SPIE PRESS | Field Guide

Field Guide to

Adaptive Optics

Second Edition

Robert K. Tyson
Benjamin W. Frazier

Field Guide to

Adaptive Optics

Second Edition

Robert K. Tyson
Benjamin W. Frazier

SPIE Field Guides
Volume FG24

John E. Greivenkamp, Series Editor

SPIE
PRESS

Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Tyson, Robert K., 1948-
Field guide to adaptive optics / Robert K. Tyson, Benjamin W. Frazier. -- 2nd ed.
p. cm. -- (SPIE field guide series ; v. FG24)
Includes bibliographical references and index.
ISBN 978-0-8194-9017-9
1. Optics, Adaptive. 2. Optical detectors. 3. Optical measurements. I. Frazier, Benjamin W. (Benjamin West) II. Title.
TA1522.T93 2012
621.36'9--dc23

2011050389

Published by

SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1.360. 676.3290
Fax: +1.360.647.1445
Email: books@spie.org
Web: http://spie.org

Printed in the United States of America.
First printing

Introduction to the Series

Welcome to the *SPIE Field Guides*—a series of publications written directly for the practicing engineer or scientist. Many textbooks and professional reference books cover optical principles and techniques in depth. The aim of the *SPIE Field Guides* is to distill this information, providing readers with a handy desk or briefcase reference that provides basic, essential information about optical principles, techniques, or phenomena, including definitions and descriptions, key equations, illustrations, application examples, design considerations, and additional resources. A significant effort will be made to provide a consistent notation and style between volumes in the series.

Each *SPIE Field Guide* addresses a major field of optical science and technology. The concept of these *Field Guides* is a format-intensive presentation based on figures and equations supplemented by concise explanations. In most cases, this modular approach places a single topic on a page, and provides full coverage of that topic on that page. Highlights, insights and rules of thumb are displayed in sidebars to the main text. The appendices at the end of each *Field Guide* provide additional information such as related material outside the main scope of the volume, key mathematical relationships and alternative methods. While complete in their coverage, the concise presentation may not be appropriate for those new to the field.

The *SPIE Field Guides* are intended to be living documents. The modular page-based presentation format allows them to be easily updated and expanded. We are interested in your suggestions for new *Field Guide* topics as well as what material should be added to an individual volume to make these Field Guides more useful to you. Please contact us at **fieldguides@SPIE.org**.

John E. Greivenkamp, *Series Editor*
College of Optical Sciences
The University of Arizona

The Field Guide Series

Keep information at your fingertips with all of the titles in the Field Guide Series:

Adaptive Optics, *Second Edition*, Robert K. Tyson & Benjamin W. Frazier

Atmospheric Optics, Larry C. Andrews

Binoculars and Scopes, Paul R. Yoder, Jr. & Daniel Vukobratovich

Diffractive Optics, Yakov G. Soskind

Geometrical Optics, John E. Greivenkamp

Illumination, Angelo Arecchi, Tahar Messadi, & R. John Koshel

Image Processing, Khan M. Iftekharuddin & Abdul A. Awwal

Infrared Systems, Detectors, and FPAs, *Second Edition*, Arnold Daniels

Interferometric Optical Testing, Eric P. Goodwin & James C. Wyant

Laser Pulse Generation, Rüdiger Paschotta

Lasers, Rüdiger Paschotta

Microscopy, Tomasz Tkaczyk

Optical Fabrication, Ray Williamson

Optical Fiber Technology, Rüdiger Paschotta

Optical Lithography, Chris A. Mack

Optical Thin Films, Ronald R. Willey

Polarization, Edward Collett

Probability, Random Processes, and Random Data Analysis, Larry C. Andrews & Ronald L. Phillips

Radiometry, Barbara G. Grant

Special Functions for Engineers, Larry C. Andrews

Spectroscopy, David W. Ball

Visual and Ophthalmic Optics, Jim Schwiegerling

Wave Optics, Dan Smith

There have been a number of books and thousands of papers published with descriptions and mathematical expressions regarding adaptive optics. The material in this *Field Guide* is a summary of the methods for determining the requirements of an adaptive optics system, the performance of the system, and requirements for the components of the system. This second edition has a greatly expanded presentation of adaptive optics control system design and operation. Discussions of control models are accompanied by various recommendations for implementing the algorithms in hardware.

This book is not just another book on adaptive optics. There are already many fine volumes. This volume is intended for students, researchers, and practicing engineers who want a "go to" book when the calculation was "needed yesterday" (by a customer who won't be paying for it until the next fiscal year).

Many of the expressions are in the form of integrals. When that is the case, we show the results graphically for a variety of practical values. Some of the material in this volume duplicates similar expressions found in other volumes of the *Field Guide* series. We have attempted to remain consistent with symbols of the other volumes. In some cases, however, we chose different symbols because they are well known within the adaptive optics literature.

Descriptions of the operation of subsystems and components and specific engineering aspects remain in the citations of the Bibliography.

This *Field Guide* is dedicated to the late Horace Babcock, whose pioneering ideas created the field of adaptive optics.

Robert K. Tyson
University of North Carolina at Charlotte

Ben W. Frazier
AOA Xinetics,
Northrop Grumman Aerospace Systems

Table of Contents

Table of Contents

Table of Contents

Glossary

a	Width of segment gap
A	Structure constant at the surface
ADC	Analog-to-digital converter
A_{Infl}	Influence function amplitude
a_0	Piezoelectric constant
b	Size of mirror segment
B	Laser brightness
B_{Astro}	Astronomical brightness
$\mathbf{B}$	Poke matrix
$\mathbf{B}_{\mathrm{c}}$	Calibrated poke matrix
$\mathbf{B}_{\mathrm{r}}$	Reduced poke matrix
$\mathbf{B}_{\mathrm{s}}$	Smoothed poke matrix
c	Speed of light ($= 3 \times 10^8$ m/s)
c_{a}	Interactuator coupling
c_{nm}	Fourier–Legendre coefficient
C_n^2	Atmospheric turbulence structure constant
CCD	Charge-coupled device
d	Size of subaperture (in object space)
d	Separation of the membrane and the addressing electrode
d_0	Characteristic distance of a laser guide star
D	Aperture diameter
DAC	Digital-to-analog converter
$\mathbf{d}$	Vector of wavefront disturbances
$\mathbf{e}$	Vector of calculated wavefront errors
E	Pulse energy of laser
e_{n}	Read-noise in electrons per pixel
f	Focal length
f_{BW}	Closed-loop bandwidth
f_{c}	Crossover frequency
f_{G}	Greenwood frequency
F	Focal length of the system
F_{Rayleigh}	Return flux for Rayleigh guide star
F_{Sodium}	Return flux for sodium guide star
g_1	Loop gain
g_2	Leak gain
G	Gain
$\mathbf{G}$	Generalized system plant model

Glossary

h	Planck's constant ($= 6.626 \times 10^{-34}$ J · s)
h	Altitude
H_T	Height of the tropopause
H–V	Hufnagel–Valley
$I(r)$	Intensity distribution
I_{Aper}	Intensity at the circular aperture
I_d	Dark current
I_{Gap}	Diffracted energy from gaps
I_J	Johnson noise current
I_n	Noise current
I_p	Photocurrent
I_s	Shot noise current
I_0	On-axis intensity
$\mathbf{I}$	Identity matrix
J_1	Bessel function
k	Wavenumber
k	Sample time index
k_B	Boltzmann constant ($= 1.38 \times 10^{-23}$ J/K)
K	Aperture shape parameter for beam propagation
K_g	Increase in error at the null
$\mathbf{K}$	Controller model
l_0	Inner scale of turbulence
L	Propagation distance
L_T	Thickness of the tropopause
L_0	Outer scale of turbulence
m	Azimuthal index for Zernike polynomials
m	Number of wavefront measurements in influence matrix
m_v	Visual magnitude of a star
M	Magnification
M^2	Beam quality
$\mathbf{M}$	Modal feedback matrix
$\mathbf{M}_r$	Reduced modal feedback matrix
MTF	Modulation transfer function
n	Radial index for Zernike polynomials

Glossary

$\mathbf{n}$	Vector of measurement noise
n_a	Number of actuators
n_B	Number of detected background photo-electrons per subaperture
n_m	Number of measurements
n_p	Number of detected photoelectrons per subaperture
n_s	Number of slaved actuators
n_v	Number of active actuators
n_R	Rayleigh scattering density
N_{Act}	Number of actuators
N_D	Number of pixels in a subaperture
NEA	Noise equivalent angle
NEP	Noise equivalent power
N_p	Photon count
N_{Zern}	Number of Zernike modes
$\mathbf{O}$	Offload matrix
OPD	Optical path difference
p	Curvature sensor image plane offset
P	Optical power
P_n	Legendre polynomial of order n
$\mathbf{P}$	Error propagation covariance matrix
$\mathbf{P}_p$	Piston projection matrix
$\mathbf{P}_t$	Tilt projection matrix
$\mathbf{P}_w$	Waffle projection matrix
PSD	Position sensing detector
PSF	Point spread function
q	Electron charge ($= 1.6 \times 10^{-19}$ C)
$\mathbf{Q}_d$	Atmospheric disturbance covariance
$\mathbf{Q}_n$	Measurement noise covariance
r	Radial coordinate
r_a	Actuator pad radius
r_c	Interactuator spacing
r_m	Mirror radius
r_s	Radius of supporting ring
r_0	Coherence length of the atmosphere
RMS	Root mean square

Glossary

$\mathbf{r}$	Vector coordinate in the wavefront
R	Radius of circle
R	Photodiode responsivity
$\mathbf{R}$	Reconstructor
$\mathbf{R}_{\mathrm{r}}$	Reduced reconstructor
RSS	Root sum square
$\mathfrak{R}$	Zernike radial polynomial
s	Shear distance
s	Summing index in the Zernike radial polynomial
s	Laplace transform variable
s_{Act}	Distance between actuators
S	Strehl ratio
$\mathbf{S}$	Sensitivity function
$\mathbf{S}_{\mathrm{d}}$	Slope discrepancy matrix
$\mathbf{S}_{\mathrm{l}}$	Slave logic matrix
$S_{\mathrm{w/jit}}$	Strehl ratio including effects of jitter
SLC	Strategic laser communications
SNR	Signal-to-noise ratio
SVD	Singular value decomposition
t	Thickness of the bimorph
t_{d}	Latency
T	Matrix transpose operator
T	Transmission of the optics
$\mathbf{T}$	Complementary sensitivity function
T_{A}	Transmission of the atmosphere
T_{m}	Tension of the membrane
T_{s}	Sampling time
$\mathbf{u}$	Vector of actuator commands
v_{G}	Wind velocity at low altitude
v_{T}	Wind velocity at the tropopause
v_{w}	Wind velocity as a function of altitude
$v(z)_{\mathrm{Bufton}}$	Wind velocity of the Bufton model
V	Applied voltage
w	Gaussian beam radius
w_0	Gaussian beam waist
$W(x, y)$	Wavefront

Glossary

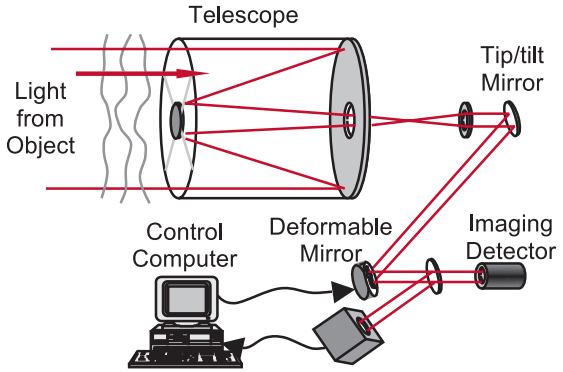
W	Wind velocity aloft
$\mathbf{W}$	Weighting matrix
WCE	Wavefront control experiment
WFS	Wavefront sensor
x, y	Cartesian coordinates
$\mathbf{y}$	Corrected output wavefront
z	Coordinate along propagation path
z	Altitude (if propagation path is vertical)
z	Z-transform variable
α_{jit}	Root-mean-square average jitter
$(\Delta\phi)^2$	Wavefront error variance (distance squared)
δ	Hartmann spot shift
$\Delta\lambda$	Spectral bandwidth
ε_0	Permittivity
ϕ	Phase
ϕ	Deformable mirror influence function (surface deflection)
ϕ	Wind direction relative to the telescope aperture
$\Phi(r)$	Phase
η	Detector efficiency
κ	Fitting constant
λ	Wavelength
λ_{LGS}	Laser wavelength
ν	Spatial frequency
ν_{G}	Wind velocity at low altitude
θ	Angular size of reference source
θ_{d}	Half-angle beam divergence
θ_0	Isoplanatic angle
ρ	Discrete-time equivalent latency
ρ_{Col}	Sodium column abundance
σ_{fitting}	Wavefront fitting error
σ_{Na}	Resonant backscatter cross section
σ_{R}	Rayleigh scattering cross section
σ_{Temp}	Temporal wavefront error
σ_{Tilt}	Wavefront tilt
σ_{WFS}	Wavefront sensor measurement error

Glossary

σ^2	Wavefront error variance (radians squared)
σ_I^2	Intensity variance
σ_n^2	Measurement error variance
σ_χ^2	Log-amplitude variance
σ_ϕ	Root-mean-square wavefront error over a subaperture
τ	Actuator rise time
Ω_N	Nyquist frequency
ζ	Zenith angle

Conventional Adaptive Optics System

A **conventional (linear) adaptive optics system**, whether it is used for imaging or for laser beam propagation, consists of three principal subsystems: a **wavefront sensor** to detect the optical disturbance, an active mirror or **deformable mirror** to correct for the optical disturbance, and a **control computer** to monitor and decode the sensor information for the active mirror. The first figure shows an adaptive optics imaging system.



The next figure shows an adaptive optics laser projection system.

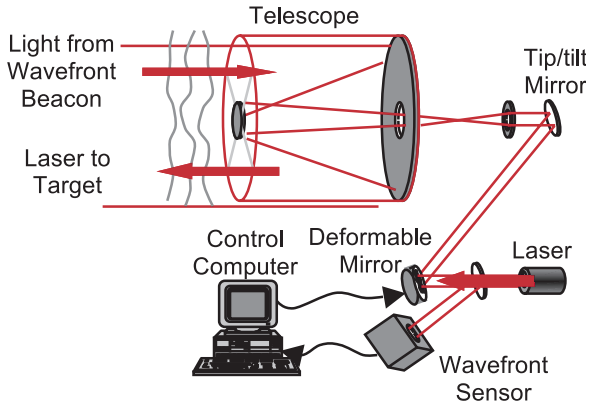
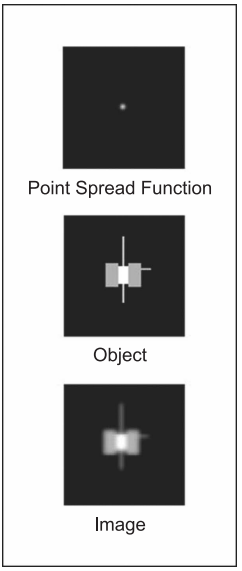


Image Spread with Atmospheric Turbulence

Linear systems theory shows how an image is composed of an object convolved with the **point spread function** (PSF) of the imaging system. **Atmospheric turbulence** degrades the PSF and smears the image.



The PSF is the image of a point source of light. The imaging process experiences **diffraction**, and the object is convolved with the PSF. The resultant image is a blurred version of the true object.

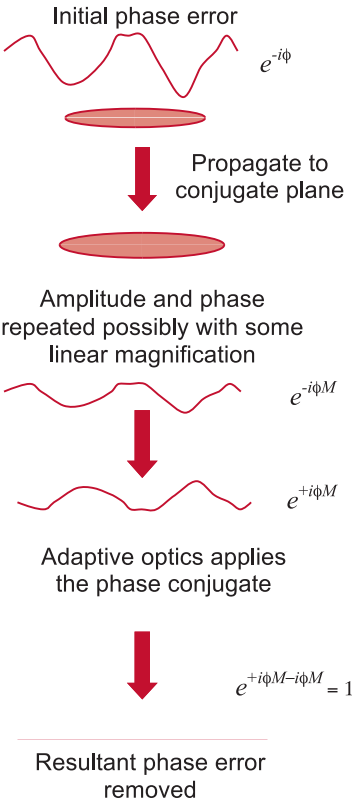
Adding **aberrations** to the optical system results in a broadening of the PSF and increased blurring.

Adaptive optics can compensate for the aberrations and reduce blurring.

What is adaptive optics? Answer for the common man: Atmospheric turbulence screws up the image. Adaptive optics unscrews it.

The Principle of Phase Conjugation

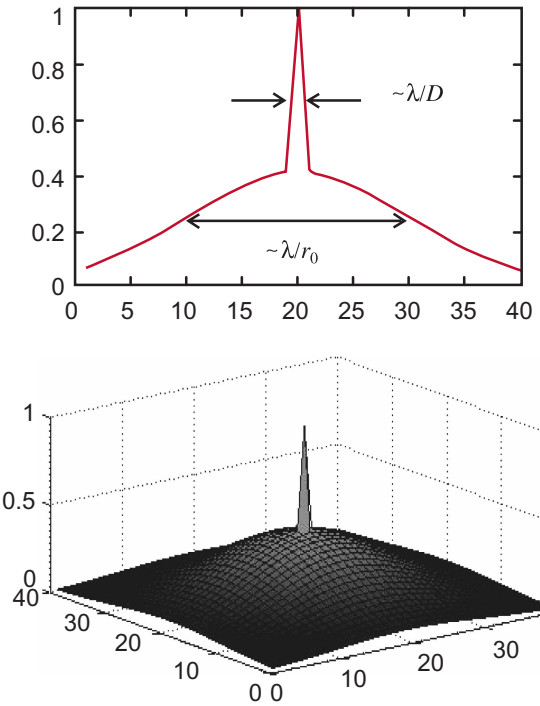
All systems of **adaptive optics** generally use the principle of **phase conjugation**. An optical beam is made up of both an amplitude A and a phase ϕ component and is described mathematically by the electric field $A\exp(-i\phi)$. Adaptive optics reverses the phase to provide compensation for the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



While **Horace Babcock** is generally thought to be the "inventor" of adaptive optics with his paper "The possibility of compensating **astronomical seeing**" [*Publ. Astron. Soc. Pac.* **65**, 229 (1953)], his exact idea was never put into practice. It wasn't until the late 1960s and early 1970s that technological developments in electro-optics made a working adaptive optics system possible.

Point Spread Function for an Astronomical Telescope

For a partially compensated astronomical telescope, the **point spread function** is limited by the **diffraction** of the optics and the **atmospheric turbulence**. For a short exposure or tilt-corrected image, the PSF has a diffraction-limited central core with an angular width proportional to λ/D, where D is the telescope pupil diameter. The uncompensated aberrations reveal a **halo** surrounding the core that has a width with an angular size of roughly λ/r_0, where r_0 represents the strength of atmospheric turbulence.



Fried's Coherence Length

Fried's coherence length is a widely used descriptor of the level of **atmospheric turbulence** at a particular site. For a fixed wavelength λ, **astronomical seeing** is given by the angle λ/r_0. For a known structure constant profile [$C_n^2(z)$, where z is the altitude] and a **flat-Earth assumption**, the coherence length for a plane wave is given by

$$r_0 = \left[0.423k^2 \sec\zeta \int_{\text{Path}} C_n^2(z)dz\right]^{-3/5}$$

where $k = 2\pi/\lambda$, ζ is the **zenith angle** (0 deg is straight overhead), and the integral is over the path to the ground-based telescope from the source of light.

Similarly, the coherence length for a spherical wave is given by

$$r_0 = \left[0.423k^2 \sec\zeta \int_{\text{Path}} C_n^2(z)\left(\frac{z}{L}\right)^{5/3} dz\right]$$

where L is the altitude of the source. Under turbulence, the resolution is limited by Fried's coherence length rather than the diameter of the telescope. Since r_0 ranges from under 5 cm with poor seeing to more than 20 cm with good seeing, even in the best conditions, a large-diameter telescope without **adaptive optics** does not provide any better resolution than a telescope with a smaller diameter.

Fried's coherence length is alternatively called "the Fried parameter," "Fried's parameter," "the seeing cell size," or simply "coherence length." These terms are interchangeable and widely used in the literature.

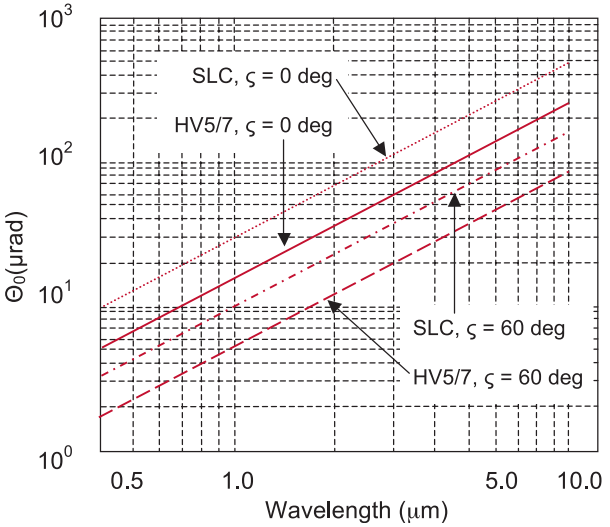
Isoplanatic Angle

Light traveling from a **wavefront beacon** should traverse the same atmosphere as the light from the object of interest. When the angular difference between the paths results in a mean-square **wavefront error** of 1.0 rad^2, the angular difference is called the **isoplanatic angle**. For a given structure constant profile [$C_n^2(z)$ where z is the altitude], and a **flat-Earth assumption**, the isoplanatic angle is given by

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

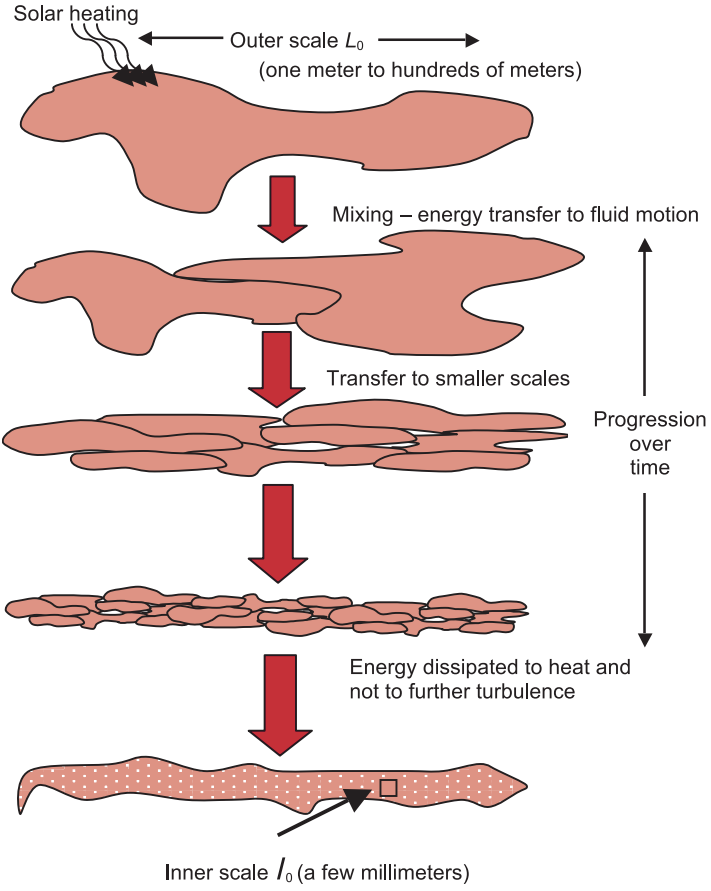
where $k = 2\pi/\lambda$, ζ is the **zenith angle**, and the integral is over the path from the ground-based telescope to the source of light above the surface.

The graph illustrates the isoplanatic angle versus wavelength for the **Hufnagel–Valley (H-V) model** and the **Strategic Laser Communication (SLC) model** of turbulence.



Kolmogorov Model

Large masses of heated air transfer kinetic energy to smaller and smaller scales. The largest size is called the **outer scale** L_0. When the scale size becomes small enough, the kinetic energy of the air is dissipated as heat. This small size is called the **inner scale** l_0.



Atmospheric Turbulence Models

One of the most widely used models for the **atmospheric turbulence** structure constant as a function of altitude is the H-V model:

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

where h is the altitude in kilometers, and C_n^2 is in units of $m^{-2/3}$. The parameters A and W are adjustable for local conditions. For the most common **H-V 5/7 model** (leading to $r_0 = 5$ cm and $\theta_0 = 7$ μrad), the structure constant at the surface A is 1.7×10^{-14}, and the wind velocity aloft W is 21.

For conditions other than the 5/7 model, we can calculate A and W from

$$A = 1.29 \times 10^{-12}\, r_0^{-5/3} \lambda^2 - 1.61 \times 10^{-13}\, \theta_0^{-5/3} \lambda^2 - 3.89 \times 10^{-15}$$

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

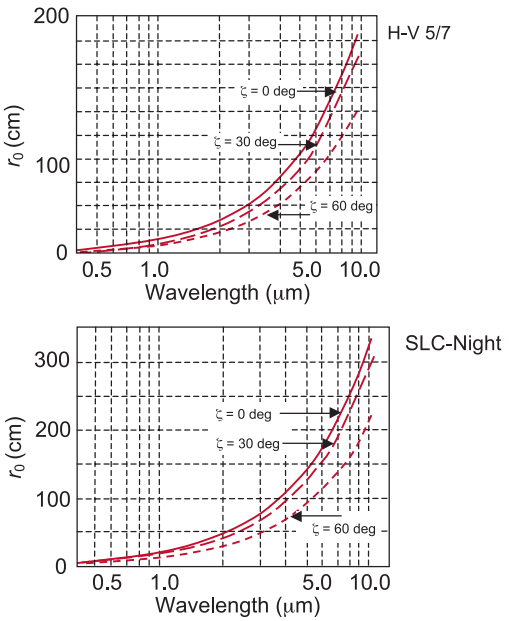
where the **coherence length** r_0 is in centimeters and the **isoplanatic angle** θ_0 is in microradians.

Other models are layered, such as the **SLC-Night model**:

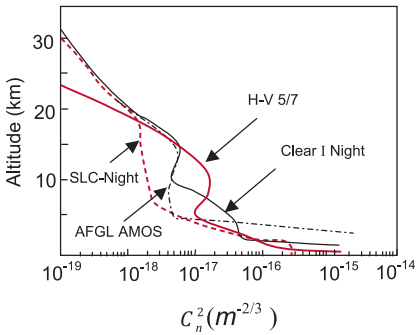
Altitude (above ground)	C_n^2
$h \le 18.5$ m	8.40×10^{-15}
$18.5 < h \le 110$ m	$2.87 \times 10^{-12}\, h^{-2}$
$110 < h \le 1500$ m	2.5×10^{-16}
$1500 < h \le 7200$ m	$8.87 \times 10^{-7}\, h^{-3}$
$7200 < h \le 20{,}000$ m	$2.00 \times 10^{-16}\, h^{-0.5}$

Coherence Length for Various Wavelengths and Turbulence Models

The **coherence length** r_0 is shown for two profiles, the **H-V 5/7** and the **SLC-Night**, for wavelengths from 0.4 μm to 10.6 μm and three **zenith angles**.



Turbulence models are site-specific because climate and geography play an important role in the C_n^2 profile.

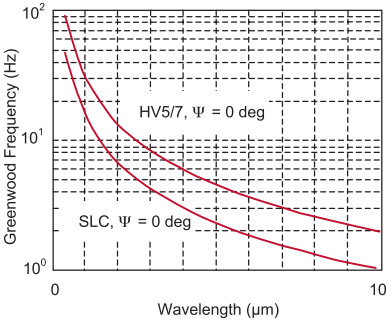


Greenwood Frequency

The **Greenwood frequency** is a measure of the rate at which **Kolmogorov atmospheric turbulence** changes with time. For the **flat-Earth assumption**,

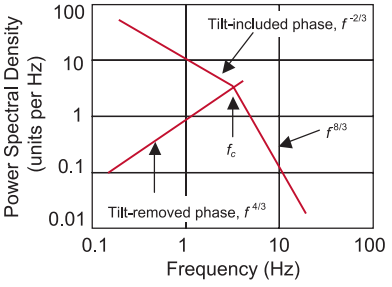
$$f_G = 2.31\,\lambda^{-6/5}\left[\sec\zeta\int_{\text{Path}} C_n^2(z)v_w(z)^{5/3}dz\right]^{3/5}$$

where ζ is the **zenith angle**, and v_w is the wind velocity as a function of altitude.



The **temporal power spectrum** exhibits behavior in regimes related to the modes of optical turbulence. For low frequencies, those less than the **crossover frequency** f_c, the **tilt**-included phase spectrum follows a (−2/3) power law. When tilt is removed the spectrum follows a (+4/3) power law. For frequencies above the crossover, the Kolmogorov spectrum is followed with the (−8/3) power law. The crossover frequency can be approximated from the turbulence-weighted wind velocity v and the aperture diameter D:

$$f_c \simeq 0.7\frac{v}{D}.$$



The Greenwood frequency determines how quickly the adaptive optics system must respond in order to compensate for turbulence. As a rule-of-thumb, the **closed-loop bandwidth** should be at least 10 times the Greenwood frequency.

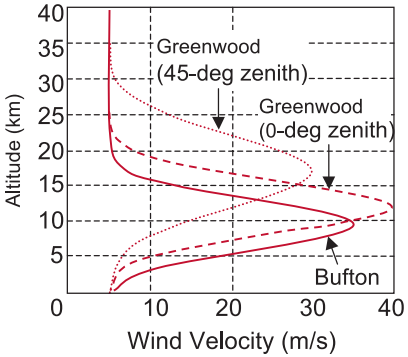
Wind Models

For calculations involving the **atmospheric wind profile**, we can use a general **Gaussian model** developed by Greenwood:

$$v_{\mathrm{w}}(z) = v_{\mathrm{G}} + v_{\mathrm{T}} \exp\left[-\left(\frac{z\cos\zeta - H_{\mathrm{T}}}{L_{\mathrm{T}}}\right)^2\right] \times \left[\sin^2\phi + \cos^2\phi\cos^2\zeta\right]^{1/2},$$

where the wind velocity v_{W} is in m/s. In the expression, v_G is the wind velocity at low altitude, v_{T} is the wind velocity at the **tropopause**, ζ is the **zenith angle**, H_{T} is the height of the tropopause, L_{T} is the thickness of the tropopause layer, and ϕ is the wind direction relative to the telescope azimuth. The wind velocity at the tropopause typically ranges from 23 m/s to 46 m/s. The **Bufton wind model** assumes a 0-deg zenith angle and wind direction and applies the values: v_{G} = 5 m/s; v_{T} = 30 m/s; H_{T} = 9.4 km; L_{T} = 4.8 km.

$$v(z)_{\mathrm{Bufton}} = 5 + 30\exp\left[-\left(\frac{z-9.4}{4.8}\right)^2\right].$$



Three realizations of the wind model: (1) Greenwood: v_{G} = 5 m/s; v_{T} = 35 m/s, H_{T} = 12.0 km, L_{T} = 5 km; ζ = 45; (2) Greenwood: ζ = 0 deg; and (3) Bufton.

Scintillation

While a beam propagates, the effect of phase **aberrations** manifests into amplitude or **intensity variations**. The **scintillation** is easily seen as the twinkling of stars in the night sky. The **amplitude fluctuation**, represented by the **log-amplitude variance** σ_χ^2, is a function of wavelength, propagation length L, and the strength of **atmospheric turbulence** C_n^2:

$$\sigma_\chi^2 \sim \left(\frac{2\pi}{\lambda}\right)^{7/6} L^{11/6} C_n^2$$

Weak turbulence:

$$\sigma_\chi^2 < 0.3$$

Moderate turbulence:

$$0.3 < \sigma_\chi^2 < 1.0$$

Strong turbulence:

$$\sigma_\chi^2 > 1.0$$

The variance in intensity that results from this variance in log-amplitude is given by

$$\sigma_I^2 = A\left[\exp\left(4\sigma_\chi^2\right) - 1\right]$$

where the variable A represents the aperture averaging factor. For weak turbulence,

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

where $k = 2\pi/\lambda$ and D is the aperture diameter. The effect of adaptive optics on the intensity variance is approximated by

$$\sigma_I^2(\text{compensated}) = \frac{\sigma_I^2(\text{uncompensated})}{N_{\text{Zern}}}$$

where N_{Zern} is the number of fully compensated **Zernike modes**.

Zernike Polynomials

Optical phase can be represented by a two-dimensional surface over the aperture. The deviation from flat (or some other reference surface) is the **wavefront error** sensed by the **wavefront sensor**. A very useful infinite-series representation of the wavefront is the **Zernike polynomial series**. Radial (index n) and azimuthal (index m) polynomials are preceded by Zernike coefficients A_{nm} and B_{nm} that completely describe the wavefront up to the order specified by the largest n or m. The series is written as

$$\Phi(r,\theta) = A_{00} + \frac{1}{\sqrt{2}}\sum_{n=2}^{\infty} A_{n0}\Re_n^0\left(\frac{r}{R}\right) + \sum_{n=1}^{\infty}\sum_{m=1}^{n}\left(A_{nm}\cos m\theta + B_{nm}\sin m\theta\right)\Re_n^m\left(\frac{r}{R}\right)$$

where the **azimuthal polynomials** are sines and cosines of multiple angles, and the radial polynomial is

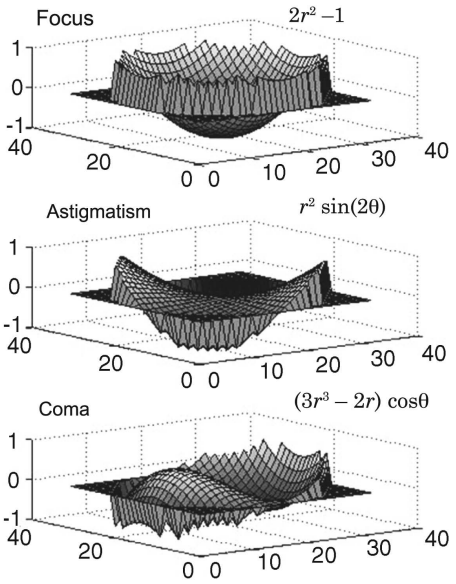
$$\Re_n^m\left(\frac{r}{R}\right) = \sum_{s=0}^{\frac{n-m}{2}}(-1)^s\frac{(n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!}\left(\frac{r}{R}\right)^{n-2s}$$

The series is especially useful in adaptive optics because the polynomials are orthogonal over a circle of radius R, common to many optical system geometries. For R normalized to unity, the first few radial terms are given here.

	n=0	**n=1**	**n=2**	**n=3**	**n=4**
m=0	1		$2r^2-1$		$6r^4-6r^2+1$
m=1		r		$3r^3-2r$	
m=2			r^2		$4r^4-3r^2$
m=3				r^3	
m=4					r^4

Zernike Polynomials (cont.)

A few terms are graphed to visualize their relationship to **third-order optical aberrations**.



Zernike polynomials associated with low-order modes.

Reflecting telescopes of the **Cassegrain** design have a central obscuration, which requires an extremely large number of Zernike coefficients—more than can be adequately described. A set of annular Zernike polynomials can be obtained from **Gram–Schmidt orthogonalization**; this series is generally used for optical systems with central obscurations.

Legendre Polynomials

Many modern solid state lasers have rectangular beam footprints over which Zernike polynomials are not orthogonal and therefore do not adequately describe the wavefront. The Legendre polynomial series is orthogonal over the range $[-1,1]$ and is often used to describe wavefronts for these beams.

In one dimension, the generating function for a **Legendre polynomial** is

$$P_n(x) = \sum_{k=0}^{n/2} (-1)^k \frac{(2n-2k)!}{2^n k!(n-k)!(n-2k)!} x^{n-2k}$$

This forms a polynomial series containing either all even powers of x or all odd powers of x, depending on the order n.

The Legendre series can be extended to two dimensions by multiplication with a second series of order m with dependence on y. A wavefront surface $[W(x,y)]$ can be computed from the coefficients (c_{nm}) for each polynomial

$$W(x,y) = \sum_{n=0}^{N} \sum_{m=0}^{M} c_{nm} P_n(x) P_m(y)$$

A series expanded in this fashion is known as a Fourier–Legendre series, and the coefficients can be determined from

$$c_{nm} = \left(\frac{2n+1}{2}\right)\left(\frac{2m+1}{2}\right) \times \int_{x=-1}^{1} \int_{y=-1}^{1} W(x,y) P_n(x) P_m(y)\, dx\, dy$$

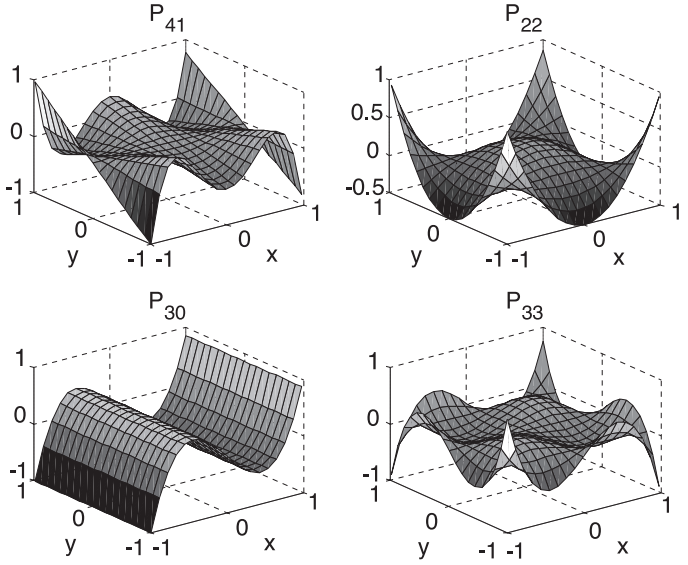
The integrals are replaced with summations for discretely sampled wavefronts.

The following table lists the first five Legendre polynomials along a single dimension, and the figure shows several

Legendre Polynomials (cont.)

surfaces corresponding to two-dimensional Legendre polynomials.

n	Polynomial
0	1
1	x
2	$\frac{1}{2}\left(3x^2-1\right)$
3	$\frac{1}{2}\left(5x^3-3x\right)$
4	$\frac{1}{2}\left(3x^2-1\right)$
5	$\frac{1}{8}\left(63x^5-70x^3+15x\right)$



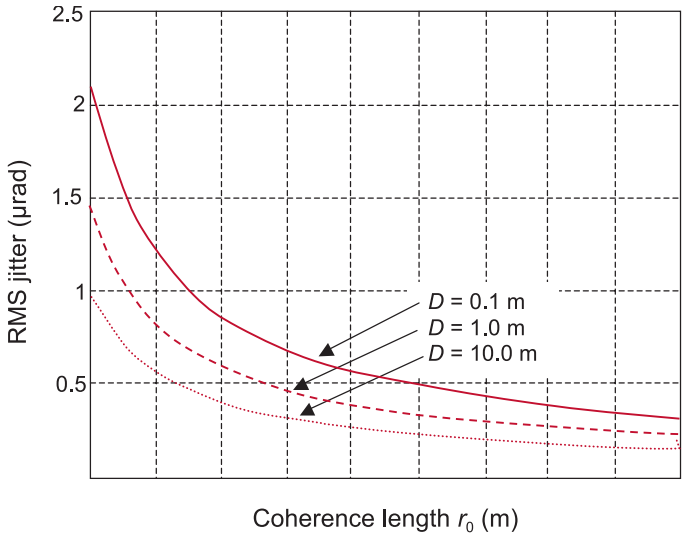
Angle of Arrival (Tilt) Fluctuations (Image Motion)

Turbulence causes **jitter** in a beam or an image. The root-mean-square (RMS) average jitter is computed from the expression

$$\alpha_{\text{jit}} = \sqrt{0.182\lambda^2 D^{-1/3} r_0^{-5/3}}$$

where D is the aperture diameter. The result is a random beam direction that varies slightly around its average and causes a correspondingly random image position. If there is any optical system that imparts **magnification** on the beam, the jitter at the image plane (or more generally, the compact portion of the beam) must be multiplied by the magnification. Using the Kolmogorov atmospheric turbulence model, the tilt fluctuations have a spectrum that is characterized by a single frequency. Greenwood calculated this frequency from beam parameters and the path of the light through the turbulence, which is strongly dependent on the wind velocity profile v_{w}^2. The Greenwood frequency for atmospheric tilt is calculated from

$$f_{G-T} = 0.331 D^{-1/6} \lambda^{-1} \sec^{1/2} \zeta \left[\int_0^L C_n^2(z) v_{\text{w}}^2 dz \right]^{1/2}.$$



Modulation Transfer Function

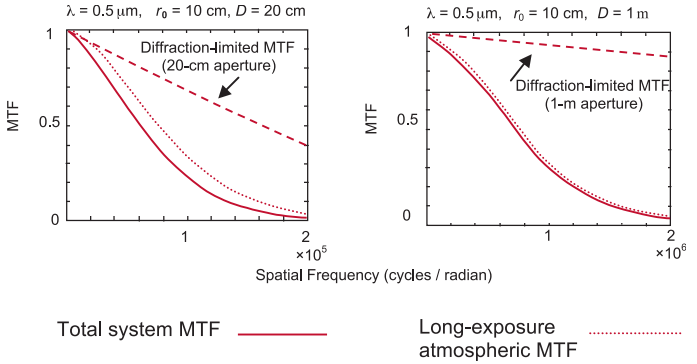
The **modulation transfer function** (MTF) is a measure of the **spatial frequency response** of an imaging system. Assuming linear response, the MTF can be composed of separate MTFs of various components or features of the system that are multiplied to find the total system MTF. The MTF for a diffraction-limited circular aperture of diameter D is

$$M_{\mathrm{DL}}(\nu) = \frac{2}{\pi}\left\{\cos^{-1}\left(\frac{\nu\lambda}{D}\right) - \left(\frac{\nu\lambda}{D}\right)\left[1 - \left(\frac{\nu\lambda}{D}\right)^2\right]^{1/2}\right\}$$

where the spatial frequency ν is in units of cycles/rad. The MTF of the uncompensated atmosphere has various forms. The **short-exposure MTF** does not include temporal beam wander (**tilt**), which smears an image over time. The **long-exposure MTF** includes the effect of beam tilt and results in an MTF that is driven by r_0 and the MTF of the atmosphere:

$$M_{\text{Long-exp}}(\nu) = \exp\left[-3.44\left(\frac{\nu\lambda}{r_0}\right)^{5/3}\right]$$

Note the improvement in imaging quality (increase in MTF) with the larger diameter, whereas in both cases, the atmosphere dominates the **image degradation**.



System Performance Estimation

One method of determining how well an adaptive optics system performs is to evaluate the **residual errors** of the system components. Because a **deformable mirror** cannot exactly match the shape of **Kolmogorov turbulence**, **fitting error** results. Similarly, because a **control system** cannot respond instantaneously to the disturbance, a **temporal error** results. When the source of the wavefront measurement (the **wavefront beacon**) is positioned away from the object of interest or an outgoing laser target position, the **wavefront sensor** measures slightly different turbulence, termed **isoplanatic error**. Noise within the wavefront sensor makes exact measurement impossible, and **sensor noise error** results.

Assuming that these errors are uncorrelated and are essentially Gaussian random variables, their variances can be added to determine a **system error**. With σ^2 in units of square radians, the system error is

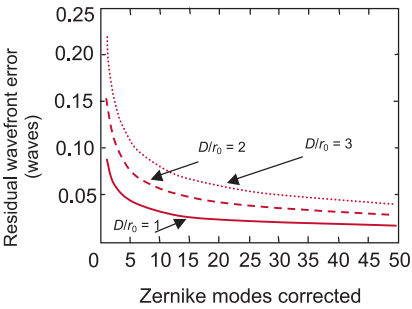
$$\sigma^2_{\text{system}} = \sigma^2_{\text{fitting}} + \sigma^2_{\text{temporal}} + \sigma^2_{\text{isoplanatic}} + \sigma^2_{\text{sensor noise}}$$

When these parameters are correlated, the analysis becomes more complex. For example, sensor noise is often a function of integration time of the wavefront sensor, which is an important contributor to the **closed-loop bandwidth** and the temporal error.

Hindsight is 20/20. If the **Hubble Space Telescope** had had an adaptive optics system on board, it would not have been necessary to correct the primary mirror aberration with a NASA-termed “emergency-servicing” mission.

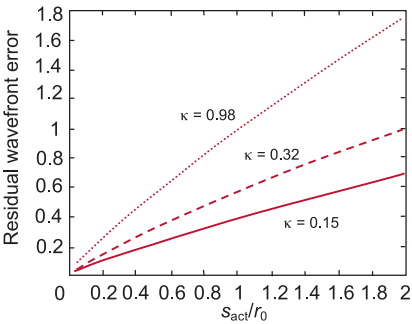
Modal and Zonal Fitting Error

When a typical **deformable mirror** cannot exactly fit the spatial structure of stochastic atmospheric turbulence, modes are used and applied to the deformable mirror, causing the **residual wavefront error** to be reduced. Where N_{Zern} is the number of completely corrected **Zernike modes**, and D is the aperture diameter, the RMS **modal wavefront error** is found from



$$\sigma_{\text{fitting}}(\text{waves}) = \frac{\sqrt{0.2944\, N_{\text{Zern}}^{-\sqrt{3}/2} \left(\frac{D}{r_0}\right)^{5/3}}}{2\pi}$$

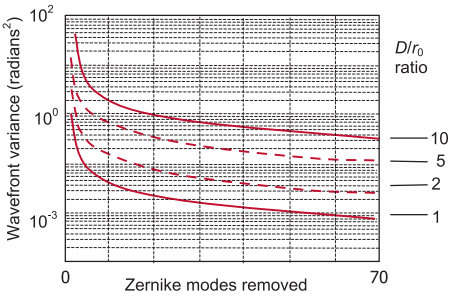
If many **actuators** across a continuous faceplate are used, the residual wavefront error can be reduced. The distance between actuators, in the same space as the measurement of r_0, is s_{Act}. **Fitting constant** κ is related to the stiffness of the **deformable mirror** faceplate as follows:



$$\sigma_{\text{fitting}}(\text{waves}) = \frac{1}{2\pi}\left[\kappa\left(\frac{s_{\text{act}}}{r_0}\right)^{5/3}\right]^{1/2}$$

Partial Correction

To evaluate **fitting error**, determine the residual **wavefront variance** after the adaptive optics system removes **Zernike modes** from the disturbed wavefront.



When Kolmogorov turbulence is assumed, calculate the wavefront variance as a function of the number of modes removed and the D/r_0 ratio. A few terms are

$$\sigma^2_{0 \text{ modes removed}} = 1.0299\left(\frac{D}{r_0}\right)^{5/3}$$

$$\sigma^2_{1 \text{ mode (1-axis tilt) removed}} = 0.582\left(\frac{D}{r_0}\right)^{5/3}$$

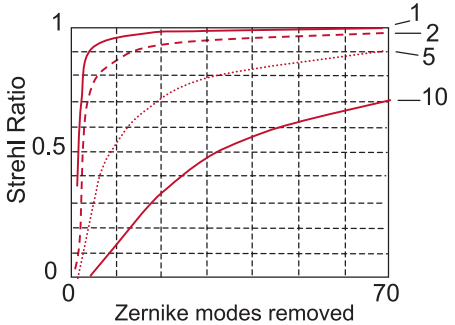
$$\sigma^2_{2 \text{ modes (2-axis tilt) removed}} = 0.134\left(\frac{D}{r_0}\right)^{5/3}$$

$$\sigma^2_{3 \text{ modes (defocus and tilt) removed}} = 0.111\left(\frac{D}{r_0}\right)^{5/3}$$

For a larger number of modes, where N_{Zern} is the total number of Zernike modes completely removed,

$$\sigma^2_{N \text{ modes removed}} = 0.2944 N_{\text{Zern}}^{-\sqrt{3}/2}\left(\frac{D}{r_0}\right)^{5/3}$$

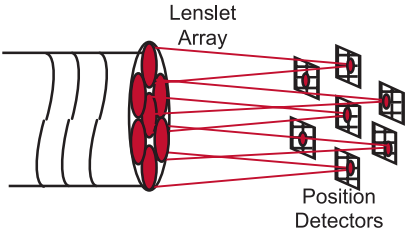
Great improvements are seen by correction of only a few Zernike modes, 95% of the energy of the **aberrations** in a Kolmogorov wavefront is contained within the first 13 modes.



Shack–Hartmann Wavefront Sensor and Error

The **Shack–Hartmann wavefront sensor** is a pupil plane measurement of local wavefront slopes (the first **derivative** of the wavefront) within a **subaperture** defined typically by a **lenslet array**. The positions of the Hartmann spots on the detector(s) are proportional to **wavefront tilt** or **slope**.

Shack–Hartmann **sensor error** is dependent on signal-to-noise ratio, size of subapertures, gaps between pixels in the detector focal plane, and the finite size of the reference source.



Variables include:

SNR = signal-to-noise ratio
K_g = increase in error at the null due to gaps in the detector focal plane
D = size of subaperture (in object space)
r_0 = coherence length of the atmosphere (in object space)
θ = angular size of the reference source
σ_ϕ = root-mean-square wavefront error in radians measured over a subaperture

$$\sigma_\phi = \frac{\pi^2 K_g}{4\mathrm{SNR}}\left[\left(\frac{3}{2}\right)^2 + \left(\frac{\theta d}{\lambda}\right)^2\right]^{1/2}, \quad r_0 > d$$

$$\sigma_\phi = \frac{\pi^2 K_g}{4\mathrm{SNR}}\left[\left(\frac{3d}{2r_0}\right)^2 + \left(\frac{\theta d}{\lambda}\right)^2\right]^{1/2}, \quad r_0 < d$$

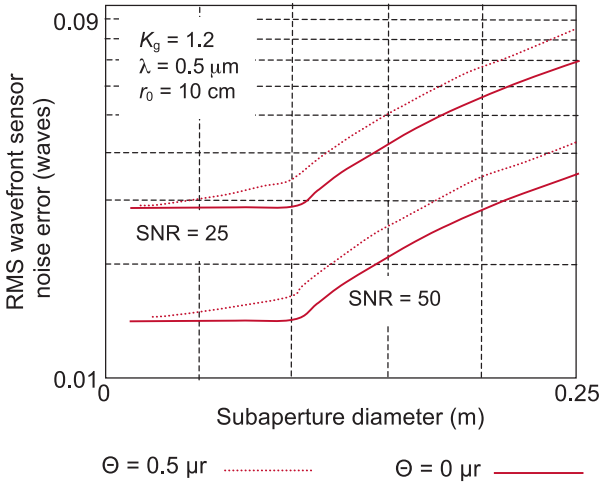
Shack–Hartmann Wavefront Sensor and Error (cont.)

For the conditions where $\lambda = 0.5$ μm and $r_0 = 10$ cm, the error is shown in the figure below.

For a conventional **CCD camera** used for wavefront **slope** measurement, the **signal-to-noise ratio** (SNR) is

$$\frac{n_p}{\left\{n_p + N_D\left[n_B^2 + \left(\frac{e_n}{G}\right)^2\right]\right\}^{1/2}},$$

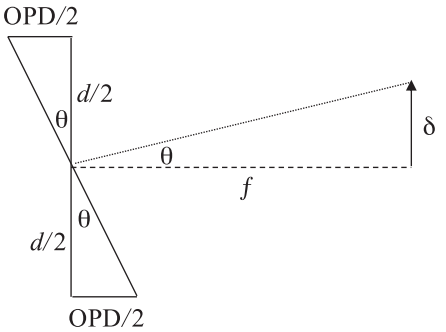
where n_p is the number of detected photoelectrons per **subaperture** (sum of all pixels); N_D is the number of pixels in a subaperture (e.g., $N_D = 4$ for quadcell); n_B is the number of detected background photoelectrons per subaperture; e_n is the read-noise in electrons per pixel; and G is the gain ($G = 1$ for a nonintensified CCD).



Shack–Hartmann Lenslet Array Selection

Tip and tilt do not adversely affect image quality, as they only translate the image. Therefore, the lenslet diameter for a Shack–Hartmann lenslet array should be chosen such that the only significant aberration across each lenslet is local tilt. Higher-order aberrations distort the spot and introduce errors in the measurement.

Using the geometry in the figure below, we can determine the relationship needed to perform system design trades. Here δ is the spot shift, f is the focal length of the lenslet, d is the diameter of the lenslet, and θ is the average tilt over the subaperture. The dashed line represents the surface normal, and the dotted line represents the tilted input beam.

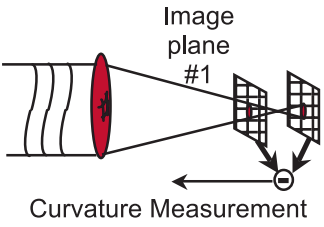


From this geometry, we can relate both the optical path difference (OPD) and δ to θ by

$$\delta = \mathrm{OPD}\frac{f}{d} = \mathrm{OPD}(f/\#)$$

For a given OPD at the wavefront sensor (WFS), we can trade off the lenslet array $f/\#$ and camera pixel size to optimize the sensor resolution, dynamic range, and noise propagation. We can use this same scale factor to convert spot shift measurements back to OPD when analyzing the system performance.

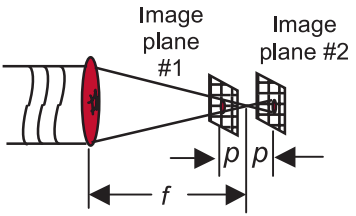
Curvature Wavefront Sensor and Error



The **curvature sensor** is an image-plane measurement of local wavefront curvature [the second **derivative** of the wavefront: $\nabla^2\Phi(\mathbf{r})$] deduced from two specific out-of-focus images. A point-by-point subtraction of the images is proportional to the wavefront curvature term minus the derivative of the wavefront at the edge $\frac{d\Phi(\mathbf{r})}{d\mathbf{n}}$:

$$I_1(\mathbf{r}) - I_2(\mathbf{r}) = C\left[\nabla^2\Phi(\mathbf{r}) - \frac{d\Phi(\mathbf{r})}{d\mathbf{n}}\right]$$

To provide an accurate measurement of wavefront curvature, the **blur** from the turbulence must be small compared to the area where the curvature measurement is taken. With p being the offset from the focal plane of the system and f as the focal length of the system, a conservative estimate of the blur requirement leads to



$$p \geq \frac{\lambda f^2}{\lambda f + r_0^2}.$$

Variance of a single curvature measurement:

$$\sigma^2_{\text{Curv. Sens.}} = \frac{p^2}{f^4 N_\text{p}}$$

where N_p is the photon count. Because the curvature sensor directly measures the Laplacian of the wavefront, **bimorph mirrors** are generally used for closed-loop compensation, as they possess Laplacian influence functions.

Pyramid Wavefront Sensor and Error

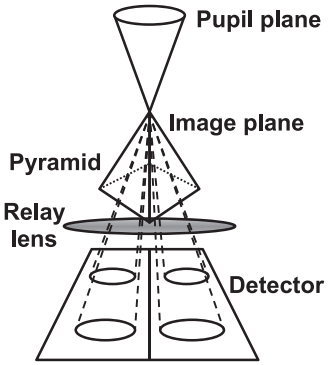
The **pyramid wavefront sensor** is a pupil-plane wavefront sensor.

Another pupil-plane wavefront sensor uses a pyramidal prism in the image plane to create four subbeams that are then optically relayed to a detector. The intensity at position $\vec{r}(x,y)$ in each of the subbeams in the detector plane ($I_{0,0} > I_{0,1} > I_{1,0} > I_{1,1}$) is used to find the x and y wavefront slopes at $\vec{r}(x,y)$:

$$S_x(\vec{r}) = \frac{I_{0,0}(\vec{r}) - I_{1,0}(\vec{r}) + I_{0,1}(\vec{r}) - I_{1,1}(\vec{r})}{I_t} \quad \text{and}$$

$$S_y(\vec{r}) = \frac{I_{0,0}(\vec{r}) + I_{1,0}(\vec{r}) - I_{0,1}(\vec{r}) - I_{1,1}(\vec{r})}{I_t}$$

where I_t is the average intensity over the detector plane.



One advantage of the pyramid technique over a Shack–Hartmann sensor is that the spatial resolution of the sensor is the size of the detector pixel in contrast to the larger lenslet subaperture size of the Shack–Hartmann. The wavefront error variance associated with the pyramid technique is approximated by

$$\sigma^2 = \frac{1}{2N} + \frac{2(e_n)^2}{N} \quad \text{rad}^2$$

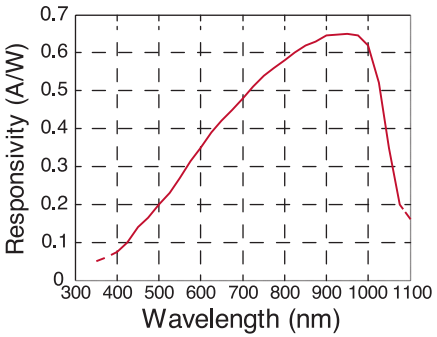
where N is the number of photons per pixel and e_n are the read-noise electrons.

Photodiodes

Photodiodes are p–n junctions and have a **response time** that is a combination of drift time, diffusion time and resistor–capacitor (RC) time. The drift time is the time required to collect carriers in the depletion region, the diffusion time is the time required for collection of carriers in the undepleted bulk region, and the RC time is the response time due to the equivalent RC circuit.

Photodiodes can be operated in either photovoltaic or photoconductive mode. The photovoltaic mode has low noise, as it uses a zero-volt bias, which does not induce a dark current. Its response time is dominated by the drift time, as the carriers are not excited by a bias voltage. This mode is used for extremely low-light-level and low-frequency applications. The photoconductive mode uses a reverse bias voltage that increases the width of the depletion region and decreases the junction capacitance, increasing the response speed and linearity. This mode has a constant dark current and increased noise, and its response time is dominated by the diffusion time, as the carriers in the undepleted region have farther to travel.

The **responsivity** R of a photodiode is the ratio of the photocurrent to the incident power and is shown below for silicon. InGaAs is generally used for IR wavelengths.



Photodiode Noise

Photodiode noise comes from **shot noise** and **Johnson noise**. Shot noise depends on the variance of the photocurrent and dark current:

$$I_s = \sqrt{2q(I_p + I_d)f}$$

Here q is the electron charge (1.6×10^{-19} C), I_p is the photocurrent, I_d is the dark current, and f is the noise measurement bandwidth. Shot noise dominates when the photodiode is operated in photoconductive mode.

Johnson noise is generated by thermal carriers and is given as

$$I_J = \sqrt{\frac{4k_B T_K f}{R_p}}$$

Here k_B is the Boltzmann constant (1.38×10^{-23} J/K), T_K is the temperature in Kelvin, and R_p is the parallel or shunt resistance. Johnson noise dominates when the photodiode is operated in photovoltaic mode, as this mode has no dark current and therefore very low shot noise.

The total noise current I_n is the root-sum-square (RSS) of the noise sources:

$$I_n = \sqrt{I_s^2 + I_J^2}$$

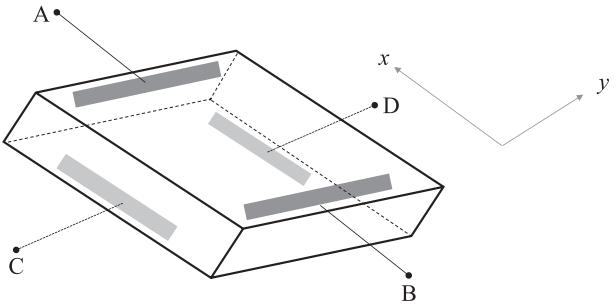
Johnson and shot noise are both white noise processes; however, $1/f$ noise is also present where the power spectrum is inversely proportional to the frequency. $1/f$ noise is commonly referred to as pink or flicker noise and has negligible magnitude for frequencies above 1 Hz, so it is typically ignored for noise performance analysis. This type of noise has been widely observed in many dynamic systems, and there is currently no universal theory for its origin.

Noise equivalent power (NEP) is the ratio of the total noise current to the responsivity and represents the level of incident intensity that generates a photocurrent equal to the noise current:

$$\text{NEP} = \frac{I_n}{R}$$

Lateral-Effect Position-Sensing Detectors

Lateral-effect position-sensing detectors (PSDs) are continuous single-element photodiodes; **duo-lateral PSDs** are the most commonly used PSDs. These devices have two resistive layers on either side of the photodiode, with contacts on both ends, as shown in the figure.



The position output of a duo-lateral PSD is given as

$$x = \frac{A - B}{A + B}, \qquad y = \frac{D - C}{C + D}$$

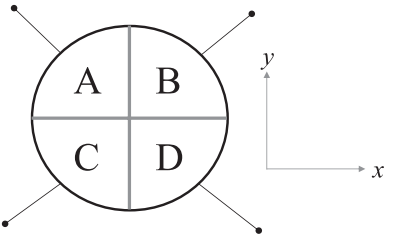
The current on either side of the photodiode is the same, meaning that $A + B = C + D$.

Duo-lateral PSDs have excellent linearity and can accurately measure position out to the edge of the device. However, the noise is dependent on the SNR of the system, and duo-lateral PSDs have lower resolution and accuracy than quadcells.

Because the quad cell and lateral-effect sensors are fundamentally different, their biasing and calculation circuits are different and incompatible with one another.

Quad Cells

Quad cells or segmented PSDs consist of four individual photodiodes that share a common substrate, as shown the figure.



The position output of this quad cell is given as

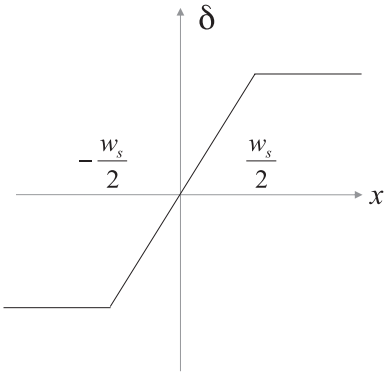
$$x = \frac{(B+D)-(A+C)}{A+B+C+D}, \qquad y = \frac{(A+B)-(C+D)}{A+B+C+D}$$

Generally, the segments are the anodes, and the cathode is on the back of the device.

Quad cells are capable of achieving higher resolution and accuracy than lateral-effect PSDs because the responsivities of the individual segments can be matched to each other and the noise is not dependent on the SNR of the system. They are excellent for null-seeking applications but are limited with the use of position offsets because the linear dynamic range is determined by the spot width w_s.

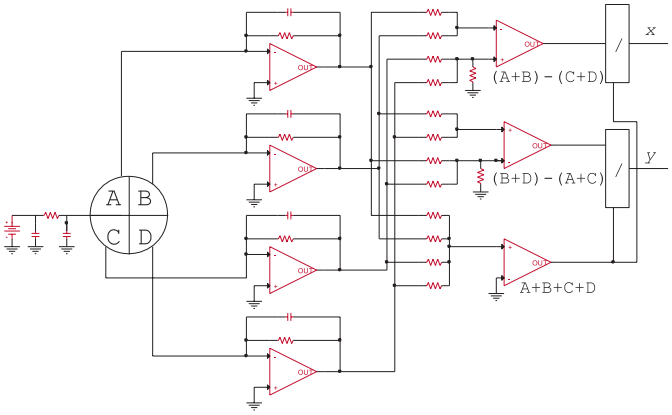
When the spot falls off the segments, the measured position saturates, as the linear dynamic range has been exceeded. The same issue occurs if the spot is contained in a single segment. For this reason, the spot width when using a quadcell is usually larger than when using a lateral effect PSD, with the optimum **spot size** for maximum linear dynamic range equal to the size of a segment. This is demonstrated in the figure on the next page, which shows the output measured position δ versus actual spot position x. The linear dynamic range of the quad cell is limited to $\pm w_s/2$.

Quad Cells (cont.)



Early high-speed Shack–Hartmann WFSs utilized avalanche photodiodes arranged in quad cell configurations for each subaperture. With advances in CCD technology, modern Shack–Hartmann WFSs generally use a larger number of pixels per subaperture. Quad cells are still used in beam control systems for jitter detection and control.

Quad cells are typically used with a simple operational amplifier (op-amp) circuit as shown in the following figure:



Noise Equivalent Angle

Noise equivalent angle (NEA) converts detector noise and algorithmic effects into an equivalent jitter angle based on the SNR. For a quad cell device, the NEA is given as

$$\mathrm{NEA} = \frac{\pi}{\mathrm{SNR}}\left(\frac{\lambda}{D_\mathrm{r}}\right)\sqrt{\left(\frac{3}{16}\right)^2 + \left(\frac{na}{8}\right)^2}$$

where D_r is the receiver diameter and na is the angular extent of the target divided by the diffraction angle of the receiver.

NEA is used extensively in assessment of tracking, jitter, and pointing performance.

High-order aberrations distort the spot and can artificially bias the output of a PSD. For this reason, atmospheric tracking and jitter loops see significant improvement when high-order wavefront correction is enabled.

Tracking an object and compensation of optical jitter may seem like similar goals but in fact are very different and have conflicting performance specifications. The object being tracked usually has dynamics that define its motion, so the track loop needs to estimate where the object is and where it will be next to follow it accurately. Jitter correction tries to reject disturbance-induced jitter by driving measured tilt to zero (or a static offset). The differences between tracking and disturbance rejection will be more clearly shown in the section on sensitivity functions.

NEA is usually defined mathematically for position estimates from standard centroid algorithms. When other algorithms such as correlation or edge detection are used, the NEA calculations are typically modified by an empirically derived scale factor.

The Strehl Ratio: Laser Beam Propagation to the Far Field with Wavefront Error

The **on-axis intensity** I_0 of a uniform circular beam after it propagates a distance L is

$$I_0 = \frac{\pi^2 (D/2)^4}{\lambda^2 L^2} I_{\text{Aper}}$$

where $L >> D$, the diameter of the aperture, and I_{Aper} is the intensity at the circular aperture (W/m^2). With **aberrations** represented by a **wavefront error** variance $(\Delta\phi)^2$ (units of optical path distance), the reduction of on-axis intensity, the **Strehl ratio**, is approximately

$$S \simeq \exp\left[-\left(\frac{2\pi}{\lambda}\right)^2 (\Delta\phi)^2\right]$$

To use this approximation, the wavefront error variance should be less than a quarter-wave.

The on-axis intensity with aberrations is then

$$I_0 = S \frac{\pi^2 (D/2)^4}{\lambda^2 L^2} I_{\text{Aper}}$$

Astronomical Strehl ratios without adaptive optics are typically very small. With adaptive optics, the Strehl ratio can be improved by orders of magnitude. For a well-conditioned beam in **weak turbulence**, the Strehl ratio without adaptive optics can be 20%, but is improvable to 90% or better with adaptive optics.

The definition of Strehl ratio in the classic text *Principles of Optics* (Born and Wolf, 1975) does not include **defocus** or **tilt** (jitter) terms in the definition. However, for adaptive optics system calculations, the absorption of these effects into the overall Strehl ratio is convenient.

Strehl Ratio

When **beam jitter** is present, the optic axis is swept over a small cone, and the average intensity in the center of the beam is reduced. When the jitter is assumed to be Gaussian, where α_{jit} is the RMS single-axis beam jitter (in radians), the intensity is multiplied by the factor

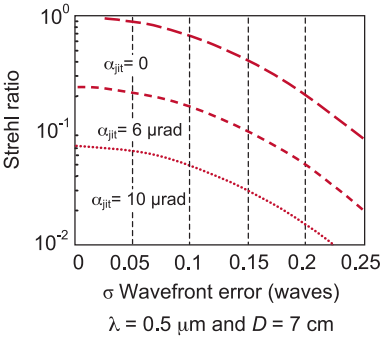
$$\frac{1}{\left[1+\left(\frac{2.22\,\alpha_{jit}D}{\lambda}\right)^2\right]}$$

to find the further reduction in **on-axis intensity**. Sometimes the combined effects of **wavefront error** and beam or image jitter are combined into the **Strehl ratio**. Rewriting the **wavefront error variance** in radians squared gives

$$\sigma^2 = \left(\frac{2\pi}{\lambda}\right)^2 (\Delta\phi)^2$$

$$S_{w/jit} = \frac{e^{-\sigma^2}}{\left[1+\left(\frac{2.22\,\alpha_{jit}D}{\lambda}\right)^2\right]}$$

Because the **wavefront variance** can be considered a Gaussian variable uncorrelated between various sources, spatial and temporal effects can be efficiently combined using this general definition of Strehl ratio.



In 1902, K. Strehl coined the term Definitionshelligkeit, which means **definition of brightness** in German and is now known as the Strehl ratio.

Laser "Brightness"

A common description of a laser beam, especially a high-power beam, is to describe its **brightness**. This is a measure of on-axis power with the explicit propagation distance removed. Combining the **effects of jitter** with the **Strehl ratio**, the brightness B is

$$B = \frac{\pi D^2 P}{4\lambda^2 \left[1 + \left(\frac{2.22\,\alpha_{\text{jit}} D}{\lambda}\right)^2\right]} TKS$$

where K is an aperture shape parameter ($K = 1$ for a circular aperture), T is the transmission of all optics in the system, and P is the optical power (watts). Brightness is expressed in watts per steradian (W/sr).

Because brightness is not dependent on propagation distance to the target, it precisely describes the laser system, the optical system, the beam jitter, and higher-order aberrations but is independent of the scenario. This has been useful for trade studies for medium- to high-power laser weapon programs.

The brightness can be used to calculate the energy density on a target by dividing brightness by the square of the distance from the beam propagation aperture to the target.

For historical reasons, the radiometric terminology used by the astronomical community differs from the terminology that has been introduced by radiometric standards organizations.

Brightness (W/sr) in standard notation is **intensity**.
Intensity (W/m^2) in standard notation is **irradiance**.

The notation common to the astronomical community is used throughout this *Field Guide*.

Laser Beam Quality

Beam quality is generally identified by the symbol M^2, and the ISO standard defines it relative to the beam parameter product, which is the product of the beam radius measured at the beam waist and the beam divergence half-angle. Beam quality is then the actual beam parameter product divided by the beam parameter product for a diffraction-limited Gaussian beam of the same wavelength. With these definitions, M^2 is given as

$$M^2 = \theta_d \frac{\pi w_0}{\lambda}$$

where w_0 is the **beam radius at the beam waist** and θ_d is the **half-angle beam divergence**.

Beam quality is a unitless ratio and is often referred to as "times diffraction limit" because the spot size in the farfield is M^2 times larger than that of the diffraction-limited beam.

Laser brightness is inversely proportional to the square of the beam quality for a given power level:

$$B \propto \frac{1}{(M^2)^2}$$

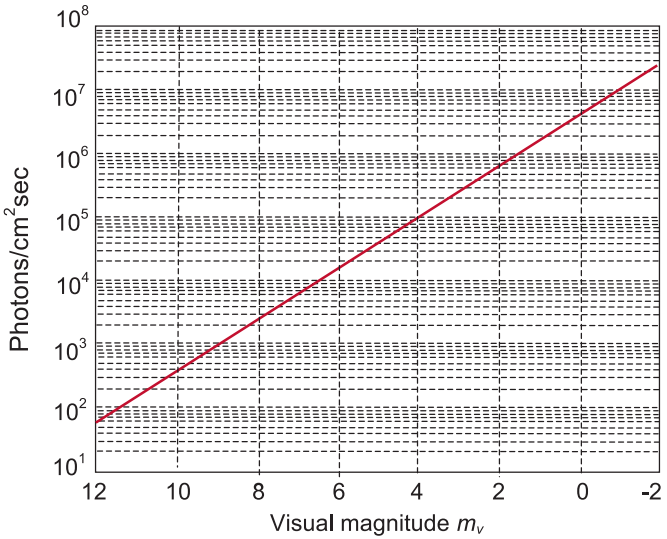
Power in the bucket (PIB) is also commonly used to describe laser performance. In this case, the measured output is integrated over a defined aperture or "bucket" that represents the desired collection area. The bucket size and shape are generally application or target specific, so, while PIB is widely used, it does not provide a standarized representation of performance.

Beam quality provides a measure of the total power at a target, while the Strehl ratio provides the expected peak power density at the target. Beam quality, PIB, and Strehl ratio all provide metrics to score the beam. However, they are only directly comparable for a specified beam profile and bucket size and shape.

Astronomical "Brightness"

The term **brightness** represents the brightness of an object in the heavens. As the object such as a star is observed, the amount of light (number of photons) collected by an aperture (such as the human eye) per second is **astronomical brightness**. The **visual magnitude** m_v of a star is a logarithmic measure of the star's brightness in the visible spectrum. Smaller numbers represent brighter stars; negative numbers represent even brighter stars. One expression that accounts for atmospheric absorption relates visual magnitude to brightness:

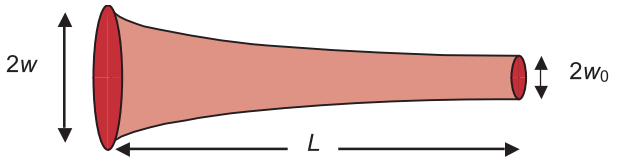
$$B_{\text{Astro}} = \left(4 \times 10^6\right) 10^{\frac{-m_v}{2.5}} \text{ photons/cm}^2\text{sec}$$



It takes about one millisecond for light to pass vertically through the Earth's atmosphere.

Spot Size for a Gaussian Beam

A laser beam spot does not have a well-defined edge. For a **Gaussian beam**, the **spot size** is defined as the diameter $2w$ where the intensity is the value $1/e^2$ or 0.135 times the intensity on the optical axis.



As a Gaussian beam with wavelength λ propagates a distance L, the spot size changes according to:

$$w_0^2 = \frac{w^2}{\left[1 + \left(\frac{\pi w^2}{\lambda L}\right)^2\right]}$$

For long propagation distances or relatively small apertures, even a diffraction-limited focused spot can become much larger than the transmission aperture.

In **atmospheric turbulence**, with a constant C_n^2 over a distance L, the beam spot size grows according to

$$w^2 = \frac{4L^2}{k^2 w_0^2} + 3.58 C_n^2 L^3 w_0^{-1/3}$$

Note that the turbulence term (the second term on the right hand side) is a function of L/w_0 to the third power. It does not take a very long propagation path through turbulence for this term to become dominant. When $L > (w_0^{5/3} k^2 C_n^2)^{-1}$, turbulence beam spreading dominates diffractive beam spreading.

Spot Size for a Uniform Circular Aperture

The **Fraunhofer diffraction pattern** of a laser beam emitted from a circular aperture with uniform (constant) amplitude and no phase **aberrations** is

$$I(r) = \left\{ \frac{2\lambda L}{\pi D r} \left[J_1 \left(\frac{\pi D r}{\lambda L} \right) \right] \right\}^2$$

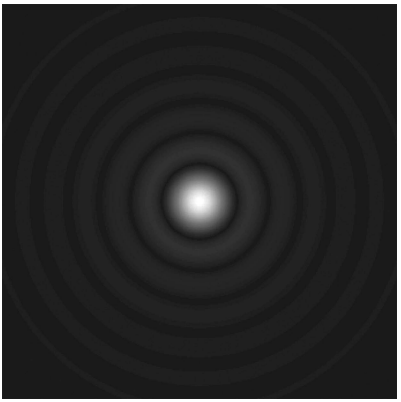
where $I(r)$ is the intensity distribution, r is the radial coordinate in the Fraunhofer plane, D is the aperture diameter, and J_1 is a **Bessel function**. The Bessel function reaches zero for a number of values of r, the first of which is

$$r = 1.22 \left(\frac{\lambda L}{D} \right)$$

This value, where the intensity goes to zero, clearly defines an "edge" of the spot. Thus the **spot size** is

$$2.44 \left(\frac{\lambda L}{D} \right)$$

This spot, called the **Airy disk**, contains 84% of the energy of the beam and can be seen in the figure along with the rings that arise from the zeros in the Bessel function.

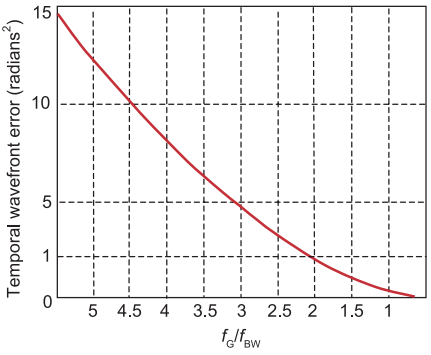


Temporal Error

The **control system** must be able to keep up with changing disturbances. For **Kolmogorov turbulence**, residual **wavefront error** is related to the **closed-loop bandwidth** f_{BW} and the **Greenwood frequency** f_G:

$$\sigma^2_{Temp} = \left(\frac{f_G}{f_{BW}}\right)^{5/3}$$

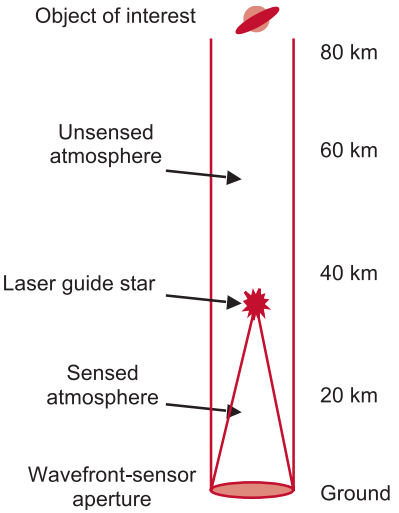
The Greenwood frequency ranges from tens to hundreds of hertz. The adaptive optics system should be capable of responding faster than the turbulence to provide adequate compensation and remain stable, with a factor of 10, as a general rule. This means that the closed-loop bandwidth should range from hundreds to thousands of hertz.



Astronomers **Foy** and **Labeyrie** independently considered using laser backscatter from the atmosphere as an artificial wavefront beacon—a laser **guide star**. However, when funding agencies were presented with the proposal, it was realized that the concept was already a few years old. Highly classified research based on ideas by Julius **Feinlieb**, Richard **Hudgin**, David **Fried**, and others had been going on within laboratories in the U.S. Air Force, Navy, and numerous defense contractors. With the collapse of the Soviet Union in 1991, much of the research was declassified and available to astronomers worldwide.

Focal Anisoplanatism (the "Cone Effect")

When an object in the sky is too dim to make **wavefront sensor** measurements and no bright object is within the **isoplanatic angle** of the object of interest, artificial **laser guide stars** can be produced. Rayleigh scattering in the lower atmosphere (16 to 20 km) can send sufficient light back toward the wavefront sensor, and the laser guide star can be placed near the object of interest.



When a laser guide star is placed a finite distance from the telescope and wavefront sensor, a portion of the atmosphere remains unsensed. The cone of atmosphere that is sensed results in a **cone effect**, which is also called **focal anisoplanatism** because the source of the wavefront is at a focus that differs from the focus of the object of interest.

The **wavefront measurement errors** caused by this effect were investigated by **Fried**, resulting in the expression

$$\sigma^2_{\text{cone}} = \left(\frac{D}{d_0}\right)^{5/3}$$

where the characteristic distance d_0 varies with the atmospheric C_n^2 profile and the altitude of the guide star z_{LGS}. When the **H-V 5/7** model of turbulence is assumed, and the altitude z_{LGS} is expressed in kilometers, then d_0 (expressed in meters) can be approximated by

$$d_0 = 0.018\, z_{\text{LGS}} + 0.39.$$

Laser Guide Stars

The **laser radar equation** for calculating the return flux (photons per square meter) for a laser of given pulse energy E is

$$F_{\text{Rayleigh}} = \eta T_{\text{A}}^2 \frac{\sigma_{\text{R}} n_{\text{R}}}{4\pi z_{\text{LGS}}^2} \frac{\Delta z \lambda_{\text{LGS}} E}{hc},$$

where

$$\Delta z = \frac{4.88 \lambda_{\text{LGS}} z_{\text{LGS}}^2}{D_{\text{Proj}} r_0}$$

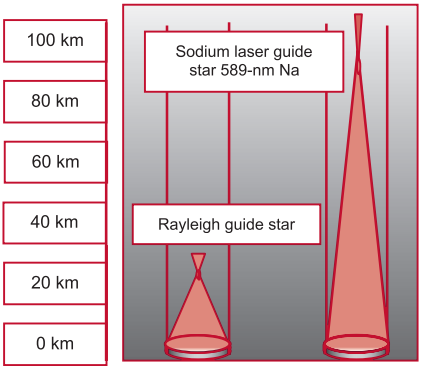
η is the detector efficiency, T_{A} is atmospheric transmission up to the guide star altitude z_{LGS}, $\sigma_{\text{R}} n_{\text{R}}$ is the cross-section-density product, λ_{LGS} is the laser wavelength, h is Planck's constant, and c is the speed of light.

In 1957, **V.P. Linnik** published "On the possibility of reducing the influence of atmospheric seeing on the image quality of stars" in the Russian journal *Optics and Spectroscopy*. This article contained the first mention of artificial **laser guide stars**, predating the invention of the laser by three years.

Volcanic aerosols can greatly enhance the backscatter useful for **Rayleigh laser guide stars**. Unfortunately, we cannot yet accurately predict an eruption. Fortunately, we cannot yet cause an eruption.

The total amount of **atomic sodium** (which is useful for laser guide stars) in the mesosphere could fit inside a large filing cabinet.

Laser Guide Stars (cont.)



To avoid the errors associated with **focal anisoplanatism**, a **laser guide star** is placed at a higher altitude. At an altitude of around 90 km, a layer of atomic sodium can produce backscatter at its resonant wavelength of 589.1583 nm.

Detected sodium-line photon flux:

$$F_{\text{Sodium}} = \eta T_{\text{A}}^2 \frac{\sigma_{\text{Na}} \rho_{\text{Col}}}{4\pi z_{\text{LGS}}^2} \frac{\lambda_{\text{LGS}} E}{hc}$$

where σ_{Na} is the resonant backscatter cross section and ρ_{Col} is the column abundance. The product of the cross section and the abundance is about 0.02.

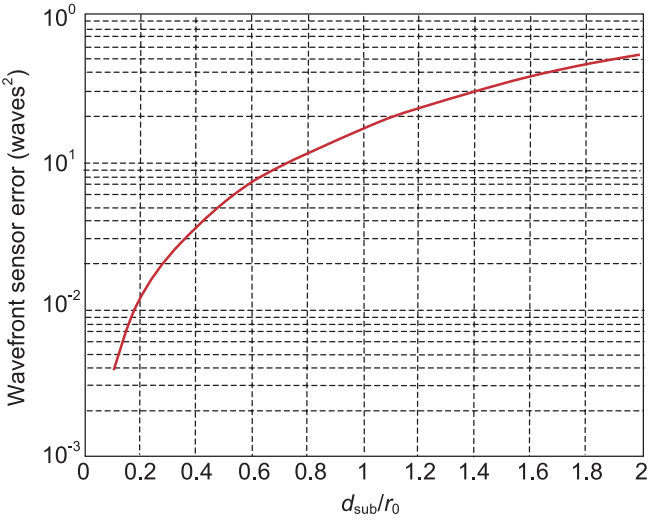
Dr. William **Happer** was the first to suggest using resonant backscatter sodium laser guide stars when he served on JASON, a group of scientists and engineers who advise agencies of the U.S. Government on matters of defense, intelligence, energy policy, and other technical issues.

Subsystem Requirements: The Wavefront Sensor

With a large number of **wavefront sensor subapertures**, irrespective of the type of sensor, the accuracy for measuring the wavefront increases. Assuming **Kolmogorov turbulence**, the **wavefront error variance** from the wavefront sensor is

$$\sigma^2_{\mathrm{WFS}} = 0.17\left(\frac{d_{\mathrm{sub}}}{r_0}\right)^{5/3}$$

for subapertures of size d_{sub}.



Further, assuming that the value of the wavefront measurement is always between $\pm 2.5\,\sigma_{\mathrm{WFS}}$, the **dynamic range** requirement for the wavefront sensor is

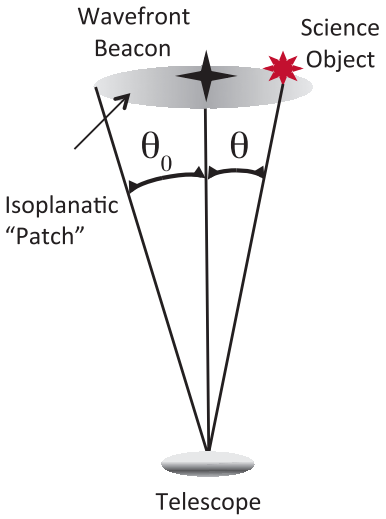
$$\mathrm{Range}_{\mathrm{WFS}} = 2.5\sqrt{0.17\left(\frac{d_{\mathrm{sub}}}{r_0}\right)^{5/3}} \quad \text{(waves)}$$

Angular Isoplanatic Error

The amount of **wavefront error** is dependent on the difference between the direction of the wavefront beacon and the science object. The variance of the wavefront (in radians squared) is

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

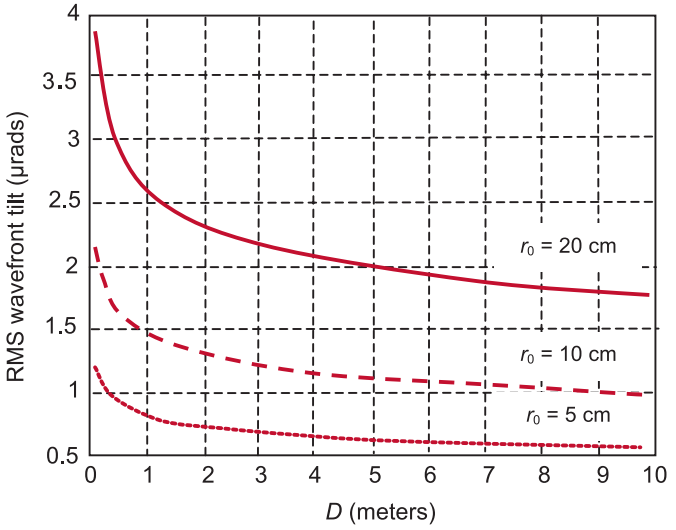
where θ is the angle between the beacon and the science object and θ_0 is the **isoplanatic angle**. The isoplanatic angle is the angle where the wavefront error between the beacon path and the science path differs by 1 radian, or about 1/6 of a wave. The edge of the isoplanatic "patch" does not have a sharp cutoff but rolls off gradually.



Subsystem Requirements: Tilt Mirror

The amount of **wavefront tilt** is dependent on the diameter of the full aperture D. The variance of the wavefront tilt (in radians) is

$$\sigma^2_{\text{Tilt}} = 0.184\left(\frac{D}{r_0}\right)^{5/3}\left(\frac{\lambda}{D}\right)^2$$



A **tilt-corrector mirror** is usually in a compact part of the beam train. The tilt angle is magnified in that part of the beam by the ratio

$$\frac{D_{\text{Telescope}}}{D_{\text{Tilt Mirror}}}$$

The angular tilt that is corrected is twice the angle through which the tilt mirror moves, thus the angular stroke required for the tilt-corrector mirror is

$$\text{Stroke}_{\text{Tilt Mirror}} = \frac{1}{2}(2.5\,\sigma_{\text{Tilt}})\left(\frac{D_{\text{Telescope}}}{D_{\text{Tilt Mirror}}}\right).$$

Subsystem Requirements: How Many Actuators? Zonal or Modal Control

For the case of zonal correction, each **actuator** and its associated wavefront sensor N are related to the **spatial-fitting error**. Inverting the expressions for fitting error for zonal correction,

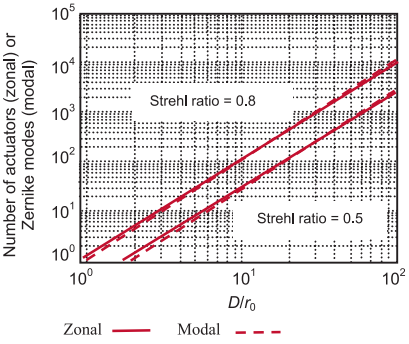
$$N_{\text{Act}} = \frac{\pi}{4}\left(\frac{D}{r_0}\right)^2 \left[\frac{\kappa}{\ln(1/S)}\right]^{6/5}$$

where N_{Act} is the total number of active actuators, D is the aperture diameter, S is the desired **Strehl ratio**, and κ is related to the mirror **influence function**. A typical value for κ is about 0.3, so the relationship reduces to

$$N_{\text{Act}} = 0.18\,[\ln(1/S)]^{-1.2}\left(\frac{D}{r_0}\right)^2$$

If each actuator is recognized as one degree of freedom for the **control system**, that degree of freedom can be translated into a **spatial mode**. By inverting the spatial fitting error for Zernike modal correction, we find that

$$N_{\text{Zern. modes}} = 0.24\,[\ln(1/S)]^{-1.15}\left(\frac{D}{r_0}\right)^{1.92}$$



Thus, the number of **actuators** is roughly equal to the number of **degrees of freedom**, which is roughly equal to the number of corrected **Zernike modes**. The values depend on actuator influence functions, the desired Strehl ratio, the completeness of compensating a specific Zernike mode, and the strength of turbulence given by the ratio D/r_0.

Subsystem Requirements: Deformable Mirror

How far must a **deformable mirror** move? The answer depends on the strength of the **aberrations**. For **Kolmogorov turbulence**, the **tilt**-corrected variance (in units per waves squared) is

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

where D is the aperture diameter, assuming that most of the disturbance is within $\pm 2.5\sigma$. Also, because of the reflection off the mirror surface, one unit of motion of the mirror results in two units of wavefront correction. The total required **stroke** of the deformable mirror is

$$\text{Stroke(waves)} = \frac{1}{2}\left(5\sqrt{0.00357\left(\frac{D}{r_0}\right)^{5/3}}\right)$$

A typical commercially available **actuator** can provide 3.5 μm of stroke, although larger actuators are available that can provide up to 30 μm of stroke. The electronic driver rather than the actuator usually limits the bandwidth; each actuator acts as a capacitor, storing charge as it is excited and discharging as it is released. The action of charging and discharging requires current; larger actuators require more current than smaller ones, and higher frequencies require more current than lower ones. The more current required, the more difficult it is to supply. A higher current can also have the effect of heating up the actuators, possibly degrading performance and reducing stroke and bandwidth.

Early deformable mirrors built for high-energy laser weapons needed to be water-cooled to avoid being damaged by the beam. Today's multilayer coatings enhance reflectivity and reduce absorption and possible damage.

Deformable Mirror Actuator Configurations

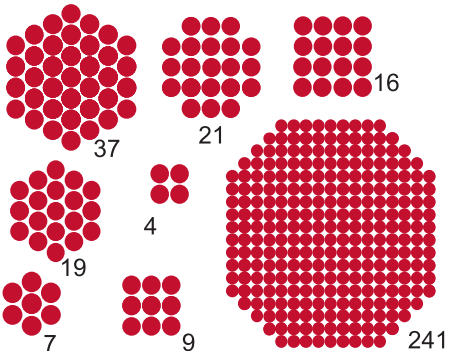
Deformable mirror actuators are typically placed into a square or a hexagonal array. Certain configurations are adjusted to fit a circular beam, and certain "canonical" numbers survive.

Square array:

$$
\begin{gathered}
4, 9, 16, \\
21 (5 \times 5 - 4 \text{ on corners}), \\
37 (7 \times 7 - 12 \text{ on corners}), \\
69 (9 \times 9 - 12 \text{ on corners}), \\
97 (11 \times 11 - 24 \text{ on corners}), \\
241 (17 \times 17 - 48 \text{ on corners}), \\
577 (25 \times 25 - 48 \text{ on corners}), \\
941 (35 \times 35 - 284 \text{ on corners}).
\end{gathered}
$$

Hexagonal array:

$$3, 7, 19, 37, 6, 91, 127, \text{or } 1 + 6 \sum_{n=1}^{N} n.$$



The hexagonal structure of actuators is the most efficient use of space. With the hexagonal configuration, the maximum number of actuators can be placed in a given area. Honeybees, which predate mammals by millions of years, did the research first.

Ferroelectric Actuators

Ferroelectric actuators are made of ceramic materials and are the most common actuators for conventional deformable mirrors. These devices provide constant strain (displacement per unit length) and can be divided into two primary types: **piezoelectric** and **electrostrictive**. The most common material used for piezoelectric actuators is lead zirconate titanate (PZT), and the most common material used for electrostrictive actuators is lead magnesium niobate (PMN).

PZT actuators use the **inverse piezoelectric effect** and have strain linearly proportional to the applied electric field. These devices have randomly oriented dipoles and must be poled prior to use. The act of poling involves heating the ceramic to the Curie temperature (where spontaneous polarization is removed and the piezoelectric effect is no longer present) and then applying a voltage to align the dipole moments. This requirement leads to creep and drift for the active device as the polarization changes. In addition, high dielectric losses with domain wall movement results in significant hysteresis, which in turn causes the actuator to heat up when operated at high frequencies.

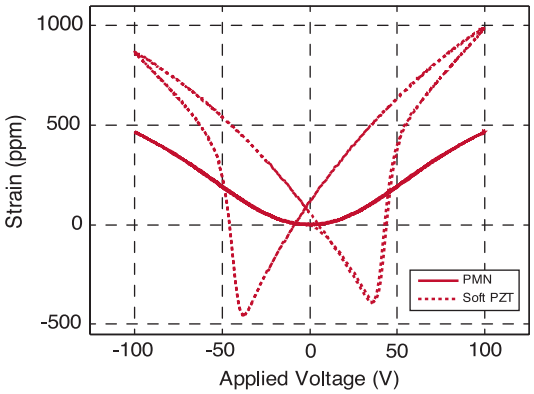
PZT can be doped with acceptors to create hard PZT or with donors to create soft PZT. The difference is that soft PZT has a higher piezoelectric constant and can provide greater strains but generates less output force. Soft PZT also has higher losses and therefore more hysteresis. Hard PZT can generate more force output and has less hysteresis, but at the cost of reduced strain.

PMN actuators use the **electrostrictive effect** and have strain proportional to the square of the applied electric field. These devices do not require poling and are thus insensitive to creep and drift, resulting in inherently higher stability. PMN actuators also have reduced hysteresis and therefore provide better phase response and allow use at higher frequencies without thermal issues.

Ferroelectric Actuators (cont.)

PMN devices have a larger modulus of elasticity than PZT devices, resulting in higher stiffness, energy density, and work output for PMN. Unfortunately, PMN has a thermally induced stroke loss of 2–3% per °C, so PZT is generally used if the deformable mirror must operate under extreme temperatures. PMN is used almost exclusively otherwise, due to the increased force output and decreased creep, drift, and hysteresis.

The figure shows **voltage–strain curves** for soft PZT and PMN actuators. The **hysteresis** is clearly shown and can be as much as 30% for PZT devices, but is typically less than 1% for PMN devices.



The strain of a ferroelectric actuator directly follows the charge on the actuator rather than the voltage. Voltage is used as the reference because it is much simpler to apply a specified voltage than to apply a specified charge.

Once the actuators are integrated into the deformable mirror, the mechanical constraint of the facesheet reduces the total achievable stroke by ~30%. The facesheets are generally designed to handle stresses up to one-half of the total actuator stroke between neighbors, so safety measures are necessary to ensure that this limit is not violated.

Electrostatic Actuators

Electrostatic actuators are used in **micro-electro-mechanical systems (MEMS)** deformable mirrors and follow the laws of general electrostatics. **Coulomb's law** is given by

$$F = \frac{Q_1 Q_2}{4\pi\varepsilon_0 d^2}$$

and indicates that the force generated by an electrostatic actuator is inversely proportional to the square of the distance d between the electrodes with charges Q_1 and Q_2.

Coulomb's law is credited to the French physicist Charles Augustin de Coulomb and was originally published in 1785. Coulomb performed much research in electromagnetic theory and is also known for developing the law of dry friction.

The amount of force that can be generated by an electrostatic actuator is significantly smaller than that generated by a ferroelectric actuator. This results in thinner facesheets and lower overall stroke for MEMS mirrors in comparison to traditional discrete actuator deformable mirrors.

Coating and polishing requirements for handling high **optical power densities** put large stresses on the facesheet. These stresses have traditionally been too extreme for the thinner facesheets found on MEMS deformable mirrors and have limited the amount of energy these devices could handle. With improvements in coating technology, the power density capability of MEMS deformable mirrors has increased, so this issue is not as significant as it once was.

Because voltage is proportional to force, high-voltage drivers are usually required for electrostatic actuators. However, there is very little current draw, and electrostatic actuators consume less total power than ferroelectric actuators. Additionally, electrostatic actuators can respond significantly faster than ferroelectric actuators. The low power consumption, high speed, and ability to achieve extremely high actuator densities are strong draws for MEMS deformable mirror technology.

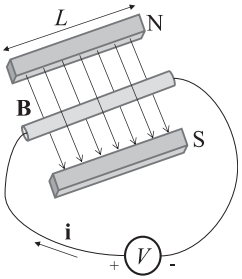
Voice Coil Actuators

Voice coil actuators provide a constant force output and operate on the principle of the **Lorentz force**. They use a moving conductor in a magnetic field to produce a force proportional to the current in the conductor.

As shown in the figure, the applied voltage V creates a current through the conductor. When the current passes through the magnetic field, a force is generated that is equal to

$$\mathbf{F} = k\mathbf{B}L\mathbf{i}N$$

where k is a constant, **B** is the **magnetic flux density magnitude**, L is the length of the coil, **i** is the magnitude of the current passing through the coil, and N is the number of conductors. Since **B** and **i** are both vectors, the force is a vector as well, and its direction is specified by the cross product of **i** and **B**, as shown in the figure.



When the coil moves through the magnetic field, it produces a **back electromotive force (EMF)** that opposes the applied voltage. When an input voltage is applied to the voice coil, it accelerates until the back EMF is equal to the input voltage and then moves with a constant velocity. This is referred to as the **limiting velocity**; the controller generally prevents reaching the limiting velocity.

Voice coils use the same electronic drivers as brushed DC motors and are generally used in steering mirrors.

Deformable Mirror Influence Function Models

Two expressions are widely used to describe the **influence function** of a continuous-faceplate **deformable mirror**.

The deflection normal to the mirror surface can have a cubic relationship:

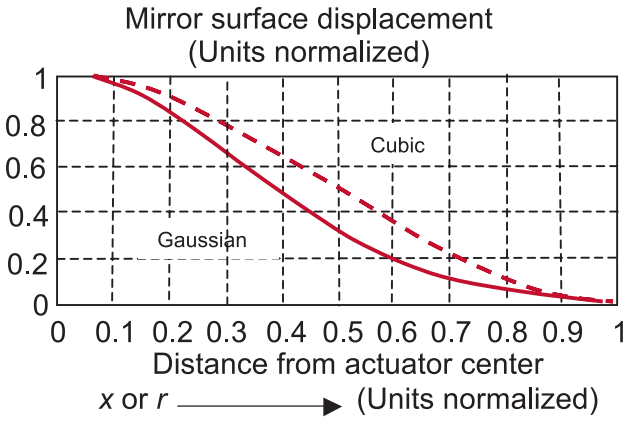
$$\phi_{\text{Cubic}}(x,y) = A_{\text{Infl}}\left[\left(1-3x^2+2x^3\right)\left(1-3y^2+2y^3\right)\right]$$

where (x,y) are Cartesian coordinates, and A_{Infl} is the amplitude of the influence function. The origin is at the actuator location.

The other expression has a Gaussian form:

$$\phi_{\text{Gaus}}(x,y) = A_{\text{Infl}} \exp\left[\frac{\ln c_a}{r_c^2} r^2\right]$$

where r is the polar radial coordinate in the mirror plane, r_c is the interactuator spacing, and c_a is the **coupling** between **actuators** expressed as a number between 0 and 1. The coupling is the movement of the surface at an unpowered actuator expressed as a fraction of the motion of its nearest-neighbor actuator.



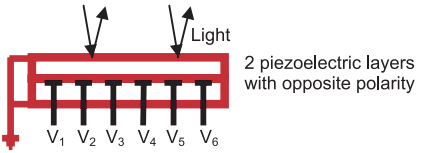
Bimorph and MEMS Mirrors

The deflection of a **bimorph mirror** at distance r from the center of an **actuator** pad is

$$\phi_{\text{Bimorph}} = \frac{V}{a_0 t^2}\left[r_a^2 - r^2 - 2r_a^2 \ln\left(\frac{r_a}{r_m}\right)\right] \text{ for } r < r_a$$

and

$$\phi_{\text{Bimorph}} = \frac{V}{a_0 t^2}\left[-2r_a^2 \ln\frac{r_a}{r_s}\right] \text{ for } r_a < r < r_m$$



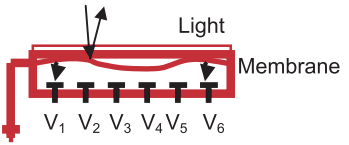
V is the applied voltage, t is the thickness of the bimorph, r_a is the actuator pad radius, r_s is the radius of the supporting ring, r_m is the mirror radius, and a_0 is the piezoelectric constant (related to the stress tensor).

The deflection of a MEMS **membrane mirror** at a distance r from the center of an actuator pad is

$$\phi_{\text{Membrane}} = \frac{\varepsilon_0 V^2}{4T_m d^2}\left[r_a^2 - r^2 + 2r_a^2 \ln\frac{r_a}{r_m}\right] \text{ for } r < r_a$$

and

$$\phi_{\text{Membrane}} = \frac{\varepsilon_0 V^2}{4T_m d^2}\left[2r_a^2 \ln\frac{r_a}{r_m}\right] \text{ for } r_a < r < r_m$$

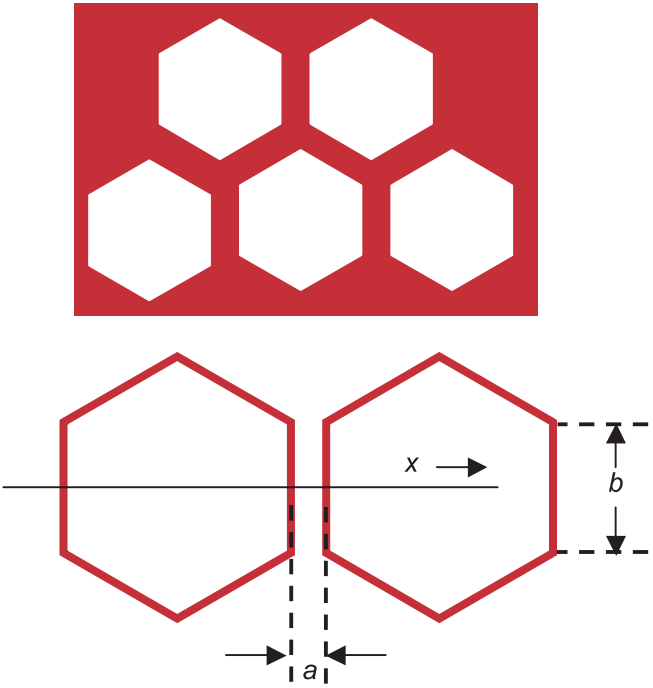


V is the applied voltage, d is the separation of the membrane and the addressing electrode, and ε_0 is the permittivity, 8.85×10^{-12} F m^{-1}.

Because the membrane is so thin, very little power is required to deflect the membrane. While voltages can be a few hundred volts, the electrical current is very small.

Segmented Deformable Mirrors

Segmented deformable mirrors exhibit a loss of light in the gaps between segments. In addition, the regular geometric pattern of segments acts as a diffraction grating.

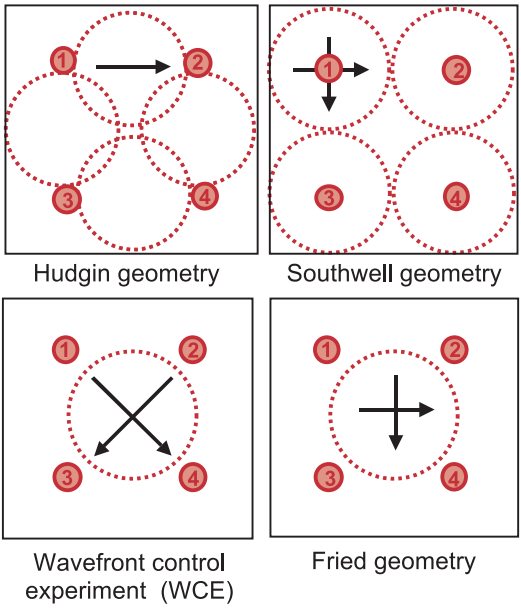


The **diffracted energy** from the effect of the gaps I_{gap} is proportional to the width of the gap a and the size of the segment b, according to the relation

$$I_{\text{gap}} \propto a^2 b^2 \left[\frac{\sin(2\pi a x/\lambda)}{(2\pi a x/\lambda)} \right]^2$$

Actuator and Wavefront Sensor Layouts

The arrangement of **actuators** and **wavefront sensor subapertures** (the registration) affects the control algorithm and the stability of the **control system**. In the figure, the numbered small circles represent actuator positions, and the dotted large circles represent the subapertures with the orthogonal **slope** measurements represented by the arrows.



Many other geometries, including random registration of the wavefront sensor and the **deformable mirror** can be used with proper calibration and mechanical stability.

The actuator spacing is also a requirement in an adaptive optics system. The **Nyquist sampling theorem** states that spatial frequencies greater than half the sampling frequency cannot be observed. This means that the actuator spacing must be less than half the smallest required spatial frequency period in order to provide the required correction.

Correctability and Flattening of a Deformable Mirror

A deformable mirror has a spatial transfer function that is dependent on the geometry of the actuators and the shape of the influence functions. Rather than using the entire frequency spectrum, we generally define the **correctability** of the deformable mirror as its ability to compensate for specific aberrations or OPDs. The correctability is calculated from finite element modeling and provided by the vendor.

Correctability is a percentage rather than an absolute value. Therefore, the magnitude of the residual surface error scales with the input OPD.

Flattening a deformable mirror is performed by commanding the actuators to produce the minimal wavefront error. The commands represent the flat file and can be generated for the deformable mirror by itself or as part of the entire system.

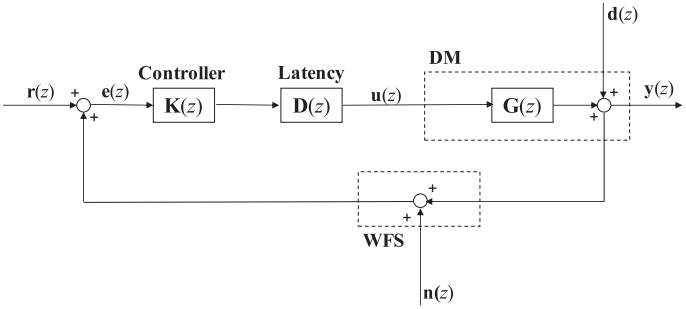
For the deformable mirror flat, the mirror is placed in front of an interferometer so that the resultant flattened wavefront error is simply the deformable mirror surface. For the system flat, the loop is closed with a reference beam, and the flattened wavefront includes the effect of optical errors along the entire path. Neither of these flattened wavefronts is a measure of correctability.

A flat file provides initial or resting conditions to the system by applying a starting figure to the deformable mirror. This figure does not affect dynamic performance of the adaptive optics system but improves tracking and jitter loop performance when the adaptive optics loop is not engaged, and is extremely useful during system alignment.

An input OPD generally cannot be reduced to the flattened residual, so this result should not be used to set performance expectations.

Adaptive Optics System Feedback Configuration

A standard configuration for **feedback control** of an adaptive optics system is shown in the figure using discrete-time subsystems.



In this block diagram, z is the Z-transform variable, **K** is the digital controller, **G** is the deformable mirror transfer function, **r** is the reference input (which is nominally **0**), **u** is the control signal, **y** is the corrected output beam, **d** is the input aberrated beam, **n** is the wavefront sensor noise input, **e** is the error signal, and **D** contains all of the system latencies.

This diagram uses a positive-feedback loop because the reconstructor is typically designed to enforce negative feedback.

Because adaptive optics systems are concerned with correcting the input aberrated beam, of primary interest is the **sensitivity function** **S**, which is the disturbance rejection function and defines the transfer function from **d** to **y**:

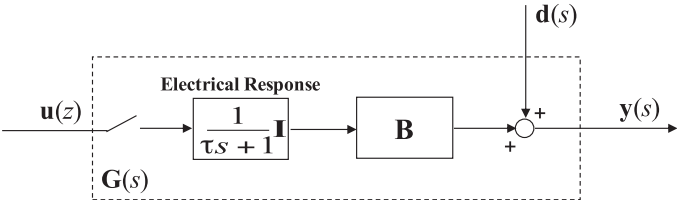
$$\mathbf{S} = [\mathbf{I} + \mathbf{GK}]^{-1}$$

The complementary sensitivity or **closed-loop transfer function** **T** from **r** to **y** is also important:

$$\mathbf{T} = [\mathbf{I} + \mathbf{GK}]^{-1}\mathbf{GK}$$

Deformable Mirror Dynamic Model

The block diagram for a deformable mirror is given in the figure.



The **deformable mirror dynamic model** contains an ADC that samples the digital command, a first-order transfer function based on the actuator time constant τ, the **poke matrix** that converts actuator space to sensor space, and a summing junction (facesheet) that provides the input for the aberrated beam. The mechanical resonance of the mirror is usually high enough that it can be ignored. The **discrete-time transfer function** is given by

$$\mathbf{G} = \mathbf{B}\frac{1 - e^{-T_s/\tau}}{z - e^{-T_s/\tau}}$$

The discrete-time state-space equations with actuator positions as states and measurements as outputs are given by

$$\mathbf{x}_{k+1} = e^{-T_s/\tau}\mathbf{x}_k + \left(1 - e^{-T_s/\tau}\right)\mathbf{u}_k$$
$$\mathbf{y}_k = \mathbf{B}\mathbf{x}_k + \mathbf{d}_k$$

When the **sampling time** T_s is long compared to the time constant τ, the exponential is small and the position update depends less on the previous position and more on the control signal, indicating a fast response relative to the sampling time. When the sampling time is short compared to the time constant, the exponential is large and the position update depends more on the previous position and less on the control signal, indicating a slow response relative to the sampling time.

Controller Dynamic Model

The standard controller of choice for an adaptive optics system is a simple **leaky integrator**, which is implemented by the digital filter

$$\mathbf{K}(z) = \frac{g_1 z}{z - g_2}\mathbf{R}$$

Here g_1 is the forward loop gain, g_2 is the leak gain, and $\mathbf{R}$ is the reconstructor, which is usually generated to provide negative feedback.

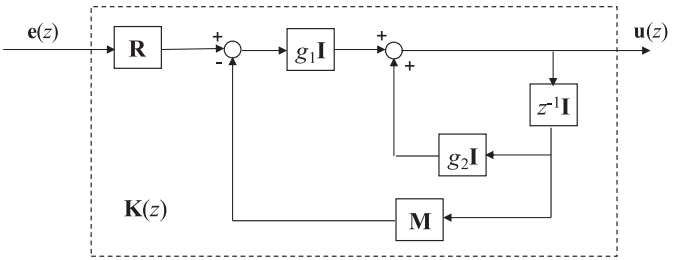
This controller is implemented by the difference equation

$$\mathbf{u}_k = g_1\mathbf{R}\mathbf{e}_k + g_2\mathbf{u}_{k-1}$$

A modal feedback term can be included to remove modes such as piston from the actuator commands by

$$\mathbf{u}_k = g_1\mathbf{R}\mathbf{e}_k + g_2\mathbf{u}_{k-1} - g_1\mathbf{M}\mathbf{u}_{k-1}$$

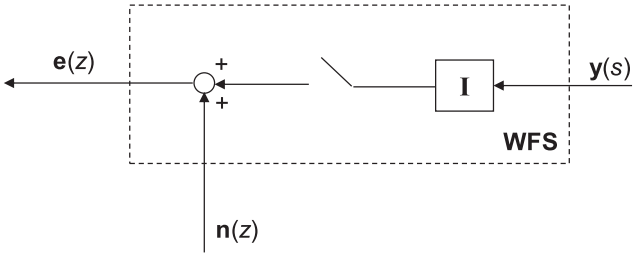
The block diagram for this **controller dynamic model** is given in the figure.



In conventional adaptive optics systems, the modal feedback matrix $\mathbf{M}$ removes only piston from the commands.

Wavefront Sensor Dynamic Model

The **wavefront sensor dynamic model** consists of an ADC, a noise input, and a first-order system that represents the camera integration time. Because the integration time is usually very small compared to the sampling time, it can be approximated by a pure delay and is typically included in the total system latency, along with the camera readout time. The resulting dynamic sensor model is shown in the figure.



The noise source includes sensor noise and algorithmic errors as well as data conversion and quantization noise. Quantization is the process of converting the intensity value from an infinite number of possible levels to a finite number of quantized levels that can be represented on a computer. Monochromatic images are typically found with resolution in 8, 10, 12, 14, or 16 bits, while color images typically use 32 bits with 1 byte (8 bits) each for the red, green, blue, and alpha channels. Adaptive optics systems generally use monochromatic cameras with ≥12 bits of resolution.

In many cases the direct wavefront sensor measurements are converted to another basis such as phase, Zernike, or Legendre polynomials. Since the error vector $\mathbf{e}$ contains the measured wavefront in the chosen representation, this model is valid as long as the reconstructor and poke matrices are generated using the same representation.

In discrete time, the **wavefront sensor output** is

$$\mathbf{e}_k = \mathbf{y}_k + \mathbf{n}_k$$

In most cases, the measurement noise $\mathbf{n}$ is assumed to be Gaussian with zero mean and variance $\sigma_n^2\mathbf{I}$.

Latency

We can combine all of the system delays into a single **latency** value t_d. This value consists of camera readout and integration time, computational processing time for the control algorithm, data transport time to and from the processor, as well as data conversion and additional processing time for the driver.

In continuous time, a **delay** is represented by

$$\mathbf{D}(s) = e^{-t_d s}\mathbf{I}$$

For first-order analytical assessment of a sampled data system, we can use the **discrete-time model** of the delay, given by

$$\mathbf{D}(z) = z^{-\rho}\mathbf{I}$$

where ρ is an integer equal to the ceiling of the latency divided by the sampling time:

$$\rho = \text{ceil}\left(\frac{t_d}{T_s}\right)$$

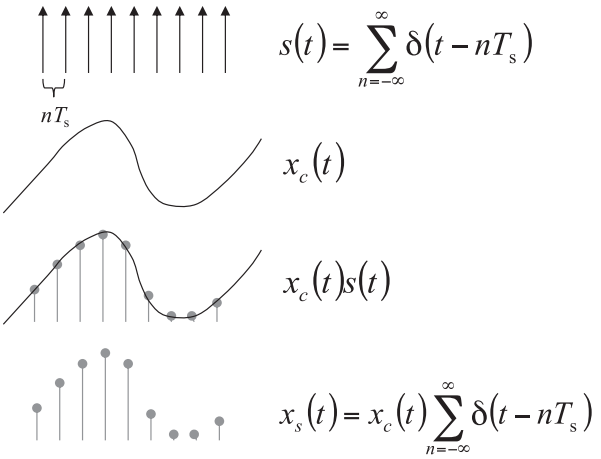
The act of sampling induces a single frame of latency for observed commands given any finite t_d. Minimal performance degradation occurs until the delay becomes large enough to push the actuator response out an additional frame.

There is no standard convention as to whether or not the default frame is included when discussing latency, so care should be taken to avoid confusion. Referring to a system with 2 frames of latency could mean the system delays induce either 1 or 2 additional frames of latency.

The discrete-time model provides a delay of an integer number of samples and does not capture transients within the sampling time that arise from the higher-order dynamics of the deformable mirror, particularly the actuator rise time. Generally, the performance specification for an adaptive optics system requires a latency value such that the actuator is able to reach some percentage (typically 60–90%) of its commanded value within one frame.

One-Dimensional Sampling

The mathematical description of **one-dimensional sampling** uses an idealized ADC that works by modulating the input signal x_c with a periodic impulse train s to create the output signal x_s that is sampled at the sampling instants nT_s as shown in the figure.



The Fourier transform of x_s is given as

$$X_s(j\Omega) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} X_c(j\Omega - k\Omega_s)$$

where $\Omega_s = 2\pi/T_s$ and is the **sampling frequency** in rad/s. This equation shows that X_s is made up of copies of X_c that are shifted by integer multiples of the sampling frequency. If the distance between copies is too small, the frequency spectra overlaps in a process known as **aliasing**.

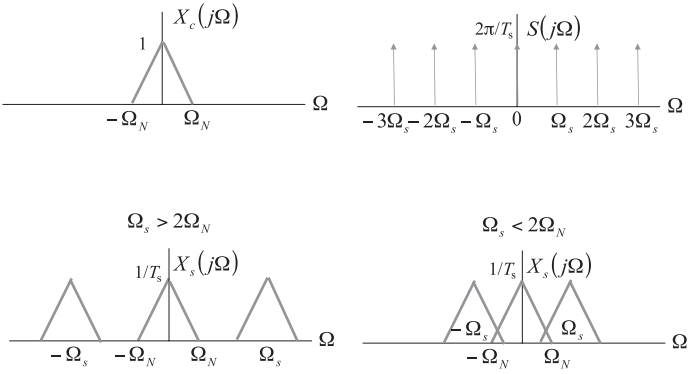
If Ω_N is the highest frequency contained in X_c, then the requirement to prevent aliasing is

$$\Omega_s > 2\Omega_N$$

From this equation, the highest frequency we can accurately represent is one-half the sampling frequency

One-Dimensional Sampling (cont.)

and is known as the **Nyquist frequency**. The effect of aliasing is shown in the figure.



The upper left-hand corner of this figure shows the initial spectrum of X, and the upper right-hand corner shows the spectrum of the pulse train S. The lower left-hand corner plot shows the sampled signal spectrum X_s without aliasing, where $\Omega_s > 2\Omega_N$, and the lower right-hand corner plot shows the sampled signal spectrum X_s with aliasing. where $\Omega_s < 2\Omega_N$.

The Nyquist frequency requirement provides an upper bound on sampling and reconstruction from an open-loop signal-processing perspective. Closed-loop systems have stability requirements that further restrict the system sampling rate.

Harry Nyquist's original work on sampling was published in 1928 and was expanded to include Claude Shannon's signal reconstruction in 1949. The Nyquist–Shannon sampling theorem is one of the foundations of information theory.

Two-Dimensional Sampling

Frequencies in two dimensions are spatial and are in units of cycles per unit length. The theory of **sampling in two dimensions** is an extension of the one-dimensional case to the spatial domain. If g_c is the input function, the sampled signal in two dimensions is given as

$$g_s(x,y) = g_c(x,y) \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \delta(x-nb)\delta(y-md)$$

where $1/b$ is the spatial sampling rate along the x axis and $1/d$ is the spatial sampling rate along the y axis. The spatially sampled signal is evaluated at the sampling points nb and md.

Real-image sensors have finite-sized pixels arranged in a uniform grid. For sampling and reconstruction analysis, the pixels are assumed to provide infinitely small point samples.

The spatial frequency domain representation of g_s is

$$G_s(f_x,f_y) = \frac{1}{bd} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} G_c\left(f_x - \frac{n}{b}, f_y - \frac{m}{d}\right)$$

The Nyquist frequency to prevent aliasing is applicable in both the x and y axes. If X_N is the peak spatial frequency of G_c along f_x and Y_N is the peak spatial frequency of G_c along f_y, aliasing occurs whenever

$$X_N > \frac{1}{2b} \text{ or } Y_N > \frac{1}{2d}$$

The spatial frequency content of even low-order optical aberrations is not strictly bandlimited, so some spatial aliasing is present in almost all cases. Aliasing can be mitigated by oversampling, but the best results are obtained by using a field stop as a spatial filter to provide a nearly bandlimited spectrum.

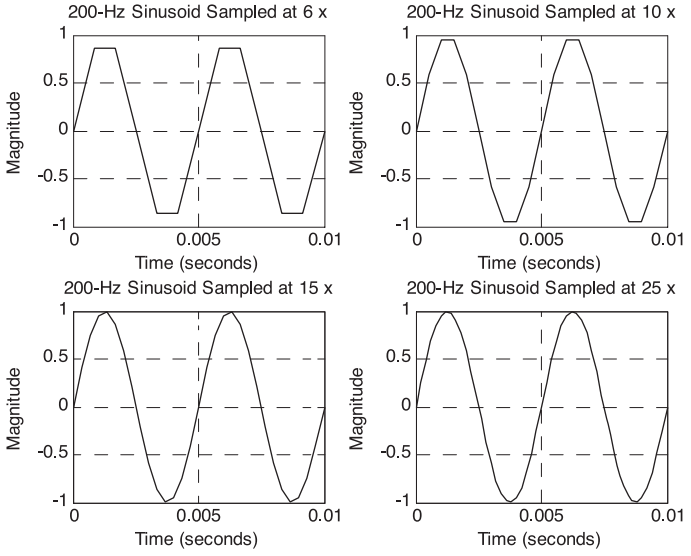
A Shack–Hartmann-type sensor measures the slope of the phase, so sampling requirements should be determined with consideration of the first derivative of the phase.

Temporal Sampling Rate Selection

For stability requirements, the ratio of the sampling frequency to the crossover frequency should be between 6 and 40:

$$6 \leq \frac{\omega_s}{\omega_c} \leq 40$$

In general, the higher the ratio, the smoother the response. The idea of smoothing can be shown by sampling a sinusoid at various multiples of its frequency, as shown for a 200-Hz sinusoid sampled at 6×, 10×, 15×, and 25× in the figure.



From this figure, we can see that the output signal is not smooth until the ratio of the sampling frequency to the frequency of interest is > 15. A general rule of thumb for digital control system design is to select a **sampling rate** that is 20–30 times greater than the target crossover frequency.

System Stability

The discrete-time open-loop transfer function L for a single channel including only the deformable mirror and controller, with modal feedback ignored, is given as

$$L(z) = \frac{1 - e^{-T_s/\tau}}{z - e^{-T_s/\tau}} \frac{g_1 z}{z - g_2}$$

This system has two poles: at $z = e^{-Ts/\tau}$ for the deformable mirror and $z = g_2$ for the controller. The discrete-time **stability** requirement for poles is that $|z_i| < 1$.

For the deformable mirror pole, we see that $|z_i| \to 1$ as $T_s/\tau \to 0$. In other words, the deformable mirror provides a stable plant until the system sampling time becomes much smaller than the time constant of the actuators, at which point it approaches marginal stability. With rise times under 100 μs, this requires sampling >> 100 kHz. Since we do not have sensors capable of reaching these high frame rates, we are guaranteed a stable plant.

If there is no leak ($g_2 = 1$), the controller pole is on the unit circle and the system is marginally stable. Because the poke matrix provides static coupling between the channels, marginal stability results in local oscillatory behavior. To account for this, we use a small amount of leak ($0.99 \le g_2 < 1.0$) to relax the pole and add robustness.

Because the reconstructor and poke matrix normalize the system, the loop gain is simply g_1. This gain should be reduced to provide the desired robustness specifications and is typically less than 0.5.

In classical control theory, stability is determined from frequency-based techniques originally developed by Harry Nyquist in the 1930s. Modern multivariable control utilizes time-based analysis, which was developed by Aleksandr Lyapunov and gained interest in the 1960s.

General Control-System Parameters

To eliminate confusion in notation, several key control system parameters are given in the following table.

Parameter	Symbol	Definition
Rise time	t_r	Time required for a step to reach 90% of the commanded value.
Settling time	t_s	Time required for a step to settle to within 5% of the final value.
Overshoot	OS	Percentage increase of the peak output response over the final value.
Crossover frequency	ω_c	Frequency where the open-loop system gain crosses unity (0 dB).
Phase margin	PM	Phase shift required to reach 180 deg of phase at ω_c. With negative feedback, this shift induces constructive interference and results in an unbounded output.
Gain margin	GM	Gain increase required to reach unity (0 dB) at the frequency where the open-loop phase reaches 180 deg. As with PM, if the system gain is above unity when the phase reaches 180 deg, the output increases and becomes unbounded.
Bandwidth	ω_b	Generally used interchangeably with crossover frequency. Sometimes defined as the frequency where the sensitivity function crosses −3 dB from below to identify the region where the compensation is working well.

Sensitivity Functions

The **sensitivity and complementary sensitivity functions** define the system performance for specified inputs. The sensitivity function **S** is responsible for compensating the input atmospheric disturbances, while the complementary sensitivity function **T** provides compensation of noise sources as well as command following for tracking.

Having a small sensitivity function improves the disturbance rejection of the loop, and having a small complimentary sensitivity function improves the noise rejection and tracking performance of the loop. However, because these functions are complementary,

$$\mathbf{S} + \mathbf{T} = \mathbf{I}$$

we have a **design trade-off** since we cannot make both functions small.

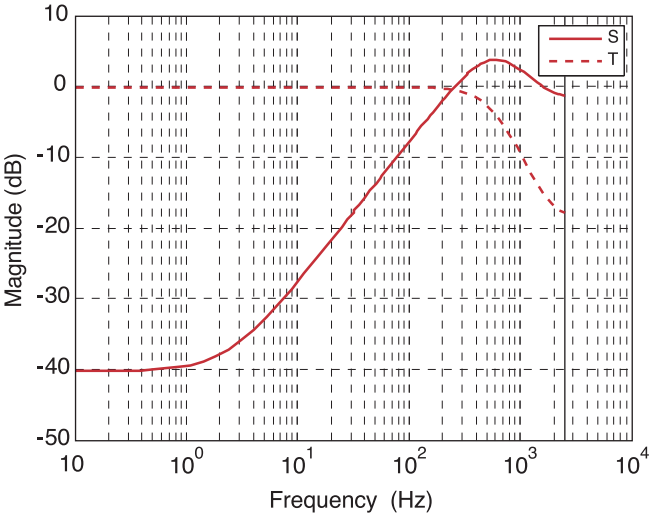
With this in mind, the standard control design aim is to make the sensitivity function small at low frequencies and the complementary sensitivity function small at high frequencies to provide good **low-frequency disturbance rejection** and **high-frequency noise rejection**. This works well in practice, as disturbances are generally confined to lower frequencies and noise is present at higher frequencies.

The single-channel sensitivity and complementary sensitivity functions are shown in the figure on the next page for an adaptive optics system with $g_1 = 0.3$, $g_2 = 0.997$, $\tau = 50$ μs, $t_d = 100$ μs, and $f_s = 5$ kHz.

This system has a sensitivity peak M_s of 3.77 dB at 602 Hz, indicating that disturbances near 602 Hz are amplified by a factor of 1.5. This loop shape is typical for most control systems and is driven by system performance and robustness. The sensitivity function provides ~40 dB of compensation (~100 × reduction) for disturbances under 10 Hz and 3 dB of compensation at ~175 Hz, indicating that disturbances under 175 Hz are well corrected.

Sensitivity Functions (cont.)

The crossover frequency from the open-loop Bode plot is 244 Hz, so the ratio of sampling rate to crossover frequency is 20.5.



Another general design rule of thumb is to keep both the sensitivity and complementary sensitivity peaks under 2 (6 dB). Restrictions on gain and phase margins (GM and PM) impose restrictions on the sensitivity peaks as well:

$$\text{GM} \geq \frac{M_s}{M_s - 1} \quad \text{PM} \geq \frac{1}{M_s}$$

Therefore, the **robustness requirement** can be specified either for PM and GM or for M_s.

Adaptive optics systems are decentralized, meaning that the system dynamics are diagonal and therefore uncoupled. This allows the dynamic and static portions to be treated separately by combining a matrix reconstructor with leaky integrators for the individual channels. In this sense, the reconstructor can be thought of as an estimator.

Bandwidth Estimation from Controller Gains

We can generate a first-order estimate of the **system crossover frequency** solely from the **loop and leak gains**, ignoring the plant dynamics. If the plant dynamics are not significant at low frequencies relative to the sampling time, this estimate is accurate.

The crossover point is defined as the frequency where the gain is unity, so we can formulate the problem for a general discrete-time transfer function as

$$\left|K\left(e^{j\omega}\right)\right|^2_{\omega=\omega_c} = 1$$

Using the single-channel controller model, substituting $z = e^{-j\omega}$, and solving for ω_c in Hz yields

$$\omega_c = \cos^{-1}\left(\frac{g_2^2 + 1 - g_1^2}{2g_2}\right)\frac{f_s}{2\pi}$$

Again, g_1 is the forward loop gain, g_2 is the leak gain, and f_s is the sampling frequency.

We can also calculate the system gain or compensation at DC (steady state) by simply setting $z = 1$ in the controller model:

$$g_{DC} = \frac{g_1}{1 - g_2}$$

The DC gain is infinite with no leak, and finite with a leak term applied, which means that the leak term rolls off the achievable steady state correction. This is consistent with the understanding that a pure integrator (leak gain = 1) in either continuous or discrete time is used for forcing the steady state error to zero.

These estimations provide $\omega_c = 240$ Hz and $g_{DC} = 40$ dB for the example used previously and are close to the simulated values.

When no leak is used, the controller becomes a standard proportional-integral (PI) controller with the integrator defined by the forward-difference transform. Application of a leak term shifts the controller so that it is no longer realizable in the traditional PI form.

Poke Matrix

The **poke matrix** captures the response of the deformable mirror as observed by the wavefront sensor and converts actuator space to sensor space. An empirical poke matrix is generated by measuring the vector of wavefront sensor outputs produced by poking a single actuator. This process is repeated for each actuator to determine a unique **influence function**:

$$\inf(a_i) = \mathbf{b}_i$$

where $\inf(a_i)$ is the influence function of the ith actuator and $\mathbf{b}_i$ is the vector of wavefront sensor measurements observed when poking the ith actuator.

To cover issues of symmetry, each actuator should be given a positive and negative poke, and the responses averaged to generate the influence function:

$$\inf(a_i) = \frac{\mathbf{b}_i^+ - \mathbf{b}_i^-}{2}$$

When generating influence functions, the measurements should be delayed until the actuator has reached its final position. At that time, multiple measurements should be taken to average for noise reduction.

The influence function for the ith actuator is the ith column in the poke matrix, normalized by the actuator command used as a poke α to maintain the correct scaling.

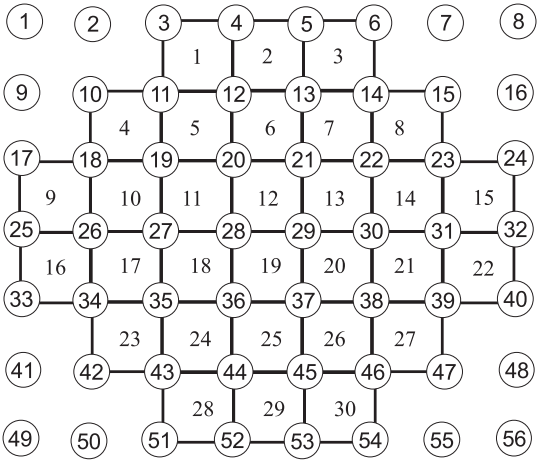
$$\mathbf{B} = \left[\frac{\mathbf{b}_1}{\alpha} \; \frac{\mathbf{b}_2}{\alpha} \; \cdots \; \frac{\mathbf{b}_n}{\alpha} \right]$$

The poke command α should be large enough to provide good SNR without saturating the commands or measurements.

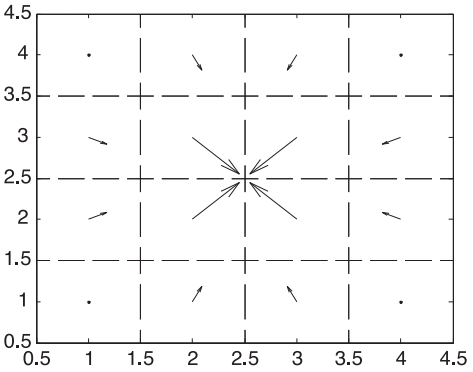
The poke matrix defines the system geometry and includes the effects of static coupling between the actuators. This is a static-system representation as there is no dynamic coupling between neighboring actuators.

Example System Geometry

To demonstrate poke matrix analysis and reconstructor generation, we define an **example system geometry** using the **Fried configuration** with 30 subapertures and 56 actuators, as shown in the following figure.

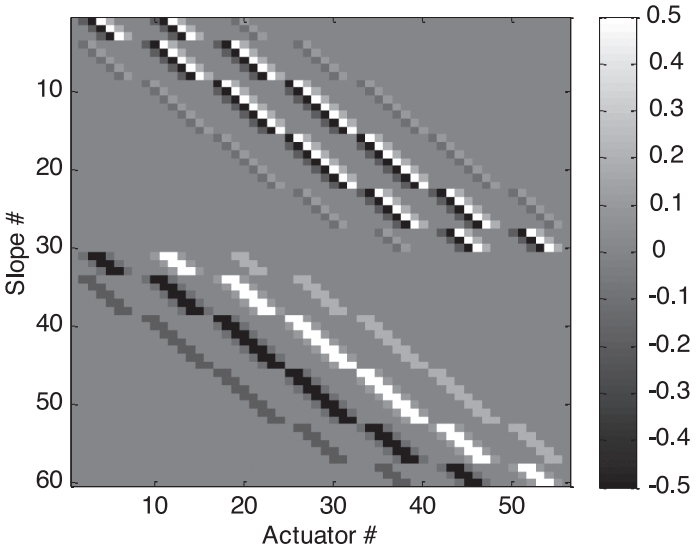


Using a cubic influence function approximation yields the following actuator influence function represented in slopes for a Shack–Hartmann sensor across a 4×4 subaperture area.



Example System Geometry (cont.)

The poke matrix generated from this influence function is shown in the figure.



The columns of the poke matrix represent the actuators, and the rows represent the slopes. In this case, the **slopes** are split so that slopes 1–30 contain the x slope values, while slopes 31–60 contain the y slope values.

When viewing a poke matrix as an image, the key is to look for clearly distinct areas of influence that form roughly diagonal blocks. The number and location of these areas depends on the geometry of the system and are not of great importance. The "sharpness" of the locations indicates a clean, low-noise response.

Singular-Value Decomposition of the Poke Matrix

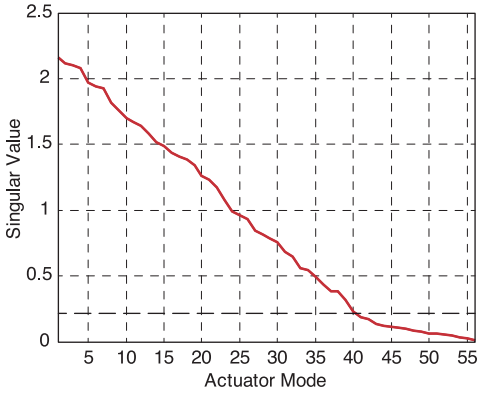
The poke matrix contains information on the actual geometry of the system, including influence function coupling and misalignment. Using a measured poke matrix to build the reconstructor provides the best system representation, as the use of an analytical poke matrix can result in gain errors due to system misalignments and actuator variations.

Of particular importance in analyzing a poke matrix is the **singular-value decomposition** (SVD), which is given by

$$\mathbf{B} = \mathbf{U}\Sigma\mathbf{V}^T$$

Here T is the matrix transpose operator, $\mathbf{U}$ is a matrix containing the left (output) singular vectors ($\mathbf{u}_i$) along its columns, $\mathbf{V}$ is a matrix containing the right (input) singular vectors ($\mathbf{v}_i$) along its columns, and Σ is a matrix containing the corresponding singular values (σ_i) along its diagonal. The right singular vectors are the deformable mirror modes, and the left singular values are the corresponding wavefront sensor modes.

A plot of the singular values determines how many modes can be well corrected. The figure shows the singular values for the example poke matrix. The dashed line indicates singular values that are less than 10% of the peak magnitude.



Actuator and Subaperture Observability

We define the observability of an actuator as the response of that actuator as seen by all of the measurement points, and the observability of a measurement point as the responses of all the actuators as seen by that point. The **actuator observability** is then the RSS of the columns of the poke matrix, while the measurement observability is the RSS of the rows of the poke matrix. This is shown by the following equations:

$$\mathrm{obs}(a_j) = \sqrt{\sum_{i=1}^{n_\mathrm{m}} (B_{ij})^2}$$

$$\mathrm{obs}(m_i) = \sqrt{\sum_{j=1}^{n_\mathrm{a}} (B_{ij})^2}$$

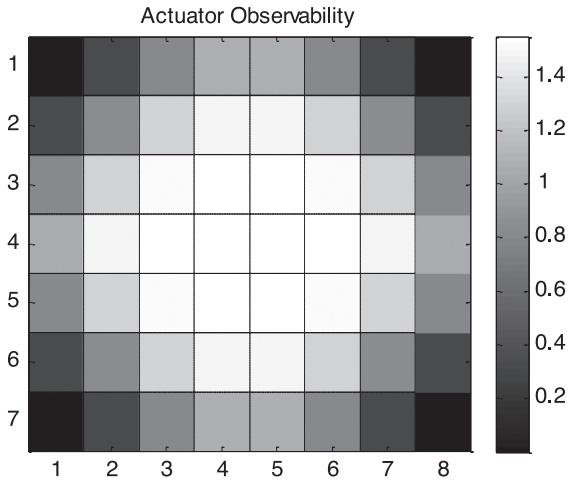
For traditional Shack–Hartmann sensors, the **subaperture observability** is the RSS of the corresponding x- and y-measurement observability.

The observability determines how responsive each actuator or measurement point is relative to the others. It can be used to identify weak or unresponsive actuators that need to be **slaved**, as well as poorly lit measurement points that should be removed from active consideration as they contribute only noise to the system.

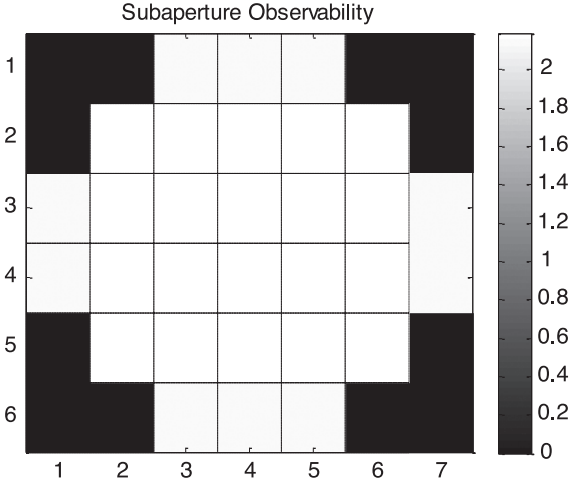
The actuator observability for the example system is shown in the figure on the next page, where we see that the observability of the central actuators is high and drops off toward the edge of the wavefront sensor grid.

In classical control theory, an observable system is one in which all of the system states can be determined from the output measurements. Poorly observable actuators (subapertures) are related to poorly observable input (output) modes.

Actuator and Subaperture Observability (cont.)



The subaperture observability for the example system is shown in the following figure. In this case, the actual difference in the observability is small, indicating a high degree of uniformity for the defined subapertures, as expected for a purely analytic poke matrix.



Identification of Actuator Locations

Actuator locations can be computed from the poke matrix using either slope or phase measurements. To maintain accuracy, either approach should be restricted to using internal actuators that are fully contained in the **subaperture grid**. Actuators that have one or more edges uncovered by subapertures have large error bars in the position estimation.

For phase measurements, we can interpolate the magnitude of the phase for each actuator poke and determine the peak. The actuator center should be at the peak location.

For slope measurements, we can use the fact that all of the slopes should be pointing toward the actuator center and look for the intersection point of the slopes. This is accomplished by first defining the equation for each slope vector:

$$y = ax + b$$

Here the slope a and slope intercept b for each equation are referenced to a common origin. The solution can then be written in terms of a least-squares equation for n slope vectors, with the intersection point defined as (x_0, y_0):

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \mathbf{A}^{-1}\mathbf{b}$$

where

$$\mathbf{A} = \begin{bmatrix} -a_1 & 1 \\ -a_2 & 1 \\ \vdots & \vdots \\ -a_n & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Actuator positions are useful in system performance analysis, especially in the face of misregistration and localized phase errors.

Poke Matrix Smoothing

Smoothing is a preprocessing step performed on the poke matrix to remove noise from subapertures beyond the analytic influence radius of the actuator. This prevents propagation of obvious noise errors into the reconstructor and can be performed by removing measurements that are far from the actuator location.

We can calculate the distance from the ith subaperture to the jth actuator using the Euclidean distance function

$$d_{ij} = \sqrt{(a_{jx} - s_{ix})^2 + (a_{jy} - s_{iy})^2}$$

where x and y represent the respective locations of the actuator a or subaperture s. This distance can be used to generate a binary mask $\mathbf{D}_\mathrm{s}$ for the poke matrix based on a smoothing distance threshold τ_s:

$$\mathbf{D}_\mathrm{s} = \begin{cases} 1 & \text{if } d_{ij} \le \tau_\mathrm{s} \\ 0 & \text{if } d_{ij} > \tau_\mathrm{s} \end{cases}$$

Masking the poke matrix with $\mathbf{D}_\mathrm{s}$ is the smoothing operation, and the smoothed poke matrix $\mathbf{B}_\mathrm{s}$ is given as

$$\mathbf{B}_\mathrm{s} = \mathrm{smooth}(\mathbf{B})$$

If the positioning relationship between actuators and subapertures is not well known, smoothing can be performed by thresholding the amplitudes of the poke responses.

The idea behind smoothing is to remove spurious artifacts that are clearly not real. This includes apparent surface responses far from the expected actuator location as well as significantly small responses.

Actuator Slaving: Active Actuator Identification

Due to mapping between the wavefront sensor and deformable mirror, some actuators may be unobservable simply by geometry. We usually want to maintain some control over these actuators to limit interactuator stresses and accomplish this by **slaving the unobservable actuators**. Prior to generating the reconstructor, we need to classify the active and slave actuators and reduce the poke matrix to contain only active actuators.

The **slave actuators** can be determined geometrically by inspection. However, slaving also provides a method for dealing with weak or malfunctioning actuators and should be calculated through their observability.

The **active actuators** are determined by thresholding the observability of each actuator by some value ε.

$$a_i = \begin{cases} \text{Active if obs}(a_i) > \varepsilon \\ \text{Slave if obs}(a_i) \leq \varepsilon \end{cases}$$

This classifies n_s slave actuators and n_v active actuators.

Actuators with low observability are classified as slaves so that we do not expend excessive control energy on unobservable or uncontrollable actuators. The slave actuators may be outside of the active sensor area or damaged in some fashion. The observability of an actuator is a function of the square of the symmetric influence function and does not delineate an actuator push or a pull.

The actuators identified as slaves for the example system are indices 1, 2, 9, 7, 8, 16, 41, 49, 50, 48, 55, and 56.

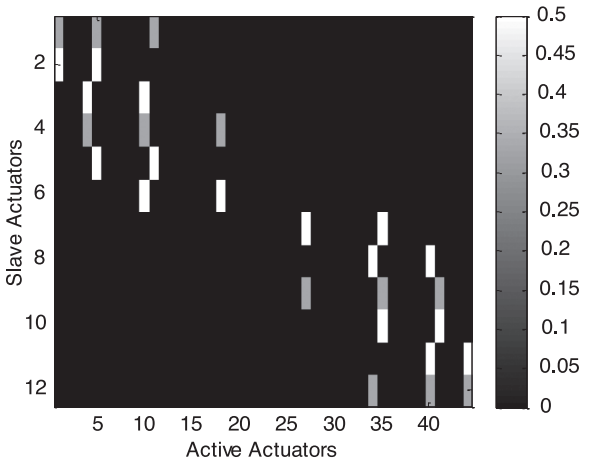
In conventional adaptive optics systems, the deformable mirror actuators overfill the optical clear aperture. The slave actuators then form an outer ring around the aperture to eliminate edge effects.

Actuator Slaving: Slave Logic

The most common approach to **actuator slaving** is to slave actuators to their nearest-neighbor active actuators. Implementing the **slave logic** as a matrix allows us to combine it with the reconstructor.

For the nearest-neighbor approach, the slave logic matrix $\mathbf{S}_\mathrm{l}$ is usually calculated through either offline analysis or brute force computations. This matrix has dimensions $n_\mathrm{s} \times n_\mathrm{v}$ and defines the desired scaling from valid actuators to slave actuators. The slave actuator commands are generated by multiplying the slave logic matrix with the active actuator commands.

The slave logic matrix for the example geometry is shown in the following image. For this approach, only valid actuators were used as nearest neighbors so that a slave actuator could not be slaved to another slave actuator. The slave command is the average of the nearest-neighbor commands.



Slaving the unobservable actuators and removing unobservable measurements ensures that the system is both controllable and observable.

Adding Slaving into the Reconstructor

Once the active and slave actuators have been determined, the poke matrix can be reduced into $\mathbf{B}_\mathrm{r}$ so that it contains only the columns and rows corresponding to active actuators and subapertures.

$$\mathbf{B}_\mathrm{r} = \mathbf{B}_\mathrm{s}(\text{active subaps, active acts})$$

The reduced reconstructor $\mathbf{R}_\mathrm{r}$ is then generated from the reduced poke matrix and can be expanded to include the slave logic $\mathbf{S}_\mathrm{l}$ to realize the full **reconstructor**. This is accomplished by partitioning the full reconstructor for active and slave actuators:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_\mathrm{r} \\ \mathbf{S}_\mathrm{l}\mathbf{R}_\mathrm{r} \end{bmatrix}$$

Partition $\mathbf{R}_1$ controls the **active actuators** and has dimensions $n_\mathrm{v} \times n_\mathrm{m}$, where n_m is the number of sensor measurements. Partition $\mathbf{R}_2$ controls the **slave actuators** and has dimensions $n_\mathrm{s} \times n_\mathrm{m}$. This allows the control algorithm to be implemented with a simple matrix vector multiplication.

The instantaneous commands for active $\mathbf{v}_k$ and slave $\mathbf{s}_k$ actuators are given by

$$\mathbf{u}_k = \begin{bmatrix} \mathbf{v}_k \\ \mathbf{s}_k \end{bmatrix} = \begin{bmatrix} \mathbf{R}_\mathrm{r}\mathbf{e}_k \\ \mathbf{S}_\mathrm{l}\mathbf{R}_\mathrm{r}\mathbf{e}_k \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{S}_\mathrm{l} \end{bmatrix} \mathbf{v}_k$$

Instead of working with partitions, in practice we want to keep the full reconstructor in the original ordering. The original ordering information must be retained when reducing the poke matrix so that the expansion can be performed correctly. In order to match the original ordering, the full reconstructor should be reordered in rows for the slaved actuators and reordered in columns for any removed subapertures. Columns for invalid subapertures should be set to $\mathbf{0}$.

Tilt Removal

Since **tip-and-tilt correction** requires a large dynamic range from the deformable mirror that could otherwise be used for higher-order correction, we generally force the reconstructor to be tilt blind. This is performed by explicitly removing the tip and tilt projections from the **poke matrix**.

For a Shack–Hartmann sensor, the **global tilt matrix** ($\mathbf{G}_t$) projects global x and y tilts onto the wavefront sensor slopes and is an $n_\mathrm{m} \times 2$ matrix:

$$\mathbf{G}_\mathrm{t} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 1 \\ \vdots & \vdots \end{bmatrix}$$

The slope ordering here follows the separation of x and y slopes as defined in the example geometry. The global tilt matrix must be normalized, so that

$$\mathbf{G}_\mathrm{t} = \frac{\mathbf{G}_\mathrm{t}}{\|\mathbf{G}_\mathrm{t}\|_2}$$

The **tilt projection matrix** $\mathbf{P}_\mathrm{t}$ is the product of the global tilt matrix and its transpose and projects the column vectors of the poke matrix onto the global tilt subspaces. This projection is then removed from the poke matrix:

$$\mathbf{P}_\mathrm{t} = \mathbf{G}_\mathrm{t}\mathbf{G}_\mathrm{t}^T$$
$$\mathbf{B}_\mathrm{t} = \mathbf{B}_\mathrm{r} - \mathbf{P}_\mathrm{t}\mathbf{B}_\mathrm{r}$$

Global tilt as defined here is based off of the average subaperture shift and represents the mean slope of the wavefront. Global tilt should not be confused with Zernike tilt (Z-tilt), which is the least-mean-squares slope of the wavefront.

Piston and Waffle Removal

Piston is an unobservable mode defined as a **global average height**; for any surface shape we can have an infinite number of piston values. Piston should always be removed, as it quickly reduces available dynamic range. The **piston basis vector** $\mathbf{b}_\mathrm{p}$ is generated by creating a vector of 1s and dividing by the square root of the number of actuators to normalize it.

The **projection matrix** $\mathbf{P}_\mathrm{p}$, which projects the actuator commands onto the basis vector $\mathbf{b}_\mathrm{p}$ is found by multiplying the basis vector by its transpose:

$$\mathbf{P}_\mathrm{p} = \mathbf{b}_\mathrm{p}\mathbf{b}_\mathrm{p}^{T}$$

Waffle is an unobservable mode in standard configurations that results in commands at the Nyquist spatial frequency. The **waffle basis vector** $\mathbf{b}_\mathrm{w}$ is defined by alternating positive and negative 1s on neighboring actuators and can be normalized by dividing by the square root of the number of actuators. As with the piston case, the waffle projection matrix $\mathbf{P}_\mathrm{w}$ is $\mathbf{b}_\mathrm{w}$ multiplied by its transpose:

$$\mathbf{P}_\mathrm{w} = \mathbf{b}_\mathrm{w}\mathbf{b}_\mathrm{w}^{T}$$

The waffle and piston bases are not necessarily orthogonal, so we use the **Gram–Schmidt orthogonalization** process, which simply subtracts the piston component (average value) from the waffle basis:

$$\mathbf{P}_\mathrm{wp} = \mathbf{P}_\mathrm{w} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{P}_\mathrm{w}(i)$$

We can remove piston, waffle, tip, and tilt in a single step to form a calibrated poke matrix $\mathbf{B}_\mathrm{c}$:

$$\mathbf{B}_\mathrm{c} = (\mathbf{I} - \mathbf{P}_\mathrm{t})\mathbf{B}_\mathrm{r}(\mathbf{I} - \mathbf{P}_\mathrm{p})(\mathbf{I} - \mathbf{P}_\mathrm{wp})$$

Tilt is a sensor mode, so it is removed on the output (left-hand) side of the matrix; piston and waffle are actuator modes and are removed on the input (right-hand) side of the matrix.

Reconstructor Generation: Least Squares

The reconstructor is essentially an inverse of the poke matrix, and generating the reconstructor is therefore an inverse problem. Approaches to **generating reconstructors** for adaptive optics systems typically follow least-squares techniques, which minimize the sum of the squares of the errors in the inversion. For a traditional **least-squares approach**, the cost function J to be minimized is given by

$$J = \left\| \mathbf{B}_c \mathbf{a} - \mathbf{e} \right\|_2^2$$

where $\mathbf{a}$ is the vector of active actuator commands and $\mathbf{e}$ is the associated vector of measured wavefront errors. The solution to this problem is the **Moore–Penrose pseudo-inverse**, which is given by

$$\mathbf{R}_r = -\left(\mathbf{B}_c^T \mathbf{B}_c \right)^{-1} \mathbf{B}_c^T$$

The negative sign enforces negative feedback and is included in all of the reconstructors described. This psuedo-inverse can also be formuated using **SVD**:

$$\mathbf{B}_c = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T$$
$$\mathbf{R}_r = -\mathbf{V} \boldsymbol{\Sigma}^{-1} \mathbf{U}^T$$

The standard least-squares approaches outlined here are sensitive to numerical issues caused by poor conditioning and only penalize the fitting error. Regularization can improve the numerical stability of the inversion process and provides a method to penalize particular actuator modes.

Least-squares techniques were originally developed by Carl Friedrich Gauss at the end of the 18th century and are still widely used in science and engineering.

Reconstructor Generation: Regularization

Using a singular system representation, the reduced reconstructor is given as

$$\mathbf{a} = \mathbf{R}_r\mathbf{e} = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T\mathbf{e} = \sum_i \frac{\left(\mathbf{u}_i^T\mathbf{e}\right)}{s_i}\mathbf{v}_i$$

From this equation, we see that small singular values result in large amplification of the corresponding input singular vector. In other words, **small singular values** induce large commands for poorly observable or poorly controllable modes that can degrade performance and cause loop instability.

To avoid issues due to small singular values, we use a **regularizing filter** $w(s^2)$ to remove the small singular values while retaining large ones.

The singular system representation with an applied filter is given as

$$\mathbf{a} = \sum_i w\left(s_i^2\right)\frac{\left(\mathbf{u}_i^T\mathbf{e}\right)}{s_i}\mathbf{v}_i$$

Two commonly used filters are the truncated SVD and the Tikhonov filter. The truncated SVD applies a simple threshold to the singular values:

$$w\left(s^2\right) = \begin{cases} 1 & \text{if } s^2 > \alpha \\ 0 & \text{if } s^2 \le \alpha \end{cases}$$

A regularized pseudo-inverse using the truncated SVD method is given as

$$\mathbf{R}_r = -\mathbf{V}\boldsymbol{\Sigma}_f^{-1}\mathbf{U}^T$$

where $\boldsymbol{\Sigma}_f^{-1}$ contains the inverses of the filtered singular values along its diagonal.

The MATLAB® command pinv (**A**, alpha) implements this method with 'alpha' used as the threshold.

The Tikhonov filter is given as

$$w\left(s^2\right) = \frac{s^2}{s^2 + \alpha^2}$$

This filter provides a damped least-squares solution, where we include a penalty on the solution in the cost

Reconstructor Generation: Regularization (cont.)

function. This penalty prevents large values of the solution such as those due to small singular values from being acceptable. The cost function for the Tikhonov filter is given by

$$J = \left\| \mathbf{B}_c \mathbf{a} - \mathbf{e} \right\|_2^2 + \alpha^2 \left\| \mathbf{a} \right\|_2^2$$

This cost function adds the sum of the squares of the solution ($\mathbf{a}$), scaled by α^2 to the cost function for the standard least-squares representation. The Tikhonov filter has the following singular system representation:

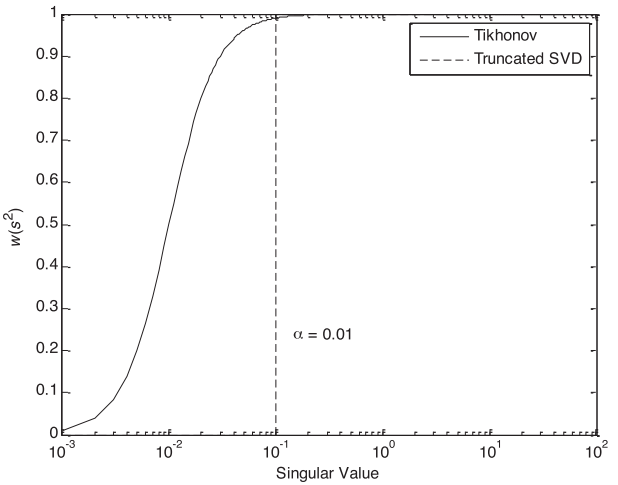
$$\mathbf{a} = \sum_i \frac{s_i \left(\mathbf{u}_i^T \mathbf{e} \right)}{s_i^2 + \alpha^2} \mathbf{v}_i$$

The regularized pseudo-inverse is given by

$$\mathbf{R}_r = - \left(\mathbf{B}_c^T \mathbf{B}_c + \alpha^2 \mathbf{I} \right)^{-1} \mathbf{B}_c^T$$

Tikhonov regularization can be thought of as effectively increasing the singular values of the low observable modes so that they have lower effective gain in the reconstructor. This technique can be extended to select specific unwanted modes to filter.

Tikhonov regularization provides a smoother roll off than the sharp clamping threshold of the truncated SVD. This can be seen in the following figure of $w(s^2)$ versus s.



T-Filter

The **T-filter** applies a spatial filter to the reconstructor to suppress waffle modes. The reconstructor is computed by a regularization or pseudo-inverse approach and then multiplied by the T-filter offline to generate the filtered reconstructor. The T-filter is found by convolving a kernel $\mathbf{K}_\mathrm{T}$ with the actuator grid to get a T-filter matrix for each actuator $\mathbf{T}_i$:

$$\mathbf{K}_\mathrm{T} = \begin{bmatrix} 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & -1 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{bmatrix}$$

$$\mathbf{A}_i = \begin{cases} 0 & \text{if } \alpha \neq i \\ 1 & \text{if } \alpha = i \end{cases} \qquad \mathbf{T}_i = \mathbf{A}_i \otimes \mathbf{K}_\mathrm{T}$$

The T-filter for each actuator needs to be converted from a matrix to a column vector ($\mathbf{T}_i \Leftrightarrow \mathbf{t}_i$). After all of the individual filters have been found, the full T-filter $\mathbf{T}_\mathrm{f}$ can be found by stacking the individual vectors and then multiplying by the transpose. The T-filter should be normalized before use:

$$\mathbf{T} = [\mathbf{t}_1 \cdots \mathbf{t}_n]$$
$$\mathbf{T}_\mathrm{f} = \mathbf{T}\mathbf{T}^T$$

The filtered reconstructor $\mathbf{R}_\mathrm{f}$ is then the product of the T-filter and the original reconstructor $\mathbf{R}$:

$$\mathbf{R}_\mathrm{f} = \mathbf{T}_\mathrm{f}\mathbf{R}$$

The T-filter provides a simple method for suppression of waffle modes in the deformable mirror by filtering the output of the reconstructor. This prevents expending a large amount of energy on high spatial frequencies.

Modal Suppression

We can extend the cost function for the Tikhonov regularization strategy to penalize particular modes and drive the solution away from those modes:

$$J = \left\| \mathbf{B}_c \mathbf{a} - \mathbf{e} \right\|_2^2 + \alpha^2 \left\| \mathbf{D} \mathbf{a} \right\|_2^2$$

Here $\mathbf{D}$ is a square weighting matrix with dimensions $n_v \times n_v$ that maps the actuator command vector $\mathbf{a}$ onto the subspaces we wish to penalize. In most cases, $\mathbf{D}$ includes the scalar gain α^2 so that modes can be weighted individually. This extension provides the following regularized pseudo-inverse:

$$\mathbf{R}_r = -\left(\mathbf{B}_c^T \mathbf{B}_c + \mathbf{W} \right)^{-1} \mathbf{B}_c^T$$

where we use a matrix $\mathbf{W}$ that acts as the square of the weighting matrix:

$$\mathbf{W} = \mathbf{D}\mathbf{D}^T$$

The filtering properties of this approach are valid only in an open-loop configuration. In closed loop, the signal is continually passed through the reconstructor, and the steady state solution approaches the standard least-squares solution. **Modal feedback** matrices are generally used to enforce the desired spatial filtering in closed loop.

For **suppression operations**, the weighting matrix $\mathbf{W}$ is not necessarily a projection matrix. A projection matrix is idempotent, meaning that

$$\mathbf{P} = \mathbf{P}^2$$

This states that a projection matrix always maps back onto itself regardless of how many times the projection is performed.

Interactuator Shear Suppression

Interactuator shear suppression penalizes the second difference or local curvature along one dimension to restrict large interactuator commands. The weighting matrix for shear can be found through convolution of the following kernels with the actuator grid to generate a set of basis matrices $\mathbf{D}_i$:

$$\mathbf{k}_1 = [1 \; -2 \; 1] \qquad \mathbf{k}_2 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$\mathbf{A}_i = \begin{cases} 0 & \forall a \neq i \\ 1 & \text{if } a = i \end{cases} \qquad \mathbf{D}_i = \mathbf{A}_i \otimes \mathbf{k}$$

The basis matrices need to be converted from matrices to column vectors ($\mathbf{D}_i \Leftrightarrow \mathbf{d}_i$). After all of the individual vectors have been found, the full basis matrix for either $\mathbf{k}_1$ or $\mathbf{k}_2$ ($\mathbf{D}_{ian}$) can be found by stacking the individual vectors:

$$\mathbf{D}_{ian} = [\mathbf{d}_1 \cdots \mathbf{d}_n]$$

Using the two convolution kernels results in separate basis matrices for the rows and columns: $\mathbf{D}_{ia1}$ and $\mathbf{D}_{ia2}$. The full basis matrix combines these two to form an $n_\mathrm{v} \times 2n_\mathrm{v}$ matrix:

$$\mathbf{D}_{ia} = [\mathbf{D}_{ia1} \; \mathbf{D}_{ia2}]$$

The weighting matrix can be found by multiplying the basis matrix by its transpose and should be normalized before use:

$$\mathbf{W}_{ia} = \mathbf{D}_{ia}\mathbf{D}_{ia}^{\mathrm{T}}$$

Because interactuator shear is inversely proportional to the square of the distance between the actuators, shear between diagonal actuators is ignored. The use of separate one-dimensional kernels for the x and y axes ensures that the penalty is only applied in those directions and not along the diagonals.

Nullspace Suppression

Since we are explicitly making certain modes unobservable in the poke matrix, the corresponding singular values are small. To counter this, we need to **suppress the nullspace** of the poke matrix. For typical problems, the nullspace of concern consists of piston, waffle, tip, and tilt. In this case, we can utilize the SVD to obtain the four smallest output singular vectors and build the **nullspace projection matrix**:

$$\mathbf{B}_\mathrm{c} = \mathbf{U}\Sigma\mathbf{V}^T$$

If the singular values are ordered from highest to lowest, the nullspace in the actuator domain is the last four columns of $\mathbf{V}$:

$$\mathbf{D}_\mathrm{n} = \mathbf{V}(:, n_a - 3 : n_a)$$

The nullspace weighting matrix is then $\mathbf{D}_\mathrm{n}$ multiplied by its transpose. Since $\mathbf{V}$ is a unitary matrix from the SVD, $\mathbf{D}_\mathrm{n}$ consists of unit vectors and does not need to be normalized:

$$\mathbf{W}_\mathrm{n} = \mathbf{D}_\mathrm{n}\mathbf{D}_\mathrm{n}^T$$

Since the calibrated poke matrix is used to generate $\mathbf{W}_\mathrm{n}$, it has dimensions $n_\mathrm{v} \times n_\mathrm{v}$.

From linear algebra, the definition of the nullspace of a matrix $\mathbf{B}_\mathrm{c}$ is the set of all vectors $\mathbf{u}$ that are solutions to the homogeneous equation $\mathbf{B}_\mathrm{c}\mathbf{u} = \mathbf{0}$.

Any modes that are unobservable, whether through system geometry or mathematical design, are part of the nullspace and should be suppressed. By using the SVD technique described here, each individual mode does not need to be explicitly defined in order to determine the system nullspace.

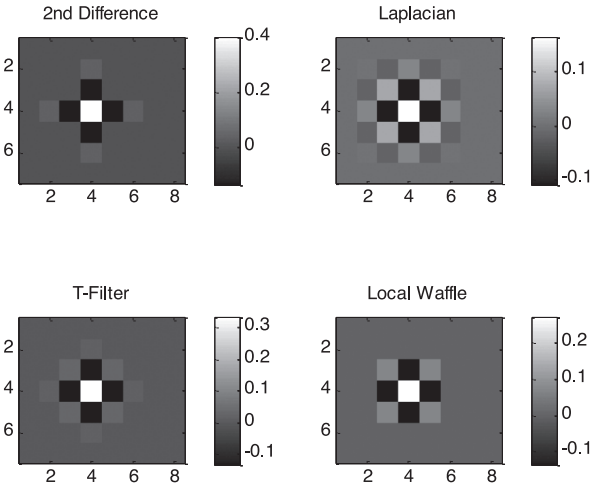
Weighting Matrices

Weighting matrices are combined with specified gains, as in

$$\mathbf{W} = \alpha_{ia}^2 \mathbf{W}_{ia} + \alpha_n^2 \mathbf{W}_n + \cdots$$

The gains for the weighting matrices generally do not need to be very large to achieve the desired effect and are usually optimized through offline analysis and simulation. For convolution-based cases, the weighting matrix forms a complete basis for the active actuator space, so care should be taken when combining them. Generally, only a single complete basis should be used.

Spatial filtering properties for several convolution kernels are shown in the figure.



The weighting matrices should constrain only active actuators so that $\mathbf{W}$ is $n_v \times n_v$. This ensures that slave actuators are not directly constrained. It is often easier to generate the weighting matrix for all of the actuators and then reduce it.

Modal Feedback

The **modal feedback** matrix is designed such that the closed-loop spatial response is equivalent to the open-loop spatial response. Therefore, the modal feedback matrix $\mathbf{M}$ is defined as

$$\mathbf{M} = \mathbf{I} + \mathbf{RB}$$

where $\mathbf{R}$ is assumed to provide negative feedback.

Typically, the modal-feedback matrix is generated from the reduced reconstructor and poke matrix so that we can separate the slaved actuators. Using the already calculated reduced matrices for the reconstructor, we can implement the reduced modal-feedback matrix as

$$\mathbf{M}_\mathrm{r} = \left(\mathbf{B}_\mathrm{c}^\mathrm{T}\mathbf{B}_\mathrm{c} + \mathbf{W}\right)^{-1}\mathbf{W}$$

The reduced modal-feedback matrix can be expanded to include the slave logic, similarly to the reconstructor expansion. The difference is that the modal-feedback matrix needs to be expanded in rows and columns:

$$\mathbf{M} = \begin{bmatrix}\mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_3 & \mathbf{M}_4\end{bmatrix} = \begin{bmatrix}\mathbf{M}_\mathrm{r} & \mathbf{0} \\ \mathbf{S}_\mathrm{l}\mathbf{M}_\mathrm{r} & \mathbf{I}\end{bmatrix}$$

- Partition $\mathbf{M}_1$ has dimensions of $n_\mathrm{v} \times n_\mathrm{v}$ and maps the feedback from active actuators to active actuators.
- Partition $\mathbf{M}_2$ has dimensions of $n_\mathrm{v} \times n_\mathrm{s}$ and maps the feedback from slave actuators to active actuators.
- Partition $\mathbf{M}_3$ has dimensions of $n_\mathrm{s} \times n_\mathrm{v}$ and maps the feedback from active actuators to slave actuators.
- Partition $\mathbf{M}_4$ has dimensions of $n_\mathrm{s} \times n_\mathrm{s}$ and maps the feedback from slave actuators to slave actuators.
- The full modal-feedback matrix has dimensions of $n_\mathrm{a} \times n_\mathrm{a}$.

Modal Feedback (cont.)

When combined with the partitioned reconstructor, the active $\mathbf{v}_k$ and slave $\mathbf{s}_k$ actuator commands are given by

$$\mathbf{u}_k = g_1 \mathbf{R} \mathbf{e}_k + (g_2 \mathbf{I} - g_1 \mathbf{M}) \mathbf{u}_{k-1}$$

$$\mathbf{u}_k = \begin{bmatrix} \mathbf{v}_k \\ \mathbf{s}_k \end{bmatrix} = g_1 \begin{bmatrix} \mathbf{R}_\mathrm{r} \\ \mathbf{S}_\mathrm{l} \mathbf{R}_\mathrm{r} \end{bmatrix} \mathbf{e}_k + \begin{bmatrix} g_2 \mathbf{I} - g_1 \mathbf{M}_\mathrm{r} & \mathbf{0} \\ -g_1 \mathbf{S}_\mathrm{l} \mathbf{M}_\mathrm{r} & g_2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{k-1} \\ \mathbf{s}_{k-1} \end{bmatrix}$$

This modal-feedback implementation provides direct feedback for the slave actuator commands through the leaky integrator. It is sometimes desirable to utilize the active actuator commands for feedback rather than the slave actuator commands. This and other modal-feedback matrix implementations can be generated to provide the desired control strategy.

As in the case for the reconstructor, reordering the matrix to switch from partitions to the original system ordering is generally the preferred approach. The modal feedback needs to be reordered in rows and columns for slaved actuators.

In traditional adaptive optics systems, the modal-feedback matrix is simply the piston projection matrix $\mathbf{P}_\mathrm{p}$.

The reconstructor and modal-feedback matrices as shown here outline the general theory for solutions to **inverse problems** in adaptive optics. Specifically, we are interested in reconstructor generation from a poke matrix with a weighted cost function. Numerous algorithms can be used to solve this problem and minimize the same cost function, but with better computational efficiency. For adaptive optics systems with large numbers of channels, sparse matrix techniques along with conjugate gradient and Fourier transform approaches have been shown to provide extremely efficient wavefront reconstruction.

Reconstructor Generation: Procedure

The reconstructor using modal feedback is based heavily on the beam control system as outlined in Hyver and Blankinship (1995). A procedure to generate this reconstructor is outlined below.

1. Smooth the poke matrix:

$$\mathbf{B}_\mathrm{s} = \mathrm{smooth}(\mathbf{B})$$

2. Reduce the poke matrix:

$$\mathbf{B}_\mathrm{r} = \mathbf{B}_\mathrm{s}(\mathrm{active\,subaps}, \mathrm{active\ acts})$$

3. Remove piston, tip, tilt and waffle:

$$\mathbf{B}_\mathrm{c} = (\mathbf{I} - \mathbf{P}_\mathrm{t})\mathbf{B}_\mathrm{r}(\mathbf{I} - \mathbf{P}_\mathrm{p})(\mathbf{I} - \mathbf{P}_\mathrm{wp})$$

4. Generate the weighting matrix for active actuators:

$$\mathbf{W} = \sum_{i=1}^{N} \alpha_i^2 \mathbf{W}_i$$

5. Generate the reduced reconstructor:

$$\mathbf{R}_\mathrm{r} = -\left(\mathbf{B}_\mathrm{c}^T \mathbf{B}_\mathrm{c} + \mathbf{W}\right)^{-1} \mathbf{B}_\mathrm{c}^T$$

6. Generate the reduced modal-feedback matrix:

$$\mathbf{M}_\mathrm{r} = \left(\mathbf{B}_\mathrm{c}^T \mathbf{B}_\mathrm{c} + \mathbf{W}\right)^{-1} \mathbf{W}$$

7. Calculate the slave logic matrix $\mathbf{S}_\mathrm{l}$.

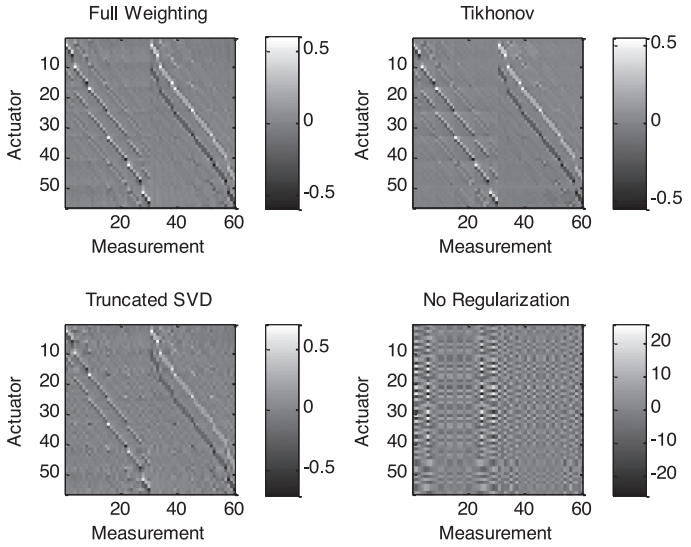
8. Expand the full-control matrices:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_\mathrm{r} \\ \mathbf{S}_\mathrm{l}\mathbf{R}_\mathrm{r} \end{bmatrix} \qquad \mathbf{M} = \begin{bmatrix} \mathbf{M}_\mathrm{r} & \mathbf{0} \\ \mathbf{S}_\mathrm{l}\mathbf{M}_\mathrm{r} & \mathbf{I} \end{bmatrix}$$

9. Reorder the control matrices to match the original numbering.

Reconstructor Comparison

The figure below shows reconstructors with actuator slaving included for the example geometry generated through complete modal weighting, general Tikhonov regularization, truncated SVD, and standard pseudo-inverse with piston, tip, tilt, and waffle removed but without regularization.



When viewing reconstructors as images, it is important to look for distinct areas that are roughly block diagonal with the pattern essentially transposed from the poke matrix. Again, the number and size of these blocks is not of great significance.

It is clear from this figure that some degree of regularization is needed, as the standard pseudo-inverse technique has high gains in the reconstructor that correspond to small singular values or poorly sensed modes. The full weighting, Tikhonov, and truncated SVD reconstructors look similar and yield effectively the same open-loop performance. The full-weighting case has the benefit of added stability from modal suppression.

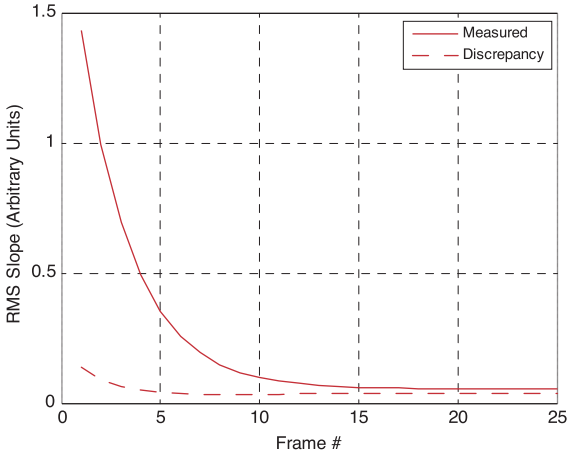
Slope Discrepancy

Slope discrepancy provides a metric for the errors in the reconstructor through the null space of the reconstructed slopes:

$$\mathbf{S}_d = (\mathbf{I} + \mathbf{BR})$$

Again, **R** is assumed to provide negative feedback.

When the slope-discrepancy projection is converted to RMS wavefront error, it provides an effective noise floor based on the system geometry. This is often shown by superimposing the slope-discrepancy RMS projection on the measured RMS wavefront error, as seen in the figure.



This figure shows the example system performance with a fully weighted reconstructor, a loop gain of 0.3, and a leak gain of 0.997 for an initial astigmatism input. The measured RMS slope error converges to the RMS slope discrepancy in ~13 frames. Because there is a leak gain, the steady state error is nonzero, as shown by the slight offset between the slope-discrepancy noise floor and the steady state RMS slope error.

Offloads and Woofer-Tweeter Systems

An offload matrix is a special control matrix that drives an active device from commands sent to a primary unit rather than from sensor measurements. **Offloads** (commonly referred to as **woofer-tweeter systems**) are generally used to preserve the dynamic range of the deformable mirror under closed-loop conditions and correct lower spatial frequencies than the primary unit can. Offloads can be handled through an implicit or explicit reconstructor.

If we have multiple offload channels, we can obtain the best performance by generating an explicit reconstructor for the offload to the sensor. This allows direct control of filtering and slaving for the offload unit. The explicit offload reconstructor can be generated by any of the techniques previously described, although we would generally match the technique used to that of the reconstructor for the primary unit.

Let $\mathbf{R}_o$ be the reconstructor for the offload unit with dimensions $n_o \times n_m$, where n_o is the number of offload channels; let $\mathbf{B}_p$ be the poke matrix for the primary unit with dimensions $n_m \times n_a$, where n_a is the number of actuators. The **offload matrix** $\mathbf{O}$ has dimensions $n_o \times n_a$ and is calculated by

$$\mathbf{O} = \mathbf{R}_o \mathbf{B}_p$$

If we have only a few offload channels, we do not need to generate an explicit reconstructor and can work with an implicit one.

Let $\mathbf{R}_p$ be the reconstructor for the primary unit with dimensions $n_a \times n_m$ and $\mathbf{B}_o$ be the poke matrix for the offload unit with dimensions $n_m \times n_o$. The offload matrix $\mathbf{O}$ is then calculated by the pseudo-inverse of the product of $\mathbf{R}_p$ and $\mathbf{B}_o$:

$$\mathbf{O} = \left[(\mathbf{R}_p \mathbf{B}_o)^T (\mathbf{R}_p \mathbf{B}_o) \right]^{-1} (\mathbf{R}_p \mathbf{B}_o)^T$$

Open-Loop Wavefront Estimation

Because wavefront measurements are not perfect, an estimator is often employed. The **open-loop minimal-variance estimator** follows the regularized least-squares techniques previously outlined and is given as

$$\hat{\mathbf{\Gamma}} = \left(\mathbf{\Gamma}^T\mathbf{\Gamma} + \sigma_n^2\mathbf{Q}_\mathrm{d}^{-1}\right)^{-1}\mathbf{\Gamma}^T$$

where $\mathbf{Q}_\mathrm{d}$ is the covariance matrix of the atmospheric turbulence, $\mathbf{\Gamma}$ is a geometry matrix defining the relationship between measurements and subapertures, and the measurement noise is assumed to be Gaussian with zero mean and variance σ_n^2. By using the inverse of the covariance matrix in this equation we are increasing the sensitivity of the estimator to expected turbulence modes.

When the geometry matrix is the poke matrix, the estimator is then the reconstructor. If the covariance matrix is represented in reduced actuator space so that $\mathbf{Q}_\mathrm{d}$ is $n_\mathrm{v} \times n_\mathrm{v}$, we can include the estimator in the reduced reconstructor by

$$\mathbf{R}_\mathrm{r} = -\left(\mathbf{B}_\mathrm{c}^T\mathbf{B}_\mathrm{c} + \mathbf{W} + \sigma_n^2\mathbf{Q}_\mathrm{d}^{-1}\right)^{-1}\mathbf{B}_\mathrm{c}^T$$

This provides the optimal constrained estimator for an open-loop system and works well in practice on closed-loop systems. However, it does not include knowledge of the system dynamics with command inputs and is not truly optimal under closed-loop conditions. When the actuator rise time and processing delay are large enough to be observed, the open-loop estimator performance is degraded.

Kalman filtering can provide better overall closed-loop performance, but this improvement may be negligible in comparison to the increased complexity, especially for extremely large systems with thousands of actuators.

Kalman Filtering

A **Kalman filter** is the minimum variance estimator for general linear systems. Consider the system with states given as the true values of the wavefront sensor measurements

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{d}_k$$
$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{n}_k$$

The disturbance and measurement noise covariances are

$$\mathbf{Q}_\mathrm{d} = \left\langle \mathbf{d}_k \mathbf{d}_k^T \right\rangle \quad \mathbf{Q}_n = \left\langle \mathbf{n}_k \mathbf{n}_k^T \right\rangle$$

We can define the steady state Kalman filter by first solving the discrete-time algebraic Riccati equation for the propagation error covariance matrix $\mathbf{P}$

$$\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}^T + \mathbf{Q}_\mathrm{d} - \mathbf{A}\mathbf{P}\mathbf{C}^T\mathbf{Q}_n^{-1}\mathbf{C}\mathbf{P}\mathbf{A}^T$$

The Kalman gain $\mathbf{K}$ is given by

$$\mathbf{K} = \mathbf{A}\mathbf{P}\mathbf{C}^T\left[\mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{Q}_n\right]^{-1}$$

The estimator is implemented by

$$\hat{\mathbf{x}}_{k+1} = \mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\mathbf{u}_k + \mathbf{K}[\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k]$$

The control signal $\mathbf{u}_k$ is calculated using the state estimate

$$\mathbf{u}_k = g_1\mathbf{R}\hat{\mathbf{x}}_k + g_2\mathbf{u}_{k-1} - g_1\mathbf{M}\mathbf{u}_{k-1}$$

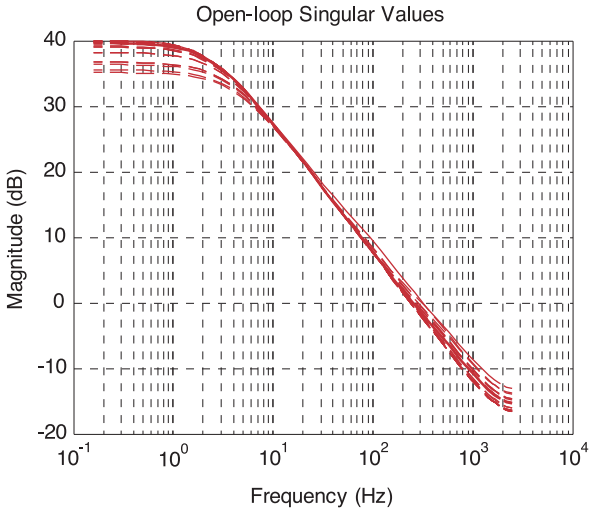
Since the optimization is done in the least-squares sense, the separation principle allows us to design the estimator and control matrices ($\mathbf{K}$, $\mathbf{R}$, and $\mathbf{M}$) independently and still provide the minimum least-squares error.

The theory behind the Kalman filter was introduced in 1960 in a paper by Rudolf E. Kalman. Subsequent developments by Richard S. Bucy have resulted in this filter being commonly referred to as the Kalman–Bucy filter. The extended Kalman filter was developed to apply the theory to nonlinear systems and is the state estimator most commonly used today for GPS and navigation systems.

Multivariable System Performance

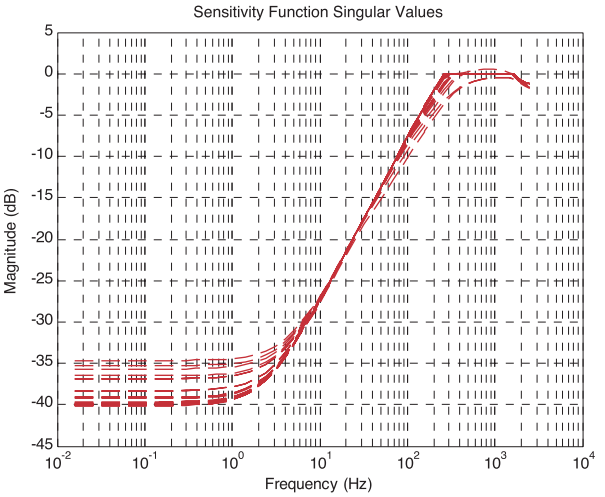
For **multivariable systems**, the concepts of bandwidth and overall performance are loose with respect to single-channel systems. They are generally specified in ranges dependent on the singular values and modes of the system. For an adaptive optics system, the coupling between modes is purely static and is captured by the reconstructor and poke matrix. This means that the system is decentralized, and the dynamics can be described with a diagonal matrix. From this, we can expect the frequency responses for each mode to have the same shape and differ by the scaling of the singular values. The example system has 60 sensor modes (local tilts), so the transfer function from disturbance inputs to corrected outputs has dimensions of 60×60. Since the singular values are widely separated, especially for unobservable modes, for clarity we limit inclusion to the first 20 modes.

The figure shows the open-loop singular values for the first 20 modes of the example system, the crossover frequency ranges from 235 to 275 Hz, and the low-frequency-gain ranges from 35 to 40 dB.



Multivariable System Performance (cont.)

The singular values for the first 20 modes of sensitivity function of the example system are shown in the following plot. The −3 dB disturbance rejection point ranges from 170 to 250 Hz, and the steady state error rejection ranges from −40 to −35 dB (100× to 56×).

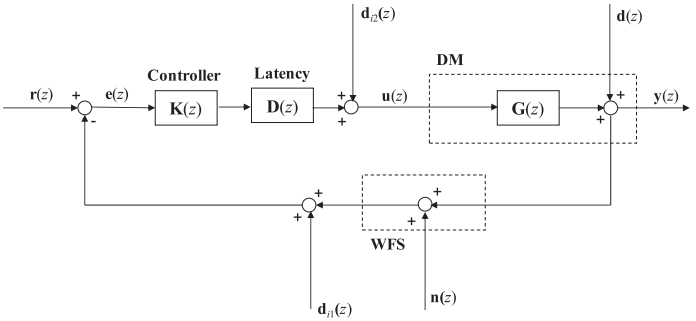


Both the open-loop and sensitivity function singular value plots match the shape of the corresponding single-channel plots as expected. Gain uncertainties due to actuator variation or system misalignment are captured by using an empirical poke matrix and do not degrade system performance. However, small uncertainties are present in the actuator time constants that can affect performance and reduce robustness.

A convenient single-valued performance metric for practical systems is the **RMS wavefront error**, although overshoot is not observable with this metric. The closed-loop time response of the RMS wavefront to a step input is commonly used to estimate an average system bandwidth.

Disturbance Injection

Disturbance injection (DI) allows us to determine the system response for specified inputs by injecting an artifical, controlled disturbance to the system through software. The DI inputs can be in either sensor space ($\mathbf{d}_{i1}$) or actuator space ($\mathbf{d}_{i2}$), as shown in the figure.



For closed-loop performance assessment, the DI inputs should be injected in sensor space $\mathbf{d}_{i1}$, as this directly measures the sensitivity function or disturbance rejection properties. For open-loop performance assessment (actuator step responses), the DI inputs should be injected in actuator space $\mathbf{d}_{i2}$. Injecting inputs in actuator space during closed-loop operation does not accurately capture a performance metric and can be additionally complicated by saturation and by limiting in the controller.

Given the multivariable nature of adaptive optics systems, the DI inputs are vectors rather than single values. A general method for testing temporal response is therefore to input a particular mode and then vary the amplitude of that mode in frequency. This allows determination of the system response to specific modes.

DI as described here works to determine local loop performance. For formal testing and performance verification, an optical disturbance should be injected, and an external unbiased truth sensor used to assess performance.

Wavefront Sensor Calibration

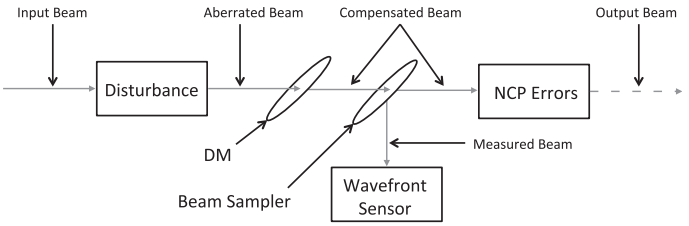
Wavefront sensor calibration is an important step to relate the local loop performance to the plane where the performance is characterized. A key issue is **non-common-path (NCP) errors**, which are errors present in one path but not in another.

In some cases, NCP errors can be measured directly through the use of a test beam. In other cases they can be estimated from analysis of system performance.

A calibration file for a wavefront sensor consists of measurement offsets to remove the phase difference between the wavefront sensor and the reference planes. For Shack–Hartmann-style sensors, this can be implemented as spot center shifts in each subaperture.

If a wavefront sensor is not properly calibrated to the reference plane, the local loop may be performing well while the observed performance at the reference plane is poor. This is part of the reason that performance characterization should be measured by an unbiased truth sensor that is separate from the system.

The figure provides an example system that shows that the output beam differs from the measured/compensated beam by the addition of the NCP errors. If the wavefront sensor is calibrated to offset the NCP errors, the output beam will be the desired clean beam.



Centroiding and Thresholding

Centroiding is generally performed by a center-of-mass algorithm

$$x_c = \frac{\sum_i \sum_j x_j I(i,j)}{\sum_i \sum_j I(i,j)} \qquad y_c = \frac{\sum_i \sum_j y_j I(i,j)}{\sum_i \sum_j I(i,j)}$$

where $I\ (i,\ j)$ is the intensity of the pixel at coordinates i, j.

Typically, a **thresholding** operation is used to eliminate poorly illuminated pixels from the measurement and can be divided into two types. When intensity gradients or fluctuations are present, the threshold value T should be found dynamically based on the illumination present in each subaperture.

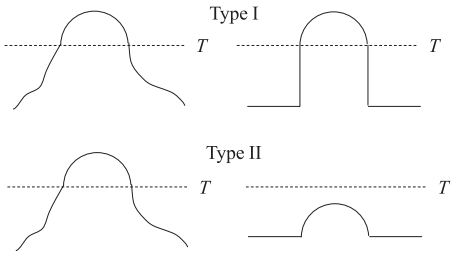
Type I Thresholding (pedestal): Values below the threshold are clipped to zero so that a pedestal term remains:

$$I_T(i,j) = \begin{cases} I(i,j) & \text{if } I(i,j) > T \\ 0 & \text{if } I(i,j) \le T \end{cases}$$

Type II Thresholding (pedestal removed): The signal is reduced by the threshold to remove the pedestal term:

$$I_T(i,j) = \begin{cases} I(i,j) - T & \text{if } I(i,j) > T \\ 0 & \text{if } I(i,j) \le T \end{cases}$$

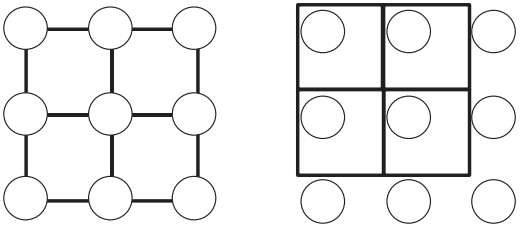
The two thresholding types are shown in the figure.



Misregistration

Misregistration is a property of adaptive optics systems with a one-to-one correspondence between measured phase points and actuators that occurs when the subapertures are translated, rotated, or otherwise misaligned relative to the actuator grid. For a Shack–Hartmann sensor, this means that the lenslet array is not properly aligned to the deformable mirror.

The figure shows the concept of misregistration using the Fried geometry. The left-hand side shows a correctly registered system with actuators on the corners of the subapertures, while the right-hand side shows a misregistered system with the actuators offset.

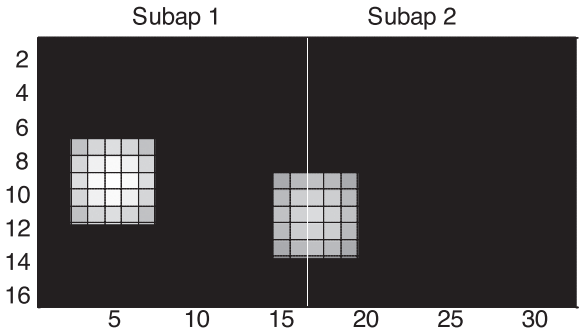


Misregistration is usually calculated by determining the actuator positions from the poke matrix and then fitting them to a grid over the subapertures. Since misregistration also makes the waffle mode visible, a common alignment technique is to put a waffle pattern on the deformable mirror, then adjust alignment to null the observed response on the wavefront sensor.

The position error of the influence functions can be mitigated by using the empirical poke matrix. However, large misregistrations reduce the observed gain of the actuators as well as increase the observability of the waffle mode; therefore, a typical specified limit on misregistration is 0.1 subapertures in translation. Regularized reconstructors are generally sufficiently robust to handle shifts of up to ± 0.1 subapertures without requiring new reconstructors.

Subaperture Spillover

Subaperture spillover occurs when illumination from one subaperture crosses the boundary and enters a neighboring subaperture. This is shown in the figure below with 16×16 pixel subapertures and a 5×5 pixel spot. The spot that should be resident in subaperture 2 is partially contained in subaperture 1.



Spillover generally occurs when the sensor dynamic range is undersized for the deformable mirror or input aberrations. It can also occur when the pixel boundaries of the subapertures are misaligned relative to the actual boundaries of the lenslets. The diffraction pattern of the spot is another potential issue, as it reduces the effective dynamic range; the example above has 5-pixel diameter spots and therefore an effective dynamic range of ±5.5 pixels rather than ±8 pixels to prevent spillover. Spillover can cause dramatic errors in the measurement, as the calculated position is pulled in the direction of the spillover. This is especially problematic when a well-lit spot crosses over into a subaperture that is poorly lit.

When realignment or increasing the sensor dynamic range is not possible, mitigation of spillover is generally accomplished by implementing **guard bands**. A guard band forces the outer row of pixels to be 0 so that they are not included in the measurement calculation.

Equation Summary

Coherence length:

$$r_0 = \left[0.423k^2 \sec\zeta \int_{\text{Path}} C_n^2(z)dz\right]^{-3/5}$$

Isoplanatic angle:

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

Greenwood frequency:

$$f_G = 2.31\lambda^{-6/5}\left[\sec\zeta \int_{\text{Path}} C_n^2(z)v_W(z)^{5/3}dz\right]^{3/5}$$

Hufnagel–Valley model:

$$C_n^2(h) = 5.94\times10^{-23}h^{10}\left(\frac{W}{27}\right)^2 \exp(-h) + 2.7\times10^{-16}\exp(-2h/3) + A\exp(-10h)$$

Gaussian wind model:

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

Bufton wind model:

$$v(z)_{\text{Bufton}} = 5 + 30\exp\left[-\left(\frac{z-9.4}{4.8}\right)^2\right]$$

Root-mean-square average jitter:

$$\alpha_{\text{jit}} = \sqrt{0.182\lambda^2 D^{-1/3} r_0^{-5/3}}$$

Equation Summary

On-axis intensity of uniform circular beam:

$$I_0 = \frac{\pi^2 (D/2)^4}{\lambda^2 L^2} I_{\text{Aper}}$$

On-axis intensity with aberrations:

$$I_0 = S \frac{\pi^2 (D/2)^4}{\lambda^2 L^2} I_{\text{Aper}}$$

Laser brightness:

$$B = \frac{\pi D^2 P}{4\lambda^2 \left[1 + \left(\frac{2.22\alpha_{\text{jit}} D}{\lambda}\right)^2\right]} TKS$$

Intensity distribution for a uniform circular aperture:

$$I(r) = \left\{\frac{2\lambda L}{\pi D r}\left[J_1\left(\frac{\pi D r}{\lambda L}\right)\right]\right\}^2$$

Strehl ratio with reduction of on-axis intensity:

$$S \cong \exp\left[-\left(\frac{2\pi}{\lambda}\right)^2 (\Delta\phi)^2\right]$$

Strehl ratio with wavefront error and jitter:

$$S_{\text{w/jit}} = \frac{e^{-\sigma^2}}{\left[1 + \left(\frac{2.22\alpha_{\text{jit}} D}{\lambda}\right)^2\right]}$$

Wavefront error variance in radians squared:

$$\sigma^2 = \left(\frac{2\pi}{\lambda}\right)^2 (\Delta\phi)^2$$

Equation Summary

Amplitude fluctuation:

$$\sigma_\chi^2 = \left(\frac{2\pi}{\lambda}\right)^{7/6} L^{11/6} C_n^2$$

Gaussian spot beam waist change:

$$w_0^2 = \frac{w^2}{\left[1 + \left(\frac{\pi w^2}{\lambda L}\right)^2\right]}$$

Gaussian growth spot with turbulence:

$$w^2 = \frac{4L^2}{k^2 w_0^2} + 3.58 C_n^2 L^3 w_0^{-1/3}$$

Zernike series:

$$\Phi(r,\theta) = A_{00} + \frac{1}{\sqrt{2}} \sum_{n=2}^{\infty} A_{n0} \Re_n^0\left(\frac{r}{R}\right) + \sum_{n=1}^{\infty} \sum_{m=1}^{n} (A_{nm} \cos m\theta + B_{nm} \sin m\theta) \Re_n^m\left(\frac{r}{R}\right)$$

Azimuthal polynomials:

$$\Re_n^m\left(\frac{r}{R}\right) = \sum_{s=0}^{\frac{n-m}{2}} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!} \left(\frac{r}{R}\right)^{n-2s}$$

Modulation transfer function for a diffraction-limited circular aperture:

$$M_{\mathrm{DL}}(\nu) = \frac{2}{\pi}\left\{\cos^{-1}\left(\frac{\nu\lambda}{D}\right) - \left(\frac{\nu\lambda}{D}\right)\left[1 - \left(\frac{\nu\lambda}{D}\right)^2\right]^{1/2}\right\}$$

Coherence length of the atmosphere with modulation transfer function:

$$M_{\text{Long-exp}}(\nu) = \exp\left[-3.44\left(\frac{\nu\lambda}{r_0}\right)^{5/3}\right]$$

Equation Summary

Modal wavefront error:

$$\sigma_{\text{fitting}}(\text{waves}) = \frac{\sqrt{0.2944 N_{\text{Zern}}^{-\sqrt{3}/2} \left(\frac{D}{r_0}\right)^{5/3}}}{2\pi}$$

Zonal fitting error:

$$\sigma_{\text{fitting}}(\text{waves}) = \frac{1}{2\pi}\left[\kappa\left(\frac{s_{\text{Act}}}{r_0}\right)^{5/3}\right]^{1/2}$$

Shack–Hartmann sensor error:

$$\sigma_\phi = \frac{\pi^2 K_g}{4\ \text{SNR}}\left[\left(\frac{3}{2}\right)^2 + \left(\frac{\theta d}{\lambda}\right)^2\right]^{1/2}, \quad r_0 > d$$

$$\sigma_\phi = \frac{\pi^2 K_g}{4\ \text{SNR}}\left[\left(\frac{3d}{2r_0}\right)^2 + \left(\frac{\theta d}{\lambda}\right)^2\right]^{1/2}, \quad r_0 < d$$

Pyramid sensor error:

$$\sigma^2 = \frac{1}{2N} + \frac{2(e_{\text{n}})^2}{N}\text{rad}^2$$

Wavefront slope measurement using a CCD camera:

$$\text{SNR} = \frac{n_{\text{p}}}{\left\{n_{\text{p}} + N_{\text{D}}\left[n_{\text{B}}^2 + \left(\frac{e_{\text{n}}}{G}\right)^2\right]\right\}^{1/2}}$$

Variance of a single curvature measurement:

$$\sigma^2_{\text{Curv. Sens.}} = \frac{p^2}{f^4 N_{\text{p}}}$$

Blur requirement:

$$p \geq \frac{\lambda f^2}{\lambda f + r_0^2}$$

Equation Summary

Temporal error:

$$\sigma^2_{\text{Temp}} = \left(\frac{f_G}{f_{BW}}\right)^{5/3}$$

Wavefront error variance assuming Shack–Hartmann wavefront sensor:

$$\sigma^2_{\text{WFS}} = 0.17\left(\frac{d_{\text{sub}}}{r_0}\right)^{5/3}$$

Wavefront measurement error:

$$\sigma^2_{\text{cone}} = \left(\frac{D}{d_0}\right)^{5/3}$$

Visual magnitude related to brightness:

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

Laser radar equation:

$$F_{\text{Rayleigh}} = \eta T_A^2 \frac{\sigma_R n_R}{4\pi z^2_{\text{LGS}}} \frac{\Delta z \lambda_{\text{LGS}} E}{hc}$$

Detected sodium-line photon flux:

$$F_{\text{Sodium}} = \eta T_A^2 \frac{\sigma_{\text{Na}} \rho_{\text{Col}}}{4\pi z^2_{\text{LGS}}} \frac{\lambda_{\text{LGS}} E}{hc}$$

Number of actuators required:

$$N_{\text{Act}} = \frac{\pi}{4}\left(\frac{D}{s_{\text{Act}}}\right)^2$$

Inverted spatial fitting error for zonal correction:

$$N_{\text{Act}} = \frac{\pi}{4}\left(\frac{D}{r_0}\right)^2 \left[\frac{\kappa}{\ln(1/S)}\right]^{6/5}$$

Equation Summary

Inverted spatial fitting error for Zernike modal correction:

$$N_{\text{Zern. modes}} = 0.24\left[\ln\left(\frac{1}{S}\right)\right]^{-1.15}\left(\frac{D}{r_0}\right)^{1.92}$$

Angular stroke for tilt-corrector mirror:

$$\text{Stroke}_{\text{Tilt Mirror}} = \frac{1}{2}(2.5\sigma_{\text{Tilt}})\left(\frac{D_{\text{Telescope}}}{D_{\text{Tilt Mirror}}}\right)$$

Variance of wavefront tilt:

$$\sigma^2_{\text{Tilt}} = 0.184\left(\frac{D}{r_0}\right)^{5/3}\left(\frac{\lambda}{D}\right)^2$$

Total required stroke for deformable mirror:

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

Sensitivity function:

$$\mathbf{S} = [\mathbf{I} + \mathbf{GK}]^{-1}$$

Complementary sensitivity function:

$$\mathbf{T} = [\mathbf{I} + \mathbf{GK}]^{-1}\mathbf{GK}$$

Standard controller in leaky integrator form:

$$\mathbf{K}(z) = \frac{g_1 z}{z - g_2}\mathbf{R}$$

Standard controller difference equation:

$$\mathbf{u}_k = g_1\mathbf{R}\mathbf{e}_k + g_2\mathbf{u}_{k-1}$$

Modal feedback difference equation:

$$\mathbf{u}_k = g_1\mathbf{R}\mathbf{e}_k + g_2\mathbf{u}_{k-1} - g_1\mathbf{M}\mathbf{u}_{k-1}$$

Equation Summary

Crossover frequency estimation for standard controller:

$$\omega_c = \cos^{-1}\left(\frac{g_2^2 + 1 - g_1^2}{2g_2}\right)\frac{f_s}{2\pi}$$

DC gain estimation for standard controller:

$$g_{DC} = \frac{g_1}{1 - g_2}$$

Actuator observability:

$$\text{obs}(a_j) = \sqrt{\sum_{i=1}^{n_m}(B_{ij})^2}$$

Measurement observability:

$$\text{obs}(m_i) = \sqrt{\sum_{j=1}^{n_a}(B_{ij})^2}$$

Piston, waffle, tip, and tilt removal:

$$\mathbf{B}_c = (\mathbf{I} - \mathbf{P}_t)\mathbf{B}_r(\mathbf{I} - \mathbf{P}_p)(\mathbf{I} - \mathbf{P}_{wp})$$

Truncated SVD reconstructor:

$$\mathbf{R}_r = -\mathbf{V}\boldsymbol{\Sigma}_f^{-1}\mathbf{U}^T$$

Tikhonov regularization reconstructor:

$$\mathbf{R}_r = -\left(\mathbf{B}_c^T\mathbf{B}_c + \alpha^2\mathbf{I}\right)^{-1}\mathbf{B}_c^T$$

Modal suppression reconstructor:

$$\mathbf{R}_r = -\left(\mathbf{B}_c^T\mathbf{B}_c + \mathbf{W}\right)^{-1}\mathbf{B}_c^T$$

Equation Summary

Modal feedback matrix:

$$\mathbf{M}_\mathrm{r} = \left(\mathbf{B}_\mathrm{c}^T \mathbf{B}_\mathrm{c} + \mathbf{W}\right)^{-1} \mathbf{W}$$

Expanded control matrices:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_\mathrm{r} \\ \mathbf{S}_\mathrm{l}\mathbf{R}_\mathrm{r} \end{bmatrix} \qquad \mathbf{M} = \begin{bmatrix} \mathbf{M}_\mathrm{r} & \mathbf{0} \\ \mathbf{S}_\mathrm{l}\mathbf{M}_\mathrm{r} & \mathbf{I} \end{bmatrix}$$

Slope discrepancy matrix:

$$\mathbf{S}_\mathrm{d} = (\mathbf{I} + \mathbf{BR})$$

Offload matrix with explicit reconstructor:

$$\mathbf{O} = \mathbf{R}_\mathrm{o}\mathbf{B}_\mathrm{p}$$

Offload matrix with implicit reconstructor:

$$\mathbf{O} = \left[(\mathbf{R}_\mathrm{p}\mathbf{B}_\mathrm{o})^T (\mathbf{R}_\mathrm{p}\mathbf{B}_\mathrm{o})\right]^{-1} (\mathbf{R}_\mathrm{p}\mathbf{B}_\mathrm{o})^T$$

Modal suppression reconstructor with estimation:

$$\mathbf{R}_\mathrm{r} = -\left(\mathbf{B}_\mathrm{c}^T \mathbf{B}_\mathrm{c} + \mathbf{W} + \sigma_n^2 \mathbf{Q}_\mathrm{d}^{-1}\right)^{-1} \mathbf{B}_\mathrm{c}^T$$

Bibliography

Born, M. and E. Wolf, *Principles of Optics*, Fifth Edition, Oxford, UK, Pergamon Press (1975).

Bufton, J. L., "Comparison of vertical profile turbulence structure with stellar observations," *Appl. Opt.* **12**, 1785 (1973).

Churnside, J. H., "Aperture averaging of optical scintillations in the turbulent atmosphere," *Appl. Opt.* **30**, 1982 (1991).

Ellerbroek, B. L., "Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques," *J. Opt. Soc. Am. A* **19**, 1803–1816 (2002).

Ellerbroek, B. L., L. Gilles, and C. R. Vogel, "Computationally efficient wavefront reconstructor for simulation of multiconjugate adaptive optics on giant telescopes," *Proc. SPIE* **4839**, 989–1000 (2003) [doi:10.1117/12.459673].

Forbes, F. F., "Bimorph PZT active mirror," *Proc. SPIE* **1114**, 146–151 (1989).

Franklin, G. F., J. D. Powell, and M. L. Workman, *Digital Control of Dynamic Systems*, Second Edition, Addison-Wesley, Reading, MA (1990).

Fried, D. L., "Focus anisoplanatism in the limit of infinitely many artificial-guide-star referrence spots," *J. Opt. Soc. Am. A* **12**, 939 (1995).

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

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

Gonzalez, R. C. and R. E. Woods, *Digital Image Processing*, Second Edition, Prentice Hall, Upper Saddle River, NJ (2002).

Goodman, J. W., *Introduction to Fourier Optics*, Second Edition, McGraw-Hill, New York (1996).

Bibliography

Goodman, J. W., *Statistical Optics*, John Wiley and Sons, New York (1985).

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

Grosso, R. P. and M. Yellin, "The membrane mirror as an adaptive optical element," *J. Opt. Soc. Am.* **67**, 399 (1977).

Hardy, J. W., *Adaptive Optics for Astronomical Telescopes*, Oxford Univ. Press, Oxford, UK (1998).

Hayes, M. H., *Statistical Digital Signal Processing and Modeling*, John Wiley and Sons, New York (1996).

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

Hyver, G. A. and R. M. Blankinship, "ALI high-power beam control," *Advances in the Astronautical Sciences* **88**, 445–469 (1995).

ISO Standard 11146, "Lasers and laser related equipment – Test methods for laser beam widths, divergence angles and beam propagation ratios," International Organization for Standardization, Geneva, Switzerland (2005).

Johnson, B. and D. V. Murphy, *Thermal Blooming Laboratory Experiment*, Part I, Lincoln Laboratory MIT Project Report BCP-2 (November 1988).

Kalman, R. E., "A new approach to linear filtering and prediction problems," *Transaction of the ASME—Journal of Basic Engineering* **82**(D), 35–45 (1960).

Lee, L. H., "Loopshaped wavefront control using open-loop reconstructors," *Optics Express* **14**(17), 7477–7486 (2006).

Miller, M. G. and P. L. Zieske, "Turbulence environment characterization," RADC-79-131, ADA072379, Rome Air Development Center, U.S. Air Force, Rome, NY (1979).

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

Bibliography

Oppenheim, A. V., R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, Second Edition, Prentice Hall, Upper Saddle River, NJ (1999).

O'Neil, P. V., *Advanced Engineering Mathematics*, Sixth Edition, Nelson, Toronto, Canada (2007).

OSI Optoelectronics, "Application Note No. 2, Silicon Photodiode Physics and Technology," OSI Optoelectronics (1982).

OSI Optoelectronics, "Application Note No. 8, Lateral Effect Photodiodes," OSI Optoelectronics (1982).

Perram, G. P., S. J. Cusumano, R. L. Hengehold, and S. T. Fiorino, *An Introduction to Laser Weapon Systems*, Directed Energy Professional Society, Albuquerque, NM (2010).

Poyneer, L. A, D. T. Gavel, and J. M. Brase, "Fast wavefront reconstruction in large adaptive optics system with use of the Fourier transform," *J. Opt. Soc. Am. A* **19**(10), 2100–2111 (2002).

Poyneer, L. A. and B. Macintosh, "Spatially filtered wavefront sensor for high order adaptive optics," *J. Opt. Soc. Am. A* **21**(5), 810–819 (2004).

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

Simon, D, *Optimal State Estimation, Kalman, H-∞, and Nonlinear Approaches*, John Wiley and Sons, Hoboken, NJ (2006).

Sinha, N. K. and B. Kuszta, *Modeling and Identification of Dynamic Systems*, Van Nostrand Reinhold, New York (1983).

Skogestad, S. and I. Postlethwaite, *Multivariable Feedback Control, Analysis and Design*, John Wiley and Sons, Hoboken, NJ (2009).

Bibliography

Taranenko, V. G., G. P. Koshelev, and N. S. Romanyuk, "Local deformations of solid mirrors and their frequency dependence," *Sov. J. Opt. Technol.* **48**, 650 (1981).

Trefethen L. N. and D. Bau III, *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia (1997).

Tyler, G., "Reconstruction and assessment of the least-squares and slope discrepancy components of the phase," *J. Opt. Soc. Am. A* **17**, 1828–1839 (2000).

Tyson, R. K., *Introduction to Adaptive Optics*, SPIE Press, Bellingham, WA (2000) [doi:10.1117/3.358220].

Tyson, R. K., *Principles of Adaptive Optics*, Third Edition, CRC Press, Boca Raton, FL (2011).

Tyson, R. K., "Adaptive optics and ground-to-space laser communications," *Appl. Opt.* **35**, 3640–3646 (1996).

Uchino, K., *Ferroelectric Devices*, CRC Press, Boca Raton, FL (2000).

Ulrich, P. B., "Hufnagel-Valley profiles for specified values of the coherence length and isoplanatic patch angle," W. J. Schafer Associates, WJSA/MA/TN-88-013, Arlington, VA (1988).

Vogel, C. R., *Computational Methods for Inverse Problems*, Society for Industrial and Applied Mathematics, Philadelphia (2002).

Vogel, C. R. and Q. Yang, "Fast optimal wavefront reconstruction for multi-conjugate adaptive optics using the Fourier domain preconditioned conjugate gradient algorithm," *Optics Express* **14**(17), 7487–7498 (2006).

Index

Index

Index

Index

Index

Robert K. Tyson is an Associate Professor of Physics and Optical Science at The University of North Carolina at Charlotte. He has a B.S. in physics from Penn State University and M.S. and Ph.D. degrees in physics from West Virginia University. He was a senior systems engineer with United Technologies Optical Systems from 1978 to 1987 and a senior scientist with Schafer Corporation until 1999. He is the author of *Principles of Adaptive Optics* [Academic Press (1991), Second Edition (1998), Third Edition, CRC Press (2011)], *Lighter Side of Adaptive Optics*, SPIE Press (2009), and *Introduction to Adaptive Optics*, SPIE Press (2000) and the editor of ten volumes on adaptive optics. He is also a Fellow of SPIE. Professor Tyson's current research interests include atmospheric turbulence studies, classical diffraction, novel wavefront sensing, and amplitude and phase manipulation techniques to enhance propagation, laser communications, and imaging.

Benjamin W. Frazier is a Principal Electro-Optical Engineer with AOA Xinetics, a small business unit of Northrop Grumman Aerospace Systems. He has B.S.E.E. and M.S.E.E. degrees from the University of North Carolina at Charlotte, where he focused on control theory for adaptive optics systems. Frazier has extensive experience with systems integration and testing of beam control systems and components, particularly deformable mirrors and wavefront control systems for high-power and solid state lasers. He currently supports multiple programs in systems engineering, integration and test, data analysis, and performance assessment.

Adaptive Optics
Second Edition

Robert K. Tyson and Benjamin W. Frazier

This Field Guide provides a summary of the methods for determining the requirements of an adaptive optics system, the performance of the system, and the requirements for the components of the system. Many of the expressions are in the form of integrals in which case the authors show the results graphically for a variety of practical values. This second edition has a greatly expanded presentation of adaptive optics control system design and operation. Discussions of control models are accompanied by various recommendations for implementing the algorithms in hardware.

SPIE Field Guides

The aim of each SPIE Field Guide is to distill a major field of optical science or technology into a handy desk or briefcase reference that provides basic, essential information about optical principles, techniques, or phenomena.

Written for you—the practicing engineer or scientist—each field guide includes the key definitions, equations, illustrations, application examples, design considerations, methods, and tips that you need in the lab and in the field.

John E. Greivenkamp
Series Editor

SPIE

P.O. Box 10
Bellingham, WA 98227-0010
ISBN: 9780819490179
SPIE Vol. No.: FG24

ISBN 978-0-8194-9017-9
9 780819 490179 90000

www.spie.org/press/fieldguides

SPIE PRESS