201,** 352-363.

Friedman, H., G. Erbert, T. Kuklo, T. Salmon, D. Smauley, G. Thompson, J. Malik, N. Wong, K. Kanz, and K. Neeb, 1995. Sodium beacon laser system for the Lick Observatory. Proc. SPIE **2534,** 150–160.

Fugate, R. Q., 1994. Observations of faint objects with laser beacon adaptive optics. Proc. SPIE **2201,** 10-21.

Fugate, R. Q., D. L. Fried, G. A. Ameer, B. R. Boeke, S. L. Browne, P. H. Roberts, R. E. Ruane, and L. M. Wopat, 1991. Measurement of atmospheric wavefront distortion using scattered light from a laser guide-star. Nature 353, 144-146.

Fugate, R. Q., C. H. Higgins, B. L. Ellerbroek, J. M. Spinhirne, B. R. Boeke, R. A. Cleis, D. W. Swindle, and M. D. Oliker, 1993. Laser beacon adaptive optics with an unintensified CCD wavefront sensor and fiber optic synthesized array of silicon avalanche photodiodes for fast guiding. *Proc. CO-16 Satellite Conference on Active and Adaptive Optics,* (ESO Conference and Workshop Proceedings No. 48) Garching, Germany, 2-5 Aug., 1993, 487–492.

Fugate, R. Q., B. L. Ellerbroek, C. H. Higgins, M. P. Jelonek, W. J. Lange, A. C. Slavin, W. J. Wild, D. M. Winker, J. M. Wynia, J. M. Spinhirne, B. R. Boeke, R. E. Ruane, J. F. Moroney, M. D. Oliker, D. W. Swindle, and R. A. Cleis, 1994. Two generations of laser-guide-star adaptive-optics experiments at the Starfire Optical Range. J. Opt. Soc. Am. **All,** 310-324.

Gaeta, C, T. R. O'Meara, W. P. Brown, M. L. Minden, P. V. Mitchell, and D. M. Pepper, 1994. Characteristics of innovative adaptive-optics servos that use membrane-based spatial light modulators. J. Opt. Soc. Am. **All,** 880-882.

Gaffard, J.-P., and C. Boyer, 1990. Adaptive optics: effect of sampling rate and time lags on the closed loop band width. Proc. SPIE **1271,** 33-50.

Gaffard, J.-P., and G. Ledanois, 1991. Adaptive optical transfer function modeling. Proc. SPIE **1542,** 34–45.

Gaffard, J.-P., P. Jagourel, and P. Gigan, 1994. Adaptive optics: description of available components at LASERDOT. Proc. SPIE 2201, 688-702.

Galvagni, J., and B. Rawal, 1991. A comparison of piezoelectric and electrostrictive actuator stacks. Proc. SPIE 1543, 296–300.

Gardner, C. S. 1989. Sodium resonance fluorescence lidar applications in atmospheric science and astronomy. Proc. IEEE 77, 408–418.

Gardner, C. S., D. G. Voelz, C. F. Sechrist, Jr, and A. C. Segal, 1986. Lidar studies of the nighttime sodium layer over Urbana, Illinois: 1. Seasonal and nocturnal variations. J. Geophys. Res. 91, 13659-13673.

Gardner, C. S., B. M. Welsh, L. A. Thompson, with prolog by R. Braham, 1990. Design and performance analysis of adaptive optical telescopes using laser guide stars. Proc. IEEE 78, 1721-1743.

Gavel, D.T., and S. S. Olivier, 1995. Performance predictions for the Keck Telescope adaptive optics system. Proc. SPIE 2534, 401–411.

Gavel, D. T., J. R. Morris, and R. G. Vernon, 1994. Systematic design and analysis of laser-guide-star adaptive-optics systems for large telescopes. J. Opt. Soc. Am. A11, 914-924.

Gavrielides, A., 1982. Vector polynomials orthogonal to the gradient of Zernike polynomials. Optics Lett 7, 526-528.

Geary, J. C, 1994. Rapid-framing CCDs with 16 output ports for laser guide star sensors. Proc. SPIE 2201, 607-612.

Geary, J. C, and G. A. Luppino, 1994. New circular radial-scan frame-storage CCDs for low-order adaptive optics wavefront curvature sensing. Proc. SPIE 2201, 588-595.

Gendron, E., 1993. Modal control optimization in an adaptive optics system. *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics* (ESO Conference and Workshop Proceedings No. 48), Garching, Germany, 2-5 Aug. 1993, 187-192.

Gendron, E., J. G. Cuby, F. Rigaut, P. Lena, J. C. Fontanella, G. Rousset, J. P. Gaffard, C. Boyer, J. C. Richard, M. Vittot, F. Merkle, and N. Hubin, 1991. Come-On Plus project: an upgrade of the Come-on adaptive optics prototype system. Proc. SPIE 1542, 297-307.

Gentles, B., M. Wells, A. J. Longmore, R. M. Myers, and R. A. Humphreys, 1995. Natural guide star AO system for the William Herschel Telescope. Proc. SPIE 2534, 62-71.

Gillett, F., and M. Mountain, 1996. The Gemini 8 m Telescopes Project — a scientific overview. Gemini Newsletter, June 1996.

Gleckler, A.D., and P. Wizinowich, 1995. W. M. Keck Observatory adaptive optics program. Proc. SPIE 2534, 386–400

Goad, L., and J. Beckers, 1989. A near infrared astronomical adaptive optics system. Proc. SPIE 1114, 73-81.

Goad, L., F. Roddier, J. Beckers, and P. Eisenhardt, 1986. National Optical Astronomy Observations (NOAO) IR Adaptive Optics Program III: criteria for the wavefront sensor selection. Proc. SPIE 628, 305–313.

Goldberg, B. A., J. Albetski, D. Biringer, P. McCormick, and A. L. Broadfoot, 1992. High-resolution spectrophotometry of extended objects using the AMOS Compensated Imaging System (CIS). *Laser Guide Star Adaptive Optics Workshop: Proceedings*, ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB, NM, March 1992, 276-288.

Gonsalves, R. A., 1982. Phase retrieval and diversity in adaptive optics. Opt. Eng. 21, 829-832.

Good, R. E., R. R. Beland, E. A. Murphy, J. H. Brown, and E. M. Dewan, 1988. Atmospheric models of optical turbulence. Proc. SPIE 928, 165–186.

Goodfriend, M. J., and K. M. Shoop, 1991. High force, high strain wideband width linear actuator using the magnetostrictive material Terfenol-D. Proc. SPIE 1543, 301-312.

Gracheva, M. E., A. S. Gurvich, and M. A. Kallistratova, 1970. Measurements of the variance of "strong" intensity fluctuations of laser radiation in the atmosphere. Izv. Vyssh. Ucheb. Zaved. Radiofiz. 13, 55-60.

Graves, J. E., F. Roddier, M. J. Northcott, J. Anuskiewicz, and G. Monnett, 1994. Adaptive optics at the University of Hawaii IV: a photon counting curvature wavefront sensor. Proc. SPIE 2201, 502-507.

Gray, P. M., M. Lloyd-Hart, J. R. P. Angel, D. W. McCarthy Jr, D. G. Sandler, T. Martinez, L. M. Close, G. Brusa, D. G. Bruns, B. A. McLeod, P. T. Ryan, T. D. Groesbeck, D. M. Wittman, B. P. Jacobsen, J. M. Hughes, M. R. Hunten, and M. Cheselka, 1995. FASTTRAC II near-IR adaptive optics system for the Multiple Mirror Telescope: description and preliminary results. Proc. SPIE 2534, 2–16.

Greenhalgh, R. J. S., L. M. Stepp, and E. R. Hansen, 1994. Gemini primary mirror thermal management system. Proc. SPIE 2199, 911–921.

Greenwood, D. P., 1977. Bandwidth specification for adaptive optics systems. J. Opt. Soc. Am. 67, 174–176.

Greenwood, D. P., 1979. Mutual coherence function of a wave-front corrected by zonal adaptive optics. J. Opt. Soc. Am. 69, 549-554.

Greenwood, D. P., and D. L. Fried, 1976. Power spectra requirements for wave-front compensative systems. J. Opt. Soc. Am. 66, 193-206.

Greenwood, D. P., and C. A. Primmerman, 1992. Adaptive optics research at Lincoln Laboratory. MIT Lincoln Laboratory Journal 5, 3-24.

Gregory, T., P. Wizinowich, D. S. Acton, A. Gleckler, P. Stomski, S. Radau, J. Maute, and M. Sirota, 1996. Keck adaptive optics: optical

bench subsystem. Opt. Soc. Am. Conf., Maui, HI, July 8-12.

Grosso, R. P., and M. Yellin. 1977. The membrane mirror as an adaptive optical component. J. Opt. Soc. Am. **67,** 399-406.

Hanbury Brown, R., and R. Q. Twiss, 1956. A test of a new type of stellar interferometer on Sirius. Nature **178,** 1046-1048.

Hanbury Brown, R., and R. Q. Twiss, 1957a. Interferometry of the intensity fluctuations in light. I. Basic theory: the correlation between photons in coherent beams of radiation. Proc. R. Soc. London **A242,** 300–324.

Hanbury Brown, R., and R. Q. Twiss, 1957b. Interferometry of the intensity fluctuations in light. II. An experimental test of the theory for partially coherent light. Proc. R. Soc. London **A243,** 291-319.

Hanbury Brown, R., and R. Q. Twiss, 1958a. Interferometry of the intensity fluctuations in light. III. Applications to astronomy. Proc. R. Soc. London **A248,** 199-221.

Hanbury Brown, R., and R. Q. Twiss, 1958b. Interferometry of the intensity fluctuations in light. IV. A test of an intensity interferometer on Sirius A. Proc. R. Soc. London **A248,** 222-237.

Hanbury Brown, R., J. Davis and L. R. Allen, 1967. The stellar interferometer at Narrabri Observatory. I. A description of the instrument and the observational procedure. Mon. Not. R. Astron. Soc. **137,** 375-392.

Hanbury Brown, R., J. Davis and L. R. Allen, 1974. The angular diameters of 32 stars. Mon. Not. R. Astron. Soc. **167,** 121-136.

Haniff, C. A., C. D. Mackay, D. J. Titterington, D. Sivia, J. E. Baldwin, and P. J. Warner, 1987. The first images from optical aperture synthesis. Nature **328,** 694-696.

Happer, W., G. J. MacDonald, C. E. Max, and F. J. Dyson, 1994. Atmospheric-turbulence compensation by resonant optical backscattering from the sodium layer in the upper atmosphere. J. Opt. Soc. Am. **A11,** 263-276.

Hardy, J. W., 1975a. Analog Data Processor. U.S. Patent 3,921,080, 18 Nov., 1975.

Hardy, J. W., 1975b. Real time wavefront correction system. U.S. Patent 3,923,400, 2 Dec. 1975.

Hardy, J. W., 1978. Active optics: a new technology for the control of light. Proc. IEEE **66,** 651–697.

Hardy, J. W., 1981. Solar isoplanatic patch measurements. *Proc. Sacramento Peak National Observatory Conference: Solar Instrumentation — What's Next?* ed. R. B. Dunn, Sunspot, NM, March, 1981,421-433.

Hardy, J. W., 1982. Active optics — don't build a telescope without it! Proc. SPIE 332, 252-259.

Hardy, J. W., 1987. Laser wavefront compensation system. Proc. SPIE **739,** 118–123.

Hardy, J. W., 1989. Instrumental limitations in adaptive optics for astronomy. Proc. SPIE **1114,** 2-13.

Hardy, J. W., 1993. Compensated Imaging System, *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics* (ESO Conference and Workshop Proceedings No. 48), Garching, Germany, 2-5 Aug, 1993, 107–114.

Hardy, J. W., and A. J. MacGovern, 1987. Shearing interferometry: a flexible technique for wavefront measurement. Proc. SPIE **816,** 180–195.

Hardy, J. W., and E. P. Wallner, 1994. Wavefront compensation using active lenses. Proc. SPIE **2201,** 77-87.

Hardy, J. W., J. Feinlieb, and J. C. Wyant, 1974. Real time phase correction of optical imaging systems. Optical Society of America (OSA) Topical Meeting on Optical Propagation through Turbulence, Univ. of Colorado, Boulder, CO., 9–11 July, 1974.

Hardy, J. W., J. E. Lefebvre, and C. L. Koliopoulos, 1977. Real-time atmospheric compensation. J. Opt. Soc. Am. **67,** 360-369.

Harrington, P. M., and B. M. Welsh, 1994. Frequency-domain analysis of an adaptive optical system's temporal response. Opt. Eng. 33, 2336-2342.

Hartmann, J., 1900. Bemerkungen uber den Bau und die Justirung von Spektrographen. Zeit. f. Instrument **20,** 47.

Hartmann, J., 1904. Objectiv unter suchungen. Zeit. f. Instrument **24,** 1.

Harvey, J. W., and J. B. Breckenridge, 1973. Solar speckle interferometry. Astrophysics **182,** L137-L139.

Harvey, J. W., and M.. Schwarzschild, 1975. Photoelectric speckle interferometry of the solar granulation. Astrophysics **196,** 221-226.

Hayes, C. L., R. A. Brandewie, W. C. Davis, and G. E. Mevers, 1977. Experimental test of an infrared phase conjugation adaptive array. J. Opt. Soc. Am. **67,** 269-277.

Herrmann, J., 1980. Least-squares wave front errors of mimimum norm. J. Opt. Soc. Am. **70,** 22-35.

Hinkley, E. D., ed. 1976. *Laser Monitoring of the Atmosphere.* Springer-Verlag, Berlin.

Hofmann, K.-H., and G. Weigelt, 1986. High angular resolution shearing spectroscopy and triple shearing interferometry. Appl. Opt. 25, 4280–4287.

Hogbom, J. A., 1974. Aperture synthesis with a non-regular distribution of interferometer baselines. Astronom. J. Suppl. Series 15, 417–426.

Hogge, C. B., and R. R. Butts, 1982. Effects of using different wavelengths in wave-front sensing and correction. J. Opt. Soc. Am. **72.** 606-609.

Holmes, J. F., and V. S. R. Gudimetla, 1983. Strehl's ratio for a two-wavelength continuously deformable optical adaptive transmitter. J. Opt. Soc. Am. **73,** 1119–1122.

Horwitz, B. A., 1990. Multiplex techniques for real-time shearing interferometry. Opt. Eng. **29,** 1223—1232.

Horwitz, B. A., 1991. High precision, wide dynamic range WCE wavefront sensor. Proc. SPIE 1543, 449-459.

Horwitz, B. A., and J. T. Watson, 1991. Real-scene phasing sensor approach; a modest proposal. Proc. SPIE 1543, 432-437.

Hubin, N., J. L. Beuzit, E. Gendron, and L. Demailly, 1993. ADONIS: a user-friendly adaptive optics system for the ESO 3.6 m telescope. *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics* (ESO Conference and Workshop Proceedings No. 48), Garching, Germany, 2-5 Aug., 1993, 71-76.

Hubin, N., B. Theodore, P. Petitjean, and B. Delabre, 1994. The adaptive optics system for the VLT. Proc. SPIE 2201, 34–45.

Hudgin, R. A., 1976. A new turbulence sensor using atmospheric dispersion. Proc. SPIE 75, 55-61.

Hudgin, R., 1977a. Wave-front reconstruction for compensated imaging. J. Opt. Soc. Am. 67, 384-387.

Hudgin, R., 1977b. Wave-front compensation error due to finite corrector element size. J. Opt. Soc. Am. 67, 393-395.

Hudgin, R. A., and S. G. Lipson, 1975. Analysis of a monolithic piezoelectric mirror. J. Appl. Phys. 46, 510-512.

Hufnagel, R.E., 1974. Variations of atmospheric turbulence. Digest of Technical Papers, Topical Meeting on Optical Propagation through Turbulence, University of Colorado, Boulder, CO, July 9-11, 1974, Paper Wa1.

Hufnagel, R. E., and N. R. Stanley, 1964. Modulation transfer function associated with image transmission through turbulent media. J. Opt. Soc. Am. 54, 52-61.

Hulburd, W. G, and D. G. Sandler, 1990. Segmented mirrors for atmospheric compensation. Opt. Eng. 29, 1186–1190.

Hulburd, W. G, T. K. Barrett, L. Cuellar, and D. G. Sandler, 1991. High bandwidth long stroke segmented mirrors for atmospheric compensation. Proc. SPIE 1543, 64-75.

Humphreys, R. A., C. A. Primmerman, L. C. Bradley, and J. Herrmann, 1991. Atmospheric-turbulence measurements using a synthetic beacon in the mesospheric sodium layer. Optics Lett. 16, 1367-1369.

Humphreys, R. A., L. C. Bradley, and J. Herrman, 1992. Sodium-layer synthetic beacons for adaptive optics. MIT Lincoln Laboratory Journal 5, 45-66.

Hunt, B. R., 1979. Matrix formulation of the reconstruction of phase values from phase differences. J. Opt. Soc. Am. 69, 393-399.

Hutchin, R. A. 1991. History and physical principles of Lodestar Technology. Lecture given to Optical Society of America, 8 Nov., 1991, unpublished.

Hutter, D. J., 1994. USNO astrometric interferometer. Proc. SPIE 2200, 81-88.

Ichikawa, K., A. W. Lohmann, and M. Takeda, 1988. Phase retrieval based on the irradiance transport equation and the Fourier transform method: experiments. Appl. Optics 27, 3433-3436.

Ishimaru, A., 1978. *Wave Propagation and Scattering in Random Media,* Vol. 2. Academic Press, New York.

Iye, Masanori, 1996. Subaru instrumentation program. Subaru Newsletter 1(3), July 1996.

Jacobsen, B., T. Martinez, R. Angel, M. Lloyd-Hart, S. Benda, D. Middleton, H. Friedman, and G. Erbert, 1994. Field evaluation of two new continuous-wave dye laser systems optimized for sodium beacon excitation. Proc. SPIE 2201, 342-351.

Jacquinot, P., and B. Roizen-Dossier, 1964. *Apodization. Progress in Optics III,* North-Holland Publishing, Amsterdam and Elsevier, New York, 29-186.

Jaffe, B., W. R. Cook Jr, and M. Jaffe, 1971. *Piezoelectric Ceramics.* Academic Press, New York.

Jagourel, P., P. Y. Madec, and M. Sechaud, 1990. Adaptive optics: a bimorph mirror for wavefront correction. Proc. SPIE 1237, 394–405.

Jelonek, M. P., R. Q. Fugate, W. J. Lange, A. C. Slavin, R. E. Ruane, and R. A. Cleis, 1992. Sodium laser guide star experiments at the Starfire Optical Range. *Laser Guide Star Adaptive Optics Workshop: Proceedings,* ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB, NM, March 1992, 213-226.

Jelonek, M. P., R. Q. Fugate, W. J. Lange, A. C. Slavin, R. E. Ruane, and R. A. Cleis, 1994. Characterization of artificial guide stars generated in the mesospheric sodium layer with a sum-frequency laser. J. Opt. Soc. Am. A11, 806–812.

Jennison, R. C, 1958. A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distribution of small angular extent. Mon. Not. R. Astron. Soc. 118, 276-284.

Jeys, T. H., 1991. Development of a mesospheric sodium laser beacon for atmospheric adaptive optics. MIT Lincoln Laboratory Journal 4, 133.

Jeys, T. H., 1992. Mesospheric laser beacon. *Laser Guide Star Adaptive Optics Workshop: Proceedings,* ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB, NM, March 1992, 196-212.

Jeys, T. H., A. A. Brailove, and A. Mooradian, 1989. Sum frequency generation of sodium resonance radiation. Appl. Optics 28, 2588-2591.

Jeys, T. H., R. M. Heinrichs, K. F. Wall, J. Korn, T. C. Holtaling, and E. Kibblewhite, 1992. Optical pumping of mesospheric sodium. *Laser Guide Star Adaptive Optics Workshop: Proceedings,* ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB, NM, March 1992, 238-254.

Johnson, P., R. Trissel, L. Cuellar, B. Arnold, and D. Sandler, 1991. Real time wavefront reconstruction

for a 512 subaperture adaptive optical system. Proc. SPIE 1543, 460–471.

Johnston, D. C, and B. M. Welsh, 1994. Analysis of multiconjugate adaptive optics. J. Opt. Soc. Am. A11, 394-408.

Jorgenson, M. B., and G. J. M. Aitken, 1992. Neural network prediction of turbulence induced wavefront degradations with application to adaptive optics. Proc. SPIE **1706,** 113.

Jorgenson, M. B., and G. J. M. Aitken, 1993. Wavefront prediction for adaptive optics. *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics* (ESO Conference and Workshop Proceedings No. 48), Garching, Germany, 2-5 Aug., 1993, 143-148.

Kane, T. J., B. M. Welsh, C. S. Gardner, and L. A. Thompson, 1989. Wave front detector optimization for laser guided adaptive telescopes. Proc. SPIE 1114, 160–171.

Karr, T. J., 1991. Temporal response of atmospheric turbulence compensation. Appl. Optics **30,** 363-364.

Kern, P., F. Rigaut, P. Lena, F. Merkle, and G. Rousset, 1990. Adaptive optics prototype system for IR astronomy II: first observing results. Proc. SPIE **1237,** 345-355.

Kibblewhite, E. J., R. Vuilleumier, B. Carter, W. J. Wild, and T. H. Jeys, 1994. Implementation of CW and pulsed laser beacons for astronomical adaptive optics systems. Proc SPIE **2201,** 272-283.

Kibblewhite, E. J., W. J. Wild, B. Carter, M. Chun, F. Shi, M. Smutko, and V. Scor, 1994. A description of ChAOS. Proc. SPIE **2201,** 458-467.

Kim, Y., and D. L. Jaggard, 1988. Band-limited fractal model of atmospheric refractivity fluctuation. J. Opt. Soc. Am. A5, 475-480.

Kimata, M., M. Denda, N. Yutani, S. Iwade, N. Tsubouchi, M. Daido, H. Furukawa, R. Tsunoda, and T. Kanno, 1987. 256 x 256 Element platinum silicide Schottky-barrier infrared charge-coupled device image sensor. Opt. Eng. 26, 209–215.

King, H. C, 1955. *The History of the Telescope.* Charles Griffin [1979 edition, Dover Publications, New York.]

Kirchhoff, G., 1883. Zur Theorie der Lichtstrahlen. Wiedemann Ann. d. Physik (2) 18, 663.

Knox, K. T., and B. J. Thompson, 1974. Recovery of images from atmospherically degraded short-exposure images. Astrophys. J. **193,** L45-L48.

Koechlin, L., and Y. Rabbia, 1985. Mesures de diametres a l'interferometre optique du CERGA. Developpements et resultats recents. Astron. Astrophys., **153,** 91-98.

Koechlin, L., D. Bonneau, and F. Vakili, 1979. Detection d'un effet de phase a l'interferometre du CERGA, application au mouvement orbital de Capella. Astron. Astrophys. 80, L13–L14.

Kokorowski, S. A., 1979. Analysis of adaptive optical elements made from piezoelectric biomorphs. J. Opt. Soc. Am. **69,** 181-187.

Koliopoulos, C. L., 1980. Radial grating lateral shear heterodyne interferometer. Appl. Optics **19,** 1523-1528.

Kolmogorov, A. N., 1941. Dissipation of energy in locally isotropic turbulence. Doklady Akad. Nauk SSSR **32,** 16. (Translation in *Turbulence, Classic Papers on Statistical Theory,* eds. S. K. Friedlander and L. Topper, Interscience, New York, 1961.)

Kuo, B. C, 1987. *Automatic Control Systems.* Prentice-Hall, Englewood Cliffs, NJ.

Kupke, R., F. Roddier, and D. C. Mickey, 1994. Curvature-based wavefront sensor for use on extended patterns. Proc. SPIE **2201,** 519-527.

Labeyrie, A., 1975. Interference fringes obtained on Vega with two optical telescopes. Astrophys. 196, L71-L75.

Lawrence, G. N., C. Huang, E. H. Levy, and R. S. McMillan, 1991. High accuracy image centroiding with a moving Ronchi ruling. Opt. Eng. **30,** 598-608.

Lawrence, R. S., 1976. A review of the optical effects of the clear turbulent atmosphere. Proc. SPIE **75,** 2-8.

Lawrence, R. S., and J. W. Strohbehn, 1970. A survey of clear-air propagation effects relevant to optical communications. Proc. IEEE 58, 1523-1545.

Lee, R. W., and J. C. Harp, 1969. Weak scattering in random media, with applications to remote probing. Proc. IEEE 57, 375–406.

Leighton, R. B., 1956. Concerning the problem of making sharper images of the planets. Sci. Am. 194, 157–166.

Lena, P. J., 1994. Astrophysical results with the Come-On + adaptive optics system. Proc. SPIE 2201, 1099–1109.

Levine, B. M., J. R. Janesick, and J. C. Shelton, 1994. Development of a low noise high frame rate CCD for adaptive optics. Proc. SPIE **2201,** 596-606.

Lewis, R. R., R. A. Benfield, M. R. H. Knowles, and A. J. Kearsley, 1993. High power copper vapour lasers for laser guide star applications, *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics* (ESO Conference and Workshop Proceedings No. 48), ed. F. Merkle, Garching, Germany, 2-5 Aug., 1993, 499-502.

Lim, Y. S., and M. Subramanian, 1972. Improved image position sensor for high resolution optical tracking. Appl. Optics 11, 890–894.

Linfoot, E. H., 1956. Transmission factors and optical design. J. Opt. Soc. Am. **46,** 740-752.

Linfoot, E. H., 1964. *Fourier Methods in Optical Image Evaluation.* Focal Press, London.

Linnik, V. P., 1993. On the possibility of reducing the influence of atmospheric seeing on the image quality of stars. Original 1957 article translated and reprinted in ESO Conference and Workshop Proceedings No. 48, ed. F. Merkle, Garching, Germany, 535-537.

Liu, Y. C., and A. Lohmann, 1973. High resolution image formation through the turbulent atmosphere. Opt. Commun. 8, 372-377.

Livingston, W. C, 1987. Astronomical detectors for .001 to 300,000 A, an overview. Optics (Paris) **18,** 187-192.

Lloyd-Hart, M., R. Dekany, D. Sandler, D. Wittman, R. Angel, and D. McCarthy, 1994. Progress in diffraction-limited imaging at the Multiple Mirror Telescope with adaptive optics. J. Opt. Soc. Am. **All.** 846-857.

Lloyd-Hart, M. et al., 1995. Adaptive optics experiments at the MMT using sodium laser guide stars. Astrophys. J. **439,** 455.

Lloyd-Hart, M., J. R. P. Angel, D. G. Sandler, T. D. Groesbeck, T. Martinez, and B. P. Jacobsen, 1997. Design of the 6.5-m MMT adaptive optics system and results from its prototype system FASTTRAC II. Proc. SPIE **2871,** 880-889.

Loos, G. C, and C. B. Hogge, 1979. Turbulence of the upper atmosphere and isoplanatism. Appl. Optics **18,** 2654-2661.

Love, G D., 1993. Liquid-crystal phase modulator for unpolarized light. Appl. Optics **32,** 2222-2223.

Lutomirski, R. F., W. L. Woodie, and R. G. Buser, 1977. Turbulence-degraded beam quality: improvement obtained with a tilt-correcting aperture. Appl. Optics 16, 665-673.

Lyot, B., 1930. La couronne solaire etudiee en dehors des eclipses. C. R. Acad. Sci. (Paris) **191,** 834.

Lyot, B., 1931. Photographie de la couronne solaire en dehors des eclipses. C. R. Acad. Sci. (Paris) **193,** 1169.

Ma, S., P. Guthals, P. Hu, and B. Campbell, 1994. Atmospheric-turbulence compensation with self-referenced binary holographic interferometry. J. Opt. Soc. Am. **All,** 428-433.

Maciejowski, 1989. *Multivariate Feedback Design.* Addison-Wesley, Reading, MA.

Mackay, C. D., J. E. Baldwin, J. Rogers, and G. Cox, 1994. Avalanche photodiodes and area CCDs for fast guiding and wavefront sensing applications. Proc. SPIE **2201,** 613-618.

Mahajan, V. N., 1976. Optical wavefront correction in real time. Proc. SPIE **75,** 109–118.

Mahan, A. I., 1962. Astronomical refraction — some history and theories. Applied Optics 1, 497–511.

Malbet, F., M. Shao, and J. Yu, 1994. Active optics and coronography with the Hubble Space Telescope. Proc. SPIE **2201,** 1135–1144.

Marechal, A., 1947. Rev. d'Optique 26, 257.

Marth, H., M. Donat, and C. Pohlhammer, 1991. Latest experience in design of piezoelectric driven fine steering mirrors. Proc. SPIE **1543,** 248-261.

Martin, H. M., and D. S. Anderson, 1995. Techniques for the optical fabrication of a 2-mm thick adaptive secondary mirror. Proc. SPIE 2534, 134-139.

Martin, H. M., J. H. Burge, D. A. Ketelson, and S. C. West, 1997. Fabrication of the 6.5 m primary mirror for the Multiple Mirror Telescope conversion. Proc. SPIE **2871,** 399-404.

Mason, W. P., 1950. *Piezoelectric Crystals and their Application to Ultrasonics,* Von Nostrand, Princeton, NJ.

Max, C. E., K. Avicola, H. D. Bissinger, J. M. Brase, D. T. Gavel, H. W. Friedman, J. R. Morris, S. S. Olivier, J. T. Salmon, and K. E. Waltjen, 1992. Feasibility experiment for sodium-layer laser guide stars at the Lawrence Livermore National Laboratory. *Laser Guide Star Adaptive Optics Workshop: Proceedings,* ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB, NM, March 1992, 535-541.

Max, C. E., K. Avicola, J. M. Brase, H. W. Friedman, H. D. Bissinger, J. Duff, D. T. Gavel, J. A. Horton, J. R. Morris, S. S. Olivier, R. W. Presta, D. A. Rapp, J. T. Salmon, and K. E. Waltjen, 1994. Design, layout, and early results of a feasibility experiment for sodium-layer guide star adaptive optics. J. Opt. Soc. Am. **All,** 813-824.

Mcalister, H. A., W. G. Bagnuolo, T. Ten Brummelaar, W. I. Hartkopf, N. H. Turner, A. K. Garrison, W. G. Robinson, and S. T. Ridgeway, 1994. CHARA array. Proc. SPIE **2200,** 129-139.

McCall, S. L., T. R. Brown, and A. Passner, 1977. Improved optical stellar image using a real-time phase correction system: initial results. Astrophys. J. **211,** 463-468.

McCarthy, D. W., B. A. McLeod, and D. Barlow, 1990. An infrared array camera for interferometry with the cophased Multiple Mirror Telescope. Proc. SPIE **1237,** 496-507.

McClure, R. D., W. A. Grundmann, W. N. Rambold, J. M. Fletcher, E. H. Richardson, J. R. Stillburn, R. Racine, C. A. Christian, and P. Waddell, 1989. An image stabilization high-resolution camera for the Canada France Hawaii Telescope. Pub. Astr. Soc. Pac. **101,** 1156.

McKechnie, T. S., 1991. Light propagation through the atmosphere and the properties of images formed by large ground-based telescopes. J. Opt. Soc. Am. A8, 346-365.

McKnight, D. J., K. M. Johnson, and R. A. Serati, 1994. 256 x 256 Liquid-crystal-on-silicon spatial light modulator. Appl. Optics **33,** 2775-2784.

Measures, R. M., 1984. *Laser Remote Sensing: Fundamentals and Applications,* Wiley, New York.

Medbury, J. D., and A. A. Gupta, 1991. Stable controller design of a six degree-of-freedom magnetically suspended fine-steering mirror. Proc. SPIE **1543,** 276-285.

Megie, G, F. Bos, J. E. Blamont, and M. L. Chanin, 1977. Simultaneous nighttime lidar measurements of atmospheric sodium and potassium. Planet. Space Sci. 26, 27-35.

Mehta, N. C, and C. W. Allen, 1994. Dynamic compensation of atmospheric turbulence with far-field optimization. J. Opt. Soc. Am. **A11,** 434-443.

Menikoff, A., 1987. Theory of reconstructors using resistor networks. Itek Optical Systems, unpublished report.

Merkle, F., 1988. Synthetic aperture imaging with the European Very Large Telescope. J. Opt. Soc. Am. A5, 904-913.

Merkle, F., G. Gehring, F. Rigaut, P. J. Lena, G. Rousset, J. Fontanella, and J. Gaffard, 1991. Adaptive optics system tests at the ESO 3.6 m telescope. Proc. SPIE 1542, 308-318.

Merritt, P. H., L. Sher, and R. E. Walter, 1992. Stability measures of adaptive optics control systems. *Laser Guide Star Adaptive Optics Workshop: Proceedings,* ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB., NM, March 1992, 352-363.

Michelson, A. A., 1890. On the application of interference methods to astronomical methods. Phil. Mag. 30, 1-20.

Michelson, A. A., 1891. Measurement of Jupiter's satellites by interference. Nature **45,** 160-161.

Michelson, A. A., and F. G. Pease, 1921. Measurement of the diameter of α Orionis with the interferometer. Astrophys. J. **53,** 249-259.

Mikesell, A. H., 1955. The scintillation of starlight. Publ. U.S. Naval Obs. 17, 139-191.

Miller, L., W. P. Brown, Jr, J. A. Jenney, and T. R. O'Meara, 1974. Imaging through turbulence with coherent adaptive optical techniques. Optical Society of America (OSA) Topical meeting on Optical Propagation through turbulence, Univ. of Colorado, Boulder, CO, July 1974, Paper ThB2.

Miller, N., J. S. Fender, and D. W. Tyler, 1993. A methodology for the adaptive optics design for the advanced electro-optical system (AEOS) *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics,* (ESO Conference and Workshop Proceedings No 48), Garching, Germany, 2-5 Aug., 1993, 83-88.

Modisett, D. L., and H. M. Martin, 1995. Optical measurements of a prototype adaptive-optic secondary mirror. Proc. SPIE **2534,** 140-149.

Morris, J. R., 1994. Efficient excitation of a mesospheric-sodium laser guide star by intermediate duration pulses. J. Opt. Soc. Am. **A11,** 832-845.

Morris, S., T. Davidge, J. Hutchings, P. Stetson, D. Durand, D. Crabtree, and P. Cote, 1996. Science drivers for adaptive optics on Gemini North. Gemini Newsletter, June 1996.

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

Murphy, D. V., C. A. Primmerman, B. G. Zollars, and H. T. Barclay, 1991. Experimental demonstration of atmospheric compensation using multiple synthetic beacons. Optics Letts 16, 1797–1799.

Myers, R. M., A. P. Doel, C. N. Dunlop, J. V. Major, R. M. Sharpies, and A. J. A. Vick, 1994. An astronomical AO system for use on a 4 m class telescope at optical wavelengths. Proc. SPIE **2201,** 437–446.

Myers, R. M., A. J. Longmore, R. A. Humphreys, G. F. Gilmore, B. Gentles, M. Wells, and R. W. Wilson, 1995. The UK Adaptive Optics Programme. Proc. SPIE **2534,** 48-52.

Nakajima, T., 1994. Planet detectability by an adaptive optics stellar coronagraph. Astrophys. J. **425,** 348-357.

Nakajima, T., S. T. Durrance, D. A. Golimowski, and S. R. Kulkarni, 1994. A coronagraphic search for brown dwarfs around nearby stars. Astrophys. J. **428,** 797-804.

Nakajima, T., B. R. Oppenheimer, S. R. Kulkarni, D. A. Golimowski, K. Matthews, and S. T. Durrance, 1995. Discovery of a cool brown dwarf. Nature 378, 463–465.

Neal, D.R., M. E. Warren, J. K. Gruetzner, T. G. Smith, and R. R. Rosenthal, 1994. A multi-tiered wavefront sensor using binary optics., Proc. SPIE **2201,** 574-585.

Neal, D. R., J. D. Mansell, J. K. Gruetzner, R. Morgan, and M. E. Warren, 1995. Specialized wavefront sensors for adaptive optics., Proc. SPIE **2534,** 338-348.

Nelson, J. E., and T. S. Mast, 1990. The construction of the Keck Observatory, in *Advanced Technology Optical Telescopes IV; Proceedings of the Meeting,* Society of Photo-Optical Instrumentation Engineers, Tucson, AZ, 12-16 Feb., 1990, Part 1, 47-55.

Nightingale, N. S., 1991. A new silicon avalanche photodiode photon counting detector module for astronomy. Exper. Astron. 1, 407-422.

Nisenson, P., and R. Barakat, 1987. Partial atmospheric correction with adaptive optics. J. Opt. Soc. Am. A4, 2249-2253.

Nisenson, P., D. C. Ehn, and R. V. Stachnik, 1976. Astronomical speckle imaging. Proc. SPIE 75, 83-88.

Nisenson, P., J. Apt, R. Goody, and C. Papaliolios, 1983. Speckle inaging for planetary research. Icarus **53,** 465-478.

Noll, R. J., 1976. Zernike polynomials and atmospheric turbulence. J. Opt. Soc. Am. 66, 207–211.

Noll, R. J., 1978. Phase estimates from slope sensors. J. Opt. Soc. Am. 68, 139-140.

Nyyssonen, D., and J. M. Jerke, 1973. Lens testing with a simple wavefront shearing interferometer. Appl. Optics **12,** 2061-2070.

Obukhov, A. M., 1949. Structure of the temperature field in a turbulent flow. Izv. Akad. Nauk SSSR, Ser Geograf. Geofiz. **13,** 58.

O'Byrne, J., 1995. Adaptive optics at the Anglo-Australian Telescope. Pub. Astron. Soc. Australia **12,** 106–116.

Ochs, G. R., Ting-i Wang, R. S. Lawrence, and S. F. Clifford, 1976. Refractive-turbulence profiles measured by one-dimensional spatial filtering of scintillations. Appl. Optics **15,** 2504–2510.

Olivier, S. S., and D. T. Gavel, 1994. Tip-tilt compensation for astronomical imaging. J. Opt. Soc. Am. **All,** 368-378.

Olivier, S. S., J. An, K. Avicola, H. D. Bissinger, J. M. Brase, H. W. Friedman, D. T. Gavel, E. M. Johansson, C. E. Max, K. E. Waltjen, W. A. Fisher, and W. Bradford, 1994. Performance of adaptive optics at Lick Observatory. Proc SP1E. **2201,** 1110–1120.

Olivier, S. S., J. An, K. Avicola, H. D. Bissinger, J. M. Brase, H. W. Friedman, D. T. Gavel, C. E. Max, J. T. Salmon, and K. E. Waltjen, 1995. Performance of laser guide star adaptive optics at Lick Observatory. Proc. SPIE **2534,** 26-37.

O'Malley, M., and E. O'Mongain, 1992. Charge-coupled devices: frame adding as an alternative to long integration times and cooling. Opt. Eng. **31,** 522-526.

O'Meara, T. R., 1977. The multi-dither principle in adaptive optics. J. Opt. Soc. Am. **67,** 306-315.

O'Neill, E. L., 1963. *Introduction to Statistical Optics.* Addison-Wesley, Reading, MA.

Orias, G, A. W. Hoffman, and M. F. Casselman, 1986. 58 x 62 InSb focal plane array for infrared astronomy, in *Instrumentation in Astronomy VI; Proceedings of the Meeting,* Society of Photo-Optical Instrumentation Engineers, Tucson, AZ, 4-8 Mar., 1986, Part 2, 408-417.

Osterbrock, D. E. 1984. *James E. Keeler: Pioneer American Astrophysicist.* Cambridge University Press, New York.

Owens, J. C, 1967. Optical refractive index of air: dependence on pressure, temperature and composition. Appl. Optics 6, 51-59.

Pan, X., M. Shao, and M. M. Colavita, 1994. Binary star astronomy with optical interferometry, Proc. SPIE **2200,** 360-371.

Parenti, R. R., and R. J. Sasiela, 1994. Laser-guide-star systems for astronomical applications. J. Opt. Soc. Am. **All,** 288-309.

Parulski, K. A., I. J. D'Luna, B. L. Benjamin, and P. R. Shelley, 1992. High-performance digital color video camera. J. Electr. Imaging 1, 35–45.

Paxman, R. G, and J. R. Fienup. 1988. Optical misalignment sensing and image reconstruction using phase diversity. J. Opt. Soc. Am. A5, 914–923.

Pearson, J. E., 1976. Atmospheric turbulence compensation using coherent optical adaptive techniques. Appl. Optics **15,** 622-631.

Pearson, J. E., W. B. Bridges, S. Hansen, T. A. Nussmeier, and M. E. Pedinoff, 1976. Coherent optical adaptive techniques: design and performance of an 18-element visible multidither COAT system. Appl. Optics **15,** 611–621.

Pearson, T. J., and A. C. S. Readhead, 1984. Image formation by self-calibration in radio astronomy. *Ann. Rev. Astron. Astrophys.* **22,** 97-130.

Pease, F. G., 1930. The new fifty-foot stellar interferometer. Sci. Am. **143,** 290-294.

Pedersen, O., 1974. *Early Physics and Astronomy,* Revised Edition, 1993. Cambridge University Press, New York.

Pennington, T. L., D. W. Swindle, M. D. Oliker, B. L. Ellerbroek, and J. M. Spinhirne, 1995. Performance measurements of Generation III wavefront sensors at the Starfire Optical Range. Proc. SPIE **2534,** 327-337.

Peters, W. N., R. A. Arnold, and S. Gowrinathan, 1974. Stellar interferometer for figure sensing of orbiting astronomical telescopes. Appl. Optics 13, 1785–1795.

Peterson, B. A., S. C'Odorico, M. Tarenghi, and E. J. Wampler, 1991. The NTT provides the deepest look into space. ESO The Messenger No. 64, 1-3.

Plante, R. L., 1991. Large Active Mirror Program (LAMP). Proc. SPIE **1543,** 146-160.

Racine, R., and B. L. Ellerbroek, 1995. Profiles of night-time turbulence above Mauna Kea and isoplanatism extension in adaptive optics. Proc. SPIE **2534,** 248-257.

Racine, R., and R. D. McClure, 1989. An image stabilization experiment at the Canada France Hawaii Telescope. Pub. Astr. Soc. Pac. **101,** 731.

Reasenberg, R. D., 1990. IOTA Interferometer Project: plans, engineering and laboratory results, Proc. SPIE **1237,** 128-137.

Reskinoff, H. L., 1993. Wavelets and adaptive signal processing. Opt. Eng. **31,** 1229-1234.

Reynolds, G, 1986. A review of partially-filled, synthetic aperture imaging systems. Proc. SPIE **643,** 141.

Ribak, E., C. Schwartz, and S. G. Lipton, 1993. Bimorph adaptive mirrors: construction and theory. *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics* (ESO Conference and Workshop Proceedings No. 48), Garching, Germany, 2-5 Aug., 1993, 313-318.

Richard, J.-C, L. Bergonzi, and M. Pemonier, 1990. 604 x 288 Electron-bombarded CCD image tube for 2D photon counting. Proc. SPIE **1338,** 241-254.

Rigaut, F., and E. Gendron., 1992. Laser guide stars in adaptive optics: the tilt determination problem. Astron. Astrophys. **261,** 677-684.

Rigaut, F., G. Rousset, P. Kern, J. C. Fontanella, J. P. Gaffard, F. Merkle, and P. Lena, 1991. Adaptive optics on a 3.6 m telescope: results and performances. Astron. Astrophys. **250,** 280.

Rigaut, F., R. Arsenault, J. Kerr, D. Salmon, M. Northcott, Y. Dutil, and C. Boyer, 1994. The Canada-France-Hawaii adaptive optics Bonnette II: simulations and control. Proc. SPIE **2201** 149-160.

Rimmele, T., O. von der Luhe, P. H. Wiborg, A. L. Widener, R. B. Dunn, and G. Spence, 1991. Solar feature correlation tracker. Proc. SPIE **1542,** 186-193.

Rimmer, M. P., 1974. Method for evaluating lateral shear interferograms. Appl. Optics **13,** 623-629.

Roddier, C, F. Roddier, M. J. Northcott, J. E. Graves, and K. Jim, 1996. Adaptive optics imaging of GG Tau: optical detection of the circumbinary ring. Astrophys. J. May 20, 1996.

Roddier, F., 1979. Rotation-shearing Interferometry. *Proc. IAU Colloquium No. 50,* High Angular Resolution Stellar Interferometry, Paper 32.

Roddier, F., 1981. *The Effects of Atmospheric Turbulence in Optical Astronomy.* Progress in Optics XIX, ed. E. Wolf. North-Holland, New York, 283-376.

Roddier, F., 1988a. Curvature sensing and compensation: a new concept in adaptive optics. Appl. Optics **27,** 1223-1225.

Roddier, F. 1988b. Interferometric imaging in optical astronomy. Physics Reports (Review section of Phys. Letts) **170, 97–166.**

Roddier, F.,1990a. Variations on a Hartmann theme. Opt. Eng. **29,** 1239-1242.

Roddier, F., 1990b. Wavefront sensing and the irradiation transport equation. Appl. Optics **29,** 1402-1403.

Roddier, F., and P. Lena, 1984a. Long-baseline Michelson interferometry with large ground-based telescopes operating at optical wavelengths. I: General formalism, interferometry at visible wavelengths. Optics (Paris) **15,** 171-182.

Roddier, F., and P. Lena, 1984b. Long-baseline Michelson interferometry with large ground-based telescopes operating at optical wavelengths II: Interferometry at infrared wavelengths, optics (Paris) **15,** 363-374.

Roddier, F., and C. Roddier, 1986. NOAO Infrared Adaptive Optics Program II: Modeling atmospheric effects in adaptive optics systems for astronomical telescopes. Proc. SPIE **628,** 298-304.

Roddier, F., C. Roddier, and N. Roddier, 1988. Curvature sensing: a new wavefront sensing method. Proc. SPIE **976,** 203-209.

Roddier, F., L. Cowie, J. E. Graves, S. Songaila, D. McKenna, J. Vernin, M. Azouit, J. L. Caccia, E. Limburg, C. Roddier, D. Salmon, S. Beland, D. Cowley, and S. Hill, 1990. Seeing at Mauna Kea: a joint UH–UN–NOAO–CFHT study. Proc. SPIE 1236,485-491.

Roddier, F., M. Northcott, and J. E. Graves, 1991. A simple low-order adaptive optics system for near-infrared applications. Pub. Astr. Soc. Pac. **103,** 131-149.

Roddier, F., J. E. Graves, D. McKenna, and M. Northcott, 1991. The University of Hawaii adaptive optics system, I. General approach. Proc. SPIE, **1542,** 248-253.

Roddier, F., J. Anuskiewicz, J. E. Graves, M. J. Northcott, and C. Roddier, 1994. Adaptive optics at the University of Hawaii, I: Current performance at the telescope. Proc SPIE **2201,** 2-9.

Roddier, F., C. Roddier, J. E. Graves and M. J. Northcott, 1995. Adaptive optics imaging of proto-planetary nebulae Frosty Leo and the Red Rectangle. Astrophys. J. **443,** 249-260.

Roddier, N., 1990. Atmospheric wavefront simulation using Zernike polynomials. Opt. Eng. **29, 1174–1180.**

Roetling, P. G., E. A. Trabka, and R. E. Kinzly, 1968. Theoretical prediction of image quality. J. Opt. Soc. Am. **58,** 342-346.

Roggeman, M. C, 1991. Limited degree-of-freedom adaptive optics and image reconstruction. Appl. Optics **30, 4227–4233.**

Roggeman, M. C, and C. L. Matson, 1992. Power spectrum and Fourier phase spectrum estimation by using fully and partially compensating adaptive optics and bispectrum postprocessing. J. Opt. Soc. Am. A9, 1525-1535.

Rogstad, D. H., 1968. A technique for measuring visibility phase with an optical interferometer in the presence of atmospheric seeing. Appl. Optics 7, 585-588.

Roland, J. J., J.-P. Gaffard, P. Jagourel, C. Boyer, and P. Gosselin, 1994. Adaptive optics: a general purpose system for astronomy. Proc. SPIE **2201,** 58-76.

Rose, A., 1948. The sensitivity performance of the human eye on an absolute scale. J. Opt. Soc. Am. **38,** 196.

Rousset, G., J.-C. Fontanella, P. Kern, P. Gigan, F. Rigaut, P. Lena, C. Boyer, P. Jagourel, J.-P. Gaffard, and F. Merkle, 1990. First diffraction-limited astronomical images with adaptive optics. Astron. Astrophys. **230,** L29-L32.

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

Rousset, G., J. L. Beuzit, N. Hubin, E. Gendron, C. Boyer, P. Y. Madec, P. Gigan, J. C. Richard, M. Vittot, J. P. Gaffard, F. Rigaud, and P. Lena, 1993. The Come-On-Plus adaptive optics system: results and performance. *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics,* (ESO Conference and Workshop Proceedings No. 48), Garching, Germany, 2-5 Aug., 1993, 65-70.

Rousset, G, J. L. Beuzit, N. Hubin, E. Gendron, P. Y. Madec, C. Boyer, J. P. Gaffard, J. C. Richard, M. Vittot, P. Gigan, and P. J Lena, 1994. Performance and results of the Come-On + adaptive optics system at the ESO 3.6 m telescope. Proc. SPIE **2201,** 1088-1098.

Salinari, P., C. Del Vecchio, and V. Biliotti, 1993. A study of an adaptive secondary mirror. *Proc. ICO-16 Satellite Conference on Active and Adaptive*

Optics (ESO Conference and Workshop Proceedings No. 48), Garching, Germany, 2-5 Aug., 1993, 247-253.

Sandler, D. G., 1992a. A multiple spot laser beacon for high-order wavefront control: theory and experiment. *Laser Guide Star Adaptive Optics Workshop Proceedings*, ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB, NM, March 1992, 164–195.

Sandler, D. G., 1992b. A new 500-channel adaptive optic system for laser guide star experiments. *Laser Guide Star Adaptive Optics Workshop Proceedings*, ed. R. Q. Fugate, Phillips Laboratory, Kirtland AFB, NM, March 1992, 686-717.

Sandler, D. G., L. Cuellar, M. Lefebvre, T. Barrett, R. Arnold, P. Johnson, A. Rego, G. Smith, G. Taylor, and B. Spivey, 1994. Shearing interferometry for laser-guide-star atmospheric correction at large *D/r*. J. Opt. Soc. Am. **All,** 858-873.

Sandler, D. G., S. Stahl, J. R. P. Angel, M. Lloyd-Hart, and D. McCarthy, 1994. Adaptive optics for diffraction-limited infrared imaging with 8-m telescopes. J. Opt. Soc. Am. **All.** 925-945.

Sandler, D. G., M. Lloyd-Hart, T. Martinez, P. M. Gray, J. R. P. Angel, T. K. Barrett, D. G. Bruns, and S. M. Stahl, 1995. 6.5-m MMT infrared adaptive optics system: detailed design and progress report. Proc. SPIE 2534, 372-377.

Sandler, D. G., S. Stahl, and J. R. P. Angel, 1995. Adaptive optics system for direct imaging of of extra- solar planets from the ground. Proc. SPIE **2534,** 378-385.

Sasian, J. M., 1992. Image plane tilt in optical systems. Opt. Eng. **31,** 527-532.

Sasiela, R. J., 1988. A Unified Approach to Electromagnetic Wave Propagation in Turbulence and the Evaluation of Multiparameter Integrals. Massachusetts Institute of Technology, Lincoln Laboratory Technical Report 807.

Sasiela, R. J., 1994. Wave-front correction by one or more synthetic beacons. J. Opt. Soc. Am. **All,** 379-393.

Sasiela, R. J., and J. G. Mooney, 1985. An optical phase reconstructor based on using a multiplier-accumulator approach. Proc. SPIE **551,** 170-176.

Sasiela. R. J., and J. D. Shelton, 1993a. Transverse spectral filtering and Mellin transform techniques applied to the effect of outer scale on tilt and tilt anisoplanatism. J. Opt. Soc. Am. **A10,** 646–660.

Sasiela. R. J., and J. D. Shelton, 1993b. Mellin transform methods applied to integral evaluation: Taylor series and asymptotic approximations. J. Math. Phys. 34, 2572-2617.

Sawicki, R. H., and W. Sweatt, 1987. Continuously deformable mirrors. Proc. SPIE 779, 23-27.

Schloerb, F. P., 1990. Image reconstruction techniques for the infrared-optical array. Proc. SPIE **1237,** 154–165.

Schwartz, C, G. Baum, and E. N. Ribak, 1994. Turbulence-degraded wavefronts as fractal surfaces. J. Opt. Soc. Am. **All,** 444–451.

Schwartz, C, E. N. Ribak, and S. G. Lipson, 1994. Bimorph adaptive mirrors and curvature sensing. J. Opt. Soc. Am. **All,** 895-902.

Sechaud, M., N. Hubin, L. Brixon, R. Jalin, R. Foy, and M. Tallon, 1988. Laser backscattered reference star for atmospheric wavefront disturbances measurement. *Proc. ESO Conference on Very Large Telescopes and Instrumentation*, ed. M.-H. Ulrich, Garching, Germany.

Seifert, W., Mitsch, H. Nicklas, and G. Rupprecht, 1994. FORS — a workhorse instrument for the ESO VLT. Proc. SPIE **2198,** 213-218.

Shack, R. V., 1956. Outline of practical characteristics of an image-forming system. J. Opt. Soc. Am. 46, 755-757.

Shack, R. V., and B. C. Platt, 1971. Production and use of a lenticular Hartmann screen. J. Opt. Soc. Am. 61, 656 (abstract only).

Shao, M., and M. M. Colavita, 1992. Long-baseline optical and infrared stellar interferometry. Ann. Rev. Astron Astrophys. **30,** 457–498.

Shao, M. and D. H. Staelin, 1977. Long baseline optical interferometry for astronomy. J. Opt. Soc. Am. 67, 81-86.

Shao, M., and D. H. Staelin, 1980. First fringe measurements with a phase-tracking stellar interferometer. Appl. Optics **19,** 1519-1522.

Shao, M., M. Colavita, and D. H. Staelin, 1986. The Mark III astrometric interferometer. Proc. SPIE **628,** 250-254.

Shapiro, J. H., 1971a. Optimal power transfer through atmospheric turbulence using state knowledge. IEEE Trans. Commun. Technol. **19,** 410-414.

Shapiro, J. H., 1971b. Reciprocity of the turbulent atmosphere. J. Opt. Soc Am. 61, 492-495.

Shapiro, J. H., 1974a. Optimum adaptive imaging through atmospheric turbulence. Appl. Optics 13, 2609-2613.

Shapiro, J. H., 1974b. Normal-mode approach to wave propagation in the turbulent atmosphere. Appl. Optics **13,** 2614-2619.

Shapiro, J. H., 1976. The seeing limit can resolve the isoplanatic patch. J. Opt. Soc. Am. 66, 469–477.

Shelton, J. C, and S. L. Baliunas, 1993. Results of adaptive optics at Mount Wilson Observatory. Proc. SPIE **1920,** 371.

Shelton, J. C, S. L. Baliunas, J. Russell and R. A. Donahue, 1993. Adaptive optics at Mount Wilson Observatory. *Proc. ICO-16 Satellite Conference on Active and Adaptive Optics*, (ESO Conference and Workshop Proceedings No. 48), Garching, Germany, 2-5 Aug., Aug., 1993, 53-58.

Shelton, J. C, T. Schneider, D. McKenna, and S. L. Baliunas, 1995. First tests of the Cassegrain adaptive optics system of the Mount Wilson 100-inch telescope. Proc. SPIE **2534,** 72-79.

Siegman, A. E., 1971. *An Introduction to Lasers and Masers*. McGraw-Hill, New York.

Simons, D., and F. Gillett, 1996. Gemini Phase II Instrumentation. Gemini Newsletter, June 1996.

Smithson, R. C, 1987. A segmented active mirror for solar observations. Proc. SPIE **779,** 18–22.

Smithson, R. C, M. L. Peri, and R. S. Benson, 1988. Quantitative simulations of image correction for astronomy with a segmented active mirror. Appl. Optics **19,** 1615-1620.

Smutko, M. F., M. Chun, F. Shi, V. Scov, W. Wild, J. Larkin, and E. Kibblewhite, 1996. An overview of the Chicago adaptive optics system. Poster presented at the American Astronomical Society San Antonio Meeting, Jan., 1996.

Sobel, D., 1995. *Longitude.* Walker, New York.

Southwell, W. H., 1977. Wavefront analyser using a maximum likelihood algorithm. J. Opt. Soc. Am. **67,** 396-399.

Southwell, W. H., 1980. Wave-front estimation from wave-front slope measurements. J. Opt. Soc. Am. **70,** 998.

Southwell, W. H., 1982. What's wrong with cross coupling in modal wavefront estimation? Proc. SPIE 365, 97-104.

Sparrow, C. M., 1916. On spectroscopic resolving power. Astrophys. J. **44,** 76-86.

Stachnik, R. V., P. Nisenson, D. C. Ehn, R. H. Hudgin, and V. E. Schirf, 1977. Speckle image reconstructions of solar features. Nature **266,** 149-151.

Stephan, E., 1874. Sur l'extreme petitesse du diametre apparent des etoile fixes. C. R. Acad. Sci. (Paris) 78, 1008-1112.

Stepp, L., N. Roddier, and D. Dryden, 1991. Active optics system for the 3.5-meter structured mirror. Proc. SPIE **1542,** 175-185.

Stepp, L. M., E. W. Huang, and M. K. Cho, 1994. Gemini primary mirror support system. Proc. SPIE **2199,** 223-238.

Stomski, P., P. Wizinowich, D. S. Acton, W. Lupton, J. Gathright, A. Conrad, H. Lewis, and T. Gregory, 1996. Keck adaptive optics: user interface and supervisory control subsystem. Opt. Soc. Am. Conf., Maui, HI, 8-12 July, 1996.

Stone, J., P. H. Hu, S. P. Mills, and S. Ma, 1994. Anisoplanatic effects in finite-aperture optical systems. J. Opt. Soc. Am. A**11,** 347-357.

Strehl, K., 1895. Zeits. Instrum. **15,** 364.

Strehl, K., 1902. Zeits. Instrum. **22,** 214.

Streibl, N., 1984. Phase imaging by the transport equation of intensity. Optics Comm. **49,** 6-10.

Strohbehn, J. W., 1968. Line-of-sight wave propagation through the turbulent atmosphere. Proc. IEEE **56,** 1301-1318.

Strohbehn, J. W., 1971. Optical propagation through the turbulent atmosphere. Progress in Optics IX, ed. E. Wolf. North-Holland, New York. 75-122.

Strohbehn, J. W., 1978. Modern theories in the propagation of optical waves in a turbulent medium, in *Laser Beam Propagation in the Atmosphere, Topics in Applied Physics,* Vol. 25, ed. J. W. Strohbehn. Springer-Verlag, Berlin, 45-106.

Sutton, E. C, 1979. Results and future uses of heterodyne spatial interferometry at 11 microns. *Proc. IAU Coll. 50, High Angular Resolution Stellar Interferometry,* Aug.-Sept. 1978, Paper 16.

Sutton, E. C, J. W. V. Storey, A. L. Betz, C. H. Townes, and D. L. Spears, 1977. Spatial heterodyns interferometry of VY Canis Majoris, Alpha Orionis, Alpha Scorpii and R Leonis at 11 microns. Astrophys. J. **217,** L97–L100.

Sutton, E. C, J. W. V. Storey, C. H. Townes, and D. L. Spears, 1978. Variations in the spatial distribution of 11 micron radiation from Omicron Ceti. Astrophys. J. **224,** L123-L126.

Swenson, G. W., 1986. Radio-astronomy precedent for optical interferometer imaging. J. Opt. Soc. Am. A3, 1311-1319.

Swenson, G. W., C. S. Gardner, and R. H. T. Bates, 1986a. Optical synthesis telescopes. Proc. SPIE **643,** 129-140.

Swenson, G. W., C. S. Gardner, and R. H. T. Bates, 1986b. On prospects for an extremely large optical/infrared array telescope. Proc. SPIE **628,** 277-280.

Swift, C, E. Bliss, and D. Lenz, 1990. Deformable mirror for zig zag solid-state lasers. Opt. Eng. **29,** 1199-1203.

Takami, H., and M. Iye. 1994. Membrane deformable mirror for SUBARU adaptive optics. Proc. SPIE **2201,** 762-767.

Tarbell, T. D., and R. C. Smithson, 1981. A simple image motion compensation system for solar observations. *Proc. Sac. Peak Nat. Obs. Conference: Solar Instrumenation — What's Next?* ed. R. B. Dunn, Sunspot NM, Mar. 1991, 491-500.

Tarengi, M., and R. N. Wilson, 1989. ESO new technology telescope. Proc. SPIE **1114,** 302.

Tatarski, V. I., 1961. *Wave Propagation in a Turbulent Medium.* McGraw-Hill, New York.

Tatarski, V. I., 1971. The Effect of the Turbulent Atmosphere on Wave Propagation. NTIS Report, U.S. Dept. of Commerce, Springfield, VA.

Teague, M. R., 1982. Irradiance Moments. J. Opt. Soc. Am. **72,** 1199.

Teague, M. R., 1983. Deterministic phase retrieval. J. Opt. Soc. Am. **73,** 1434.

Thompson, A. R., J. M. Moran, and G. W. Swenson, Jr, 1986. *Interferometry and Synthesis in Radio Astronomy,* Wiley, New York.

Thompson, L. A., 1994. UnISIS: University of Illinois Seeing Improvement System, an adaptive optics instrument for the Mt. Wilson 2.5 m telescope. Proc. SPIE **2201,** 1074-1087.

Thompson, L. A., and R. M. Castle, 1992. Experimental demonstration of a Rayleigh-scattered laser guide star at 351 nm. Optics Lett. **17,** 1485-1487.

Thompson, L. A., and C. S. Gardner, 1987. Experiments on laser guide stars at Mauna Kea Observatory for adaptive optics in astronomy. Nature **328, 229–231**.

Thompson, L. A., and C. S. Gardner, 1989. Excimer laser guide star techniques for adaptive imaging in astronomy. **Proc. SPIE 1114,** 184-190.

Thompson, L. A., and H. R. Ryerson, 1984. An active mirror stabilizing instrument system **(ISIS)** for use on Mauna Kea. Proc. SPIE **755,** 560-568.

Thompson, L. A., and X. Yao-Heng, 1995. Laser beacon system for the **UnISIS** adaptive optics system at the Mt. Wilson 2.5 meter telescope. Proc. SPIE **2534, 38–47**.

Thompson, L. A., R. Castle, and D. L. Carroll, 1991. Laser guide stars for adaptive optics systems: Rayleigh scattering experiments. Proc. SPIE **1542,** 110-119.

Thorburn, W. G., and L. Kaplan, 1991. A low voltage electrodistortive mirror system for wavefront control. Proc. SPIE **1543,** 52-63.

Timoshenko, S., and S. Woinowsky-Krieger, 1959. *Theory of Plates and Shells,* Second Edition. McGraw-Hill, New York.

Titterton, P., 1973. Scintillation. J. Opt. **Soc.** Am. 63, **439–444**.

Townes, C. H., W. C. Douchi, B. Sadoulet, and E. C. Sutton, 1986. Long baseline spatial interferometer for the **IR**. Proc. SPIE 628, 281-284.

Tyler, G. A., 1984. Turbulence-induced AO performance degradation: evaluation in the time domain. J. Opt. Soc. Am. **A1,** 251-262.

Tyler, G. A., 1994a. Bandwidth Considerations for tracking through turbulence. J. Opt. Soc. Am. **A11,** 358-367.

Tyler, G. A., 1994b. Merging: a new method for tomography through random media. J. Opt. Soc. Am. **A11,** 409-424.

Tyler, G. A., 1994c. Rapid evaluation of d_0: the effective diameter of a laser-guide-star adaptive-optics system. J. Opt. Soc. Am. **A11,** 325-338.

Tyler, G. A., 1994d. Wave-front compensation for imaging with off-axis guide stars. J. Opt. Soc. Am. **A11,** 339-346.

Tyler, G. A., and D. L. Fried, 1982. Image position error associated with a quadrant detector. J. Opt. Soc. Am. **72,** 804-808.

Tyson, J. A., 1990. Progress in low-light-level charge-coupled device imaging in astronomy. J. Opt. Soc. Am. A7, 1231-1236.

Urn, G. S., B. F. **Smithgall,** and C. L. O'Brien, 1982. Minimum variance estimation of wavefront aberration. Proc. SPIE **351, 96–102**.

Underwood, I., D. G. Vass, A. O'Hara, D. C. Burns, P. W. McOwan, and J. Gourlay, 1994. Improving the performance of liquid-crystal-over-silicon spatial light modulators: issues and achievements. **Appl.** Optics **33,** 2768-2774.

Ungar, P. J., D. S. Weiss, E. Riis, and S. Chu, 1989. Optical molasses and multi-level atoms: theory. J. Opt. Soc. Am. B6, 2058-2083.

Uscinski, B. J., 1977. *The Elements of Wave Propagation in Random Media.* McGraw-Hill, New York.

Valley, G. C, 1980. Isoplanatic degradation of tilt correction and short-term imaging systems. Appl. Optics 19, 574-577.

Valley, G. C, and S. M. **Wandzura,** 1979. Spatial correlation of phase-expansion coefficients for propagation through atmospheric turbulence. J. Opt. Soc. Am. 69, **712–717**.

Valley, S. L. (Ed.), 1965. *Handbook of Geophysics and Space Environments.* McGraw-Hill, New York.

Veldkamp, W. B., and G. J. Swanson, 1983. Developments in fabrication of binary optical elements. Proc. SPIE **437,** 54-59.

Vernin, J., and F. Roddier, 1973. Experimental determination of two-dimensional **spatiotemporal** power spectra of stellar light scintillation. Evidence for a multi-layer structure of the air turbulence in the upper troposphere. J. Opt. Soc. Am. 63, 270-273.

von der Luhe, O., 1988. Wavefront error measurement technique using extended incoherent light sources. Opt. Eng. **27,** 1078-1087.

von der Luhe, **O.,** A. L. Widener, Th. Rimmele, G. Spence, R. B. Dunn, and P. Wiborg, 1989. Solar feature correlation tracker for ground-based telescopes. Astron. Astrophys. **224,** 351-360.

Vural, K., 1987. Mercury cadmium telluride short-and-medium wavelength infrared staring focal plane arrays. Opt. Eng. 26, 201-208.

Wallner, E. P. 1976. The effects of Atmospheric dispersion on compensated imaging. Proc. SPIE **75, 119–125**.

Wallner, E. P., 1977. Minimizing atmospheric dispersion effects in compensated imaging. J. Opt. **Soc.** Am. 67, **407–409**.

Wallner, E. P., 1982. Comparison of wavefront sensor configurations using optimal reconstruction and correction. Proc. SPIE **351,** 42-53.

Wallner, E. P., 1983. Optimal wave-front correction using slope measurements. J. Opt. Soc. Am. **73,** 1771-1776.

Wallner, E. P., 1984. Comparison of diffractive and refractive effects in two-wavelength adaptive transmission. J. Opt. Soc. Am. A1, 785-787.

Wallner, E. P., 1985. Private communication.

Wallner, E. P., 1994. Optimizing the locations of **multiconjugate** wavefront correctors. Proc. SPIE **2201, 110–116**.

Wallner, E. P., and W. B. Wetherell, 1980. Atmospheric dispersion correctors with broad spectral bandpass for large telescopes, in *Optical and Infrared Telescopes for the 1990s,* ed. A. Hewitt. (KPNO Conference Proceedings), Vol. II.

Wang, C. P., 1975. Isoplanicity for imaging through turbulent media. Optics **Commun. 14,** 200-204.

Wang, J. Y., and J. K. Markey, 1978. Modal compensation of atmospheric turbulence phase distortion. J. Opt. Soc. Am. **68,** 78-87.

Weigelt, G., 1977. Modified astronomical speckle interferometry, speckle masking. Optical Commun. **21,** 55-59.

Welsh, B. M., 1991. Imaging performance analysis of AO telescopes using laser guide stars. Proc. SPIE **1542,** 88-99.

Welsh, B. M., and C. S. Gardner, 1989a. AO systems using slope sensors. J. Opt. Soc. Am. A6, 1913-1923.

Welsh, B. M., and C. S. Gardner, 1989b. Non linear resonant absorption effects on the design of resonance fluorescence lidars and laser guide stars. Appl. Optics **28,** 4141-4153.

Welsh, B. M., and C. S. Gardner, 1991. Effects of turbulence-induced anisoplanatism on the imaging performance of adaptive-astronomical telescopes using laser guide stars. J. Opt. Soc. Am. A8, 69-80.

Wild, W., B. Carter, and E. Kibblewhite, 1993. Laboratory tests of the Wavefront Control Experiment. Proc. SPIE **1920,** 263.

Wild, W., E. Kibblewhite, Fang Shi, B. Carter, G. Kelderhouse, R. Vuilleumier, and H. Manning, 1994. Field tests of the Wavefront Control Experiment. Proc. SPIE 2201, 1121-1134.

Wild, W., E. Kibblewhite, and V. Scor, 1994. Quasi-hexagonal deformable mirror geometries. Proc. SPIE **2201,** 726-739.

Wilson, R. N., F. Franza, P. Giordano, L. Noethe, and M. Tarenghi, 1988. Active optics: the NTT and the future. ESO The Messenger (El Mensajero) No. 53, 1-7.

Wilson, R. N., F. Franza, and L. Noethe, 1991. Set-up and performance of the optics of the ESO New Technology Telescope (NTT) in the observatory. J. Modern Optics 38, 219-243.

Wilson, R. W. and C. R. Jenkins, 1996. Adaptive optics for astronomy: theoretical performance and limitations. Mon. Not. R. Astron. Soc. **268,** 39-61.

Wirth, A., and R. Ruquist, 1985. Adaptive optical correction for extended source imaging. Proc. SPIE **551,** 127-130.

Withoft, C. G., 1990. Wavefront sensor noise reduction and dynamic range expansion by means of optical image intensification. Opt. Eng. **29,** 1233-1238.

Wizinowich, P., M. Lloyd-Hart, B. McLeod, D. Colucci, R. Delany, R. Angel, D. McCarthy, B. Hulburd, and D. Sandler, 1991. Neural network adaptive optics for the Multiple Mirror Telescope. Proc. SPIE **1542,** 148-158.

Wizinowich, P., D. S. Acton, A. Gleckler, T. Gregory, P. Stomski, K. Avicola, J. Brase, H. Friedman, D. Gavel, and C. Max, 1996. W. M. Keck Observatory adaptive optics facility. Opt. Soc. Am. Conf. Maui, HI, 8-12 July, 1996.

Wolfe, W. L., and G. J. Zissis (Eds.), 1985. *The Infrared Handbook.* Environmental Research Institute of Michigan, Ann Arbor, 20-22.

Woolf, N. J., 1982. Seeing and the design and location of a 15 meter telescope. Proc. SPIE 332, 193-197.

Wu, S-T., 1987. Infrared properties of nematic liquid crystals: an overview. Opt. Eng. **26,** 120-128.

Wyant, J. C, 1973. Double frequency grating lateral shear interferometer. Appl. Optics **12,** 2057-2060.

Wyant, J. C, 1974. White light extended source shearing interferometer. Appl. Optics 13, 200-202.

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

Wyngaard, J. C, Y. Izumi, and S. A. Collins Jr, 1971. J. Opt. Soc. Am. **61,** 1646.

Wyrowski, F., 1992. Modulation schemes of phase gratings. Opt. Eng. 31, 251-257.

Yaglom, A. M., 1949. Doklady Akad. Nauk SSSR **69,** 743.

Yellin, M., 1976. Using membrane mirrors in adaptive optics. Proc. SPIE 75, 97-102.

Young, D. M., 1972. *Iterative Solutions of Large Linear Systems.* Academic Press, New York.

Yura, H. T., 1973. Short-term average optical beam spread in a turbulent medium. J. Opt. Soc. Am. 63, 567-572.

Yura, H. T., 1976. An elementary derivation of phase fluctuations of an optical wave in the atmosphere. Proc. SPIE 75, 9-15.

Zollars, B. G., 1992. Atmospheric-turbulence compensation experiments using synthetic beacons. The Lincoln Laboratory Journal 5, 67-92.

Index