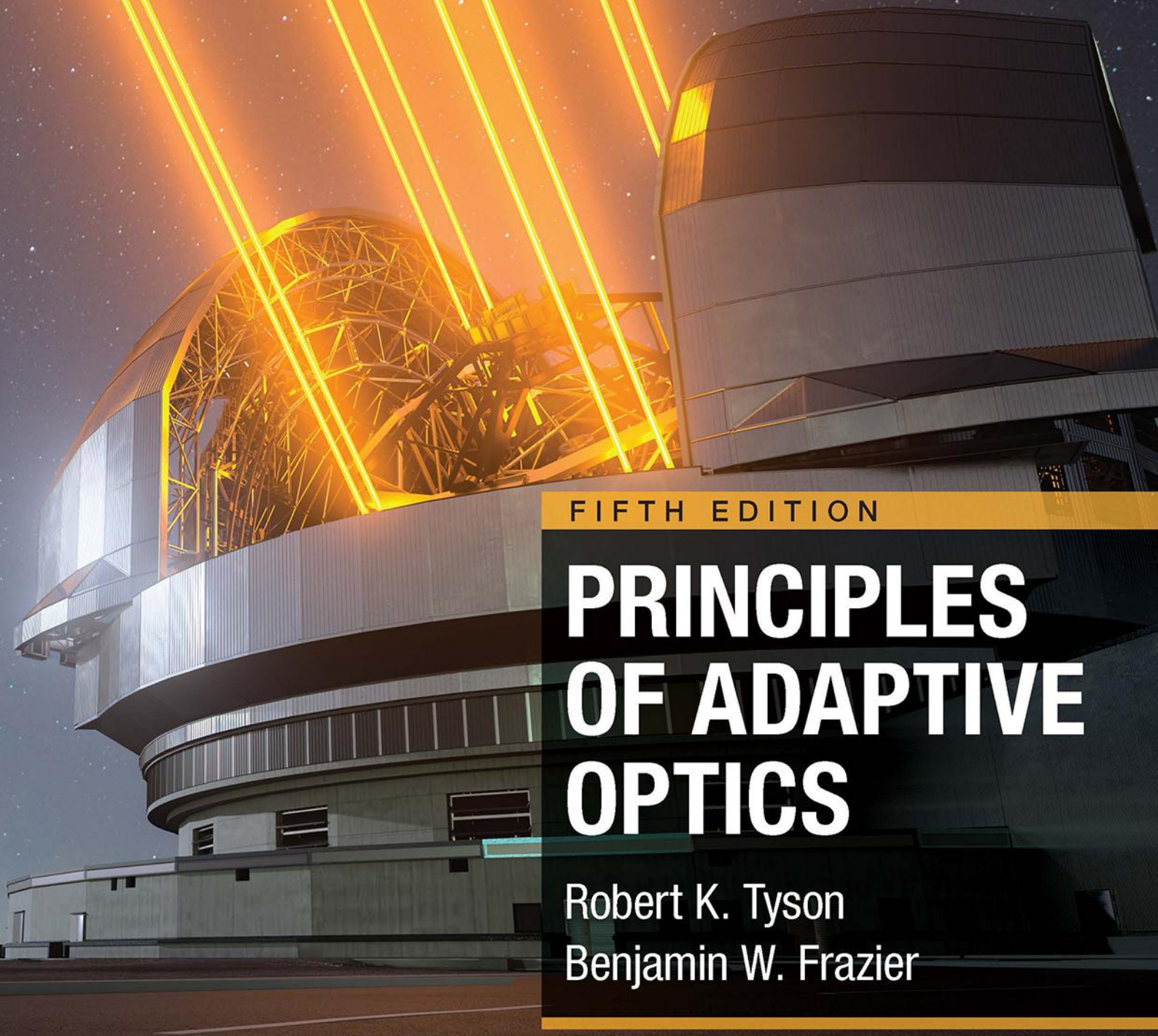

FIFTH EDITION

PRINCIPLES OF ADAPTIVE OPTICS

Robert K. Tyson
Benjamin W. Frazier

CRC Press
Taylor & Francis Group

Principles of Adaptive Optics

Principles of Adaptive Optics

Fifth Edition

Robert K. Tyson
Benjamin W. Frazier

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Cover: Artist's Impression of the ELT: A Sharp Eye on the Sky

This image shows an artist's impression, based on the final design for the telescope, of ESO's Extremely Large Telescope (ELT), which will be the biggest 'eye on the sky' when it achieves first light later this decade. The telescope uses lasers to create artificial guide stars to measure how much the light is distorted by turbulence in the Earth's atmosphere. The deformable M4 mirror adjusts its shape in real time to compensate for these changes in the atmosphere, helping the ELT produce images 16 times sharper than the Hubble Space Telescope.

Credit: ESO (https://elt.eso.org/public/images/ELT4k-12-Night4-cc/)

Fifth edition published 2022
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

Third edition published by CRC Press 2010
Fourth edition published by CRC Press 2016

Library of Congress Cataloging in Publication Data
A catalog record has been requested for this book

ISBN: 978-0-367-67603-2 (hbk)
ISBN: 978-0-367-69048-9 (pbk)
ISBN: 978-1-003-14019-1 (ebk)

DOI: 10.1201/9781003140191

Typeset in Palatino
by SPi Technologies India Pvt Ltd (Straive)

Contents

Foreword

The technology of adaptive optics has advanced enormously from the days, half a century ago, when a computer that filled a small room drove a deformable mirror weighing many kilograms through hundreds of wires protruding from the back. Now, miniaturized actuators with large stroke, controlled over a simple USB connection from a computer smaller than a shoe box, have enabled the production of deformable mirrors that exceed the performance of their earlier counterparts by every metric of the canonical SWaP-C acronym: Size, Weight, and Power draw, and (not at all insignificantly) Cost. Novel methods of converting optical phase aberration into measurable irradiance changes, accompanied by new computational techniques for inverting the optically implemented functions, have advanced the scientific reach of adaptive optics into territory unimagined when the first adaptive optics systems were constructed. It has become almost *de rigueur* that any new large astronomical telescope, if it is to be built on the ground, include adaptive optics in its design from the outset. And even some of those now planned for space will include adaptive optics to exquisite precision as they collect data that characterize potentially habitable exoplanets.

Not a single exoplanet had yet been confirmed when *Principles of Adaptive Optics* first appeared in 1991. And yet, despite the enormous scientific and technological progress since then, in many directions, the physical and engineering principles of real-time wavefront control remain unchanged. This book, covering those principles, remains an essential and comprehensive reference that also offers concrete real-world guidance to every practitioner of the art from an author who has been there, done that. It is the text I have recommended for years to the graduate students who take my course on *Adaptive Optics and Wave Propagation through Random Media*. In this new edition, Ben Frazier joins Bob Tyson as a co-author in a broad update undertaken to reflect the latest advances in adaptive optics technology and methods. It is rare indeed, particularly in such a dynamic field, that a textbook demonstrates the staying power of *Principles of Adaptive Optics*, now in its fifth edition thirty years after its publication.

Michael Hart
The University of Arizona

Preface to the Fifth Edition

When the first edition of *Principles of Adaptive Optics* was written more than thirty years ago, trips to the library were still the best way to research a technical topic. Looking for keywords such as *adaptive optics*, *wavefront*, or *deformable mirrors* in a limited set of international journals from 1962 to 1978 produced about 60 articles. Today, a larger, but still limited, set of journals has over 15,000 articles since 1962 with over 3,300 published in the last five years. Thus, a book on Adaptive Optics has had to, itself, adapt. Covering everything on the topic would produce an Encyclopedia, rather than a book of Principles. To that end, we have looked closely at the specific topics that were not covered fully in previous editions. A topic that needed improvement is wavefront reconstruction and controls, which is a very active area of research and technological advancement.

By the term *we* used in the previous paragraph, I, Bob Tyson, am pleased to add a co-author to the *Principles* series. Dr. Ben Frazier, a former colleague at The University of North Carolina at Charlotte, has added his expertise to overcome my shortcomings and to produce what we humbly call, "The best *Principles of Adaptive Optics* yet."

As part of the development of this book, we have included a series of examples and end-of-chapter design exercises running throughout that reinforce the concepts from each chapter and culminate in the complete design of an adaptive optics system. To perform the simulations required by the design, we tested many simulation frameworks before settling on HCIPy, an open-source, object-oriented framework written in Python for performing end-to-end simulations of high-contrast imaging instruments (https://docs.hcipy.org) (Por et al., 2018). "High Contrast Imaging for Python (HCIPy): an open-source adaptive optics and coronagraph simulator," *Proc. SPIE*, Vol. 10703, 1070342). HCIPy was chosen because (1) it is written in Python and does not require expensive third-party tools or libraries; (2) it is open source; (3) it contains most of the desired elements; and, most importantly, (4) it comes with a number of reasonably well-documented tutorials that allow development of models without requiring a user to become an expert in the code itself. While custom code was developed in MATLAB® and Python for many of the examples throughout the book, the design exercises and closed-loop adaptive optics simulations all leverage the HCIPy framework.

We sincerely thank the thousands of researchers and authors who have contributed to the material in this book. We note also the few authors who gave us fits because they have changed their names over the course of their publishing life. Whether the purpose was social, professional, political, tribal, or avoiding the authorities, we wish you success beyond simply being *et al.*

Robert K. Tyson
The Villages, Florida

Benjamin W. Frazier
Ellicott City, Maryland

List of Abbreviations

AC	Alternating Current
ADC	Analog-to-Digital Converter
ANN	Artificial Neural Network
AO	Adaptive Optics
AOSLO	Adaptive Optics Scanning Laser Opthalmoscope
AR	Anti-reflection
BER	Bit Error Rate
BIBO	Bounded Input Bounded Output
BLIP	Background Limited In Performance
BLPG	Buried Long-Period Grating
BLUE	Best Linear Unbiased Estimator
BWM	Beam Walk Mirror
CCD	Charge Coupled Device
CFS	Complex Field Sensor
CG	Conjugate Gradient
CGI	Coronagraphic Instrument
CGNR	Conjugate Gradient on the Normal Equation to minimize the Residual
CMOS	Complementary Metal–Oxide–Semiconductor
COAT	Coherent Optical Adaptive Technique
COIL	Chemical Oxygen Iodine Laser
CoG	Center of Gravity
COTS	Commercial Off-The-Shelf
CPCM	Common Path/Common Mode
CPU	Computational Processing Unit
CT	Continuous Time
DAC	Digital-to-Analog Converter
DC	Direct Current
DFWM	Degenerate Four-Wave Mixing
DH	Digital Holography
DM	Deformable Mirror
DMD	Digital Mirror Device
DOF	Degrees of Freedom
dOTF	Differential Optical Transfer Function
DR	Dynamic Range
D-SABRE	Distributed-Spline-Based Aberration Reconstruction
DSP	Digital Signal Processing
DT	Discrete Time
EBCCD	Electron Bombarded Charge Coupled Device
EMCCD	Electron Multiplication Charge Coupled Device
EOF	Empirical Orthogonal Functions
ESO	European Southern Observatory
EUV	Extreme Ultraviolet
ExAO	Extreme Adaptive Optics
FD-OCT	Frequency-Domain Optical Coherence Tomography
FEM	Finite Element Method
FFT	Fast Fourier Transform
FPGA	Field Programmable Gate Array
FSM	Fast Steering Mirror
FWM	Four-Wave Mixing
GDT	Gradient Descent Tomography
GLAO	Ground-Layer Adaptive Optics
GM	Gain Margin
GPU	Graphical Processing Unit
GS	Gerchberg–Saxton
HEL	High Energy Laser
HV	Hufnagel–Valley
HWFS	Holographic Wavefront Sensor
IFFT	Inverse Fast Fourier Transform
IMC	Internal Model Control
IR	Infrared
LCOS	Liquid Crystal on Silicon
LC-SLM	Liquid Crystal Spatial Light Modulator
LGS	Laser Guide Star
LOCI	Locally Optimized Combination of Images
LOS	Line-Of-Sight
LQG	Linear Quadratic Gaussian
LQR	Linear Quadratic Regulator
LSB	Least Significant Bit
LSPV	Least-Squares Principal Value
LTAO	Laser Tomography Adaptive Optics
LTI	Linear Time-Invariant
LWIR	Long-Wave Infrared
MAP	Maximum a Posteriori
MCAO	Multi-Conjugate Adaptive Optics
MEMS	Micro-electromechanical System
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean Squared Error
MOAO	Multi-Object Adaptive Optics
MOS	Metal–Oxide–Semiconductor
MPM	Monolithic Piezoelectric Mirror
MSE	Mean Squared Error
MTF	Modulation Transfer Function
MVM	Matrix Vector Multiplication
NEA	Noise Equivalent Angle
NEP	Noise Equivalent Power
NGS	Natural Guide Star
NLO	Non-linear Optics
NSO	National Solar Observatory
OAM	Orbital Angular Momentum
OCT	Optical Coherence Tomography

OOK	On-Off-Keying
OPD	Optical Path Difference
OTF	Optical Transfer Function
PCA	Principal Component Analysis
PCM	Phase Conjugate Mirror
PCO	Post-processing Congruence Operation
P-Cured	Preprocessed Cumulative Reconstructor with Domain reconstruction
PD	Proportional Derivative
PDF	Probability Density Function
PFC	Predictive Fourier Control
PID	Proportional Integral Derivative
PLZT	Lead-Lanthanum-Zirconate-Titanate
PM	Phase Margin
PMN	Lead-Magnesium-Niobate
PMN-PT	Lead-Magnesium-Niobate–Lead-Titanate
PMT	Photomultiplier Tube
PPM	Pulse Position Modulation
PSD	Power Spectral Density
PSF	Point Spread Function
PVDF	Polyvinylidene Fluoride
PWM	Pulse Width Modulation
PZT	Lead-Zirconate-Titanate
QKD	Quantum-Key-Distribution
RF	Radio Frequency
RLS	Recursive Least Squares
RMS	Root-Mean-Square
ROC	Region of Convergence
RSS	Root Sum of Squares
SBS	Stimulated Brillouin Scattering
SCAO	Single Conjugate Adaptive Optics
SGD	Stochastic Gradient Descent
SISO	Single Input Single Output
SLM	Spatial Light Modulator
SLO	Scanning Laser Opthalmoscope
SNR	Signal-to-Noise Ratio
SPGD	Stochastic Parallel Gradient Descent
SQR	Signal-to-Quantization Ratio
SRI	Self-Referencing Interferometer
SRS	Stimulated Raman Scattering
STED	Stimulated Emission Depletion
SVD	Singular Value Decomposition
SWAP	Size, Weight, And Power
TD-OCT	Time-Domain Optical Coherence Tomography
TEFD-OCT	Time-Encoded Frequency Domain Optical Coherence Tomography
TSVD	Truncated Singular Value Decomposition
UAV	Unmanned Aerial Vehicle
WFS	Wavefront Sensor
ZOH	Zero-Order Hold

1

History and Background

1.1 Introduction

Adaptive optics is a scientific and engineering discipline whereby the performance of an optical signal is improved by using information about the environment in which it passes. The optical signal can be a laser beam or the light that eventually forms an image. The principles that are used to extract that information and apply the correction in a controlled manner make up the content of this book.

Various deviations from this simple definition make innovation a desired and necessary quality of the field. Extracting the optical information from beams of light coming from galaxies light-years away is a challenge (Babcock 1992, Merkle and Beckers 1989). Applying a correction to a beam that has the power to melt most man-made objects pushes the state of technology. Adaptive optics is growing; it is changing. It has become a necessary technology in optical systems constrained by dynamic aberrations.

Developments in adaptive optics have been evolutionary. There is no *inventor* of adaptive optics. Hundreds of researchers and technologists have contributed to the development of adaptive optics, primarily over the past forty years. There have been great strides and many, many small steps. The desire to propagate an undistorted beam of light or to receive an undistorted image (Beckers 1993), has made the field of adaptive optics a scientific and engineering discipline in its own right.

Adaptive optics has clearly evolved from the understanding of wave propagation. The basic knowledge of physical optics is aided by the developments of new materials, electronics, and methods for controlling light. Adaptive optics encompasses many engineering disciplines. This book will focus on the *principles* that are employed by the technology community that requires the use of adaptive optics methods. That community is large. The field could encompass many of the topics normally found in optics, electronic controls, and material science texts. This book is intended to condense the vast array of literature and provide a means to use the principles that have been developed over the years.

If one considers that everything having to do with actively controlling a beam of light is adaptive optics, then the field is, indeed, very broad. However, the most commonly used restriction to that definition leads to the approach that adaptive optics deals with *the control of light in a real-time, closed-loop fashion*. Therefore, *adaptive optics* is a subset of the much broader discipline, *active optics*. The literature often confuses and interchanges the usage of these terms,[1] and instances in this book show that discussion of *open loop* approaches to some problems often should be considered (Roggemann and Welsh 1996).

Other restrictions on the definition of *adaptive optics* have been seen. For the purposes of this book, adaptive optics includes more than *coherent, phase-only* correction. There are a number of techniques that make use of intensity correction for the control of light, and other techniques that make multiple corrections over various pupil conjugates. Incoherent imaging is definitely within the realm of adaptive optics.

The existence of the animal visual system is an example of adaptive optics being in use for much longer than recorded history. The eye is capable of adapting to various conditions to improve its imaging quality. The active focus "system" of the eye–brain combination is a perfect example of adaptive optics. The brain interprets an image, determines the correction, either voluntary or involuntary, and applies that correction through biomechanical movement of parts of the eye such as the lens or cornea. When the lens is compressed, the focus is corrected. The eye–brain can also track an object's direction. This is a form of tilt-mode adaptive optics system. The iris can open or close in response to light levels, which demonstrates adaptive optics in an intensity control mode; and the muscles around the eye can force a "squint" that, as an aperture stop, is an effective spatial filter and phase controlling mechanism. This is both *closed-loop* and *phase-only* correction.

The easiest answer to the layman's question, "What is adaptive optics?" is straightforward, albeit not completely accurate: "It's a method of automatically keeping the light focused when it gets out of focus." Every sighted person understands when something is not in focus. The image isn't clear; it's not sharp; it's fuzzy. If we observe light that is not in focus, we can either move to where the light is in focus or not move and apply a correction to bring the beam into focus. This is the principle that our eyes go through constantly. The adaptive process of focus sensing is a learned process. The correction (called *accommodation*) is a learned response. When the correction reaches its biophysical limit, we require outside help, i.e., corrective lenses. The constant adjustment

DOI: 10.1201/9781003140191-1

of our eyes is a closed-loop adaptive process. It is performed with optics. It is, therefore, adaptive optics.

1.2 History

A number of review authors (Hardy 1978, Pearson 1980) cite Archimedes' destruction of the Roman fleet in 215 BC as an early use of adaptive optics. As the attacking Roman fleet approached the army defending Syracuse, soldiers lined up so that they could focus sunlight on the sides of the ships. By polishing their shields or some other "burning glass" and properly positioning each one, hundreds of beams were directed toward a small area on the side of a ship. The resultant intensity was apparently enough to ignite the ship and defeat the attackers. The "burning glass" approach by Archimedes was clearly innovative; however, details of the scientific method are scarce (Claus 1973, Stavroudis 1973). Whether a feedback loop or phase control was used was never reported. The survival of Syracuse was.

The use of adaptive optics has been limited by the technology available. Even the greatest minds of physical science failed to see its utility. Isaac Newton, writing in *Opticks* in 1730, saw no solution to the problem of the atmospheric turbulence limitations of astronomy (Newton 1979):

> If the Theory of making Telescopes could at length be fully brought into Practice, yet there would be certain Bounds beyond which Telescopes could not perform. For the Air through which we look upon the Stars, is in perpetual Tremor; 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.

In 1953 Horace Babcock proposed using a deformable optical element, driven by a wavefront sensor, to compensate for atmospheric distortions that affected telescope images (Babcock 1953). This appears to be the earliest reference to use of adaptive optics as we define the field today. Linnik described how a beacon placed in the atmosphere could be used to probe the disturbances (Linnik 1957). Although Linnik's paper is the first reference to what we now refer to as "guide stars," his concept precedes the invention of the laser and was, until the early 1990s, unknown to Western scientists developing laser guide stars.

Development of the engineering disciplines involved in adaptive optics has taken a course common to technology development. As problems arose, solutions were sought. Directed research within the field was often aided by innovations and inventions on the periphery. The need to measure the extent of the optical problem and to control the outcome was often dependent upon the electronics or computer capability of the day, the proverbial *state-of-the-art*.

Other methods that do not employ real-time wavefront compensation, such as speckle interferometry (Labeyrie 1970, Roddier and Roddier 1986) or hybrid techniques combining adaptive optics and image post-processing, have been successful (Roggemann and Welsh 1996).

An important review article gives an excellent account of the history of active and adaptive optics, describing the state-of-the-art as it was in 1978 (Hardy 1978). Developments of the first three decades are described in detail through this book and reviewed in (Babcock 1990, Hardy 1991, Greenwood and Primmerman 1993, Benedict et al. 1994, Tyson and Schulte in den Bäumen 1990). By the 1980s, over 1 billion U.S. dollars had been spent on adaptive optics systems by the U.S. defense industry. In 1991, much of the U.S. military work in adaptive optics was declassified (Fugate et al. 1991, Primmerman et al. 1991). The research involving laser guide stars (Feinleib 1982) was published (Lukin 1980, 1981; Emaleev and Lukin 1982) and discussed to enhance the research work developing in the astronomical community (Foy and Labeyrie 1985).

Using the broad definition: "Adaptive optics is a system technology that achieves the desired goal of adding or removing aberrations to or from light wavefronts to improve image clarity and resolution for delivery to a vision system, biological or machine," scientific journals and technical societies continually publish new techniques and results. Besides technology development and field demonstrations, the first results for an infrared adaptive optics system on an astronomical telescope were presented in 1991 (Merkle and Hubin 1991). Results from other systems around the world continue to be presented. One example of the exciting improvements in astronomical imaging is shown in Figure 1.1.

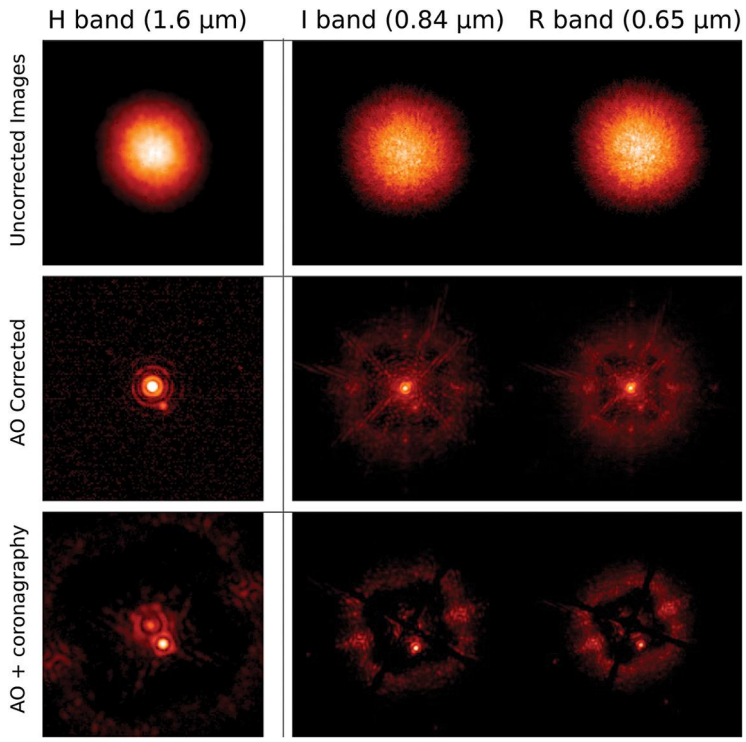

FIGURE 1.1
High resolution images of a low-mass companion star shown without adaptive optics, with adaptive optics, and with AO and coronagraphy in three bands. The companion star is about 4000 times fainter than the parent star 0.24 arcsec apart. (Images were taken with the SPHERE-SAXO system (Spectro-Polarimetric High-contrast Exoplanet Research – Sphere AO for eXoplanet Observation) on the 8 meter Very Large Telescope (VLT) in Chile. Photos courtesy of ONERA, Chatillon, France, and ESO, Garching bei München, Germany.)

While large telescopes still require complex one-off expensive designs to image extrasolar planets and study black hole physics, many other applications have come into the forefront. The 2020 Nobel Prize in Physics was awarded to theorist Roger Penrose and astronomers Reinhard Genzel and Andrea Ghez for their discovery of a black hole at the center of the Milky Way galaxy.

> The accomplishments of the observational team led by Genzel and Ghez were enabled by infrared telescopes– and, in particular, by the emergence of adaptive optics. The latter technique dramatically sped up the process of making the intricate observations of stellar orbits needed to infer the presence of the supermassive black hole at the galactic center.
>
> (Wills 2020)

Well into the 21st century, adaptive optics technology has branched from high-energy lasers and large astronomical telescopes into the commercial sector and the biomedical sector. With that, the higher volume and simplifications has brought down the cost. In addition to retinal imaging and other biomedical applications (discussed in Chapter 4), it has recently been reported that one brand of DVD player has an adaptive optics element in it to correct for distortions from the disc. With over a million systems per month in production, the cost of each adaptive optics "system" is less than $1 U.S. dollar (BBC 2011). The history of adaptive optics continues to be written (Palca 2013) with the future incredibly bright (pun intended).

1.3 Physical Optics

The principles of adaptive optics are based on the premise that one can change the effects of an optical system by adding, removing, or altering optical elements. For most optical systems of interest, diffraction effects are harmful. That is, they make the propagation of a beam of light or the image of some object *different*. When a beam is propagated, either collimated or focused, we normally want *all* of the light to reach the receiver in the best condition. Similarly, for an imaging system, we want the image to be the best reproduction of the object that we can get. Diffractive effects degrade the image as they degrade the propagation process. We cannot get rid of diffractive effects; they are inherent in Maxwell's laws. The best that we can do is reach the limit of diffraction.

When mechanical or thermal defects degrade the image or propagation process, we can try to alter the optical system to compensate for the defects even though we cannot get rid of them.

1.3.1 Propagation with Aberrations

One way to describe the propagation of light as it applies to adaptive optics is through mathematical formalism. Literally hundreds of volumes have been written describing optical propagation. The classic text (Born and Wolf 1975) describes it in a manner that clearly shows the effects of the phase of an optical field in a pupil plane on the resultant field in an "image" plane.

For a beam of coherent light at wavelength λ, the intensity of light at a point P on the image, or focal, plane a distance z away is given by

$$I(P)=\left(\frac{Aa^2}{\lambda R^2}\right)^2\left|\int_0^1\int_0^{2\pi} e^{i\left[k\Phi-\upsilon\rho\cos(\theta-\psi)-\frac{1}{2}u\rho^2\right]}\rho d\rho d\theta\right|^2 \tag{1.1}$$

The circular pupil has radius a and coordinates ρ, θ; the image plane has polar coordinates r, ψ; the coordinate z is normal to the pupil plane; R is the slant range from the center of the pupil[2] to point P; $k = 2\pi/\lambda$; and $k\Phi$ is the deviation in phase from a perfect sphere about the origin of the focal plane. See Figure 1.2. To simplify the presentation, normalized coordinates in the focal plane are used:

$$u=\frac{2\pi}{\lambda}\left(\frac{a}{R}\right)^2 z \tag{1.2}$$

$$\upsilon=\frac{2\pi}{\lambda}\left(\frac{a}{R}\right) r. \tag{1.3}$$

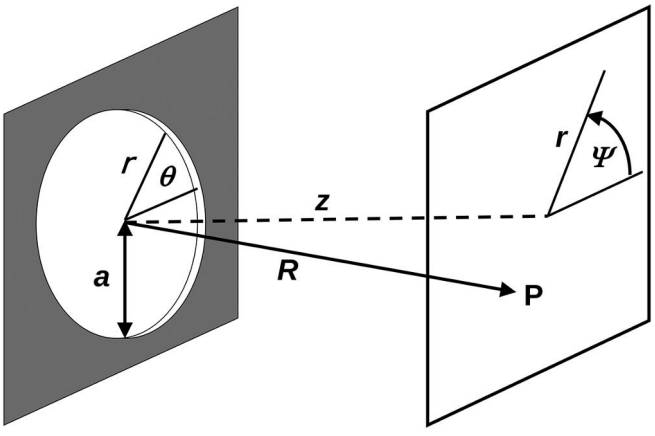

FIGURE 1.2
The coordinate system for the diffraction calculation with aberration Φ.

The uniform electric field amplitude in the pupil plane is A/R, and, therefore, the intensity in the pupil plane, $I_{z=0}$, is A^2/R^2.

From the standpoint of adaptive optics, the most important quantity in the above expressions is Φ, which is often, mistakenly, called simply *phase*. It is more commonly referred to as the *wavefront*. Φ represents all the aberrations that are present in the optical system prior to its propagation to point P.

If no aberrations are present, the intensity is a maximum on-axis ($r = 0$), which is called the *Gaussian image point*.

$$I_{\Phi=0}\left(P_{r=0}\right)=\pi^2\left(\frac{Aa^2}{\lambda R^2}\right)^2=\left(\frac{\pi^2a^4}{\lambda^2R^2}\right)I_{z=0}. \tag{1.4}$$

The *Strehl ratio*, S, also called the normalized intensity, is the ratio between the intensity on-axis of an aberrated beam and the intensity on-axis of an unaberrated beam. If the tilt (distortion) aberration is not removed, the *axis* of this definition would be normal to the plane of that tilt, rather than parallel with the z-axis. Static tilt should be removed when the Strehl ratio is used as a figure of merit for the quality of beam propagation.

Combining the preceding expressions, the Strehl ratio becomes

$$S=\frac{I(P)}{I_{\Phi=0}}=\frac{1}{\pi^2}\left|\int_0^1\int_0^{2\pi} e^{i\left[k\Phi-vp\cos(\theta-\psi)-\frac{1}{2}u\rho^2\right]}\rho d\rho d\theta\right|^2. \tag{1.5}$$

For small aberrations, when tilt[3] is removed and the focal plane is displaced to its Gaussian focus, the linear and quadratic terms in the exponential of Eq. (1.5) disappear. If the remaining aberrations, which are now centered about a sphere with respect to point P, are represented by Φ_P, the Strehl ratio simplifies to

$$S=\frac{1}{\pi^2}\left|\int_0^1\int_0^{2\pi} e^{jk\Phi_P}\rho d\rho d\theta\right|^2, \tag{1.6}$$

which shows how the wavefront affects the degradation of the propagation. If the beam at the pupil is unaberrated, $\Phi_P = 0$, the Strehl ratio reduces to unity, $S = 1$; that is, the intensity at focus is diffraction-limited. Equation (1.4) shows that the absolute intensity is enhanced by the factor proportional to the square of the *Fresnel number*, $a^2/\lambda R$. A larger aperture, a smaller wavelength, or a shorter propagation distance will increase the maximum intensity in the focal plane. All systems with *any* aberration at all, i.e., $\Phi_P > 0$, will have a Strehl ratio less than 1.

If the wavefront aberration is small, its variance can be directly related to the Strehl ratio. The wavefront variance $(\Delta\Phi_P)^2$ can be found from

$$\left(\Delta\Phi_P\right)^2 = \frac{\int_0^1\int_0^{2\pi}\left(\Phi_P - \bar{\Phi}_P\right)^2 \rho d\rho d\theta}{\int_0^1\int_0^{2\pi}\rho d\rho d\theta}, \tag{1.7}$$

where $\bar{\Phi}_p$ is the average wavefront. It can be shown (Born and Wolf 1975) that the Strehl ratio for $\Delta\Phi_P < \lambda/2\pi$ is just

$$S = 1 - \left(\frac{2\pi}{\lambda}\right)^2 \left(\Delta\Phi_P\right)^2 \cong \exp\left[-\left(\frac{2\pi}{\lambda}\right)^2 \left(\Delta\Phi_P\right)^2\right], \tag{1.8}$$

which gives a simple method of evaluating the propagation quality of a system by considering only the small variance of the wavefront. The square root of the variance, formally, the standard deviation of the wavefront, $\Delta\Phi_P$, is often called the *root-mean-square phase error*, *rms phase error, phase error*, or *wavefront error*. These terms are used interchangeably in the literature. The fact that quality of beam propagation can be directly related to the *rms phase error* is a very powerful result.

1.3.2 Imaging with Aberrations

The physical processes behind imaging combine the process of propagation with the effects of lenses, mirrors, and other imaging optics. The process that combines these effects results from an examination of Kirchoff's formula for propagation, which is a more generalized form of Eq. (1.1). The field at point P, in the x–y plane $U(P)$, is given by (Born and Wolf 1975)

$$U\left(x,y\right) = \frac{-iA}{2\lambda}\iint_{Ap}\frac{e^{ik(r+s)}}{rs}\left[\cos\left(n,r\right) - \cos\left(n,s\right)\right]dS, \tag{1.9}$$

where the coordinates are defined in Figure 1.3 and the integral is over the aperture. A is the amplitude of the field with a plane wave represented by $U(z)_{\text{plane}} = Ae^{\pm ikz}$ and a spherical wave represented by $U(z)_{\text{sph}} = A/re^{\pm ik}$. The source is in the x_0–y_0 plane with a distance z_0 to the x'–y' pupil plane. The distance z represents the distance from the pupil to the image plane.

For propagation distances that are large, i.e., $z >$ the largest of x, y, x', y', x_0, or y_0, Eq. (1.9) reduces to the Fresnel integral,

$$U\left(x,y\right) = C'\iint_{Ap} e^{if\left(x',y'\right)}dx'dy' \tag{1.10}$$

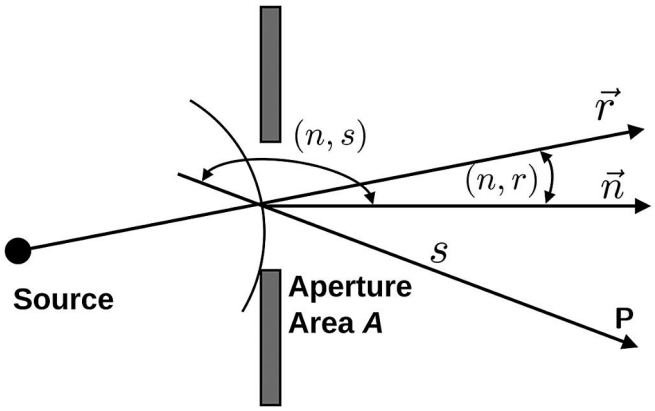

FIGURE 1.3
The geometry of the Kirchoff formula.

where the term C' is not a function of the pupil coordinates,

$$C' \equiv \frac{-A}{2\lambda}\left[\cos\left(n,r\right) - \cos\left(n,s\right)\right]\frac{e^{ik(z_0+z)}}{zz_0} \exp\left[\frac{ik}{2z_0}\left(x_0^2 + y_0^2\right)\right] \times \exp\left[\frac{ik}{2z}\left(x^2 + y^2\right)\right], \tag{1.11}$$

and the function in the exponent is

$$f\left(x',y'\right) = \frac{1}{2}\left(x'^2 + y'^2\right)\left(z_0^{-1} + z^{-1}\right) - x'\left(\frac{x_0}{z_0} + \frac{x}{z}\right) - y'\left(\frac{y_0}{z_0} + \frac{y}{z}\right). \tag{1.12}$$

In many beam propagation or astronomical imaging applications, the propagation distances are quite large. When this distance satisfies the conditions,

$$z_0 \gg \frac{\left(x'^2 + y'^2\right)\max}{\lambda} \text{ and } z > \frac{\left(x'^2 + y'^2\right)\max}{\lambda} \tag{1.13}$$

the Kirchoff integral reduces to the Fraunhofer integral,

$$U\left(x,y\right) = C'\iint_{Ap}\exp\left\{-ik\left[\left(\frac{x_0}{z_0} + \frac{x}{z}\right)x' + \left(\frac{y_0}{z_0} + \frac{y}{z}\right)y'\right]\right\}dx'dy' \tag{1.14}$$

When combining terms in C' and terms in the kernel that relate to the source in the x_0–y_0 plane, the integral takes the form of a Fourier transform,

$$U\left(x,y\right) = C'\iint_{-\infty} U\left(x',y'\right)_{Ap}\exp\left\{\frac{ik}{z}\left(xx' + yy'\right)\right\}dx'dy', \tag{1.15}$$

which can be written in a shortened notation,

$$U\left(x,y\right) \propto F\left[U\left(x',y'\right)\right]. \tag{1.16}$$

In the preceding equation, F represents the Fourier transform operation and $U(x', y')$ is the *pupil function*, or *aperture function*. When the optical system contains aberrations Φ, the pupil function can be written as

$$U(x',y') = A(x',y')e^{ik\Phi(x',y',x_0,y_0)} \tag{1.17}$$

which uses the fact that aberrations are derived from all elements of the optical system, including the physical pupil in the (x', y') plane.

When no aberrations are present and the aperture is a simple geometrical shape, Eq. (1.15) can be solved exactly. For a rectangular aperture of sides $2a$ and $2b$, after propagation distance L, the normalized field that results from Fraunhofer diffraction is

$$U(x,y) = \frac{4L^2 \sin\left(\frac{kax}{L}\right)\sin\left(\frac{kby}{L}\right)}{k^2xy}. \tag{1.18}$$

An important result is the Fraunhofer diffraction from a circular aperture. In polar coordinates, the normalized field becomes

$$U(r,\theta) = \frac{\lambda L}{\pi ar} J_1\left(\frac{2\pi ar}{\lambda L}\right), \tag{1.19}$$

where J_1 is the first-order Bessel function of the first kind. The intensity of light from the diffracted field is the square of the magnitude of these fields, that is, $I = |U|^2$.

The pupil function of an imaging system is modified by the effect of lenses and other focusing optics. A lens transmission function is given in one form in (Gaskill 1978).

$$T_{\text{lens}} = \exp\left[\frac{-i\pi}{\lambda f}\left(x^2 + y^2\right)\right], \tag{1.20}$$

where f is the focal length of the lens. The process of adding lens transmission functions and intermediate propagations to the integral equations reduces Eq. (1.15) to

$$U(x,y) = C'\iint U(x',y')\exp\left[-ik\left(\frac{1}{z} - \frac{1}{f}\right)(xx' + yy')\right]dx'dy' \tag{1.21}$$

which shows the retention of the Fourier transform properties of the imaging system with f as its effective focal length. By introducing a coordinate change in the image plane,

$$\xi = \frac{kx}{f} \tag{1.22}$$

and

$$\eta = \frac{ky}{f}, \tag{1.23}$$

the Fraunhofer diffraction becomes

$$U(\xi,\eta) = C'\iint U(x',y')e^{-i(\xi x' + \eta y')}dx'dy', \tag{1.24}$$

which says that the Fraunhofer diffraction pattern, at focal distance f, is the Fourier transform of the pupil function.

For pupils with and without aberrations, this diffraction pattern is called the *point spread function* (PSF). The new coordinates ξ, η have the units of inverse distance and are called *spatial frequencies*. This is analogous to the spectra calculated in the time domain where frequency is inverse time. In imaging applications, spatial frequencies are usually two-dimensional (2D).

Some important physical results can be derived from Eq. (1.24). If the pupil function is spatially limited, such as that with a finite aperture, the Fourier transform operation cuts out the high spatial frequency of objects. The high frequencies contain the fine detail of the object. Therefore, an imaging system that can pass these high frequencies, such as one with a large aperture, is "better" insofar as it can image with high resolution. Also, an aberration-free pupil will not produce an image where the high frequency characteristics of the aberrations screen the high frequency content of the object. This is the principal need for adaptive optics in imaging systems. The fact that image forming quality can be specified by a function that describes how much spatial frequency content is passed to the image is an important result of Eq. (1.24).

The field of the image for a coherent imaging system[4] can be expressed as the convolution of the geometrical image with the point spread function (Gaskill 1978),

$$U_{\text{im}} = W\left[\left(U_{\text{geom im}}\right) ** \text{PSF}\right]. \tag{1.25}$$

The geometrical image is the distribution of the object that has been adjusted for imaging system magnification, shifting along the plane, and geometrical (ray trace) shadows and differences of intensity. In many discussions of imaging in the literature (Goodman 1968), the geometrical image and the object are used interchangeably. This is not a problem until exact calculations are performed, since the principles of imaging are the same. The constant W takes into account the radiance of the object and the constant transmission or absorption of elements within the system. The 2D convolution process represented by ** accounts for the diffraction effects

of the optics. The convolution operation can be Fourier-transformed on both sides (Goodman 1968),

$$F\left(U_{\text{im}}\right) = W \cdot \left[F\left(U_{\text{geom im}}\right) F\left(\text{PSF}\right) \right] \tag{1.26}$$

The Fourier transform of the PSF is, itself, a measure of the imaging quality of a system. The *optical transfer function* (OTF) represents how each spatial frequency in an object field is *transferred* to the image (Hopkins 1955, 1957),

$$\text{OTF} = F\left(\text{PSF}\right) \tag{1.27}$$

One definition of the *resolution* of the system is the spatial frequency at which the OTF vanishes. In many circumstances, this is more descriptive of the optical system than the simple Rayleigh criterion of resolution (Born and Wolf 1975).[5]

Since one observes the intensity distribution of images rather than the electric fields, the magnitudes of the fields are important. The objects and images are formed from Fourier components of the intensity. Considering one frequency component, ξ for instance, the intensity of the object I_{obj} can be represented by a constant and a frequency component,

$$I_{\text{obj}} = b_0 + b_1 \cos 2\pi\xi x. \tag{1.28}$$

Similarly, the intensity of the image can be written as

$$I_{\text{im}} = c_0 + c_1 \cos 2\pi\xi x. \tag{1.29}$$

The modulation, or contrast, M is the relationship between the peaks and valleys of the intensity at the designated frequency,

$$M = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \tag{1.30}$$

When substituting the expressions for the individual intensities, the modulation M becomes normalized to the constant level of intensity. The results are

$$M_{\text{obj}}\left(\xi\right) = \frac{b_1\left(\xi\right)}{b_0\left(\xi\right)} \quad \text{and} \quad M_{\text{im}}\left(\xi\right) = \frac{c_1\left(\xi\right)}{c_0\left(\xi\right)}. \tag{1.31}$$

It can be shown (Gaskill 1978) that the ratio of these modulations is the magnitude of the OTF,

$$\left|\text{OTF}\left(\xi\right)\right| = \frac{M_{\text{im}}}{M_{\text{obj}}}. \tag{1.32}$$

The magnitude of the OTF is the ratio of how much modulation in the object was transferred to modulation in the image. This function, rightly called the *modulation transfer function* (MTF), is a very useful measure of the ability of an imaging system to transfer spatial detail. The MTF contains no phase information, since it is only the magnitude of the OTF. A perfect system has an MTF = 1 at all frequencies, which requires an infinitely large aperture. The MTF of realistic systems is less than 1.0 for all spatial frequencies greater than zero.

1.3.3 Representing the Wavefront

A number of mathematical constructs are used to describe the phase of a beam. The deviation of phase from a reference sphere (Born and Wolf 1975) is the wavefront. In adaptive optics, the wavefront is usually the quantity that we want to change to alter the propagation characteristics of the beam.

The *wavefront* is a 2D map of the phase at an aperture or any other plane of concern that is normal to the line of sight between the origin of the beam and the target. In an imaging system, the plane of concern would be normal to the line of sight between the object and the image. By this definition, tilt of the beam would be considered part of the wavefront. In many instances, the tilt and the piston[6] are of interest for adaptive optics and are included. The wavefront is positive in the direction of propagation.

1.3.3.1 Power Series Representation

One way to represent the 2D wavefront map is by a power series in polar (ρ, θ) coordinates,

$$\Phi\left(\rho,\theta\right) = \sum_{n,m=0}^{\infty} S_{n,m1}\, \rho^n \cos^m\theta + S_{n,m2}\, \rho^n \sin^m\theta. \tag{1.33}$$

In this series, the primary, or Seidel, aberrations (Born and Wolf 1975) are explicit. A coordinate transformation to Cartesian coordinates is easily made ($r = \rho \cos\theta$, $y = \rho \sin\theta$). Table 1.1 shows the first few Seidel terms.

1.3.3.2 Zernike Series

The commonality of circular apertures, telescopes, and lenses makes the treatment in polar coordinates very attractive. The power series representation is, unfortunately, not an orthonormal set over a circle. One set of polynomials that is orthonormal over a circle, introduced by Zernike (1934), has some very useful properties. The series, called a *Zernike series*, is composed of sums of power series terms with the appropriate normalizing factors. A detailed description of the Zernike series is given in (Born and Wolf 1975), and an analysis of Zernike polynomials and atmospheric turbulence

TABLE 1.1

Representation of Seidel Terms

n	*m*	Representation	Description
0	0	1	piston
1	1	$\rho \cos\theta$	tilt, distortion
2	0	ρ^2	focus, field curvature, "sphere"
2	2	$\rho^2 \cos^2\theta$	astigmatism, cylinder
3	1	$\rho^3 \cos\theta$	coma
4	0	ρ^4	spherical aberration

including their Fourier transforms was done by Noll (1976).[7] Roddier (1990) shows how Zernike polynomials can be used in modeling to describe the phase aberrations of the atmosphere. Winker expands the analysis technique to include the effects of a finite outer scale of turbulence (Winker 1991), while Boreman and Dainty (1996) present a generalization that includes non-Kolmogorov turbulence (Golbraikh and Kopeika 2008). Other non-Kolmogorov models are used that better support local propagation measurements (Zilberman et al. 2008, Lawrence et al. 1970, Belen'kii 1995).

The general Zernike series contains all aberration terms, including piston and tilt:

$$\Phi(\rho,\theta) = A_{00} + \frac{1}{\sqrt{2}}\sum_{n=2}^{\infty} A_{n0} R_n^0\left(\frac{\rho}{R'}\right) + \sum_{n=1}^{\infty}\sum_{m=1}^{n}\left[A_{nm}\cos m\theta + B_{nm}\sin m\theta\right] R_n^m\left(\frac{\rho}{R'}\right) \tag{1.34}$$

for $n - m$ = even. R' is the radius of the circle over which the polynomials are defined. The radial polynomial R_n^m is defined as

$$R_n^{\pm m}\left(\frac{\rho}{R'}\right) = \sum_{s=0}^{\frac{n-m}{2}} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!}\left(\frac{\rho}{R'}\right)^{n-2s}, \tag{1.35}$$

where these radial terms contain mixtures of the Seidel terms.

Although the Zernike series appears rather complicated and unwieldy, it does have a number of features useful to adaptive optics. It transforms easily under rotations about the Cartesian axes, it includes a polynomial for each pair of radial (n) and azimuthal (m) order (Born and Wolf 1975), and the coefficients of the series can be used for aberration balancing. This is helpful for simple systems when defocus, for instance, can be used to correct for some amounts of spherical aberration (Born and Wolf 1975). The amounts of the various modes needed are related to the normalization constant of the radial polynomial (Buchdahl 1960, Nijboer 1947, Woodruff 1975). If all the modes are counted up through radial order n, the relationship between the total number of Zernike modes Z_m and the number of the radial order is found from

$$Z_m = \frac{1}{2}(n+1)(n+2). \tag{1.36}$$

Table 1.2 shows the first few radial Zernike terms.

Another useful property of the Zernike series is the simple manner in which the rms wavefront error can be calculated. If all the coefficients of the series that represent a wavefront are known, the geometric sum of the non-piston terms yields the wavefront variance,

$$(\Delta\Phi)^2 = \sum_{n=1}^{\infty}\sum_{m=0}^{n}\frac{A_{nm}^2 + B_{nm}^2}{2(n+1)}. \tag{1.37}$$

This variance can be used to calculate the Strehl ratio directly. In the special case where the wavefront is symmetric about the meridional plane ($B_{nm} = 0$), the power-series expansion coefficients can be calculated from the Zernike coefficients (Tyson 1982). Similarly, the Zernike coefficients can be calculated from the power-series coefficients (Conforti 1983). Since we can calculate the Zernike coefficients A_{nm} and B_{nm} by simple integrals, the Zernike series provides a manageable representation of the wavefront for computational purposes or determining the effects of primary aberrations on circular beams. If the circular geometry is maintained, Zernike polynomials can be scaled, translated, and rotated over the full aperture with a single matrix calculated from the Fourier properties of Zernike polynomials and Bessel function integrals (Tatulli 2013).

1.3.3.3 Zernike Annular Polynomials

Some optical systems, Cassegrain and Newtonian telescopes, for instance, have a centrally obscured beam. A wavefront for these systems is not easily represented

TABLE 1.2

Radial Zernike Terms

Radial Term	Description
$R_0^0 = 1$	piston
$R_2^0 = 2\left(\frac{\rho}{R'}\right)^2 - 1$	Zernike focus contains piston
$R_4^0 = 6\left(\frac{\rho}{R'}\right)^4 - 6\left(\frac{\rho}{R'}\right)^2 + 1$	Zernike spherical aberration contains focus and piston
$R_3^1 = 3\left(\frac{\rho}{R'}\right)^3 - 2\left(\frac{\rho}{R'}\right)$	Zernike coma contains tilt

using a power series or Zernike polynomials, since the obscuration is represented by very high spatial frequencies in the radial direction. A large number of terms of the series are required to represent even primary aberrations. Mahajan discusses a series that is orthonormal over an annulus (Mahajan 1981). By using the Gram–Schmidt orthogonalization process, a series of polynomials based on the Zernike polynomials is obtained. This series includes the quantity representing the inner radius of the optical wavefront and uses terms similar to the radial Zernike polynomials. Converting Zernike circular polynomials to other geometries results in errors as their coefficients lack the physical significance of balanced aberrations. The errors are reported for annular pupils (as in many reflecting telescopes), hexagonal pupils (as in large segmented telescopes like Keck), elliptical pupils (as in the mammalian eye), rectangular pupils (as in smart phone cameras), and square pupils (as in some high-power lasers) (Dai and Mahajan 2008).

1.3.3.4 Lowest Aberration Modes

Besides piston, which is only a uniform shift in the entire wavefront, tilt and focus are the primary aberrations that mostly affect propagation or an image. According to the rigorous theory, these are not even "aberrations" (Born and Wolf 1975). They have a fundamental geometrical significance. Repeating Eq. (1.1) for the intensity of light at a point P in a focal plane,

$$I(P)=\left(\frac{Aa^2}{\lambda R^2}\right)^2\left|\int_0^1\int_0^{2\pi} e^{i\left[k\Phi-v\rho\cos(\theta-\psi)-\frac{1}{2}u\rho^2\right]}\rho d\rho d\theta\right|^2, \quad (1.38)$$

one can see that the exponent contains three basic terms, the wavefront contribution $k\Phi$, the tilt contribution $v\rho\cos(\theta-\psi)$, and the focus contribution $u\rho^2$. If we determine that the wavefront Φ contains tilt in the x direction of a magnitude K_x, we can perform a coordinate transformation $\Phi=\Phi'+K_x\rho\sin\theta$. With the tilt term separated into its x–y components, the focal length represented by f, and the aperture radius, a, the exponent becomes

$$k\Phi'+kK_x\rho\sin\theta-\frac{ka}{f}r\rho\sin\psi\sin\theta$$
$$-\frac{ka}{f}r\rho\cos\psi\cos\theta-\frac{1}{2}kz\left(\frac{a}{f}\right)^2\rho^2. \quad (1.39)$$

Removing the tilt from the higher-order wavefront term Φ' and transforming according to $x'=x-(R/a)K_x$, $x'=y$, $z'=z$, the exponent becomes

$$k\Phi-\frac{ka}{f}\rho'\cos(\theta-\psi')-\frac{1}{2}kz\left(\frac{a}{f}\right)^2\rho^2, \quad (1.40)$$

which is the same form as Eq. (1.1). The distribution of light in the image plane will be the same as the original untilted wavefront. The centroid of the image will shift, however, by an amount equal to fK_x/a.

A similar transformation can be made for a focal shift. If the wavefront contains terms $K_z\rho^2$, then the distribution of light will not change in the focal plane, but the focal distance f will shift an amount proportional to K_z. This very important result is fundamental to the measurement of focus or tilt in a beam and will be discussed in detail in Chapter 5.

1.3.4 Interference

Interference occurs when two or more coherent light beams are superimposed. White light interference can occur because (incoherent) white light can be thought of as the sum of coherent components that interfere. Basic principles of optical interference can be used for practical applications like measuring wavefronts in adaptive optics (Koliopoulos 1988).

The *intensity* is the time-averaged squared magnitude of the electric field. We can begin by expressing the electric field vector of a plane wave as

$$\mathbf{E}=\frac{1}{2}\left[\mathbf{A}(\mathbf{r})e^{-i\omega t}+\mathbf{A}^*(\mathbf{r})e^{i\omega t}\right], \quad (1.41)$$

where the vector components of the amplitude are

$$A_x=a_xe^{i\mathbf{k}\cdot\mathbf{r}-\delta_x}, \quad (1.42)$$

$$A_y=a_ye^{i\mathbf{k}\cdot\mathbf{r}-\delta_y}, \quad (1.43)$$

$$A_z=a_ze^{i\mathbf{k}\cdot\mathbf{r}-\delta_z}, \quad (1.44)$$

and the phases of the components are the δ's. The magnitude of the field $|E|^2$ takes the form

$$|E|^2=\frac{1}{4}\left(\mathbf{A}^2e^{-2i\omega t}+\mathbf{A}^{*2}e^{2i\omega t}+2\mathbf{A}\cdot\mathbf{A}^*\right). \quad (1.45)$$

Averaging over a large time interval the intensity, the magnitude of the field, becomes

$$I=\left\langle\mathbf{E}^2\right\rangle=\frac{1}{2}\mathbf{A}\cdot\mathbf{A}^*=\frac{1}{2}\left(a_x^2+a_y^2+a_z^2\right). \quad (1.46)$$

If two such fields are superimposed, the vectors add:

$$\mathbf{E}=\mathbf{E}_1+\mathbf{E}_2. \quad (1.47)$$

The magnitude of the sum of the two fields becomes

$$\left\langle\mathbf{E}^2\right\rangle=\left\langle\mathbf{E}_1^2\right\rangle+\left\langle\mathbf{E}_2^2\right\rangle+2\left\langle\mathbf{E}_1\cdot\mathbf{E}_2\right\rangle, \quad (1.48)$$

and the intensity of the two superimposed fields is

$$I = I_1 + I_2 + 2\langle \mathbf{E}\cdot\mathbf{E}\rangle = I_1 + I_2 + \left(a_{x1}a_{x2} + a_{y1}a_{y2} + a_{z1}a_{z2}\right)\cos\delta, \quad (1.49)$$

where δ is the phase difference between the two fields. Without loss of generality, but for simplifying the presentation, the light can be treated as transverse and linearly polarized, i.e., $a_{yi} = a_{zi} = 0$. The intensities of the individual beams are, from Eq. (1.41),

$$I_1 = \frac{1}{2}a_{x1}^2, \quad (1.50)$$

$$I_2 = \frac{1}{2}a_{x2}^2, \quad (1.51)$$

and the intensity of the superimposed beams is

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\delta. \quad (1.52)$$

The maximum intensity occurs when $\cos\delta = 1$, i.e., $\delta = 0, 2\pi, 4\pi, \ldots$, and the minimum occurs when the cosine term is zero, i.e., $\delta = \pi, 3\pi, 5\pi, \ldots$.

For the special, but not so uncommon, case $I_1 = I_2$, the intensity of the superimposed beams is

$$I = 4I_1\cos^2\left(\delta/2\right), \quad (1.53)$$

with the maximum intensity $I_{\max} = 4I_1$ and the minimum intensity $I_{\min} = 0$.

Measuring the interference pattern is equivalent to measuring the spatial coherence function of an optical field. The Van Cittert–Zernike theorem (Roggemann and Welsh 1996) states that the spatial coherence properties of an optical field are a Fourier transform of the irradiance distribution of the source. The fields of optical and radio interferometry are based upon this theorem.

The application of these principles of interference is also fundamental to the development of wavefront control. Adaptive optics is, if nothing more, an engineering field where controlling the phase δ at one place leads to managing the optical intensity I at another place.

EXAMPLE 1.1 FITTING A WAVEFRONT

If we now assume that we know the shape of the wavefront, either in terms of primary aberration polynomials, Zernike polynomials, or any other description, it would be necessary to fit that wavefront to a correction device to close the loop on an adaptive optics system. We discuss the wavefront distortion from atmospheric turbulence and other sources in the next chapter, but let's assume that the wavefront can be accurately measured, as we will show that it can in Chapters 5 and 6. Now, the question arises: How do we fit the wavefront?

We draw from a paraphrased historical discussion remembered by one of the authors (RKT) which took place around 1979 at Pratt & Whitney Aircraft Rocket Support Division in West Palm Beach, Florida, where multikilowatt carbon dioxide lasers were being made and tested. To correct the distorted beam leaving the 2-meter aperture before sending it downrange to a calorimeter on a sled of a railroad car, the deformable mirror needed to be constructed. Keeping in mind that the high energy laser was powerful enough to melt any metallic surface, (that was the point), the mirrors in the optical train needed active cooling. The molybdenum deformable mirror was cooled with water flowing directly through the mirror substrate. Piezoelectric stacks of actuators, each with very high voltage supplying them, were attached to the back of the roughly 25-centimeter clear aperture optical surface. I remember the deformable mirror system being heavy enough that the only way to position it was using a forklift.

The design was mechanical-limited. Actuators strong enough to deform the surface had a minimum size and could be placed about 2 centimeters apart.

The first engineering discussion came from the electromechanical engineers: "How many actuators do you need?". The optical design engineers answered, "As many as you can put on." The circular argument continued from the optical engineers, "For the next generation of actuators, how small can you make them?". "Maybe 1.5 centimeters in diameter," the electrical engineers decided.

The optical engineers then asked, "How big a mirror can you make?". The mechanical engineers responded, "How big do you want?".

Optical engineers now say, "We need enough actuators to fit the wavefront we're trying to correct."

Mechanical, electrical, thermal engineers, "bean counters," and senior program managers ask firmly, "How many is that?".

Optical engineers answer, "We don't know."

1.4 Radiometry

Radiometry is the measurement of radiation, or more specifically for the field of adaptive optics, it is the measurement of *optical radiant energy*. Rather than an extended treatise on radiometry of all forms, this discussion will

apply to light at the optical wavelengths: ultraviolet, visible, and infrared. Radiometry is included in the generation, transmission, and detection of light and involves the measurement of light in space, direction, and time.

A target or a celestial object may be active, that is, emitting radiation by virtue of its temperature or some atomic excitation, or it can be passive by reflecting light from an active or another passive source. Among the target field, there may be other sources of light which we can regard as background or noise. The units of radiometry begin with *energy* expressed in joules (*J*), which can actually be directly measured with detectors that count how many photons hit it. The *power* or *radiant flux* is the rate of increase of energy or energy per unit time, expressed in watts ($W = J/s$). In radiometry, radiance is the radiant flux emitted, reflected, transmitted, or received by a given surface, per unit solid angle per unit projected area. *Spectral radiance* is the radiance of a surface per unit wavelength. The SI unit of radiance is the watt per steradian per square meter ($W{\cdot}sr^{-1}{\cdot}m^{-2}$), while that of spectral radiance in wavelength is the watt per steradian per square meter per nanometer ($W{\cdot}sr^{-1}{\cdot}m^{-2}{\cdot}nm^{-1}$) (McCluney 1994). We suggest that you refer to the many discussions of radiometric terms that follow these basics. It is beyond the scope of a small section of an adaptive optics book to discuss flux area density, irradiance, radiant exitance, flux solid angle density, and dozens of others (Wolfe 1998, Grant 2011).

To confine our discussion to adaptive optics, we will converge to *luminance* which is expressed in power per unit area and unit solid angle weighted by the spectral response of the detector. They fall under the expression of lumens per square meter per steradian ($lm/m^2/sr$) or candelas per square meter (cd/m^2). Candela is luminous power per unit steradian emitted by a point source in a particular direction. The lumen (lm) gives the total luminous flux of a light source by multiplying the intensity (in candela) by the angular span over which the light is emitted. Furthermore, *illuminance* is power per unit area weighted by the spectral response or lumens per square meter (lm/m^2). For example, the illuminance of a visible magnitude 6 star is 10^{-8} lm/m^2. The planet Venus is about 1.3×10^{-4} lm/m^2, the full moon is 1 lm/m^2, and street lighting is roughly 100 lm/m^2.

1.4.1 Solid Angle

As seen in the previous section, the notion of solid angle is very important to the concept of measuring radiation and enclosing it within a spatial dimension. Solid angles, represented in steradians (sr) and typically using the variable Ω, generalize an angle from 2D to 3D. A 2D angle (degrees, arcminutes, arcseconds, radians) is a set of unit vectors from the origin to points on a unit circle. The magnitude of the angle is its arc length along the circle. In three dimensions, the unit vectors point to positions on a unit sphere. A solid angle is a section of the unit sphere and its magnitude is the area. The value of the solid angle over all 3D space is simply 4π steradians.

1.4.2 Radiative Transfer

In addition to measuring the radiation at the source or target, adaptive optics is very concerned about the transfer of radiation from one point to another. It is during the transfer that refraction, absorption, and scattering distort the light and the information it carries.

In the simplest of terms, we observe radiation passing through a small area element in every spatial direction. This describes the amount of energy radiated in each of the two senses (forward and reverse) in each spatial direction, per unit time, per unit area of the source, per unit area of the receiver at a distance, per unit wavelength interval. If the spectral radiance is I_λ, the fundamental form of the differential equation shows that as radiation travels, it loses energy to absorption and gains energy by emission processes, and redistributes energy by scattering.

$$\frac{1}{c}\frac{\partial}{\partial t}I_\lambda + \Omega \cdot \nabla I_\lambda + \left(k_{\lambda,\text{scat}} + k_{\lambda,\text{abs}}\right)I_\lambda = j_\lambda + \frac{1}{4\pi}k_{\lambda,\text{scat}}\int_\Omega I_\lambda d\Omega \quad (1.54)$$

Here, c is the speed of light, j_λ is the emission coefficient, $k_{\lambda,\,\text{scat}}$ and $k_{\lambda,\,\text{abs}}$ are the scattering and absorption opacities, respectively, and the integral term represents radiation scattered from other directions onto the surface. In later chapters, we will present many equations to describe the flow of radiation either in a beam or over large solid angles with specific regard to amplitude and phase of the electromagnetic field.

1.5 Terms in Adaptive Optics

A number of terms are used throughout the adaptive optics community. Some of the terms have evolved from electrical or mechanical engineering. Some derive from the terminology of special military applications (U.S. Military Standards 1995). Some terms uniquely apply to adaptive optics.

Many authors use various definitions for the same or similar terms. This is not unusual in a rapidly evolving international field. To maintain consistency in this book, some terms are defined as they are generally applied in the adaptive optics engineering community involved with both astronomy and propagation. Other definitions will be found in the context with which they are used in later chapters.

The difference between *active* and *adaptive* optics was discussed in Section 1.1. It requires a rather broad definition of *open* and *closed* loop. If an optical system or an electro-optical system employs feedback in any fashion, it can be deemed *closed loop*. This applies to both positive and negative feedback, and to optical, electrical, mechanical, or any other method of closing the loop between *information* and *compensation*.

If the optical information is gathered at the receiver or target (in beam propagation) or the image plane (in an imaging application), the system is considered *target loop*. If the information is intercepted before the target or image plane or the application of some correction, the short-circuited loop is called *local loop*.

If optical information is received by the adaptive optics system before a propagating beam reaches the target, it is considered *outgoing wave*. If the information about the propagation is conveyed from the target back to the correction system, it is called *return wave*. Rough, extended objects that are targets of a propagating beam present special problems with speckled return. These speckles can be mitigated with a process called *speckle-average phase conjugation* (Vorontsov et al. 2009).

If the *compensation* requires the movement of a macroscopic mass, it is deemed *inertial*. If the *compensation* alters the state of matter, rather than grossly moving it, it is called *non-inertial*. If the compensation is done by spatially dividing the region of correction and treating each region independently (possibly with cross-coupling), it is called *zonal correction*. Conversely, if the compensation is done by dividing the region of correction in another manner, such as mathematically decomposing the correction into normal modes, it is called *modal correction*.

Understandably, some of these definitions have gray areas. For instance, acousto-optic variance of the index of refraction moves mass on a molecular level, but because it is not a gross macroscopic application, it is non-inertial. Similarly, there might be no "target" defined in the case of adaptive optics inside a laser resonator. This target loop might then be called simply *output loop*. Finally, high energy lasers on moving platforms use *inertial* in the navigation sense; an *inertial* reference unit determines orientation and position in an *inertial*, i.e., non-accelerating frame.

Spot Size: When a beam strikes an opaque surface, the spot of light can take on almost any shape depending upon the apertures and the wavefront of the beam. If the beam is circular and has a sharp edge, it is easy to specify the spot radius or diameter. Gaussian intensity profiles do not have sharp cutoffs at the edge. The *edge* is usually specified as the point where the intensity reaches 13.5% of the maximum. This is the $1/e^2$ point of the beam intensity, also called the beam *waist*. A beam shape that has an easily defined edge is the Fraunhofer diffraction pattern of a circular aperture (Eq. (1.19)). Since the Bessel function reaches its first zero at a well-known value, $J_1(3.8) = 0.0$, the "edge of the circular spot" is defined as the first dark ring, a region of destructive interference. The "spot" in this case contains 84% of the energy. See Figure 1.4.

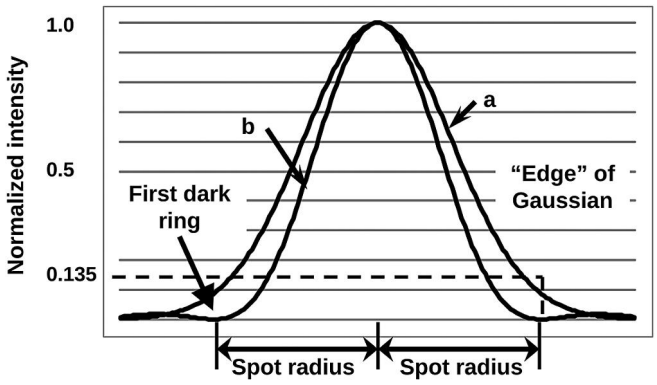

FIGURE 1.4
The ***spot size*** is defined differently for various intensity profiles. (a) Gaussian profile, (b) far field spot of a uniform beam diffracted by a circular aperture.

Other beam spots may be non-symmetric, greatly distorted, or have numerous pockets of high and low intensity (like the speckle pattern on a rough diffuse surface). By taking the area of the smallest spot that has 84% of the energy and computing its equivalent circle diameter, an approximation for the *spot size* is achieved.

Beam Divergence: Using Huygens' wavelet concept (Bickel and Hausler 1988, Born and Wolf 1975), we can see that a plane wave passing through an aperture will begin to diverge from its collimated form as soon as it leaves the aperture. The amount of divergence is dependent upon the initial beam amplitude and phase, the distribution over the aperture, the propagation medium that scatters and diffracts the beam, and one's definition of the "edge" of the beam. The *beam divergence* is usually expressed in terms of an angle between the axis of propagation and the line joining the edge of the beam at different points of the beam path. The solid angle defined by the beam edge over the path of the beam is occasionally used to express beam divergence.

Beam Quality: The use of the Strehl ratio is a fundamental description of the amount of intensity reduction due to aberrations. Another quantity, the *beam quality*, is used in a similar manner. A number

of definitions for beam quality have been used in the optics community. The most common definition states that beam quality is the square root of the inverse of the Strehl ratio,

$$BQ = \sqrt{1/S}. \tag{1.55}$$

For a Gaussian beam, the beam quality is equivalent to the linear increase in beam waist due to beam divergence. A beam with radius M times larger than an ideal Gaussian beam will always diverge at an angle M times larger than an ideal Gaussian beam. This leads to the M^2 metric, which is popular with laser systems because it allows the beam quality to be included in the propagation equations in a trivial manner; we can treat it as an embedded Gaussian and simply scale the spot widths by $1/M$ (Siegman 1998). In general, the linear size growth of the diffraction spot is the beam quality.

Some investigators have redefined the diffraction limit of a beam to better describe a particular system. For instance, the Strehl ratio relates the intensity of the aberrated beam to the intensity of the unaberrated beam. If a beam is not uniform when it is unaberrated, such as an annular beam or any other obscured beam, the unaberrated beam intensity will not be as high as a similar, equal-area unobscured beam. The addition of phase aberrations would normally reduce the on-axis intensity of the beam. However, in some cases, the addition of phase aberrations could increase the on-axis intensity of a strangely obscured beam. This results in a Strehl ratio greater than one, and a subsequent beam quality less than one. By replacing the on-axis intensity of the obscured but unaberrated beam, in the definition of Strehl ratio, the beam quality can be viewed as a single number relating phase aberrations, alone, to propagated intensity. Beam quality should always be related to the unaberrated beam of the same size and shape. This is an important point as it is a non-trivial matter to either measure the aberrated beam or define the unaberrated beam parameters. There are standards devoted to this practice (ISO 2013a, 2013b), but beam quality metrics are still poorly understood and can differ greatly even within the same organization. The performance of a system can be vastly overestimated by carelessly (or intentionally) defining beam quality. For more details on the trials and tribulations with beam quality metrics, see (Ross 2013).

Jitter: *Jitter* is the dynamic tilt of a beam. It is usually expressed in terms of angular variance or its square root, the rms deviation of an angle. The dynamics are often expressed in terms of a power spectral density (PSD) like any mechanical vibration. If the jitter is very slow, usually slower than the response of the system under consideration, it is called *drift*. Since tilt alters the direction and not the shape of a propagated beam, it does not have any effect on the Strehl ratio according to its formal definition. However, if a system is constructed to maintain a beam on a receiver, i.e., a physical spot in space, jitter will cause the beam to sweep across the spot. The time average intensity on the target is reduced by an amount related to the jitter.

The on-axis far-field intensity, I_{ff}, for a circular aperture with reductions due to diffraction expressed as variance of the wavefront, $\sigma^2 = (k\Delta\Phi)^2$, jitter, and transmission losses for m optical elements is given by

$$I_{\text{ff}} \cong \frac{I_0 TK \exp\left[-\sigma^2\right]}{1+\left(2.22\alpha_{\text{jit}} D/\lambda\right)^2}, \tag{1.56}$$

where

$$T = \prod_{i=1}^{m} T_i \tag{1.57}$$

is the transmission of m optics, K is an aperture shape correction described in (Holmes and Avizonis 1976), D is the aperture diameter, λ is the wavelength, and α_{jit} is the *one-axis* rms jitter. The intensity without aberrations or jitter is found from substitution in Eq. (1.4),

$$I_0 = \frac{\pi D^2 P}{\left(2\lambda L\right)^2}, \tag{1.58}$$

where L is the propagation distance and P is the uniformly distributed input power into a circular aperture. As a consequence of Babinet's principle, the intensity of an annular circular aperture with an obscuration ratio ε becomes (Born and Wolf 1975)

$$I_0\big|_{obsc} = \frac{\pi D^2 P}{\left(2\lambda L\right)^2}\left(1-\varepsilon^2\right). \tag{1.59}$$

Power-in-the-Bucket: Knowing the on-axis intensity is sufficient to describe the far-field effects when the beam has a relatively simple analytic form for the intensity distribution. For greatly

aberrated beams or those with complicated apertures, the intensity on-axis is insufficient to describe the far-field effects. A measurement of the total power in the focal or target plane or over a portion of the focal plane is needed. Integrating the power deposited in a circle on the focal plane results in a quantity called *total integrated power*. If one places a "bucket" of a specified radius in the focal plane, the "bucket" would catch that amount of power; thus the term *power-in-the-bucket*. For the perfect circular aperture, the power-in-the-bucket follows (Born and Wolf 1975)

$$\mathrm{PB} = 1 - J_0^2\left(\frac{2\pi ab}{\lambda L}\right) - J_1^2\left(\frac{2\pi ab}{\lambda L}\right), \tag{1.60}$$

where a is the aperture radius and b is the bucket radius.

Brightness: It is often necessary to express the propagation capability of a system in terms independent of the propagation distance. The preceding expression for far-field intensity I_{ff}, when multiplied by the square of the propagation distance, L^2, results in a quantity expressed in units of power per solid angle. This term, called *brightness*, is a description of the propagating system, rather than the effect of the propagation itself:

$$\text{Brightness} \approx \frac{\pi D^2 PTK \exp\left[-\sigma^2\right]}{4\lambda^2\left[1+\left(2.22\alpha_{\mathrm{jit}} D/\lambda\right)^2\right]}. \tag{1.61}$$

Astronomical Brightness: Astronomical brightness, the number of photons reaching the Earth's surface, in a given area in unit time, depends upon the magnitude of the star. It is defined for a visible passband by

$$B_{\mathrm{asto}} = \left(4\times 10^6\right) 10^{-m_v/2.5} \text{ photons/cm}^2 - \text{sec}, \tag{1.62}$$

where m_v is the visual magnitude of the observed star. The limit of vision of an unaided human eye in a dark location is roughly visual magnitude of 6. A visual magnitude of 14 is roughly the brightness of a sunlit geosynchronous satellite.

Seeing: Astronomers use the term *seeing* to describe the condition of the turbulence in the atmosphere. It is based on the ability to resolve two-point objects when observed through the atmosphere. It is essentially the same as the Rayleigh criterion (Born and Wolf 1975, Rayleigh 1879) for resolving two point objects. That is, the full-width-half-maximum (FWHM) of the point spread function stated as an angle is the *seeing*. It is normally expressed in arc seconds (or arc minutes) recognizing that 1 arcsec = 4.8 μradians. Uncompensated atmospheric seeing can be as low as 0.45 arcsec (Ardeberg and Anderson 1993) or, sometimes, as high as 2.0 arcsec (Roddier et al. 1990). For surveillance of low earth orbit space objects, the seeing should be less than 0.02 arcsec (Massie et al. 1992). For astronomy, good seeing is 0.1–0.5 arcsec. Adaptive optics is used to improve seeing.

Fluence: Using lasers as weapons or simply in material processing, it is important to know the *fluence* on the target. The intensity (W/area) by itself does not easily translate to laser damage. The absorption of energy over a period of time is more relevant. The fluence can be expressed as energy per unit area in its simplest form. Thus, multiplying the intensity of a beam on target by its dwell time on the target results in the useful quantity, normally expressed in units of joules/area and very commonly J/cm^2.

EXAMPLE 1.2 ESTIMATING FLUENCE WITHOUT TURBULENCE

Using Eq. (1.61) where Brightness is calculated for a laser system,

$$\text{Brightness} \approx \frac{\pi D^2 PTK \exp\left[-\sigma^2\right]}{4\lambda^2\left[1+\left(2.22\alpha_{\mathrm{jit}} D/\lambda\right)^2\right]},$$

we can substitute parameters of the CO_2 laser system from Pratt & Whitney. With a power of 1 kW, a transmission aperture of 2 m diameter, and a wavelength of 10.6 μm, we can assume transmission T of unity, aperture factor K of unity, and phase losses (the exponential term) equal to 0.9. Assuming that no jitter is evident, the Brightness is 2.5 × 10^{13}W/sr. This is a characteristic of the laser system. Now if it is focused at 1.0 km, the intensity on the target is

$$I_{\mathrm{Tgt}} = \frac{2.5\times 10^{13}\,\mathrm{W/sr}}{\left(10^3\,\mathrm{m}\right)^2}\left(\frac{\mathrm{m}}{100\mathrm{cm}}\right)^2 = 2.5\times 10^3\,\mathrm{W/cm^2}.$$

To find the fluence, we multiply by the laser dwell time of (say) 0.5 s, and find the fluence on target is 1.25 J/cm^2. Do not look into this laser with your one remaining eye.

AO SYSTEM DESIGN EXERCISE – TELESCOPE SPECIFICATION

The chapters will be summarized with an example of the consequences of the discussion in each chapter, so that the design of a complete single conjugate adaptive optics system runs throughout the book. Here we will start with the basic specifications of the telescope itself. We will design an adaptive optics system for a 3.7 m diameter primary mirror that has a central obscuration ratio (for the secondary mirror) of 20%. Figure 1.5 shows the aperture for the telescope with obscurations from the secondary mirror and the spiders holding the secondary mirror, and Figure 1.6 shows the ideal (unaberrated) PSF. Future chapters will build upon these specifications to demonstrate the need for (and impact of) adaptive optics.

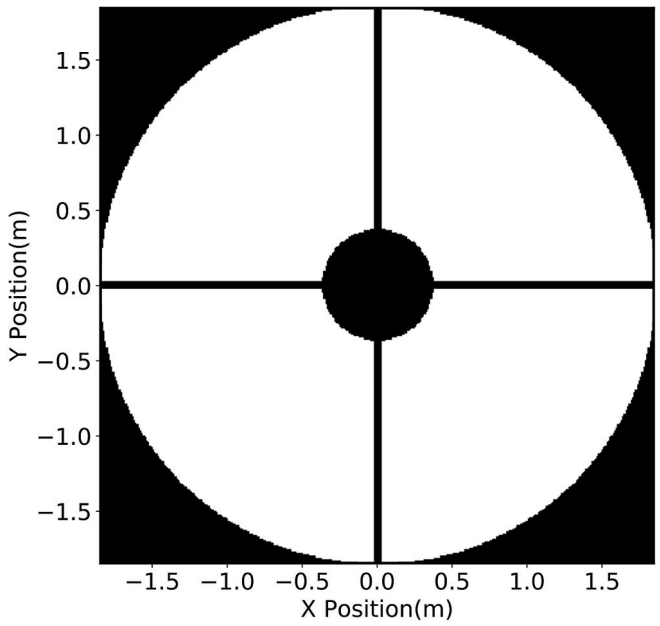

FIGURE 1.5
Telescope aperture for a 3.7 m primary mirror with a 20% central obscuration ratio.

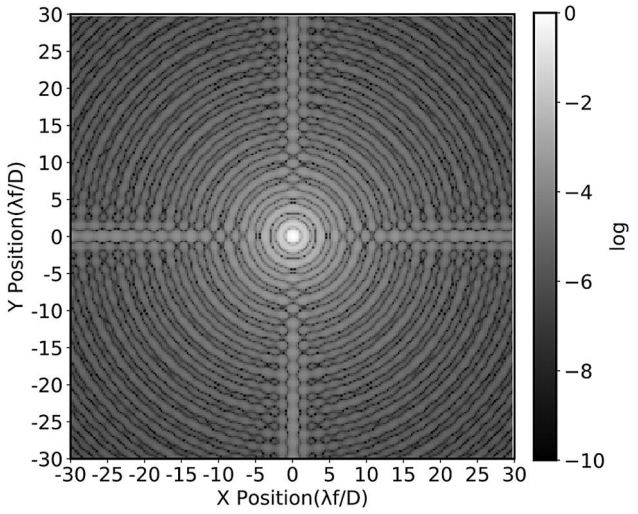

FIGURE 1.6
Ideal PSF for the example telescope, with units specified in terms of spatial resolution ($\lambda f/D$).

1.6 Questions and Problems

Problem 1.1 Considering historical evidence of optical manufacturing capability in ancient times, is Archimedes' "burning glass" impossible, feasible, likely, or found on the second floor of the British Museum?

Problem 1.2 From Eq. (1.1), show why tip/tilt and focus are not considered aberrations in the classical sense.

Problem 1.3 Calculate the Fraunhofer diffraction pattern for an annular aperture with outside radius R and inside radius (radius of the obscuration) εR, with $0 \leq \varepsilon < 1$.

Problem 1.4 Prove Eq. (1.37).

Problem 1.5 Verify that the central spot of Figure 1.4 contains about 84% of the total energy.

Problem 1.6 Explain why one should be careful when describing a highly aberrated wavefront with a least-squares fit to only a few Zernike terms.

Notes

1 Wilson *et al*. 1987 differentiate them by bandwidth. They refer to systems operating below 1/10 Hz as *active* and those operating above 1/10 Hz as *adaptive*. This definition is used widely in the astronomy community.
2 R is also the radius of the wavefront sphere that would perfectly focus at P.
3 For purposes here, tilt is a component of Φ linear in the coordinates $\rho \cos\theta$ or $\rho \sin\theta$ in the pupil plane.
4 Incoherent and white light imaging systems are treated in a similar manner, but the exact form is different.
5 Lord Rayleigh proposed that two equal, but displaced, objects are "resolved" when the principal intensity maximum of one coincides with the first intensity minimum of the other (Rayleigh 1879).

6 Piston is the constant retardation or advancement of the phase over the entire beam.
7 There is a difference in the normalization between the series described by Born and Wolf and that of Noll. Each of Noll's polynomials should be multiplied by the factor $(2(n + 1))^{-1/2}$ to get the Born and Wolf normalization.

References

Ardeberg, A., and T. Anderson, 1993. "*Active optics at work*," *Proc. ICO-16 Satellite Conf. on Active and Adaptive Optics, ESO Conf. and Workshop Proc.* **48**, 433.

Babcock, H. W., 1953. "The possibility of compensating astronomical seeing," *Publ. Astron. Soc. Pac.* **65**, 229–236.

Babcock, H. W., 1990. "Adaptive optics revisited," *Science* **249**, 253–257.

Babcock, H. W., 1992. "*Astronomical background for adaptive optics*," *Proc. Laser Guide Star Adaptive Optics Workshop* **1**, U. S. Air Force Phillips Laboratory, Albuquerque, NM.

BBC News: Science and Environment, 2011. "*Adaptive optics come into focus*," http://www.bbc.co.uk/news/science-environment, 12500626, February 18 2011.

Beckers, J. M., 1993. "Adaptive optics for astronomy: Principles, performance, and applications." *Annu. Ref. Astrophys.* **31**, 13–62.

Belen'kii, M. S., 1995. "Influence of stratospheric turbulence on infrared imaging," *J. Opt. Soc. Am. A* **12**, 2517–2522.

Benedict, Jr., R., J. B. Breckinridge, and D. L. Fried, 1994. "Atmospheric compensation technology: Introduction," *J. Opt. Soc. Am. A* **11**, 257–262.

Bickel, G., and G. Hausler, 1988. "Optical simulation of Huygens's principle," *J. Opt. Soc. Am. A* **5**, 843–848.

Boreman, G. D., and C. Dainty, 1996. "Zernike expansions for non-Kolmogorov turbulence." *J. Opt. Soc. Am. A* **13**, 517–522.

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

Buchdahl, H. A., 1960. "Optical aberration coefficients," *J. Opt. Soc. Am.* **50**, 539–544.

Claus, A. C., 1973. "On Archimedes' burning glass," *Appl. Opt.* **12**, A14.

Conforti, G., 1983. "Zernike aberration coefficients from Seidel and higher-order power-series coefficients," *Opt. Lett.* **8**, 7–9

Dai, G., and V. N. Mahajan, 2008. "Orthonomal polynomials in wavefront analysis: Error analysis," *Appl. Opt.* 47(19), 3433–3445.

Emaleev, O. N., and V. P. Lukin, 1982. "Correction of angular displacements of optical beams," *Sov. J. Quantum Electron.* **12**(11), 1470–1474.

Feinleib, J., 1982. *Proposal 82-P4*, Adaptive Optics Associates, Cambridge, MA.

Foy, R., and A. Labeyrie, 1985. "Feasibility of adaptive optics telescope with laser probe," *Astron. Astrophys.* **152**, L29.

Fugate, R. Q., D. L. Fried, G. A. Ameer, et al., 1991. "Measurement of atmospheric wavefront distortion using scattered light from a laser guide star," *Nature* **353**, 144–146.

Gaskill, J. D., 1978. *Linear Systems, Fourier Transforms, and Optics*, Wiley, New York.

Golbraikh, E., and N. Kopeika, 2008. "Turbulence strength parameter in laboratory and natural optical experiments in non-Kolmogorov cases," *Opt. Commun.* **242**, 333–338.

Goodman, J. W., 1968. *Introduction to Fourier Optics*, McGraw-Hill, New York.

Grant, B. G., 2011. *Field Guide to Radiometry*, SPIE Press, Bellingham, WA.

Greenwood, D. P., and C. A. Primmerman, 1993. "The history of adaptive-optics development at the MIT Lincoln Laboratory," *Proc. SPIE* **1920**, 220–235.

Hardy, J. W., 1978. "Active optics: A new technology for the control of light," *Proc. IEEE* **66**, 651–697.

Hardy, J. W., 1991. "Adaptive optics – A progress review," *Proc. SPIE* **1542**, 2–19.

Holmes, D. A., and P. V. Avizonis, 1976. "Approximate optical system model," *Appl. Opt.* **15**, 1075–1082.

Hopkins, H. H., 1955. "Interferometric methods for the study of diffraction images," *Optica Acta* **2**, 23–29.

Hopkins, M. M., 1957. "The numerical evaluation of the frequency response of optical systems," *Proc. Phys. Soc. London, Sect. B* **70**, 1002–1004.

ISO, 2013a. Lasers and laser-related equipment – Test methods for laser beam widths, divergence angles and beam propagation ratios – Part 1: Stigmatic and simple astigmatic beams, Standard Number 11146-1:2005.

ISO, 2013b. Lasers and laser-related equipment – Test methods for laser beam widths, divergence angles and beam propagation ratios – Part 2: General astigmatic beams, Standard Number 11146-2:2005.

Koliopoulos, C. L., 1988. "Avoiding phase-measuring interferometry's pitfalls," *Photonics Spectra* 169–176.

Labeyrie, A., 1970. "Attainment of diffraction-limited resolution on large telescope by Fourier analyses speckle pattern in stars images," *Astron. Astrophys.* **6**, 85–87.

Lawrence, R. S., G. R. Ochs, and S. F. Clifford, 1970. "Measurements of atmospheric turbulence relevant to optical propagation," *J. Opt. Soc. Am. A* **60**, 826–830.

Linnik, V. P., 1957. "On the possibility of reducing the influence of atmospheric seeing on the image quality of stars," *Opt. Spectrosc.* **3**, 401–402.

Lukin, V. P., 1980. "Correction of random angular displacements of optical beams," *Sov. J. Quantum Electron.* **10**(6), 727–732.

Lukin, V. P., 1981. "Comparative characteristics of some correction algorithms," *Sov. J. Quantum Electron.* **11**(10), 1311–1315.

Mahajan, V. N., 1981. "Zernike annular polynomials for imaging systems with annular pupils," *J. Opt. Soc. Am.* **71**, 75, erratum 1408.

Massie, N. A., Y. Oster, G. Poe, L. Seppala, and M. Shao, 1992. "Low-cost, high-resolution, single-structure array telescopes for imaging of low-Earth-orbit satellites," *Appl. Opt.* **31**, 447–456.

McCluney, W. R., 1994. *Introduction to Radiometry and Photometry*, Artech House, Boston, MA.

Merkle, F., and J. M. Beckers, 1989. "Application of adaptive optics to astronomy," *Proc. SPIE* **1114**, 36–42.

Merkle, F., and N. Hubin, 1991. "Adaptive optics for the European Very Large Telescope," *Proc. SPIE* **1542**, 283–292.

Newton, I., 1979. *Opticks*, based on 4th Ed., 1730; Dover, New York.

Nijboer, B. R. A., 1947. "The diffraction theory of optical aberrations; Part II: Diffraction pattern in the presence of small aberrations," *Physica* **23**, 605–620.

Noll, R. J., 1976. "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.* **66**, 207–211.

Palca, J., 2013. "For sharpest views, scope the sky with quick-change mirrors," National Public Radio, http://www.npr.org/2013/06/24/190986008.

Pearson, J. E., 1980. "*Recent advances in adaptive optics*," Paper TLFF4, *Conf. on Lasers and Electro-optical Systems*, San Diego, CA.

Primmerman, C. A., D. V. Murphy, D. A. Page, B. G. Zollars, and H. T. Barclay, 1991. "Compensation of atmospheric turbulence optical distortion using a synthetic beacon," *Nature* **353**, 141–143.

Rayleigh, L., 1879. "Investigation in optics, with a special reference to the spectroscope: On the accuracy required in optical surfaces," *Philos. Mag.* **5**, 8–10.

Roddier, N., 1990. "Atmospheric wavefront simulation using Zernike polynomials," *Opt. Eng.* **29**, 1174–1180.

Roddier, F., J. E. Graves, and E. Limburg, 1990. "Seeing monitor based on wavefront curvature sensing," *Proc. SPIE* **1236**, 475–480.

Roddier, F., and C. Roddier, 1986. "National Optical Astronomy Observatories (NOAO) infrared adaptive optics program II: Modeling atmospheric effects in adaptive optics systems for astronomical telescopes," *Proc. SPIE* **628**, 299–305.

Roggemann, M. C., and B. Welsh, 1996. *Imaging through Turbulence*, CRC Press, Boca Raton, FL.

Ross, T. S., 2013. *Laser Beam Quality Metrics*, SPIE Press, Bellingham, WA.

Siegman, A. E., 1998. "How to (Maybe) Measure Laser Beam Quality", *DPSS (Diode Pumped Solid State) Lasers: Applications and Issues, OSA Trends in Optics and Photonics* **17**, MQ1.

Stavroudis, O. N., 1973. "Comments on: On Archimedes' burning glass," *Appl. Opt.* **12**, A16.

Tatulli, E., 2013. "Transformation of Zernike coefficients: A Fourier-based method for scaled, translated, and rotated wavefront apertures," *J. Opt. Soc. Am. A* **30**(4), 726–732.

Tyson, R. K., 1982. "Conversion of Zernike aberration coefficients to Seidel and higher-order power-series aberration coefficients," *Opt. Lett.* 7, 262–264.

Tyson, R. K., and J. Schulte in den Bäumen, eds., 1990. *Adaptive Optics and Optical Structures*, *Proc. SPIE* **1271**.

U.S. Military Standards 1995. 490 and 490a, superceded by Mil. Std. 961, Air Force Material Command Standardization Office.

Vorontsov, M. A., V. V. Kolosov, and E. Polnau, 2009. "Target-in-the-loop wavefront sensing and control with a Collett-Wolf beacon: Speckle-average phase conjugation," *Appl. Opt.* **48**(1), A13.

Wills, S., 2020. "Infrared imaging, adaptive optics spurred Nobel-worthy discovery," www.osa-opn.org/home/newsroom/2020/october.

Wilson, R. N., F. Franza, and L. Noethe, 1987. "Active optics I: A system for optimizing the optical quality and reducing the costs of large telescopes," *J. Mod. Opt.* **34**(4), 485–509.

Winker, D. M., 1991. "Effect of finite outer scale on the Zernike decomposition of atmospheric optical turbulence," *J. Opt. Soc. Am. A* **8,** 1568–1573.

Wolfe, W. L., 1998. *Introduction to Radiometry*, Tutorial Texts in Optical Engineering – TT29, SPIE Press, Bellingham, WA.

Woodruff, C. J., 1975. "A comparison, using orthogonal coefficients, of two forms of aberration balancing," *Optica Acta* **22**, 933–941.

Zernike, F., 1934. "Diffraction theory of the knife-edge test and its improved form, the phase contrast method" *Physica* **1**, 689–704.

Zilberman, A., E. Golbraikh, and N. S. Kopeika, 2008. "Propagation of electromagnetic waves in Kolmogorov and non-Kolmogorov atmospheric turbulence: Three-layer altitude model," *Appl. Opt.* **47**(34), 6385–6391.

2

Sources of Aberrations

Unwanted variations of intensity in an image or during beam propagation create the need for adaptive optics. As shown in Chapter 1, it is the deviation of the phase from the reference sphere, known as the *wavefront*, that is the principal cause of the intensity variations that can be treated by adaptive optics. There are many sources of wavefront error.

Astronomers are mostly concerned about the turbulence in the atmosphere that degrades an image. Engineers working to propagate a beam and maximize its useful energy at a receiver must be concerned with errors introduced by lasers themselves, the optics that direct them, and the propagation medium they encounter. This chapter will discuss the many sources of phase aberrations addressed by adaptive optics systems. They include linear effects due to turbulence, optical manufacturing, and misalignments, as well as errors that result from non-linear thermal effects and fluid properties. The minimization of these effects is always a consideration in developing any optical system from the ground up. The real-time compensation for these disturbances is the realm of adaptive optics.

2.1 Atmospheric Turbulence

Naturally occurring small variations in temperature (<1°C) cause random changes in wind velocity (eddies), which we view as turbulent motion in the atmosphere. The changes in temperature give rise to small changes in the atmospheric density and, hence, to the index of refraction. These index changes, on the order of 10^{-6}, can accumulate. The cumulative effect can cause significant inhomogeneities in the index profile of the atmosphere. The wavefront of a beam will change in the course of propagation. This can lead to beam wander, intensity fluctuations (scintillations), and beam spreading.

These small changes in the index of refraction act like small lenses in the atmosphere. They focus and redirect waves and eventually, through interference, cause intensity variations. Each of these "lenses" is roughly the size of the turbulence eddy that caused it. The thin lens model is a useful approximation, but it is not completely accurate because there are few discontinuities in the atmosphere.

The most common effects of turbulence can be seen manifested in the twinkling and quivering of stars. *Twinkling* is the random intensity variation of the light from a star because of the random interference between waves from the same star passing through slightly different atmospheric paths. The average position of the star also shows a random *quiver* because the average angle-of-arrival of light from the star is affected by the changing index of refraction along its path through the atmosphere. A third effect known a long time ago by astronomers was the apparent *spreading* of the star image due to turbulence. The aberrations introduced by the optics did not account for the large spot image of the star, a point object. Turbulence produces random higher-order aberrations that cause the spreading.

These three primary effects of turbulence will be examined here, since they are subject to compensation by adaptive optics. Other atmospheric effects, such as aerosol scattering and molecular absorption, will only be introduced when they affect the performance of the adaptive optics (Avila 2011). The control of high spatial and temporal changes of a wavefront can reduce the phase changes caused by the random temperature fluctuations. This is the principal reason for using adaptive optics.

Volumes have been written on this topic (Chernov 1960; Ishimaru 1977; Strohbehn 1978; Tatarskii 1961), and hundreds of papers also have been written examining the theoretical basis for describing the effects of turbulence and the experimental verification of those theories (Fante 1980). A 1986 book by Lukin, updated and translated from Russian (Lukin 1995), is a comprehensive theoretical examination of adaptive optics-related turbulence phenomena.

Turbulence, as it applies to adaptive optics, can be summarized in basic principles that relate the physical turbulence phenomena with optical propagation and phase effects. Propagation through atmospheric turbulence is not completely understood, especially in terms of small-scale phenomena. The theories of turbulence are based on statistical analyses, since the complexity of the real atmosphere is beyond the capabilities of deterministic prediction or numerical analysis. The emphasis on a statistical description of atmospheric turbulence has resulted in a number of very useful theories and scaling laws that describe the average effects on gross properties, such as total beam wander, beam spread, and scintillation (Fante 1979).

DOI: 10.1201/9781003140191-2

2.1.1 Refractivity

The atmosphere, composed mostly of nitrogen and oxygen, is transparent to visible light. Because it is matter, and not vacuum, the speed of light slows down in the atmosphere. Described in detail in the next section, the amount of velocity reduction is expressed as the dimensionless quantity, the index of refraction, usually denoted by n with the definition related to the velocity reduction $n = c/v$, where v is the velocity of light inside the medium and c is the velocity of light in vacuum. The index of refraction of air depends upon temperature, which relates to density. It is the temporally dynamic small variations of temperature and density throughout the atmosphere along an optical beam path that cause local beam wander and large-scale phase variations in a beam of light.

2.1.2 Statistical Representations

Areas of high- and low-density air are moved around by random winds. This process can be described by statistical quantities. Kolmogorov (1961) studied the mean-square velocity difference between two points in space separated by a displacement vector **r**. The structure tensor D_{ij} is defined as

$$D_{ij} = \left\langle \left[\upsilon_i(\mathbf{r}_1 + \mathbf{r}) - \upsilon_i(\mathbf{r}_1) \right] \left[\upsilon_j(\mathbf{r}_1 + \mathbf{r}) - \upsilon_j(\mathbf{r}_1) \right] \right\rangle, \tag{2.1}$$

where υ_i and υ_j are the different components of the velocity and the brackets represent an ensemble average. It is not a simple equation to evaluate with real velocity descriptors; however, if three assumptions about the atmosphere are made, the structure tensor can be simplified. First, if the atmosphere is locally homogeneous (velocity depends on vector **r**), second, if the atmosphere is locally isotropic (velocity depends only on the magnitude of **r**), and third, if the turbulence is incompressible ($\nabla \cdot \mathbf{v} = 0$), the tensor now becomes a single structure function (Batchelor 1953),

$$D_\upsilon = \left\langle \left[\upsilon_r(r_1 + r) - \upsilon_r(r_1) \right]^2 \right\rangle. \tag{2.2}$$

If the separation r is small (within the inertial subrange of turbulence), the structure function takes on a 2/3 law dependence on r,

$$D_\upsilon = C_\upsilon^2 r^{2/3}. \tag{2.3}$$

The constant C_υ^2 is the velocity structure constant, which is a measure of the energy in the turbulence. This form of the structure function is valid when the value of r is above the smallest eddy size l_0 and below the largest eddy size L_0. The small eddy, called the *inner scale* or the *microscale*, is the size below which viscous effects are important and energy is dissipated into heat. The inner scale is related to the rate of dissipation of turbulent kinetic energy ε and the kinematic viscosity ν by Ochs and Hill (1985)

$$l_0 = 7.4\left(\nu^3 / \varepsilon\right)^{1/4}. \tag{2.4}$$

The large eddy, called the *outer scale*, is the size above which isotropic behavior is violated. One can approximate l_0 as a few millimeters near the surface to centimeters or more in the troposphere. The outer scale L_0 ranges from a few meters to hundreds of meters. Near the ground, below one kilometer, $L_0 \approx 0.4h$ where h is the height above the ground (Tatarskii 1961). In the free atmosphere, the outer scale is normally on the order of tens of meters, but it can reach hundreds of meters (Barat 1982).

Through the concept of passive additives, Tatarskii (1961) and Corrsin (1951) related the velocity structure to the index of refraction structure. This quantity D_n is much more important when we must deal with propagation issues,

$$D_n(r) = C_n^2 r^{2/3}; l_0 < r < L_0, \tag{2.5}$$

where C_n^2 is the refractive index structure constant, a measure of the strength of turbulence. Effects of finite inner scale and behavior of the structure function when r is near l_0 are discussed in the literature (Whitman and Beran 1988).

The atmosphere can be assumed to be made up of a mean index of refraction $\langle n(\mathbf{r}) \rangle$ and a fluctuating index part $n_1(\mathbf{r})$. The covariance of the refractive index field B_n becomes

$$B_n = \left\langle n_1(\mathbf{r} + \mathbf{r}_1) n_1(\mathbf{r}_1) \right\rangle. \tag{2.6}$$

The power spectral density (PSD) is the Fourier transform of the covariance,

$$\Phi_\mathbf{n}(\mathbf{K}) = \frac{1}{(2\pi)^3} \int \mathbf{d}^3\mathbf{r} \mathbf{B}_\mathbf{n}(\mathbf{r}) \mathbf{e}^{-\mathbf{i}\mathbf{K}\cdot\mathbf{r}}, \tag{2.7}$$

where **K** is the 3D spatial wave number. Using Kolmogorov's inertial subrange expression, Eq. (2.5), changing coordinates to spherical terms **K** = (K, θ, $\varnothing$), and carrying out the ensemble averaging (Strohbehn 1978), the PSD becomes

$$\Phi_n(K) = \frac{5}{18\pi} C_n^2 K^{-3} \int_{l_0}^{L_0} dr \sin(Kr) r^{-1/3}. \tag{2.8}$$

If we let the integral limits diverge, $l_0 \to 0$, and $L_0 \to \infty$, the integral results in the *Kolmogorov spectrum*:

$$\Phi_n(K) = 0.033 C_n^2 K^{-11/3}. \tag{2.9}$$

The convenience of a zero inner scale and infinite outer scale was used for performing the integration. Tatarskii described a spectrum that could be used for finite inner scales (Tatarskii 1961, Lukin 1995),

$$\Phi_n(K) = 0.033 C_n^2 K^{-11/3} \exp\left(\frac{-K^2}{(5.92/l_0)^2}\right). \tag{2.10}$$

A spectrum useful for cases with finite outer scales is the *Von Karman spectrum* (Tatarskii 1961),

$$\Phi_n(K) = \frac{\Gamma(11/6)\pi^{-9/2}}{\Gamma(1/3)8}\delta^2 L_0^3 \left(1 + K^2/K_0^2\right)^{-11/6}. \tag{2.11}$$

In this expression the constant term is evaluated,

$$\frac{\Gamma(11/6)\pi^{-9/2}}{\Gamma(1/3)8} \cong 2.54 \times 10^{-4}, \tag{2.12}$$

$K_0 = 2\pi/L_0$, and the variance of the refractivity fluctuations is

$$\delta^2 = \frac{C_n^2}{1.9 K_0^{2/3}}. \tag{2.13}$$

Each of these spectra is valid only for regions in the inertial subrange,

$$\frac{2\pi}{L_0} \le K \le \frac{2\pi}{l_0}. \tag{2.14}$$

2.1.3 Refractive Index Structure Constant

The *refractive index structure constant*, C_n^2, is a measure of the strength of turbulence. It is, by no means, constant. It varies with seasons as well as daily and hourly. It varies with geographic location and with altitude. Figure 2.1 shows measured variations and two well-known numerical models to approximate the average. It can be perturbed very easily by artificial means such as aircraft. There have been many measurements of C_n^2 (Dayton et al. 1992, Hufnagel 1989, Vedrenne et al. 2007, Ko et al. 2017). No theoretical model is accurate for the many cases of turbulence. Based on experimental observations, Hufnagel suggested

$$C_n^2 \left\{ \left[\left(2.2 \times 10^{-53}\right) h^{10} \left(W/27\right)^2 \right] e^{-h/1000} + 10^{-16} e^{-h/1500} \right\} \exp\left[r(h,t) \right], \tag{2.15}$$

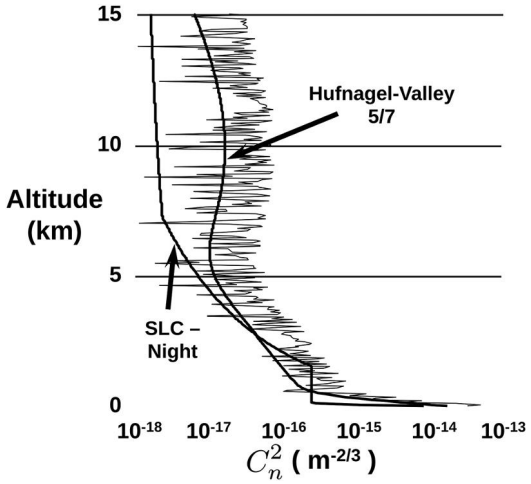

FIGURE 2.1
One measurement of the refractive index structure constant C_n^2 as a function of altitude. The Hufnagel–Valley 5/7 model and the SLC–Night model are shown for comparison.

where h is the height in meters above sea level, W is the wind correlating factor defined as

$$W = \left[(1/15\text{km}) \int_{5\text{km}}^{20\text{km}} \upsilon^2(h) dh \right]^{1/2}, \tag{2.16}$$

and $r(h, t)$ is a zero mean homogeneous Gaussian random variable. $\upsilon(h)$ is the wind speed at height h. The units of C_n^2 are $m^{-2/3}$ (Hufnagel 1989).

The dependence of wind speed on altitude is required for a rigorous application of Eq. (2.15). A wind model commonly used and applicable to calculations throughout this book was developed by Bufton (1973):

$$\upsilon(z) = 5 + 20 \exp\left\{ -\left[(z - 9.4)/4.8 \right]^2 \right\}, \tag{2.17}$$

where z is in kilometers and the wind velocity υ is in meters per second.

In his study of anisoplanatism discussed in Chapter 3, Valley (1980) altered this model slightly. Ulrich added a term to account for the boundary layer with the resultant Hufnagel–Valley Boundary (HVB) model (Ulrich 1988) expressed as

$$C_n^2 = 5.94 \times 10^{-23} z^{10} e^{-z} \left(\frac{W}{27}\right)^2 + 2.7 \times 10^{-16} e^{-2z/3} + A e^{-10h}, \tag{2.18}$$

where the height above mean sea level z is in kilometers, h is the height above the surface of the site in kilometers, and C_n^2 is in $m^{-2/3}$. W is an adjustable parameter related to the upper atmosphere wind velocity. A is a scaling constant (Ulrich 1988; Valley 1980). If the site is at sea level, both W and A can be adjusted to correspond to specific integrated values of the coherence length r_o

and the isoplanatic angle θ_o.[1] For the case of r_o = 5 cm, θ_o = 7 μrad and λ = 0.5 μm (the so-called HV 5/7 model), the parameters are $A = 1.7 \times 10^{-14}$ and W = 21. With the wavelength given in micrometers, the coherence length r_o given in centimeters, and the isoplanatic angle θ_o given in microradians, the HVB parameters can be found for any site conditions (Tyson 1996) using

$$W = 27\left(75\theta_0^{-5/3}\lambda^2 - 0.14\right)^{1/2} \tag{2.19}$$

and

$$A = 1.29\times10^{-12} r_o^{-5/3}\lambda^2 - 1.61\times10^{13}\theta_o^{-5/3}\lambda^2 - 3.89\times10^{-15}. \tag{2.20}$$

Another model commonly used to calculate parameters associated with atmospheric turbulence is the SLC-Night model, named after the Submarine Laser Communications program for which it was developed. Where h is the altitude above ground (in meters), the model is provided in Table 2.1.

Using these models must come with a number of concerns. While the statistics of the models usually represents the correct average spectrum, for very small scales and large scales, they sometimes break down and provide inaccurate phase screens at various altitudes. It was recently reported that the sparse spectrum model of the atmosphere can generate 2D Monte Carlo random wavefronts with the most common power-law spectra including the Kolmogorov spectrum (Charnotskii 2013).

Direct measurements of C_n^2 have been reported. Using slope detection and ranging (SLODAR), Zuraski used two or more natural guide stars or two or more Rayleigh beacons and also used the overlap to directly measure the C_n^2 profile near the ground (Zuraski et al. 2020). Ullwer et al. made use of the Weather Research and Forecast Model to calculate a worldwide dataset of profiles with a 1.1 km altitude resolution (Ullwer et al. 2019). Yatcheva et al. used a combination of six devices, e.g., scintillometers and sonic anemometers, to characterize C_n^2 (Yatcheva et al. 2015) and Farley et al. reduced 10,000 C_n^2 profiles from ESO Paranal into 18 profile clusters to place a range on Fried's coherence length and isoplanatic angles (Farley et al. 2018; see Section 2.1.4).

TABLE 2.1

SLC-Night Model for C_n^2

Altitude	C_n^2
$h \le 18.5$	8.40×10^{-15}
$18.5 < h \le 110$	$2.87 \times 10^{-12} h^{-2}$
$110 < h \le 1500$	2.5×10^{-16}
$1500 < h \le 7200$	$8.87 \times 10^{-7} h^{-3}$
$7200 < h < 20{,}000$	$2.00 \times 10^{-16} h^{-0.5}$

2.1.4 Turbulence Effects

Turbulence causes high spatial frequency beam spreading, low spatial frequency beam wander, and intensity variations. Beam spread is produced by eddies that are smaller than the beam size. Wander is produced by eddies that are larger than the beam size (Strohbehn 1978). The Kolmogorov spectrum suggests that intensity variations are produced by eddies with sizes on the order of $\sqrt{\lambda L}$, where L is the propagation distance.

2.1.4.1 Fried's Coherence Length

In a study of an optical heterodyne communications receiver, Fried found that the maximum diameter of a collector that is allowed before atmospheric distortion seriously limits performance is r_0, where the *coherence length*[2] is

$$r_0 = \left[0.423k^2 \sec(\beta)\int_0^L C_n^2(z)\,dz\right]^{-3/5}. \tag{2.21}$$

where L is the path length, β is the zenith angle, and C_n^2 can vary with altitude z (Fried 1965a).

The phase structure function for Kolmogorov turbulence can be found for a number of cases in terms of this parameter (Fried 1965b). For plane waves,

$$D_\phi = 6.88\left(\frac{r}{r_0}\right)^{5/3}. \tag{2.22}$$

The radial variable r is normal to the propagation direction. For spherical waves, the coherence length is slightly modified,

$$r_{0\text{sph}} = \left[0.423k^2 \sec(\beta)\int_0^L C_n^2(z)\left(\frac{z}{L}\right)^{5/3} dz\right]^{-3/5}. \tag{2.23}$$

Equation (2.22) can be generalized for non-Kolmogorov turbulence such as that described by Karhunen–Loève functions which do not have a closed analytic form (Roddier 1990, Anzuola and Gladysz 2017). Nicholls et al. (1996) discuss a generalized model for the structure function that takes the form

$$D_n(r) = \gamma_\eta\left(\frac{r}{R_0}\right)^{\eta-2}. \tag{2.24}$$

When $\eta = 11/3$, $R_0 = r_o$, and $\gamma_\eta = 6.88$, Kolmogorov turbulence is assumed. If η is less than 11/3, some of the energy for phase aberrations shifts from lower to higher order modes, requiring modifications to adaptive optics.

For a constant C_n^2 (i.e., for a horizontal path length), the coherence length for plane waves reduces to

$$r_0\big|_{\text{pl}} = 1.68\left(C_n^2 L k^2\right)^{-3/5}. \tag{2.25}$$

For spherical waves, the coherence length is

$$r_0\big|_{\text{sph}} = 3.0\left(C_n^2 L k^2\right)^{-3/5}. \tag{2.26}$$

In astronomy, starlight is a plane wave when it enters the atmosphere, so the plane wave description is assumed. Typical nighttime median r_0 at wavelength λ can be approximated by Fried and Mevers (1974)

$$\left[r_0\right]_{\text{median}} = .114\left(\frac{\lambda}{5.5\times10^{-7}}\right)^{6/5}\left(\sec\beta\right)^{-3/5}, \tag{2.27}$$

where λ is given in meters.

Fried's coherence length has been used to describe various atmospheric turbulence phenomena. The coherence length can be measured indirectly (Atad et al. 1990, Scheglov 1981). By using a two-channel Hartmann sensor (Section 5.3.2) as a differential image motion monitor (Forbes et al. 1988, Sarazin and Roddier 1990), atmospheric seeing can be measured. Because r_0 incorporates the turbulence strength $\text{FWHM}_{\text{core}} = \sqrt{\left(1.22\lambda/D\right)^2 + \left(2.7\alpha_{\text{jit}}\right)^2}$, the wavelength, and the propagation path into one parameter, it is widely used for scaling laws and descriptions of atmospheric phenomena. For instance, the spectrum of turbulence in the spatial domain (with spatial frequency ξ expressed in inverse meters) is expressed in terms of coherence length (Noll 1976). With r_0 in meters, the coherence length is

$$\Phi\left(\xi\right) = \left(0.023/r_0^{5/3}\right)\xi^{-11/3}. \tag{2.28}$$

This spatial frequency spectrum can be used to describe all the aberration modes of atmospheric turbulence in terms of Zernike coefficients and other polynomials (Roddier 1990). Also, the spectrum can be integrated to give an overall wavefront variance,

$$\sigma^2 = \int \Phi\left(\xi\right) d^2\xi. \tag{2.29}$$

It is important to point out that a measurement of r_0 is a snapshot in time. It may vary over very short time scales. For system design purposes, the median seeing for a site can be measured and used in modeling and design. For example, the median seeing at the Mauna Kea, Hawaii site has been reported to be 0.55 arcsec (Gilles et al. 2018) and 0.475 arcsec (Ono et al. 2016).

2.1.4.2 Scintillation

Intensity variations are usually expressed as the fluctuation of the log of the amplitude (*log–amplitude fluctuations*). The Kolmogorov spectrum suggests that log–amplitude fluctuations are produced by eddies with sizes on the order of $\sqrt{\lambda L}$. L is the propagation distance. Assuming the Von Karman spectrum and $l_0 \ll \sqrt{\lambda L}$, the log amplitude variance σ_χ^2 is given by Tatarskii (1961)

$$\sigma_\chi^2 = 0.307 k^{7/6} L^{11/6} C_n^2 \tag{2.30}$$

for a plane wave, and

$$\sigma_\chi^2 = 0.124 k^{7/6} L^{11/6} C_n^2 \tag{2.31}$$

for a spherical wave. These formulae break down for long paths and strong turbulence (Primmerman et al. 1995) because perturbation theory was used to derive them. Tatarskii used a technique derived by Rytov (1937) to develop this theory (Tatarskii 1961). Because of the limitations of the *Rytov approximation*, these equations hold only for $\sigma_\chi^2 < 0.3$.

Using the simple multiple lens model for the atmosphere, one can calculate the amplitude variance. Since each eddy of diameter l acts like a lens, it will have a focal length of $f \approx l/\Delta n$. In the geometric optics regime $l > \sqrt{\lambda L}$ and $l_0 < \sqrt{\lambda L}$, the log–amplitude variance is (Clifford 1978)

$$\sigma_\chi^2 \approx k^{7/6} L^{11/6} C_n^2, \tag{2.32}$$

which is within a multiplicative constant of the more rigorous Tatarskii method.

The scintillation effects are most prominent when a small or point receiving aperture is used, such as in an optical communications receiver, a small optical tracker, or the human eye. The most common technique for modeling the propagation of light through atmospheric turbulence is the generation of static phase screens at various points along the beam path (with the correct C_n^2 for the altitude). With phase screens placed too close to the imaging or receiving aperture, seeing effects appear to be such that the entire field is within the isoplanatic angle. If the phase screens are too close to the object or transmitter, severe anisoplanatism will result as high scintillation and seeing effects near the aperture are not seen. This is the classic *shower door effect*. A person in the shower can place their eye near the rippled glass and clearly see the scene outside, while an observer like the Peeping Tom across the room, cannot clearly see the subject inside. This has been quantified. The total number and placement of phase screens in the propagation model should be such that each phase screen must account

for no more than 10% of the total scintillation (Bos and Roggemann 2012).

In some cases, scintillation can be interpreted in a centroid tracker as a tilt or jitter error when no phase tilt actually exists (Holmes 2009). For larger apertures, the scintillation effects are averaged over the extent of the aperture and the focused irradiance varies less. Scintillation is observed as a variation in the irradiance, σ_I^2, expressed as

$$\sigma_I^2 = A\left[\exp\left(4\sigma_X^2\right) - 1\right], \tag{2.33}$$

where A is an *aperture averaging factor* (Churnside 1991, Angel 1992).

Churnside presents a number of expressions for the aperture averaging factor under various conditions (Churnside 1991). For weak turbulence conditions, and a small l_0, the aperture averaging factor is

$$A = \left[1 + 1.07\left(\frac{kD^2}{4L}\right)^{7/6}\right]^{-1}. \tag{2.34}$$

For other conditions, the literature should be consulted (Peleg and Moloney 2006).

Because atmospheric turbulence is dynamic, scintillation is dynamic. Tatarskii gives the temporal power spectrum of scintillation in terms of the frequency f, and the zenith angle β (Tatarskii 1961).

$$\begin{aligned} P(f) = {} & 8.27\sec^{7/3}\beta k^{2/3}\int_0^L \frac{C_n^2(z)z^{4/3}}{\upsilon_w(z)}dz \\ & \times\int_0^\infty\left[x^2 + \frac{f^2}{f_o^2(z)}\right]^{-11/6}\sin\left(x^2 + \frac{f^2}{f_o^2(z)}\right)dx. \end{aligned} \tag{2.35}$$

The frequency $f_o(z)$ is a function of the wind velocity $\upsilon_w(z)$

$$f_o(z) = \left[\frac{2k}{z\sec\beta}\right]^{1/2}\upsilon_w(z). \tag{2.36}$$

The integral over the dummy variable x in Eq. (2.35) can be approximated by Tyson (1996)

$$\int[\cdots]dx = \exp\left[-1.8\left(\frac{f}{f_o} - 0.5\right)^{1.9}\right] \text{for} \frac{f}{f_o} \geq 0.5 \tag{2.37}$$

$$\int[\cdots]dx = 1 \text{ for} \frac{f}{f_o} < 0.5. \tag{2.38}$$

2.1.4.3 Beam Wander or Tilt

The turbulent atmosphere will cause wander of a beam as it propagates. When the wander is fast, it is often called beam *jitter*. When the wander is slow, it is called *drift*. It is caused by the dynamic tilt of the wavefront. Sasiela gives a general description of the wavefront variance:

$$\begin{aligned} \sigma_{wF}^2 = {} & 0.2073k^2\int_0^L dzC_n^2(z)\int_{-\infty}^{\infty}d\vec{\kappa}\vec{\kappa}^{-11/3} \\ & \cos^2\left[\frac{\vec{\kappa}^2(z-L)}{2k}\right]\prod_N^{i=0}F_i(\vec{\kappa},z), \end{aligned} \tag{2.39}$$

where the z integral is over the atmospheric path, $\vec{\kappa}$ is a 2D spatial frequency, N is the number of Zernike modes included, and $k = 2\pi/\lambda$ (Sasiela 1994). The filters F_i for the Zernike modes (Sasiela and Shelton 1993) are given by

$$F_{\text{even }m,n}(\kappa) = 2(n+1)\left[\frac{2J_{n+1}\left(\frac{\kappa D}{2}\right)}{\kappa D/2}\right]^2\cos^2(m\phi), \tag{2.40}$$

$$F_{\text{odd }m,n}(\kappa) = 2(n+1)\left[\frac{2J_{n+1}\left(\frac{\kappa D}{2}\right)}{\kappa D/2}\right]^2\sin^2(m\phi), \tag{2.41}$$

$$F_{m=0,n}(\kappa) = (n+1)\left[\frac{2J_{n+1}\left(\frac{\kappa D}{2}\right)}{\kappa D/2}\right]^2. \tag{2.42}$$

Using the $n = 1$, $m = 1$ terms in the filters from Eq. (2.41), Sasiela derives the expression for wavefront tilt variance σ_{tilt}^2:

$$\begin{aligned} \sigma_{\text{tilt}}^2 = {} & 0.2073k^2\int_0^L dzC_n^2(z) \\ & \times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}d\vec{\kappa}\left[\vec{\kappa}^2 + k^2\right]^{-11/6}\left(\frac{16}{kD}\right)^2\left[\frac{J_2(\kappa D/2)}{\kappa D/2}\right]^2, \end{aligned} \tag{2.43}$$

where J_2 is the second-order Bessel function of the first kind, $\vec{\kappa}$ is spatial frequency, and D is the aperture diameter. A similar expression was derived for the tilt variance for an annular aperture with a centered obscuration of diameter εD,

$$\sigma_{\text{tilt}}^2(\varepsilon D) = 0.2073k^2\int_0^L dzC_n^2(z)$$

$$\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\vec{\kappa}\left[\vec{\kappa}^2 + k^2\right]^{-11/6}\left(\frac{16}{kD\left(1-\varepsilon^4\right)}\right)\left[\frac{J_2(kD/2)}{kD/2} - \varepsilon^3\frac{J_2(k\varepsilon D/2)}{k\varepsilon D/2}\right]^2 \quad (2.44)$$

Using Eq. (2.43) and evaluating the coherence length, the variance of the tilt angle α^2 can be found:

$$\alpha_{2-\text{axis}}^2 = 0.364\left(\frac{D}{r_0}\right)^{5/3}\left(\frac{\lambda}{D}\right)^2 \text{ or } \alpha_{1-\text{axis}}^2 = 0.182\left(\frac{D}{r_0}\right)^{5/3}\left(\frac{\lambda}{D}\right)^2 \quad (2.45)$$

for 2 uncorrelated axes or each axis.

The dynamics of the tilt are manifested in their power spectra. For gradient tilt (G-tilt), (Gladysz 2016) like that measured by quadrant detectors, the one-axis tilt power spectrum takes the form (Tyler 1994)

$$P(f) = 0.155D^{-1/3}\sec\beta f^{-8/3}\int_0^L F_G\left(\frac{f}{\upsilon_w}\right)C_n^2(z)\upsilon_w^{5/3}dz, \quad (2.46)$$

where the function F_G is

$$F_G(y) = \int_0^1 dx\frac{x^{5/3}}{\sqrt{1-x^2}}J_1^2(\pi y/x). \quad (2.47)$$

For low frequencies, this reduces to the −2/3 power law expression

$$P(f)_{\text{low}} \to 0.804D^{-1/3}\sec\beta f^{-2/3}\int_0^L C_n^2(z)\upsilon_w^{-1/3}dz, \quad (2.48)$$

and for high frequencies it follows a −11/3 power law:

$$P(f)_{\text{high}} \to 0.110D^{-1/3}\sec\beta f^{-11/3}\int_0^L C_n^2(z)\upsilon_w^{8/3}dz. \quad (2.49)$$

Similar expressions follow for Zernike tilt (Z-tilt)[3]:

$$P(f) = 0.251D^{-1/3}\sec\beta f^{-14/3}\int_0^L F_Z\left(\frac{f}{\upsilon_w}\right)C_n^2(z)\upsilon_w^{11/3}dz, \quad (2.50)$$

where the function F_Z is

$$F_Z(y) = \int_0^1 dx\frac{x^{11/3}}{\sqrt{1-x^2}}J_2^2(\pi y/x). \quad (2.51)$$

For low frequencies, this reduces to the same −2/3 power law expression that was found for G-tilt,

$$P(f)_{\text{low}} \to 0.804D^{-1/3}\sec\beta f^{-2/3}\int_0^L C_n^2(z)\upsilon_w^{1/3}dz. \quad (2.52)$$

For high frequencies, the Z-tilt power spectrum follows a −17/3 power law:

$$P(f)_{\text{high}} \to 0.014D^{-1/3}\sec\beta f^{-17/3}\int_0^L C_n^2(z)\upsilon_w^{14/3}dz. \quad (2.53)$$

A single parameter can represent the spectrum for adaptive optics compensation analysis. The characteristic frequency of the tilt of atmospheric turbulence, often called the *tilt Greenwood frequency*, is expressed by Tyler (1994). For G-tilt,

$$f_T = 0.331D^{-1/6}\lambda^{-1}\sec^{1/2}\beta\left[\int_0^L C_n^2(z)v_w^2dz\right]^{1/2}. \quad (2.54)$$

For Z-tilt, the constant in Eq. (2.54) becomes 0.368.

Measurement of atmospheric tilt can be complicated by the wind-driven telescope vibrations induced in the angle-of-arrival (AOA) measurement system. Separating the two sources of AOA fluctuations is especially difficult in small lightweight telescopes. Experiments that quantify the relative sources have been successful (Tichkule and Muschinski 2014).

2.1.4.4 Higher-order Phase Variation

Turbulent eddies that are smaller than the beam size break up the beam and spread out the energy. The wavefront variations arise because the refractive index is a random function of space and time. The phase difference $\delta\phi$ between two parallel paths through the atmosphere separated by distance ρ is $\delta\phi \simeq k\rho[\Delta n(\rho)]$, where the difference in index is Δn. By averaging over many statistical realizations and assuming a path length L, the *phase structure function* can be determined,

$$D_\varphi = 1.46k^2C_n^2L\rho^{5/3}, \quad l_0 < \rho < L_0. \quad (2.55)$$

When propagation occurs over long paths and many eddies, the wave can lose *coherence*. If the wavefront is strongly aberrated before it encounters an eddy, the plane wave approximations used thus far are no longer valid. If the phase difference is $\langle(\delta\phi)^2\rangle < \pi^2$, the beam becomes coherent. If ρ_0 is the largest eddy for which a plane wave is considered coherent, from Eq. (2.55) the *coherence distance* is

$$\rho_0 \approx \left[\frac{\pi^2}{\pi^2C_n^2L}\right]^{3/5}. \quad (2.56)$$

This coherence limitation can be used to derive the total beam spread. Lutomirski and Yura (1971) used the extended Huygens–Fresnel principle to show that

the angular spread θ of a collimated uniform beam of radius α in uniform turbulence[4] is

$$\theta^2 \approx \frac{1}{k^2 a^2} + \frac{1}{k^2 \rho_0^2}. \tag{2.57}$$

For the specific case of a Gaussian beam (Ji and Li 2009),

$$I = \frac{w_o^2}{w_b^2} \exp\left(-\frac{2\rho^2}{w_b^2}\right). \tag{2.58}$$

For short propagation distances ($z \ll \pi w_0^2/\lambda$), the beam waist w_b increases from its initial size w_o according to

$$w_b^2 = w_o^2 + 2.86 C_n^2 k^{1/3} L^{8/3} w_o^{1/3}. \tag{2.59}$$

For long distances ($z \gg \pi w_0^2/\lambda$), it increases according to

$$w_b^2 = \frac{4z^2}{k^2 w_o^2} + 3.58 C_n^2 L^3 w_o^{-1/3}. \tag{2.60}$$

The Strehl ratio and imaging system spatial resolution can be computed from the phase variance. Noll (1976) shows that the uncompensated turbulence wavefront variance is

$$\sigma_{\text{unomp}} = 1.02\left(\frac{D}{r_o}\right)^{5/3}. \tag{2.61}$$

Because we are concerned with beam spread and not the wander of the beam centroid, the wavefront variance, after two-axis tilt is removed, becomes

$$\sigma^2_{\text{tiltcomp}} = 0.134\left(\frac{D}{r_o}\right)^{5/3}. \tag{2.62}$$

The Strehl ratio, S, can be calculated using Eq. (1.6), disregarding the terms that do not contribute to beam spreading. In addition to the beam spread from small turbulent eddies, large eddies jitter the beam. Over a long exposure, a beam, spread from high-order turbulence effects, will be spread more by the high-frequency jitter. The long-exposure Strehl ratio is (Gavel and Olivier 1994)

$$S = \frac{e^{-\sigma^\sigma_{\text{comp}}}}{1 + \left(2.22\alpha_{\text{jit}} D/\lambda\right)^2} \tag{2.63}$$

where σ^2_{comp} is the wavefront variance with tilt removed, and α_{jit} is the 1-axis tilt jitter. The image of a star captured with an astronomical telescope, that is, the optical system point spread function, will have a bright central core and a wider background halo like that shown in Figure 2.2.

Parenti shows a modified Strehl ratio expression that accounts for the background halo (Parenti 1992):

$$S = \frac{e^{-\sigma^\sigma_{\text{comp.}}}}{1 + \left(2.22\alpha_{\text{jit}} D/\lambda\right)^2} + \frac{e^{-\sigma^\sigma_{\text{comp.}}}}{1 + \left(D/r_o\right)^2}. \tag{2.64}$$

The full-width-half-maximum (FWHM) of the central PSF core can be approximated by

$$\text{FWHM}_{\text{core}} = \sqrt{\left(1.22\lambda/D\right)^2 + \left(2.7\alpha_{\text{jit}}\right)^2}. \tag{2.65}$$

The halo has a FWHM equal to $1.22\lambda/r_o$. Detailed studies of the core and halo and observations of hundreds of PSFs at various sites reveal some interesting results. While it was often assumed that halo amplitude was reduced and it remained the same size or shrunk in radius as adaptive optics was used to sharpen the central core, this turns out not to be the case. With detailed simulations and data from the 3.5 m telescope at the Starfire Optical Range, it was shown that with low-order tip/tilt correction, the halo had shrunk in size, but with higher-order aberration correction, the halo increased in size following an inverse relationship with adaptive optics inter-actuator spacing (Gladysz et al. 2008, 2012; Johnson et al. 2009). Another study of the PSF shape and size showed that the distance from the wavefront guide star in the field-of-view greatly affected the amount of variation. The simulations were verified by data from the Very Large Telescope (VLT) and the Large Binocular Telescope (LBT; Schreiber et al. 2012).

The resolution R of an optical system imaging through the atmosphere (using Rayleigh's criterion (Rayleigh 1879)) becomes slightly complicated by the halo being superimposed upon the image core (Parenti 1992). The resolution is

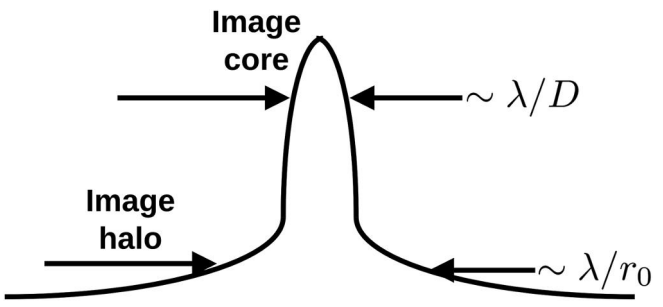

FIGURE 2.2
The point spread function through the atmosphere exhibits a diffraction-limited central core and a halo.

$$R = 1.22\left(\frac{\lambda}{D}\right)\frac{Q}{S}, \tag{2.66}$$

where S is the Strehl ratio from Eq. (2.64) and

$$Q = \sqrt{\frac{e^{-2\sigma^2_{\text{comp}}}}{1+\left(2.22\alpha_{\text{jit}}D/\lambda\right)^2} + \frac{1-e^{-\sigma^2_{\text{comp}}}}{1+\left(D/r_o\right)^2}}. \tag{2.67}$$

Turbulence is a random process that has temporal statistics as well as spatial statistics. From an adaptive optics point of view, these statistics can be important. If turbulence moves very slowly, then an adaptive optics system can view the disturbance as static. Correction would only be required in the spatial domain. If the turbulence is very fast, an adaptive optics system would only see the time-averaged disturbance. Fortunately, the actual turbulence is driven by winds and local eddies, which can be tracked by electronic means. A beam will pass through a different part of the air every time the wind moves it one beam diameter, $t \approx D/v_w$ (D is the beam diameter; v_w is the wind velocity). When a beam is much larger than the coherence length, the time of interest is the time τ it takes for the wind to transport the beam one coherence length (Zuev and Lukin 1987),

$$\tau \approx 0.53\left(\frac{r_0}{\upsilon_{\text{w}}}\right)\left(\frac{D}{r_0}\right)^{1/6} \tag{2.68}$$

When the wind velocity is high, the movement of eddies across beam edges becomes very fast. The result is high-frequency amplitude and phase fluctuations. The power frequency spectrum for higher-order wavefront variations $P_\phi(f)$ shows the $f^{-8}/^3$ power law dependence,

$$P_\phi\left(f\right)_{f\to\infty} = .0326k^2 f^{-8/3}\int_0^L C_n^2\left(z\right)\upsilon_{\text{w}}^{5/3}\left(z\right)dz. \tag{2.69}$$

An adaptive optics system with a closed-loop servo response should reject most of the phase fluctuations. Greenwood calculated the characteristic frequency f_{G} (often called the *Greenwood frequency*; Greenwood 1977).

$$f_{\text{G}} = 2.31\lambda^{-6/5}\left[\sec\beta\int_0^L C_n^2\left(z\right)\upsilon_{\text{w}}^{5/3}\left(z\right)dz\right]^{3/5}. \tag{2.70}$$

In the constant wind case, the Greenwood frequency can be approximated by

$$f_{\text{G}} = 0.43\frac{\upsilon_{\text{w}}}{r_{\text{o}}}. \tag{2.71}$$

For cases of interest, the Greenwood frequency of the atmosphere is in the range of tens to hundreds of Hertz. Beland and Krause-Polstorff (1992) present measurements that show how the Greenwood frequency can vary between sites. Mt. Haleakala in Maui, Hawaii, has an average Greenwood frequency of 20 Hz. For strong winds (jet stream, etc.) and ultraviolet wavelengths, the Greenwood frequency can reach 600 Hz.

In cases of modeling the turbulence with a finite wind, researchers use the "frozen flow hypothesis." This model assumes that the wind velocity carrying the turbulent eddies is much greater than the internal velocity of the eddies, which allows the turbulent layer to act like a "frozen" phase slab sliding across the aperture.

2.1.4.5 Phase Tears and Branch Points

Although the atmosphere is continuous, the phase of a monochromatic beam of light varies only in the region 0 to 2π radians. In the presence of strong turbulence, there can appear phase discontinuities in the phase field. The discontinuities can appear as irregular lines along the field where 2π or multiple 2π phase jumps occur along the line. They appear to be rips or *tears* in the 2D phase field. Also occurring are singular points known as *branch points* where the amplitude and phase are simply undefined in the mathematical sense (Fried 1998). For a purely monochromatic beam, these present little problem to the propagation. However, for a beam with any spectral spread or multiple wavelength beams, the 2π phase discontinuity is manifested as an absolute distance jump, or an optical path distance (OPD) jump, that affects the beams of light differently. The existence and manifestation of phase tears and branch cuts have been studied in detail (Fried 2001), but the major problem arises when one's adaptive optics system must decode the OPD across the discontinuities (Le Bigot and Wild 1999, Le Bigot et al. 1998) and determine the so-called *hidden phase* (Banakh and Falits 2003).

One method to unwrap the optical phase in the presence of discontinuities involves using a least-squares unwrapper of the field and isolating the branch points. By comparing the least-squares unwrapped phase with the non-least-squares components, one can find an unwrapped field which is a modulo-2π-equivalent to the original field (Venema and Schmidt 2008). The problem of branch points in adaptive optics is discussed in detail in the context of wavefront reconstruction in Section 9.3.10.

2.1.5 Turbulence Modulation Transfer Function

The distortion due to optics, turbulence, atmospheric absorption, and scattering were combined into a simple formulation by Lutomirski (1978). The modulation transfer function (MTF) of the system is the product of the MTF of turbulence and the MTF of optics with an attenuation by the atmospheric extinction ε,

$$M_{\text{sys}} = e^{-\varepsilon z} M_{\text{opt}} M_{\text{turb}}, \tag{2.72}$$

where z is the propagation path length. The MTF for a diffraction-limited circular aperture of diameter D is

$$M_{\text{opt}} = \frac{2}{\pi}\left\{\cos^{-1}\left(\frac{\xi F\lambda}{D}\right) - \left(\frac{\xi F\lambda}{D}\right)\left[1-\left(\frac{\xi F\lambda}{D}\right)^2\right]^{1/2}\right\}, \tag{2.73}$$

where F is the focal length of the optical system, ξ is spatial frequency, and the turbulence MTF (Goodman 1985) is

$$M_{\text{turb}} = 1, z \ll \left(0.4k^2 C_n^2 L_0^{5/3}\right)^{-1}, \tag{2.74}$$

$$M_{\text{turb}} = \exp\left(-\left(2.01\xi F\lambda / r_0\right)^{5/3}\right), z \gg \left(0.4k^2 C_n^2 L_0^{5/3}\right)^{-1}. \tag{2.75}$$

Belen'kii derived the MTF for turbulence (Belen'kii 1996) when a finite inner scale l_0 is assumed:

$$M_{\text{turb}}\left(l_0\right) = \exp\left[-\frac{\left[\left(\xi F\lambda\right)^2 + l_0^2\right]^{5/6} - l_0^{5/3}}{\left(r_o/2\right)^{5/3}}\right]. \tag{2.76}$$

2.1.6 Multiple Layers of Turbulence

While turbulence is continuous throughout the atmosphere, there are clearly some regions that have stronger turbulence and will greatly affect astronomical images. When phase conjugation is described later, it will be assumed that all the wavefront error is concentrated at a single horizontal layer. The optical conjugate of the adaptive optics sensors and correctors is placed at the position of this layer. The strongest turbulence is typically found at low altitude, often only a few meters above the ground and the telescope. The optical system is then designed to conjugate this layer of turbulence assuming that all the integrated turbulence above it is included in the ground layer, or, more simply, ignoring completely the turbulence at other altitudes.

Because atmospheric conditions change, the distance to the strongest turbulent layer may change, so the optical system conjugate pupils must be, in themselves, adaptable.

The next strongest turbulent layer is typically at the altitude of the jet stream, near 10 km. The jet stream turbulence is very high, despite the low density of air, because the large wind differences vary over only a few hundred meters. Strong eddies form. The jet stream appears in the C_n^2 models at an altitude specified by an average position. As any weather report tells us, the jet stream is highly variable in its strength, location, and duration. An adaptive optics system that has conjugate pupils at the jet stream must also be adaptable to the varying conditions. The statistics of the Strehl ratio can be determined by measurement of a few relevant turbulence parameters such as r_0, f_G, or σ_I^2 at various sites of interest (Tyler 2006).

The strength of turbulence being variable and continuous along a vertical path is not important for objects exactly on the optical axis. The combined integrated index of refraction is indifferent to the vertical position of the turbulence which caused it. On the other hand, for objects or optical paths that are off-axis, the total integrated index is strongly dependent geometrically on the altitude of the disturbance. To compensate for multiple layers of different turbulence, multiple pupil phase conjugates must be in the system, greatly complicating the design and construction. The system design consequences of multi-conjugate systems are discussed in detail in Chapter 4.

2.2 Marine Environments

2.2.1 Marine Layer

Navy and Coast Guard vessels are keenly aware of the unique atmospheric conditions near the marine atmospheric surface layer. Atmospheric extinction losses are induced by scattering and absorption both by molecules and aerosols. Aerosol extinction can be a strong function of surface height under windy conditions. The near-surface atmosphere is vertically inhomogeneous with substantial refractive index and extinction gradients. Modeling of the atmosphere in the marine layer includes examining existing databases and developing new codes that are reliable (Gathman 1983).

To illustrate the complexity, one can limit an examination to aerosol extinction alone. In the marine layer, sea-salt particles and water vapor dominate. Particles less than 0.1 μm are created by condensation and nucleation of molecules and grow up to 2.5 μm by coagulation. Particles above 2.5 μm are generated by mechanical processes, such as wind blowing over soil or sand, and sea spray. In addition, aerosols can be hygroscopic or non-hygroscopic.

While the upper atmosphere turbulence is modeled using Kolmogorov statistics, the marine layer must take into account air temperature, relative humidity, sea surface temperature, cloud cover, wind direction and wind speed (Frederickson et al. 2006). In addition, latitude, longitude, time-of-day, and day-of-year are also major elements contributing to the analysis. Optical turbulence models (estimating C_n^2) for the surface layer are

based upon Monin–Obukhov similarity theory (Monin and Obukhov 1954). This theory assumes that the mean flow and horizontally homogeneous surface layers have four independent variables: height above the surface, friction velocity, surface kinematic heat flux, and buoyancy variable (Frederickson et al. 2000). While the Monin–Obukhov theory holds under most conditions, it breaks down when the temperature difference between the air and the sea surface is large and the wind speed is low. This leads to a thermally stratified atmosphere and vertical mixing is cut off. Scaling law models have been developed for large-scale predictions (Venet 2007), but detailed models have yet to address the short time, variable local conditions (Doss-Hammel et al. 2004).

2.2.2 Underwater Effects

Transmitting information on radiative waves through water has been a problem for millennia. Acoustic signals (SONAR, etc.) and radio waves show a marked attenuation with frequency. For electromagnetic waves from ultraviolet (UV) to the far infrared (IR), there is a transmission window at blue-green colors. Open ocean has a transmission window more to the blue of the spectrum, and shore or harbor waters have a transmission window at the green color (Hardy et al. 2019). Because scattering is higher than absorption in water, collimated lasers with Mbps-class information can propagate about 20 "extinction lengths" which are each defined as $1/e$ attenuation. In clear water, this translates to about 800 m. In less clear ocean, the length is about 100 m, and in turbid water, the propagation length is only tens of meters.

With the addition of pointing and tracking systems, rudimentary adaptive optics have been demonstrated in experiments using pools and underwater devices. A laser beam was sufficiently compensated after propagating through a scattering suspension of polystyrene microspheres in distilled water with the assistance of an adaptive optics system consisting of a bimorph deformable mirror and a Shack–Hartmann wavefront sensor (Galaktionov et al. 2017).

Restaino et al. (2014) carried out a test program in a 10 m-long water tank with a 19-actuator deformable mirror, a 144 subaperture wavefront sensor, and a 30 Hz control system. The figure-of-merit was the shape of the point spread function. The conclusion was that the system successfully compensated aberrations as long as the turbulence, particulate scattering, and absorption did not overwhelm the correction system.

Other experiments have been carried out recently. Hardy et al. carried out a test program where two underwater vehicles were required to coarsely align, then acquire, track, and point such that a 1 Gbps signal could be transmitted (Hardy et al. 2019). The experiment used a 20 m pool and extrapolated data to several hundred meters in clear ocean water. There was no adaptive optics system. From a strictly acquisition point of view, undersea location determination is limited. Inertial navigation systems are only accurate to 0.001 times the distance travelled, i.e., 10 m uncertainty for every 10 km of travel. Inclinometers provide much better depth certainty and tilt information (Williams et al. 2017; Hamilton et al. 2017). A theoretical study by Baykal indicates that adaptive optics will only be marginally effective (Baykal 2020).

What is encouraging is that, even though there are no master oscillator power amplifier (MOPA) lasers in the blue-green and IR techniques cannot be used, visible wavelength cameras are extremely good.

2.3 Thermal Blooming

Whereas atmospheric turbulence occurs whenever any amount of thermal agitation is present, thermal blooming occurs whenever the atmosphere absorbs enough energy from a beam to alter the local index of refraction. The resultant self-induced distortion is called *thermal blooming*. When continuous-wave laser beams were sufficiently developed into the kilowatt range, expansion of the beam size ("blooming") was observed (Smith 1977). Both steady-state and transient thermal blooming have been investigated.

When the absorbed laser power is balanced by thermal conduction, natural convection, or forced convection due to wind or beam motion, the thermally bloomed beam *turns into the wind*. For a zero-wind case, thermal blooming, often called *thermal defocusing* (Gebhardt and Smith 1969; Richerzhagen et al. 1996), occurs, since the lowest index of refraction occurs near the center of the beam, where the beam intensity was the highest. This atmospheric negative lens causes the beam to defocus.

An important case occurs when the wind, or an artificial wind caused by beam slewing, causes the beam to take on a characteristic crescent-shaped pattern. See Figure 2.3. Light is refracted more toward the dense portion of the air, the least intense region. The kinetics of the atmospheric effects are discussed in detail by a number of authors (Gebhardt 1976; Smith 1977). The effects of convection and conduction are straightforward. The interaction with a compensated high-power beam is a subject of much study. The combined effects of blooming and turbulence are even more complex and are being investigated with detailed wave optics and fluid models. See Section 2.2.2.

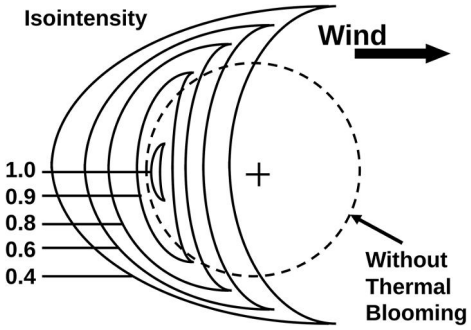

FIGURE 2.3
Computer-generated isointensity plot of a high-power beam in the atmosphere shows the effects of thermal blooming. The shape of the beam takes on a characteristic crescent shape bending into the wind.

2.3.1 Blooming Strength and Critical Power

As a beam propagates through the atmosphere, the power is absorbed. For a uniformly absorbing atmosphere, the intensity changes according to Smith (1977)

$$I(r,L) = I(r,0)\exp\left(-\alpha L - N_B e^{-r^2/a^2}\right), \tag{2.77}$$

where the linear absorption coefficient is α, the beam radius is a, the propagation path is L, and the dimensionless parameter N_B is the magnitude of the blooming distortion (Bradley and Herrmann 1971).

For a case where there is wind present (or the beam is being slewed through the air), the distortion is not symmetric. The blooming strength becomes a function of wind velocity v_w, the beam radius a, the specific heat at constant pressure C_p, and the density of the air ρ. The Bradley–Herrmann blooming distortion number N_B is (Bradley and Herrmann 1971)

$$N_B = \frac{-4\sqrt{2}P(dn/dT)k\alpha L}{\rho C_p v_w 2a}, \tag{2.78}$$

where P is the beam power. The mass density is ρ and dn/dT is the change of index of refraction with temperature. The index of refraction of air at temperature T is (Penndorf 1957)

$$(n_0)_{air}(T) = 1 + (n_{15} - 1)\left(\frac{1.0549}{1 + 0.00366T}\right), \tag{2.79}$$

where T is in degrees Celsius. The index for air at 15°C is given by Edlen (1966),

$$(n_{15} - 1)\times 10^8 = 8342.1 + \frac{2406030}{130 - \lambda^{-2}} + \frac{15996}{38.9 - \lambda^{-2}}, \tag{2.80}$$

where λ is in microns. For example, $(n_0)_{air} = 1.000276$ at 15°C and λ = 0.6328 μm. For a focused beam, the radius a is replaced by an effective radius a_{eff} over the propagation length L,

$$a_{eff} = \left(\frac{La\lambda}{\pi}\right)^{1/3}. \tag{2.81}$$

Where the absorption coefficient is not constant along the propagation path and there is a physical slewing of the beam, Karr presents a more general expression for the distortion number (Karr 1987a):

$$N_D = 8\sqrt{2}\pi\left[\frac{\partial n/\partial T}{\rho C_p}\right]\frac{P}{2\lambda a}\int_0^\infty \frac{\alpha(z)}{v_w(z) + \dot{\theta} z} \exp\left\{-\int_0^z \left[\alpha(z') + \alpha_s(z')\right]dz'\right\}dz, \tag{2.82}$$

where $\alpha(z)$ is the path-dependent absorption coefficient and $\alpha_s(z)$ is the path-dependent scattering coefficient. The angular slew rate is θ.

The effect of thermal blooming on the propagation of the beam is the physical spreading of the energy and reduction of the energy on-axis. The blooming strength (or distortion number) is roughly the number of *radians* of wavefront error induced by the blooming. Experiments and wave optics calculations (Smith 1977) verify that this reduction, which is represented by the Strehl ratio S, empirically takes the form

$$S = \frac{1}{1 + K' N_B^m}, \tag{2.83}$$

where $K' = 0.0625$ and $m = 2$ for an infinite Gaussian beam, and $K' = 0.01$ and $m = 1.2$ for a uniform beam. If the intensity on-axis is modified by the Strehl ratio for thermal blooming distortion, Eq. (2.83), the intensity becomes

$$I_{Bloom} \propto \frac{Pa^2 S}{\lambda^2 L^2} \propto \frac{P}{1 + \kappa P^m}. \tag{2.84}$$

The constant $\kappa = K'(N_B/P)^m$ absorbs the parameters affecting blooming strength. Note that the intensity of the bloomed beam is non-linear in power P. For any finite value of κ (that is, any thermal blooming), an increase of power will only result in an increase in intensity until the power reaches a certain value. See Figure 2.4. This *critical power* P_{crit} is related to the blooming strength by

$$P_{crit} = \left(\frac{-1}{\kappa(1-m)}\right)^{1/m}. \tag{2.85}$$

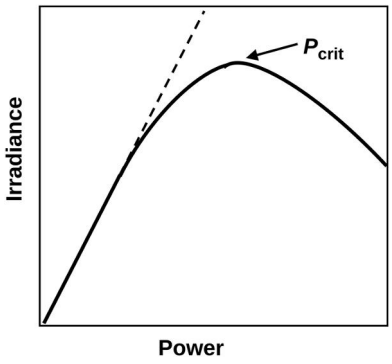

FIGURE 2.4
Power increases result in increased intensity until the critical power is reached.

Without the application of adaptive optics, the distortions due to thermal blooming can be reduced in a number of ways. An increase in wind, or slew rate, reduces the blooming strength and increases the Strehl ratio directly (Konyaev and Lukin 1985). Although this is simple, it is often impractical, since the wind on a specific day in a specific place is unpredictable and the beam propagation scenario might not allow slewing the beam in an arbitrary manner.

If the power in the beam is applied for a period of time too short for the atmosphere to develop the thermal lens, blooming could be avoided. This principle can be employed by using pulsed lasers. Blooming will occur whenever density changes are propagated transverse to the beam by acoustic waves. By properly controlling the pulse length and the pulse repetition frequency (PRF), thermal blooming often can be reduced.

Pulsing a laser to avoid thermal blooming is not possible in all cases. Reducing the pulse length without simultaneously reducing power can cause other phenomena worse than blooming. Gas breakdown is the ionization of the air path that results from intense radiation in the gas. The result is often total laser beam absorption (Smith and Meyerand 1976).

When the pulse is long compared to the time a/v_s, where a is the beam radius and v_s is the velocity of sound, the *long-time thermal blooming* strength is given by

$$N_{\mathrm{lt}} = \frac{2(-dn/dT)\alpha I_0 t L^2}{\rho C_p a^2}, \tag{2.86}$$

where $\alpha L \ll 1$. Although the expression for long-time blooming strength shows a linear increase in time, there is a limiting factor. It was shown by Ulrich that there is a time when the on-axis intensity drops, because of thermal blooming, such that the blooming can no longer continue (Ulrich 1973). This point is achieved when the intensity is about one-tenth of its initial value. The time it takes for this saturation to occur is

$$t_{\mathrm{s}} = 0.04\left[\frac{8n_0\rho C_p}{(dn/dT)v_{\mathrm{s}}^2\alpha}\right]^{-1/2}\frac{\lambda^2 L}{a\sqrt{E}}, \tag{2.87}$$

where E is the pulse energy.

When the pulse length t_p is short compared with the lens forming hydrodynamic time t_{h}, the pulse thermal blooming strength N_{p} becomes (Smith 1977)

$$N_{\mathrm{p}} = \frac{8(-dn/dT)v_s^2 E\alpha L^2 t^3}{3\pi\rho C_p t_p a^6}, \tag{2.88}$$

where E is the pulse energy and the quantity αL is assumed to be small, i.e., $\alpha L \ll 1$. Equation (2.88) shows a cubic dependence on the time t after the onset of the pulse. Short-term transient thermal blooming is often called *t-cubed* blooming.

When a number of pulses are strung together, a condition called *multiple-pulse thermal blooming* occurs (O'Neill et al. 1975; Dunphy and Smith 1977). The parameters that describe the effects of blooming for this case are the blooming strength N_{mp} and the number of pulses per flow time NP, so that

$$N_{\mathrm{mp}} = \frac{2(-dn/dT)\alpha E L^2}{\rho C_p \pi a^4} \tag{2.89}$$

and

$$\mathrm{NP} = [\mathrm{PRF}]2a/v. \tag{2.90}$$

These basic expressions can be used in Eq. (2.83) to evaluate the effects of thermal blooming in the atmosphere. The application of adaptive optics to correct for thermal blooming phenomena will require a knowledge of the temporal as well as the spatial effects of blooming distortion (Pearson 1978). Other than the pulse and multiple-pulse effects described earlier, the overall frequency for atmospheric response f_b under blooming conditions (Nichols and Duneman 1982) is $f_b = v/(2a)$, where the wind velocity v can be caused by slewing or actual wind or a combination of both, and the beam radius a must be evaluated for each case. For a collimated beam, the aperture radius is a suitable approximation, and for a focused beam, the beam radius at $0.75L$ can be used for an approximation (Akkapeddi et al. 1981).

The historical summary given here and the simple expressions can be used for order-of-magnitude analyses. The non-linear behavior of the high-power beam in the atmosphere requires both extensive analysis and supporting experimentation (Lukin et al. 1994).

2.3.2 Turbulence, Jitter, and Thermal Blooming

The combined and often closely coupled effects of atmospheric turbulence, beam jitter, and thermal blooming are the topics of much investigation. Although we can show simple rules for estimating the three effects independently, the response of the atmosphere for the three competing phenomena is extremely complicated. For example, thermal blooming strength is a function of the beam radius a and wind velocity v. Beam jitter has the effect of a false wind, which would alter the degree of blooming. If the jitter is high-frequency, the beam intensity is smeared before the atmosphere has time to respond, giving the appearance of a larger effective radius. For intermediate frequencies, the beam's own jitter might follow the *bending into the wind* phenomenon or even counteract it. All of these might be occurring simultaneously (Gebhardt 1976).

Atmospheric turbulence has the effect of inducing dynamic tilt, which has the same characteristics as complicated beam jitter. Turbulence can also induce beam spreading because of its high spatial frequency content. This can increase the apparent beam radius, thus reducing the local intensity and the magnitude of the turbulence.

Another phenomenon occurs when adaptive optics is used to try to correct for the combined effects of thermal blooming and turbulence. The errors in the wavefront that occur in the propagation path tend to defocus or spread the beam. Compensation for the negative atmospheric lens is the application of the conjugate positive lens in the beam-transmitting optics. This effectively increases the intensity which, on one hand, is the desired effect, but, on the other hand, increases the strength of thermal blooming and forces the atmosphere into a more *negative* lens. This instability (Karr 1989) along the beam path has been observed in a number of experiments (Johnson and Primmerman 1989), and it was observed as small spatial scale (high spatial frequency) phase-conjugate instabilities (PCIs) (Johnson and Schonfeld 1991; Schonfeld 1992; Schonfeld 1989; Schonfeld 1990; Barnard 1989).

Many investigators have studied the complexities of turbulence and thermal blooming (Enguehard and Hatfield 1991; Gebhardt 1990; Norton et al. 1985; Wallace et al. 1974). They have investigated simple scaling relationships (Breaux et al. 1979), detailed physics of the distortion process (Briggs 1987; Hatfield 1988), and the application of adaptive optics (Morris 1988; Schonfeld 1992; Wallace 1988). They have used four-dimensional (4D) wave-optics models (Crawford et al. 1990; Morris 1988; Myers 1988) supported by experimental verification (Johnson and Murphy 1988; Primmerman and Fouche 1976). The limitations of adaptive optics compensation for thermal blooming, especially the requirement for full-field[5] compensation, generate the most interest (Karr 1987b). The complex interaction of diffraction, turbulence, and the high-power thermal blooming phenomenon at small scales limits the applicability of phase-only correction (Myers and Allen 1985). A large study of the combined effects of thermal blooming and turbulence using theory and wave-optics simulations was conducted by Spencer (2020a, 2020b). In the study, steady-state simulations were expanded to include time-dependent simulations. The turbulence–thermal blooming interaction was characterized by the log-amplitude variance and the thermal blooming distortion number. In the simulations ($\lambda = 1.0\,\mu$m, propagation distance is 5 km, atmospheric absorption $\alpha = 5 \times 10^{-6}\,\mathrm{m}^{-1}$, atmospheric scattering $\sigma_{\mathrm{scat}} = 5 \times 10^{-5}\,\mathrm{m}^{-1}$, and the wind is 5 m/s), the power was varied up to 250 kW and the transverse wind is assumed to have cleared the source plane such that a fully formed crescent is present. The spherical wave Rytov number increased from 0 to 1. Spencer noted that the distortion number reached saturation at $\sigma^2_{\mathrm{Rytov}} = 0.6$. This supports the notion that the interaction of the two phenomena makes it very difficult for phase-only adaptive optics to provide correction in the strong turbulence regime.

2.4 Aero-Optics

Imaging from a moving aircraft presents some problems. An aircraft moving through ambient air produces turbulence in the boundary layer directly in front of the objective of an aerial camera (Whiteley and Goorskey 2013) or a beam director (Wang et al. 2012). The wavefront aberration is caused by density fluctuations of the mixing layer, dominated by the dynamic characteristics of vortices inside the flow field. In the flow field inside the boundary of the supersonic mixing layer, vortices begin rolling up, merging, and pairing as the flow continues until the vortices cease to grow due to viscous dissipation (Caulfield and Peltier 2000; Guo and Liu 2017).

An expression for the boundary layer MTF (Lei and Tiziani 1993) follows an exponential similar to the one for Kolmogorov atmospheric turbulence (Eq. (2.75)):

$$M_{\mathrm{BL}} = \exp\left[-3.44\left(\lambda\xi F / r_{\mathrm{BL}}\right)^{5/3}\right] \tag{2.91}$$

where F is the focal length of the objective and r_{BL} is the correlation length for the boundary layer. The correlation length, analogous to Fried's parameter,

depends upon the flight speed, the aircraft surface form, and the air density. At visible wavelengths, r_{BL}varies from 5 to 20 mm (Kelsall and D'Amato 1977). Many of the aero-optical wavefronts can be predicted, identified, and controlled (Hinnen et al. 2008) without the brute force application of very high-speed, very high-resolution adaptive optics (Tesch et al. 2011). There are many mechanical methods used to reduce the boundary-layer aberration (Gordeyev et al. 2007) and to characterize the remaining aberration around cylindrical turrets (Gordeyev et al. 2011), flat-windowed turrets (Porter et al. 2011), and conformal turrets (Lu et al. 2020). George et al. have made measurements of the density field and wavefront distortions in high-speed flows (George et al. 2017), and they mention mechanical methods of flow control including fairing design, boundary layer suction, blowing, window placement, and directing the pilot on aircraft maneuvering. A lot of airborne data has been collected over the past few decades and this has been greatly enhanced by the recently completed Airborne Aero-optics Laboratory (AAOL) program (Jumper 2013; Kalensky et al. 2020). While it is generally desirable to have an adequate prediction capability for the aero-optical wavefront statistics, sometimes the statistics are not stationary due to flow topology. In this case, adaptive finite impulse response prediction filters with multi-input-multi-output filters can be employed to improve the control process (Tesch and Gibson 2011).

In recent years, the use of commercial drones has increased dramatically for environmental imaging, mapping, law enforcement, disaster relief, and the delivery of consumer products. Drones have been proposed and shown to be a platform for adaptive optics artificial guide stars (Basden 2018) and as beacons for free-space optical links (Dabiri et al. 2018). Measurements of optical turbulence surrounding the payloads of professional drones have been made showing that the greater source of turbulence is through the propeller airflow and not from the increase temperature of the propulsion system motor (Garcia et al. 2019).

2.5 Non-atmospheric Sources

Although turbulence is the major source of concern for astronomers using adaptive optics, and thermal blooming is a limiting factor for high-power beam propagation, adaptive optics is often used to correct for aberrations that enter a system in other ways.

2.5.1 Optical Misalignments and Jitter

The inadvertent introduction of tilt or dynamic tilt (jitter) into a system can reduce its performance. In some cases, this performance is only slightly reduced, as in the case of an imaging system where tilt simply moves the position of the image without affecting its quality. These problems often can be accepted. On the other hand, slight misalignments in optics inside a laser resonator can greatly reduce its performance and, in many cases, eliminate the possibility of lasing altogether. Errors in setup or mounting, external error sources such as mechanical (acoustic) vibration (Parmentier and Greenberg 1973), and thermal problems can all contribute negatively to a poorly performing system. Defects such as tilt (Yura and Hanson 1987) and decenter (McKinley et al. 1988) can be studied using ABCD matrices. See Section 3.1.

Adaptive optics, in a tilt correction mode, can be used to correct for these errors (Bruno and Wirth 1996). The problem of maintaining alignment or beam pointing has evolved from early electromechanical telescope and radar control systems to the present-day scanning and control systems that operate in the megahertz range. This book treats the control of tilt as the fundamental low-order wavefront aberration. A tilted component in a system can induce other higher-order aberrations. A tilted mirror or beamsplitter intercepting a converging beam will introduce the next higher-order aberration: astigmatism. A tilted spherical optical element in a telescope will introduce coma. Severe tilts in complicated optical configurations can introduce many high-order aberrations.

The elimination of tilt, or pointing error, is the first goal of adaptive optics. Control of fast tilt (jitter) or slow tilt (drift) is often all that is needed in an adaptive optics system.

2.5.2 Platform Motion

Prior to building an adaptive optics system to remove tilt or higher order aberrations, it is necessary to identify the aberration sources and remove them before requiring active or adaptive optics. The *platform* as we use it here consists of all the mechanical structure which holds the optics, associated sensors, and supporting hardware. In a typical telescope, imaging system, or laser projection system (as if there were such a thing), there will be sources of vibrations that travel through the mechanics and end up affecting the optical beam. These are highly varied. The identity of each usually consists of a transfer function or spectrum which indicates the amplitude of the vibrational source and the frequency of the source. Although the transfer functions can be quite complicated, showing multiple resonant peaks and antiresonances, they often are identified by proper

analysis during the design phase (Skidmore et al. 2010). Sources are either structural or acoustic and include building motion, building services (air conditioning and heating, elevators), electric motors (vacuum pumps, air compressors, transformers, and heavy machinery), air currents, acoustics, and foot traffic (Thorlabs 2020). Any deflection of the platform results in a change of relative position of the optical components.

After identifying the sources for a particular location, one can begin to mechanically reduce, isolate, or damp out the sources from the operational optical elements. Because many of the optical systems are built onto a 2D table, one can concentrate on bending or torsional modes. However, for geometrically complex optical systems, the transfer functions are 4D (three space axes and one frequency axis; Bronowicki 2006). Optical mounting and vibration isolation can be an electromechanical function, but may not result in enough reduction in the induced aberrations. The solution, then, is to add adaptive optics to further optically reduce the aberrations.

2.5.3 Large Optics: Segmenting and Phasing

Striving for ever-better means of collecting faint starlight or focusing laser beams to extremely small spots requires larger and larger optics. The manufacture of larger optical elements presents new problems for investigators and new challenges for adaptive optics. Extending technology from the Keck telescopes' alignment and phasing of the thirty-six 1.8-m segments into a 10-m diameter aperture has evolved into phasing 8-m segments into behemoths as large as the 42-m diameter European Extremely Large Telescope (ELT).

In large optics, small local deviations from perfect curvature are manifested as high spatial frequency aberrations that require correction. Casting large mirrors requires extreme care (Dunkle 1989). Even after a perfect casting, figuring, and coating, large optics (mostly mirrors) are placed in relatively hostile environments. One operation that is often aberration-producing is *placing* the mirror. The effects of gravity on massive objects (thousands of pounds and many feet across) cause the mirror to sag and deform (Antsiferov et al. 1988). If this mirror is placed on a movable mount, the direction of the deformation will change with the motion. The sag cannot be calibrated out, nor can it be polished away, since it will change during operation.

The gravitational sag of a large mirror can be calculated. The deflection ω_z, normal to the surface, of a circular plate of radius a that is lying horizontally is (Barnes 1979)

$$\omega_z = \frac{3\rho g\left(1-\nu^2\right)}{16Eh^2}\left[\frac{5+\nu}{1+\nu}a^4 - \frac{2\left(3+\nu\right)}{1+\nu}a^2r^2 + r^4\right], \quad (2.92)$$

where ρ is the material density, E is Young's modulus, ν is Poisson's ratio, h is the disk thickness, and g is gravitational acceleration. To first order, the aberrations for a mirror or large lens under the influence of gravity are a constant piston, a sphere term proportional to r^2, and a spherical aberration proportional to r^4. For a 2.4-m fused-silica mirror (Hubble Space Telescope), for instance, there can be a total deflection of 3 μm and a spherical aberration component of 0.7 μm on the ground during polishing that would not be present in orbit.

Adaptive optics can be used to correct for gravity effects and other large effects. Large mirrors present an interesting problem, and they also present interesting solutions. Large mirrors can be self-correcting; that is, they can contain the necessary physical elements to adjust their figure and correct aberrations in the other parts of the system as well (Large 1985).

Large optics can be subdivided into smaller segments. These segments then can be co-aligned or *phased* to act as a larger mirror. The aberrations induced in large mirrors and the systems that are used to phase segments have been under study for a number of years (Fender 1984; Sintsov and Zapryagaev 1975). For the purpose of the study of adaptive optics, aberrations that are due to gravity (low spatial frequency) and aberrations due to local defects or deformation (high spatial frequency) can be treated in the same manner as similar aberrations on smaller optics, with only the magnification of spatial frequency taken into account.

Large mirror jitter is the same as small mirror jitter in terms of temporal frequency.[6] The spatial frequency of the aberration is usually expressed in *cycles/diameter*. Thus, a defect in a 1-m large mirror that is 1 cm across has a much higher spatial frequency content (100 cyc/diam) than the same defect in a 20-cm mirror (20 cyc/diam).

The aberrations that occur at mismatched edges have spatial frequency content that is very high. These almost-infinite spatial frequencies must be treated separately. The correction over the edges (which is usually a very small portion of the area of the aperture) can be simply disregarded, or the light that impinges on these edges can be masked out. The appearance of a sharp edge in a wavefront can disrupt an adaptive optics system. If the wavefront is measured at two points on either side of the edge and it senses a difference in phase, the control system may perceive this to be a local tilt, whereas it is actually a piston error on one or both sides of the edge. If the error is greater than one wavelength, the adaptive optics system may become unstable. These difficulties are addressed in Chapter 9.

2.5.4 Thermally Induced Distortions of Optics

When a beam of light strikes an optical surface, some of the energy is absorbed. For low-power beams, imaging

systems, and transparent glasses, the absorption is rarely noticed. High-power beams (>1.0 kW/cm^2) can induce a temperature rise in the optics that is proportional to the absorbed incident intensity (Hello and Vinet 1990). Passive or uncooled mirrors and actively cooled mirrors exhibit distortions under high-power illumination (Sheldon et al. 1982).

For a mirror modeled as a flat plate, the strain ε induced is proportional to the temperature rise T,

$$\varepsilon = \alpha T, \tag{2.93}$$

where α is the coefficient of linear thermal expansion of the material. For a plate in the x–y plane, the in-plane stress σ_{xx} and σ_{yy} is given by

$$\sigma_{xx} = \frac{E}{1-\nu^2}\left[\left(\varepsilon_{xx} + \nu\varepsilon_{yy}\right) - \alpha\left(1+\nu\right)T\right] \tag{2.94}$$

$$\sigma_{yy} = \frac{E}{1-\nu^2}\left[\left(\varepsilon_{yy} + \nu\varepsilon_{xx}\right) - \alpha\left(1+\nu\right)T\right] \tag{2.95}$$

where the strain tensors are ε, Young's modulus is E, and ν is Poisson's ratio. Myers and Allen (1985) show how the in-plane stresses result in the strain

$$\varepsilon_{xx} = \frac{\nu}{E}\left(\sigma_{xx} + \sigma_{yy}\right) + \alpha T. \tag{2.96}$$

With a uniform incident intensity, a flat circular plate of thickness h shows an out-of-plane displacement $\omega(r, \theta)$,

$$\omega\left(r,\theta\right) = \left[C_0 + C_2 r^2\right] + \left[C_1 r + C_3 r^3\right]\exp\left(i\theta\right). \tag{2.97}$$

This is a solution of Poisson's equation $\nabla^2\omega = -m_t$, where m_t is the thermal moment density,

$$m_t = \frac{-\alpha E}{1-\nu}\int_{-h/2}^{h/2} T\left(x,y,z\right) z\,dz. \tag{2.98}$$

The constants C_i are computed from the specific boundary conditions. For a plate uniformly constrained at the edges (i.e., $C_1 = C_3 = 0$), the distortion is a constant plus a function of the radial distance from the center. Thus, if uniform power P is absorbed, the temperature profile results in stresses that cause a characteristic paraboloidal bowing of the plate,

$$\omega_z\left(r\right) = P\left(C_0' + C_2' r^2\right). \tag{2.99}$$

Since *bowing distortion* is proportional to the total absorbed power, it is also called *power-induced distortion*.

If the intensity profile $I(x, y)$ on the plate is not uniform, regions of local temperature increases will result in another form of distortion. The displacement in the out-of-plane direction ω_z can be found by solving

$$\nabla^4\omega_z = \frac{P_0}{D} \tag{2.100}$$

with the appropriate boundary conditions. P_0 is the load density, and the plate modulus is given by

$$D = \frac{Eh^3}{12\left(1-\nu^2\right)}. \tag{2.101}$$

The result for a circular plate is the integral of the strain tensor ε_{zz} (Roark 1965),

$$\omega_z\left(x,y\right) = \int_{-h/2}^{h/2}\varepsilon_{zz}dz = \int_{-h/2}^{h/2}\alpha T\left(x,y,z\right)dz. \tag{2.102}$$

Since the temperature *profile* is nearly proportional to the intensity *profile*, the resultant out-of-plane thermal growth is proportional to the absorbed intensity distribution. Where α_{abs} is the absorption coefficient,

$$\omega_z\left(x,y\right) = \xi\alpha_{\text{abs}} I\left(x,y\right). \tag{2.103}$$

The proportionality symbol ξ used in much of the literature resembles a worm and has been called the *worm factor*. This constant describes the amount of thermal growth due to local intensity variations. It is a function of material parameters, cooling system design, and coolant flowrates. Measured worm factors range from 5 to 100 Å/W/cm^2.

Since this form of thermal distortion produces a map of the flux or intensity pattern, it is called *thermal mapping distortion, flux-induced distortion*, or *intensity mapping distortion*. It is proportional to the intensity *distribution* rather than the total power. Thermal mapping has an interesting property. Since the optical surface responds to the intensity of the impinging beam by distorting along the axis of the beam, reflections from that surface are advanced or retarded by the surface map. Thermal mapping translates intensity information from a beam into the phase (wavefront). A non-uniformity in the intensity results in reduction of Strehl ratio and more severe beam propagation limitations.

The thermomechanical response of a plate cannot always follow the intensity pattern exactly. Thermal diffusion tends to smooth the temperature distribution, and the mechanical response of the plate tends to smooth it more. A more exact treatment of thermal mapping would show those effects as convolutions (represented by *) among the response of the material to thermal loads R_T, response of the material to mechanical

stress R_M, and response to the absorbed intensity forcing function $\alpha I(x, y, t)$,

$$\omega_z(t) = R_T(t) * R_M(t) * \theta\xi I(x,y,t). \tag{2.104}$$

These responses can often be calculated numerically for different conditions. Both steady-state and transient solutions are often needed. For instance, a continuous beam impinging on a mirror with active cooling reaches steady-state distortion in a few seconds or less. The temperature profile, used in Eq. (2.98), for the surface of a mirror when it is illuminated by a pulsed beam is a function of the thermal diffusivity ζ and thermal conductivity k_t,

$$T(x,y,\tau) = T_0(x,y,0) + \frac{I(x,y)}{k_t}\sqrt{\frac{4\zeta\tau}{\pi}}, \tag{2.105}$$

where τ is the pulse length and T_0 is the surface temperature before the pulse (Carslaw and Jaeger 1980). Accurate prediction of the thermal deformations is crucial to the beam control of high power lasers such as those used in optical lithography machines (Haber et al. 2013).

2.5.5 Gravity Sag

Large precision optics have to be manufactured with surface figures at the subnanometer level. Telescopes with adaptive optics systems in the 21st century can have internal optics, secondary mirrors, turning flats, off-axis paraboloids, and deformable mirrors near 10 m in diameter (McGregor et al. 2012). Deformations from optical mounting stresses, annealing, and deposited coatings must be incorporated into the optical design (Robinson 2000). Included must be active compensation for the optics as they are moved in 3D space to aim the telescope (Burke and Griesmann 2012). Optics also distort from their own weight and must maintain figure as the optical elements move while the gravitational vector changes from out-of-plane to in-plane and everything in between (Gu et al. 2014).

Simple formulas from Roark would show that edge mounted optics exhibit large gravity sag throughout the motion (Young 1989). Other mounting techniques must be employed to reduce gravity sag prior to committing to the cost of manufacturing (Mehta 1983). While finite element modeling (FEM) is the principal tool for gross design, it also can be applied to the microlevel to include coating stresses and mounting geometries (Gensemer and Gross 2015).

2.5.6 Manufacturing and Micro-errors

Manufacturing of optics has always been technology-limited. The choice of glass or metal substrates has been dependent upon their ultimate performance. In 1730 Isaac Newton wrote (Newton 1979):

> But because Metal is more difficult to polish than Glass, and is afterwards very apt to be spoiled by tarnishing, and reflects not so much Light as Glass quick-silver'd over does: I would propound to use instead of the Metal, a Glass ground concave on the foreside, and as much convex on the back-side, and quick-silver'd over on the convex side. The Glass must be everywhere of the same thickness exactly.
>
> **(Prop. VII, Theor. VI)**

The manufacturing technology of the optical surfaces limits the overall performance of the optical system. With adaptive optics technology, many of the defects that arise from manufacturing can be corrected.

The materials chosen for refractive or reflective optics may have surface errors that figuring and polishing cannot remove. Generally, they are of such low magnitude or high spatial frequency that adaptive optics is useless against them. Occasionally a large systematic error may occur that produces a low-order aberration. A variation in polishing pressure, for instance, may produce such a defect that can be removed with adaptive optics when the optical element is used. Usually, these static, low-order defects can be calibrated out or compensated for in system setup, without relying on closed-loop control.

Even after polishing and coating, there may be roughness in the surface (called *microstructure*) that would affect an adaptive optics system. The adverse effects of the aberrations are usually less than the increased scattering from the optical surface. Scattered light affects sensors and can result in reduced-control system performance. The microstructure of polished optical materials is expressed in root-mean-square (rms) roughness (Elson et al. 1979). The roughness can range from about 10 Å for fused quartz to over 100 Å for materials such as molybdenum and beryllium. The fraction of incident light scattered, the total integrated scatter (TIS), from a surface with rms roughness δ is

$$\text{TIS} \cong \left(\frac{4\pi\delta}{\lambda}\right)^2, \tag{2.106}$$

where λ is the incident light wavelength. For roughnesses approaching even small fractions of a wavelength, we can see that the negative effects of scattering can be severe. Large roughnesses are also known to be extremely sensitive to incident high powers. If light is multiply reflected in the hills and valleys of the surface, absorption and *microthermal mapping* can lead to stresses that result in permanent damage to a mirror surface.

Another characteristic of the microstructure or any other random surface property is its autocovariance