GUIDELINES FOR ADAPTIVE-OPTIC CORRECTION BASED ON APERTURE FILTRATION

A Dissertation

Submitted to the Graduate School

of the University of Notre Dame

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

by

John Paul Siegenthaler

Dr. Eric J. Jumper, Director

Graduate Program in Aerospace and Mechanical Engineering

Notre Dame, Indiana

December 2008

© Copyright 2008

John Paul Siegenthaler

GUIDELINES FOR ADAPTIVE-OPTIC CORRECTION BASED ON APERTURE FILTRATION

Abstract

by

John Paul Siegenthaler

Aperture filtration refers to the effects of viewing optical wavefront distortions of infinite extent through a finite aperture. If the length-scale of the aberration is larger than this aperture, then the portion of the aberration visible in the aperture at any moment in time will not reach the full magnitude of the aberration seen in its entirety. The aperture acts as a spatial filter, mitigating the effects of large-scale wavefront distortions while having little effect on smaller-scale aberrations, with the dividing line between large-scale and small-scale being the size of the aperture itself. This dissertation presents and charts the development of a set of analytic formulas for judging and predicting the effectiveness of adaptive-optic corrective systems applied over finite apertures. This includes some simplified formulas and benchmarks as guides for the minimum requirements a system will need to meet to be effective, and the maximum degree of effectiveness such systems can reasonably achieve.

This work is dedicated to my parents. To my father, Kenneth E. Siegenthaler, who taught me to endure in the face of difficulty, and to my mother, Phyllis J. Petersen, who encouraged me to keep moving forward.

CONTENTS

Figures	vii
Acknowledgments	xvii
Nomenclature	xviii
Chapter 1: Introduction	1
1.1. Overview	1
1.2. Optical Applications and Distortions	2
1.2.1. Optical Distortions	2
1.2.2. Aero-Optics	
1.2.3. Adaptive Optics	4
1.3. Chapter Överviews	4
Chapter 2: Fundamentals of Optical Propagation and Aberrations	7
2.1. Overview	7
2.2. Propagation of Light	7
2.2.1 Near-Field Pronagation	9
2.2.2. Propagation over Intermediate Distances	13
2.2.3. Far-Field Propagation	20
2.2.5.1 di Freder Popugation	
2.3 Fluid-Ontic Interaction	33
2.3.1 Near-Field Effects	35
2.3.2. Far-field Effects	
Chapter 2. Methods of Optical Massurement and Competion	17
2 1 Overview	
2.2 Tin/Tilt A hometions	
2.2.1 Definitions of Tilt	
3.2.1. Definitions of Tilt	
3.2.2. Weasuring The	
3.3. Wavefront Aderrations	
3.3.1. Figher-Order wavefront Measurements	0U
3.3.2. Snack-Haruman Sensors	
3.3.3. Malley Probes	

3.4. Applying Correction	67
3.4.1. Tip-Tilt Correction	
3.4.2. Higher-Order Correction	69
3.5. Concerns and Limitations of Correction	
3.5.1. Spatial Resolution	
3.5.2. Temporal Resolution	
3.5.3. Latency	
3.5.4. Combined Effects	
Chapter 4: Atmospheric Propagation	86
4.1. Overview	
4.2. Atmospheric Models	87
4.2.1. Turbulence as a Separate Quality	
4.2.2. Spectral Analysis of Turbulence	
4.2.3. The Kolmogorov Spectrum	
4.3. Optical Turbulence	105
4.3.1. Coherence Length	108
4.3.2. The Greenwood Frequency	
4.4. The Limitations of C_n^2	115
Chapter 5: Aero-Optic Shear Layer Experiments at AEDC	
5.1. Overview	
5.2. The Acoustic Research Tunnel	119
5.2.1. Initial Results	120
5.2.2. Re-reduction of the AEDC Data	
5.2.3. Severity of the Distortion	126
5.2.4. Predictions and Prevailing Theory	128
Chapter 6: The Physical and Optical Mechanisms of Shear Layers	
6.1. Overview	
6.2. Shear Layer Mechanics	
6.2.1. Kelvin-Helmholtz Instability	
6.2.2. Initial Development	
6.2.3. Growth	
6.3. Weakly-Compressible Modeling	
6.3.1. Discrete Vortex Modeling	155
6.3.2. Optical Modeling	159
6.3.3. Model Results	
6.4. Conclusions	170
Chapter 7: Optical Characterization of a Shear Layer	
7.1. Overview	
7.2. Notre Dame Weakly-Compressible Shear Layer Facility	
7.2.1. Facility Design	174
7.2.2. Modifying the Facility	177

7.2.3. Flow Conditions	
7.3. Verification of the Weakly Compressible Model	
7.4. Optical Measurements	
7.4.1. Measurement System	
7.4.2. Results	
	222
Chapter 8: Effects and Requirements of Correction	
8.1. Overview	
8.2. Assumption of a Sinusoidal wavefront	
8.3. Tip-Till Correction within an Aperture	
8.3.1. Simple Simulated Correction	
8.5.2. Optical Scaling for Shear Layers	234
8.4. The Aperture Filter	
8.4.2 Two Dimensional Circular A parture Eilter	239 247
8.4.2. C Tilt Aperture Filters	
8.4.4. Diston Only Correction	
8.5. Higher Order Correction	
8.5.1 Basis Spline Correction Fitting	
8.5.2 Diston Only Sub Aparturas	
8.5.2. Fision-Only Sub-Apertures	
8.5.4 Continuous Surfaces	
8.5.5. Higher Order Correction by Simulation	
8.6. Temporal Effects	
8.6.1 Periodic Correction	
8.6.2 Latency	
8.6.3 Periodic Correction with Latency	
8.7 Using the Filter Functions	
8.8 Conclusions	289 292
0.0. Conclusions	
Chapter 9: Aperture FilterS in Application	
9.1. Overview	
9.2. T/T Correction in a Compressible Shear Layer	
9.3. Periodic Correction of a Heated Jet	
9.4. Tracking Systems	306
9.5. Application to Atmospheric Turbulence	
9.5.1. Piston and T/T Correction	
9.5.2. Latency	
9.6. Conclusions	
Chapter 10: Conclusions and Recommendations	
10.1. Context and Foundations	
10.2. Development and Understanding	
10.3. Contributions	
10.4. Recommendations for Future Work	

Appendix A: The Redesign and Construction of a New Shear Layer Facility	322
A.1. Introduction	322
A.2. Test Section	323
A.3. Diffuser	325
A.4. Sonic Throat	326
A.5. Splitter-Plate Section	327
A.6. High-Speed Inlet	328
A.7. Low-Speed Inlet	
Appendix B: Expanded Derivations of Aperture Filter Functions	
B.1. Introduction	
B.2. Uncorrected Cases	338
B.2.1. One-Dimensional Aperture	
B.2.2. Two-Dimensional Aperture	339
B.3. Piston Correction	
B.3.1. One-Dimensional Aperture	
B.3.2. Two-Dimensional Aperture	
B.4. Z-Tilt Correction	
B.4.1. One-Dimensional Aperture	
B.4.2. Two-Dimensional Aperture	
B.5. G-Tilt Correction	
B.5.1. One-Dimensional Aperture	353
B.5.2. Two-Dimensional Aperture	
B.6. Temporal Effects	358
B.6.1. Periodic Correction	359
B.6.2. Latency	
B.6.3. Periodic Correction with Latency	
B.6.4. Temporal Effects for Two-Dimensional Apertures	
B.7. Higher-Order Correction	
B.7.1. The General Case	
B.7.2. Local Piston Correction	
B.7.3. First-Order Correction	
B.7.4. Third-Order Correction	

FIGURES

Figure 2.1: A sketch of rays and wavefronts in a non-uniform medium 10
Figure 2.2: The construction of Huygens' principle
Figure 2.3: Propagation by complex transmission functions
Figure 2.4: Propagation by a complex transmission function, as found from a near-field wavefront
Figure 2.5: Propagation by a complex transmission function, through a medium with a variable index of refraction
Figure 2.6: Sinc and squared sinc functions
Figure 2.7: (a) A rectangular aperture (b) The associated far-field pattern of irradiance. 28
Figure 2.8: Bessel function-based curves
Figure 2.9: (a) A circular aperture (b) The associated far-field pattern of irradiance 30
Figure 2.10: Near-field irradiance for a top hat (a) and a Gaussian (b) beam, both of radius 0.5
Figure 2.11: An optical ray as part of a wavefront passing through a distorting flow field.
Figure 2.12: Phase varying linearly over an aperture (a) and a displaced far-field pattern (b)
Figure 2.13: A spherical phase aberration over an aperture (a) and an unfocused far-field pattern (b)

Figure 2.14: Phase variations over an aperture (a) and a divided far-field pattern (b) 43
Figure 2.15: Phase variations over an aperture (a) and a blurred and fragmented far-field pattern (b)
Figure 3.1: Notre Dame Adaptive-Optic system. (a) Full system. (b) Detail of the Tip- Tilt compensation portion of the system
Figure 3.2: G-tilt defined by averaging local tilt, Vs Z-tilt defined by a linear fit 50
Figure 3.3: Sensor regions of a quad cell
Figure 3.4: Components and arrangement of a Lateral Effect Detector sensor
Figure 3.5: A far-field irradiance pattern based on a sinusoidal wavefront (a) and this pattern if the wavefront is corrected according to Z-tilt
Figure 3.6: (a) A simulated wavefront, with corrections. (b) Uncorrected and diffraction limited far-field patterns. (c) Far-field patterns with various types of tilt correction
Figure 3.7: A Shack-Hartman sensor
Figure 3.8: Indications of convective velocity from Malley probe measurements
Figure 3.9: A schematic of the T/T corrective system within the ND AO system
Figure 3.10: (a) flat mirror reflections, (b) shaped mirror conjugate reflection, (c) half- magnitude shaped mirror correction
Figure 3.11: A model of the Xinetics deformable mirror73
Figure 3.12: A simulated wavefront produced by summation of sine waves75
Figure 3.13: Simulated Shack-Hartman reconstruction76
Figure 3.14: Simulated Shack-Hartman reconstruction77

viii

Figure 3.16: (a) A 12 th order polynomial fit to the 12-sensor reconstruction of the wavefront. (b) Residual wavefront variations after correction by the reconstruction in (a)
Figure 3.17: Time evolution of periodic correction, (a) near-field phase, (b) far-field irradiance
Figure 3.18: Time evolution of periodic correction, (a) near-field phase, (b) far-field irradiance
Figure 3.19: Time varying Strehl ratios for periodic correction
Figure 4.1: Turbulent Spectra Approximation
Figure 4.2: Structure function with $r^{2/3}$ fit over the inertial range
Figure 4.3: Fried's normalized signal-to-noise ratio. ¹³
Figure 5.1: AEDC ART facility.(Havener) ⁶
Figure 5.2: Station 1 wavefront reconstruction, from data filtered at 2.5 kHz. (Hugo) ⁸ 121
Figure 5.3: Station 2 wavefront reconstruction, from data filtered at 2.5 kHz. (Hugo) ⁸ 121
Figure 5.4: Wavefront reconstruction from AEDC data for station 1 with varying cutoff frequencies in high-pass filtering: (a) 2000 Hz, (b) 1200 Hz, (c) 750 Hz, (d) 0 Hz (unfiltered). (Fitzgerald, 2002.) ⁹
Figure 5.5: Wavefront reconstruction from AEDC data for station 2 with varying cutoff frequencies in high-pass filtering: (a) 2000 Hz, (b) 1200 Hz, (c) 750 Hz, (d) 0 Hz (unfiltered). (Fitzgerald, 2002.) ⁹
Figure 5.6: One-dimensional wavefronts and far field patterns as a function of time for station 1 (left) and 2 (right). (Cicchiello, 2001)
Figure 5.7: Strehl ratios produced by the irradiance patterns in Fig 5.6. (Cicchiello, 2001) ¹¹
Figure 5.8: (left) Wavefronts generated by computational simulation and an adiabatic heating model. (right) AEDC results for flow conditions similar to those in the computational simulation.(Fitzgerald) ¹³

Figure 6.1: Separation, shear layer, and other potentially-aberrating flows for a possible turret configuration
Figure 6.2: Strehl ratio as a function of azimuth look angle. (Demos Kryazis) ¹
Figure 6.3: The basic flow of a shear layer
Figure 6.4: Hyperbolic-tangent velocity profile approximation for shear layers
Figure 6.5: (a) Vortex line, (b) perturbed vortex line with non-zero net forces and torques.
Figure 6.6: Stability curve for a hyperbolic-tangent velocity profile. (Michalke)
Figure 6.7: Linear-fit basis of vorticity thickness
Figure 6.8: The basis of momentum thickness for a boundary layer
Figure 6.9: Sketch of spatially developing profiles and momentum thickness of a shear layer
Figure 6.10: Effects of disturbances extending into the flows of a shear layer
Figure 6.11: Vortex rows resulting from Kelvin-Helmholtz instability. (Regularized with forcing.) (Roberts, 1982)
Figure 6.12: Vortex pairing. (Winant, 1974)
Figure 6.13: Momentum thickness growth rate for incompressible flows. (Ho and Huerre, 1984) ¹¹
Figure 6.14: Representation of a two-dimensional shear layer by discrete vortices 156
Figure 6.15: Velocity field produced by the DVM model for AEDC conditions, centered at a position of $x=48.3$ cm.(Fitzgerald) ⁴²
Figure 6.16: (left) Perturbed and rolled-up vortex line from WCM simulation. (right) Pressure field associated with the simulated conditions to the left. ⁴²

Figure	6.17: (left) Perturbed and rolled-up vortex line from WCM simulation showing one large roller. (right) Pressure field associated with the simulated conditions the left. ⁴²	to 165
Figure	6.18: (a) Wavefronts generated by computational simulation and an adiabatic heating model. (b) AEDC results for flow conditions similar to those in the computational simulation. (Fitzgerald) ⁴²	166
Figure	6.19: WCM wavefronts, corresponding to AEDC ART station 2. $(Fitzgerald)^{42}$	167
Figure	6.20: Reconstructed wavefronts from measurements at station 1 of the AEDC ART facility. (Fitzgerald) ⁴²	168
Figure	6.21: Superimposing station 1 wavefronts onto WCM simulation for station 2. (Fitzgerald) ⁴²	169
Figure	7.1: Schematic of the Notre Dame Weakly-Compressible Shear Layer Facility. (First configuration.)	175
Figure	7.2: Schematic of WCSL test section	176
Figure	7.3: Sonic throat between test section and diffuser. Flow is from left to right 1	178
Figure	7.4: Settling tank interior, with baffles and sound-dampening foam	179
Figure	7.5: Schematic of a "quiet valve" segment 1	180
Figure	7.6: Test section with removable Plexiglas segments 1	182
Figure	7.7: Schematic of the Notre Dame Weakly-Compressible Shear Layer Facility. (Modified)	183
Figure	7.8: Mach number profile at the splitter plate 1	186
Figure	7.9: Streamwise velocity profile at the splitter plate	186
Figure	7.10: Static pressure along the WCSL test section	188
Figure	7.11: Test section static pressures under optical-survey run conditions	189

Figure 7.12: Condensation-based flow visualizations of a structure in the WCSL facility.
Figure 7.13: Phase-Locked Average of 200 Trigger Traces Overlaid by a Single Trace.
Figure 7.14: Average pressures associated with a coherent structure at $x = 0.5$ m 195
Figure 7.15: Pseudo-reconstruction of an average static pressure well (left) compared to WCM simulation. (right)
Figure 7.16: Malley probe beams directed through the flow and WCSL test section 198
Figure 7.17: Plan-view schematic of optical bench and sensors
Figure 7.18: WCSL facility with Malley probe setup on the optical bench 201
Figure 7.19: Calibration curves for sensors used in study of the ND WCSL facility 202
Figure 7.20: Deflection data (a) and OPD reconstruction (b) from a position of 20.4 cm.
Figure 7.21: Partitioning of reconstructed OPD with an aperture (a) and the partitioned section (b)
Figure 7.22: Applying a linear fit to the reconstructed wavefront within an aperture (a) and the wavefront after T/T correction (b)
Figure 7.23: Beam-deflection power spectra: upstream beam (a), downstream beam (b).
Figure 7.24: Phase of beam-deflection cross-correlation
Figure 7.25: Phase of beam-deflection cross-correlation
Figure 7.26: PSD of vibration and electronic noise data from the Malley probe
Figure 7.27: PSD of beam deflection, shifted by position
Figure 7.28: PSD of beam deflection, shifted by position

Figure 7.29: Frequency with position and a/x +b fit	218
Figure 7.30: Characteristic Λ and linear growth rate	220
Figure 7.31: Optical aberration from a shear layer, varying with position and aperture size.	221
Figure 8.1: Momentum thickness of forced shear layers. (Oster and Wygnanski, 1982)) ¹ 226
Figure 8.2: WCM realizations of a shear layer without forcing (left) and with forcing (right). (Nightingale, 2005) ⁵	227
Figure 8.3: Normalized beam jitter according to WCM realization of a shear layer without forcing (left) and with forcing (right). (Nightingale, 2005) ⁵	227
Figure 8.4: Segment of OPL history for 800 Hz, forced shear layer at $x = 570$ mm. (Rennie, 2006) ⁶	228
Figure 8.5: T/T correction for distortions with a long length scale.	230
Figure 8.6: T/T correction for distortions with a short length scale	230
Figure 8.7: T/T corrected <i>OPD</i> for simulated sinusoidal disturbances.	232
Figure 8.8: Non-dimensionalized results for sinusoidal disturbances.	234
Figure 8.9: Optical aberration from a shear layer, varying with position and aperture s	ize. 237
Figure 8.10: Non-dimensionalized Shear Layer Results.	238
Figure 8.11: System gain for T/T correction over a one-dimensional aperture	245
Figure 8.12: System gain for T/T correction over a one-dimensional aperture	246
Figure 8.13: System gain for T/T correction over a two-dimensional aperture	248
Figure 8.14: System gain for T/T correction over a two-dimensional aperture	249
Figure 8.15: System gain for G-tilt correction over a one-dimensional aperture	251

Figure 8.16: G-tilt correction for $A_p > \Lambda$
Figure 8.17: Far-field effects of G-tilt correction for short length scales 253
Figure 8.18: System gain for G-tilt correction over a one-dimensional aperture
Figure 8.19: System gain for G-tilt correction over a circular aperture
Figure 8.20: System gain for G-tilt correction over a circular aperture
Figure 8.21: System gain for piston correction over one-dimensional (left) and circular (right) apertures
Figure 8.22: System gain for piston correction over one-dimensional (left) and circular (right) apertures
Figure 8.23: B-spline curve fitting: A set Basis functions (left) are scaled and summed to produce a least-squares fit to a given curve or set of data points. (right)
Figure 8.24: Segmented Mirror Correction
Figure 8.25: Third-order basis function
Figure 8.25: Third-order basis function
 Figure 8.25: Third-order basis function
 Figure 8.25: Third-order basis function. Figure 8.26: Numerically-computed piston-only corrective gain with sub-apertures. Figure 8.27: Numerically-computed piston-only corrective gain scaled by sub-aperture dimensions. 275 Figure 8.28: Effects of segmented correction.
Figure 8.25: Third-order basis function. 272 Figure 8.26: Numerically-computed piston-only corrective gain with sub-apertures. 274 Figure 8.27: Numerically-computed piston-only corrective gain scaled by sub-aperture dimensions. 275 Figure 8.28: Effects of segmented correction. 276 Figure 8.29: Effects of correction with a continuous surface. 277
Figure 8.25: Third-order basis function. 272 Figure 8.26: Numerically-computed piston-only corrective gain with sub-apertures. 274 Figure 8.27: Numerically-computed piston-only corrective gain scaled by sub-aperture dimensions. 275 Figure 8.28: Effects of segmented correction. 276 Figure 8.29: Effects of correction with a continuous surface. 277 Figure 8.30: Comparing segmented and continuous surface correction for large numbers of sub-apertures. 278
Figure 8.25: Third-order basis function. 272 Figure 8.26: Numerically-computed piston-only corrective gain with sub-apertures. 274 Figure 8.27: Numerically-computed piston-only corrective gain scaled by sub-aperture dimensions. 275 Figure 8.28: Effects of segmented correction. 276 Figure 8.29: Effects of correction with a continuous surface. 277 Figure 8.30: Comparing segmented and continuous surface correction for large numbers of sub-apertures. 278 Figure 8.31: Effects of segmented correction. 279
Figure 8.25: Third-order basis function. 272 Figure 8.26: Numerically-computed piston-only corrective gain with sub-apertures. 274 Figure 8.27: Numerically-computed piston-only corrective gain scaled by sub-aperture dimensions. 275 Figure 8.28: Effects of segmented correction. 276 Figure 8.29: Effects of correction with a continuous surface. 277 Figure 8.30: Comparing segmented and continuous surface correction for large numbers of sub-apertures. 278 Figure 8.31: Effects of segmented correction. 279 Figure 8.32: Effects of segmented correction. 280

Figure	8.34: Reduction in wavefront distortions for different degrees of latency in the correction
Figure	9.1: Comparing power spectra for beam jitter, tilt removed, and predicted filter function
Figure	9.2: (a) Tilt spectra for different aperture sizes. (b) Scaling tilt spectra by Strouhal number
Figure	9.3: Time-averaged Strehl ratio as a function of time delay between constructing a perfect correction and applying it (τ_2), normalized by the time-averaged Strehl ratio for $\tau_2 = 0.0$ for three values of the period between applying corrections (τ_1). (Cicchiello, 1997) ¹
Figure	9.4: Time series of experimental wave fronts for propagation through a heated jet. (Cicchiello, 1997) ¹ (Hugo, 1996)
Figure	9.5: Strehl ratio predictions by corrective gain function and large-aperture approximation
Figure	9.6: Normalized Strehl ratio predictions (left) compared to experimental results ¹ (right)
Figure	9.7: Scoring and Tracking cameras for Boeing test at ACL. ⁴
Figure	9.8: Results of Boeing tests at ACL. ⁴
Figure	9.9: Scaled Boeing results compared with circular-aperture filter function for Z-tilt correction
Figure	A.1: Wire-frame design and photo of the second-generation WCSL facility 323
Figure	A.2: Wire-frame design and photo of test section for the second WCSL
Figure	A.3: Wire-frame design of the diffuser for the second WCSL
Figure	A.4: Sonic throat for the second WCSL
Figure	A.5: Wire-frame design and photo of splitter-plate section for uniting the high and low-speed flows

Figure A.6: Wire-frame design and photo of high-speed inlet nozzle for the second WCSL.	. 330
Figure A.7: Wire-frame design of low-speed inlet expansion for the second WCSL	. 331
Figure A.8: "Straw-box" and spacer used for inducing total pressure loss	. 332
Figure A.9: Wire-frame design and photo of low-speed section (below the high-speed inlet) for inducing a drop in total pressure	1 . 333

ACKNOWLEDGMENTS

I would like to express my profound gratitude to my research advisor, Dr. Eric J. Jumper, who offered me the opportunity to strive for a Ph.D. at the University of Notre Dame, and who has been a valuable mentor in the years required to reach this point.

I express thanks to my dissertation committee, Dr. Stanislav Gordeyev, Dr. Scott C. Morris, and Dr. John E. Renaud. Their suggestions, advice and time are greatly appreciated.

I would like to thank Asad Asghar for taking me under his wing as my time as a Notre Dame graduate student was beginning and his was ending.

I would also like to thank Dr. Thomas C. Corke and David A. Cavalieri for their advice and support of my research.

These efforts were sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under Grant Number FA9550-06-1-0160. The U.S. government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon.

NOMENCLATURE

ACL	Advanced Concepts Laboratory
AEDC	Arnold Engineering Development Center
AO	Adaptive Optics
A_p	Aperture diameter
ART	Acoustic Research Tunnel
BQ	Beam Quality
С	Speed of light in a vacuum
CCD	Charge Coupled Device
C_n^2	Atmospheric optical turbulence parameter
C_p	Specific heat at constant pressure
C_v	Specific heat at constant volume
d	Width, diameter
D_A	Structure function for property A
DM	Deformable Mirror
DVM	Discrete Vortex Model
Ε	Energy
f	Frequency
f_{3dB}	Frequency associated with -3dB gain for a system
f_C	Frequency of correction
f_D	Frequency associated with disturbances
f_G	Greenwood frequency
FSM	Fast Steering Mirror
f_x	Scaled far-field position in the x-direction.
G	Gain function
$G_{x-tr x-sc}$	Cross spectral density (as defined in Boeing tracking experiment)
G_{xx}	Autospectral density (as defined in Boeing tracking experiment)
Ι	Intensity
J_n	Bessel function of the first kind, n th order.
k	Wavenumber
K_{GD}	Gladstone-Dale constant
L	Length
M_C	Convective Mach number
n	Index of refraction
N_C	Number of corrections per cycle
ND	Notre Dame

OPD	Optical Path Difference
OPL	Optical Path Length
Р	Pressure
р	Pressure (varying)
PSD	Power Spectral Density
q	Indicator of turbulent kinetic energy
R	Gas constant
r	Radial distance
rr	Relative position vector
r_0	Fried parameter
RANS	Reynolds-Averaged Navier-Stokes
R_{ij}	Reynolds stress, element (i,j)
r_s	Distance of spacing between actuators or corrective elements
S	Distance or position along a path
SABT	Small Aperture Beam Technique
SRA	Strong Reynolds Analogy
SR	Strehl ratio
St	Strouhal number
St_X	Strouhal number based on a length scale of X
Т	Temperature
t	Time
T/T	Tip-Tilt
TKE	Turbulent Kinetic Energy
и	Velocity
U	Velocity, complex transmission function
U_∞	Free-stream velocity
U_C	Convection velocity
V	Velocity
v	Velocity (y-direction)
WCSL	Weakly Compressible Shear-Layer
x	Position along x-axis
\vec{x}	Position vector
\overline{X}	Average value of <i>X</i> .
$\langle X \rangle$	Ensemble average of quantity X
\hat{X}	Fourier transform of X
\widetilde{X}	Varying component of property X
X_{rms}	Root-mean-square of quantity X.
у	Position along y-axis
Ζ	Length along z-axis
Z.	Position along z-axis
α_x	Angle of deflection in the x-direction
γ	Ratio of specific heats (C_p/C_v)
γ^2	Coherence function (as defined in Boeing tracking experiment)
δ_{vis}	Visual thickness
Δx	Small change or increment in quantity x

Momentum thickness
Vorticity thickness
Deflection angle
Disturbance length scale
Wavelength
Dynamic viscosity
Kinematic viscosity
Density
Residual phase variance due to property <i>x</i>
Phase variance
Time interval
Viscous stress, element (i,j)
Phase angle

CHAPTER 1:

INTRODUCTION

1.1. Overview

This dissertation will explore aspects of aero-optic distortions, produced as light passes through air under compressible flow conditions due to motion of air and a physical object having high velocities relative to each other. The goal of this study is to produce guidelines and estimates of required performance for the design and construction of systems to correct for these distortions. In pursuit of this goal the effects of apertures as filters will be examined and quantified.

Finite apertures act as spatial filters. If the length-scale of the aberration is larger than this aperture, then the portion of the aberration visible in the aperture at any moment in time will not reach the full magnitude of the aberration seen in its entirety. Thus, the overall amplitude or form of the distortion seen in the aperture may depend upon the relative length scales of aberrations in the wavefront and the aperture through which these aberrations are viewed. This dissertation presents and charts the development of a set of analytic formulas for judging and predicting the effectiveness of adaptive-optic corrective systems applied over finite apertures. This includes some simplified formulas and benchmarks as guides for the minimum requirements a system will need to meet to be effective, and the maximum degree of effectiveness such systems can reasonably achieve

1

1.2. Optical Applications and Distortions

In recent years there have been increasing efforts to mount optical systems on aircraft. Airborne cameras and other sensors have been around for some time, but their performance can be improved. The increasing demands for higher-bandwidth communications are producing greater interest in other forms of transmitting and receiving airborne optical systems,¹ and clearly we are moving into the era of speed-of-light, airborne offensive and defensive laser weapon systems.²

Whether the energy or signals are being projected or received, the air that an aircraft travels through may become a source of aberration and distortion of the electromagnetic waves being transmitted. Studies and characterization of these distortions have been ongoing for decades.^{3,4,5,6} However, these studies have focused on stationary, ground-based systems. As will be shown in this dissertation, the descriptive equations, quantifying parameters, and guidelines for corrective systems produced by these studies are not always applicable to a moving system under flight conditions.

1.2.1. Optical Distortions

When visible light or other electromagnetic radiation passes though a medium, local variations in the properties of that medium can lead to differences in how that light travels along different pathways through that medium. In a traveling wavefront or beam of such radiation, this is perceived as distortions that can cause problems in trying to transmit or receive images, signals, or simply the energy itself through that medium. If the goal is to produce an image then these distortions may change the apparent shape and position of objects in that image, or degrade its overall resolution. If the goal is to transmit or receive a signal, then the energy of that signal may be diverted from the receiver or target point.⁵

1.2.2. Aero-Optics

The term "aero-optics" refers to a subset of this general problem in which the air is moving at relative velocities sufficient for compressibility effects to produce significant density fluctuations in the turbulent flow.⁷ Also of consideration in aerooptics are the presence of physical objects in the flow and the shapes of those objects. Regions of separated flow produce vortices and additional turbulence as sources of optical aberration, and even attached flow may induce forms of distortion referred to as lensing effects.^{7,8}

The classification of aero-optics as a separate class of optical problems is used to differentiate from problems of propagation through the free atmosphere. By its very nature, propagation of light through aero-optic flow fields takes place over relatively short distances measured in meters and fractions of meters, while atmospheric propagation takes place over extended distances measured in hundreds or thousands of meters. Additionally the mechanisms producing optical aberrations may be different in compressible flow near and object than what is seen in the largely incompressible free atmosphere. As will be shown in later chapters, this fact alone means that different approximations of the governing equations.

3

1.2.3. Adaptive Optics

As mentioned earlier, the goal of this study was to develop guidelines for integrating aero-optic effects. In particular, those guidelines for the design and implementation of adaptive optics (AO) systems.⁵ AO systems detect and correct optical aberrations using a variety of methods and techniques.

Significant work has already been done in characterizing atmospheric propagation and providing guidelines in the design and implementation of AO systems, but as with the case of characterizing the atmosphere's optical turbulence, the guidelines for correcting its resulting aberrations are not suited for use with aero-optic flows.

1.3. Chapter Overviews

Chapter 2 will provide background information on the fundamentals of optical propagation over short and long distances. This will include the effects of optical distortion, some differentiation in types of distortion, and how passing through a fluid medium can produce those distortions.

Chapter 3 will provide further background information concerning methods of measuring optical distortions of different types, and methods of correcting those distortions.

Chapter 4 will examine the body of work already in place concerning atmospheric propagation so that it may be compared to aero-optic conditions. This will include a description of the nature of atmospheric turbulence, the optical effects of such turbulence, the fundamental assumptions underlying these characterizations, and how those characterizations were developed.

4

Chapter 5 will review the results of an experiment performed at the Arnold Engineering Development Center (AEDC) at Arnold Air Force Base in Tennessee. This experiment measured optical effects of propagation through a compressible shear layer, and demonstrated the existence of optical distortions with a severity that could not be explained by the prevailing theories at the time. Further analysis of these results revealed aspects of these disturbances and prompted the studies described in the rest of this dissertation.

Chapter 6 will examine the physical mechanisms governing the development and behavior of shear layers, the circumstances under which such flows may form, and why they are of particular interest in the field of aero-optics. It will also describe some theory and modeling of compressible shear layers that were developed to explain the results from the AEDC experiment.

Chapter 7 will present the equipment, techniques, and results of a set of experiments performed in the Hessert Aerospace Laboratory at the University of Notre Dame. These experiments confirmed aspects of the theory and models described in chapter 6, and provided further characterization of the optical effects of compressible shear layers.

Chapter 8 will explore some of the ramifications of these results for the design of AO systems. In particular, it will examine the effects of a finite receiving aperture or transmission beam diameter upon the types of aberrations observed, and the effectiveness and limitations of basic forms of correction in such cases.

Chapter 9 will apply the principles and guidelines found in chapter 8 to specific cases and will compare these results to those found by other researchers.

Finally, chapter 10 will summarize the principles and engineering guidelines

found in this study, with suggestions for implementation and future studies.

- ⁶ Grechko, G.M., Gurvich, A.S, Kan, V., Sokolovskii, S.V., and Savchenko, S.A., "Scintillations and random refraction during occultations by terrestrial atmosphere," J. Opt. Soc. Am. A 2, 2120– 2126 (1985).
- ⁷ Gilbert, K.G., "Overview of aero-optics," in Aero-Optical Phenomena, Gilbert, K. and Otten, L., eds. (American Institute of Aeronautics and Astronautics, 1982), Vol. 80, pp. 1–9.
- ⁸ Pond, J.E., Sutton, G.W., "Adaptive Optics Improvements for Flight-Induced Aero-Optical Aberrations of a Forward-Facing Optical Dome," AIAA 2006-3073, 37th AIAA Plasmadynamics and Lasers Conference, San-Francisco.

¹ Wilson, J.R., "NASA, DOD take the next step in laser communications," Military and Aerospace Electronics, September 2003.

² Missile Defense Agency Fact Sheet, Airborne Laser, 08-FS-0004

³ Sutton, G.W., "Effects of turbulent fluctuations in an optically active fluid medium," AIAA J 1969;7(9):1737–43.

⁴ Tatarski, V.I., "Wave propagation in turbulent medium," New York: Dover, 1961

⁵ Tyson, R., "Principles of adaptive optics," San Diego: Academic Press, 1991

CHAPTER 2:

FUNDAMENTALS OF OPTICAL PROPAGATION AND ABERRATIONS

2.1. Overview

The purpose of this chapter is to provide an introduction of concepts important in optical problems. In particular, this will focus on those aspects relating to propagation of light, how aberrations may affect this propagation over long distances, and where these aberrations may come from.

2.2. Propagation of Light

Since the 19th century, it has been known that the propagation of light (i.e., electro-magnetic energy bounded on the high energy side by ultraviolet radiation and by infrared radiation on the low energy side) is governed by Maxwell's equations. As such, light propagates as waves and the propagation of those waves is subject to local variations in the medium in which it travels.¹ A surface of constant phase as light propagates through the medium is defined as its wavefront.² Huygens showed that the propagation of a wavefront can be determined by approximating the wavefront as a continuum of in-phase point sources smeared over the surface defined by the wavefront at an instant in time and tracing the interfering spherical waves emanating from these point sources to the next instant in time. Effects, consequences, and ramifications of this

interference will be discussed later in this chapter; however, Huygens further showed that such constructions prove that light propagates normal to its wavefront and that these directions can be replaced by "rays" of light. This fact is generally referred to as Huygens' Principle and is the basis of geometric optics.³

In a number of studies^{4,5,6} researchers have shown that when determining wavefront effects in the near field (i.e., over the relatively short propagation distance associated with aero-optic flows) the use of geometric optics or ray tracing introduces negligible error. More recent studies^{7,8} have revisited and confirmed this fact. Furthermore, even the constructs of geometric optics can be relaxed for aero-optics in the near field. The directional change or deviation of an optical ray is so light over distances associated with aero-optic flows that rays may be assumed to propagate in straight lines, aligned with the direction in which the overall beam is propagating. Thus, a ray represents an optical path corresponding to a beam of light with zero thickness. A collimated laser beam may be thought of as a collection of rays traveling, at least initially, in parallel.

In a vacuum or a substance with uniform optical properties, rays travel in straight lines at constant speeds. However, variations in the properties of a medium can produce regional variations in the index of refraction (n) which in turn governs the speed of light in that region. The value of n for a material is defined as the speed of light in that material, divided by the speed of light in a vacuum (c). As a result, the light passing through those regions with higher values of n will move slower than the light passing through other regions. Another way to express this is that the pathways with higher

8

average values of *n* are effectively longer, as far as the light is concerned. This effective length is called Optical Path Length (*OPL*).

As previously discussed, a wavefront is defined as a locus of points with constant phase,³ it can be thought of as a surface defining a "sheet" of light, comprised of the rays that left a light source at some common point in time. Each ray defines an optical path, containing every position that a narrow pulse of light would occupy during its time of transit. The wavefront is made of points corresponding to the leading edges of one of these rays, and the wavefront as a function of time captures all of the rays at a given moment.

To draw an analogy, if a set of athletes were to begin running from a given point or starting line when the starting gun sounded, then a path tracing the course of a single runner from start to finish would correspond to a ray. A line or curve drawn to pass through the positions occupied by each of the runners at a particular moment during the race would correspond to a wavefront.

2.2.1. Near-Field Propagation

As previously noted, inhomogeneities in a medium may produce regional variations in n and in the speed of light at that point in that medium. Details on how this occurs will be covered in section 2.3, but for now it is sufficient to note that these regional variations do occur.

When this happens, portions of a wavefront passing through regions with values of n that deviate from the average for the region will travel faster or slower than the

average speed of the wavefront, and so may lead or lag the rest of that wavefront. A sketch of this principle in action is shown in Fig. 2.1.



Figure 2.1: A sketch of rays and wavefronts in a non-uniform medium.

In more precise terms, the speed of light in a medium is determined by v = c/n. Therefore, in a period of time, Δt , a ray of light will travel a distance, Δz , such that

$$\frac{\Delta z}{\Delta t} = v = \frac{c}{n}.$$
(2.1)

From this, a final value of z for some time t can be found. However, doing this properly requires keeping track of where the ray was at each moment or time step in t, so that the local value of n as a function of position could be identified.

A more common approach is to determine the *OPL* from the initial position of the wavefront to the average position of this wavefront at the later time. The optical length of each path segment is then the physical length of that segment multiplied by the index of refraction:

$$\Delta OPL = \Delta z \cdot n \,. \tag{2.2}$$

From this,

$$OPL(x, y, z) = \int_{0}^{z} n(x, y, z') dz'.$$
 (2.3)

A wavefront can then be found as the locations in *z* as a function of *x* and *y* for which OPL(x,y,z) is a constant. The term Optical Path Difference (*OPD*) indicates the variation in *OPL* from the mean value.

$$OPD(x, y) = OPL(x, y) - OPL(x, y).$$
(2.4)

The index of refraction can also be written as an average index summed with the local variations from that average in the form $n = n + \Delta n(x,y,z)$. Using this form of expression, *OPD* corresponds to an integration of Δn , just as *OPL* corresponds to an integration of n along a path.

Consider a wavefront that begins as a planar wavefront at a point z = 0, and then propagates some distance L. The average position of this wavefront will be at z = L, but any distortions induced upon the wavefront will manifest as variations in z-position as a function of x and y, so that the wavefront will be defined by a surface $z = L + \Delta L(x,y)$. A wavefront is a locus of constant phase; therefore the *OPL* traveled by the light comprising the wavefront will also be the same value for each point on the wavefront. If we consider the *OPL* from z = 0 to z = L, the average *OPL* over this position in z should equal the constant *OPL* defining the points of the wavefront with an average position of z = L. Therefore, at some position (x,y) in the plane of z = L

$$\overline{OPL}_{z=L} = \int_{0}^{L+\Delta L(x,y)} n(x, y, z) dz$$
(2.5)

and

$$OPD_{z=L}(x, y) = OPL_{z=L}(x, y) - \overline{OPL}_{z=L} = \int_{0}^{L} n(x, y, z) dz - \int_{0}^{L+\Delta L(x, y)} n(x, y, z) dz$$
(2.6)

so

$$OPD_{z=L}(x, y) = -\int_{L}^{L+\Delta L(x, y)} n(x, y, z) dz .$$
 (2.7)

Variations over a wavefront, which are represented by ΔL in the equations above, are generally measured in microns or fractions of microns. Changes in *n* over such distances are almost always negligible, so it is reasonable to consider the value of *n* to be constant over the distance from z = L to $z = L + \Delta L$ and consider only the variations in *x* and *y*.

$$OPD_{z=L}(x, y) = -\int_{L}^{L+\Delta L(x, y)} n_{z=L}(x, y) dz.$$
 (2.8)

When dealing with air, n is generally close to a value of one, and can be expressed as

$$n(x, y) = 1 + \tilde{n}(x, y)$$
. (2.9)

From this,

$$OPD_{z=L}(x, y) = -\int_{L}^{L+\Delta L(x, y)} 1 + \tilde{n}_{z=L}(x, y) dz = -\Delta L(x, y) - \Delta L \tilde{n}_{z=L}(x, y).$$
(2.10)

The exact value of *n* varies with density and wavelength in ways that will be addressed in section 2.3, but for air under standard atmospheric conditions and wavelengths in the range of visible light, $n \approx 1.0003$. Thus, the approximation

$$OPD_{z=L}(x, y) \cong -\Delta L(x, y) \tag{2.11}$$

can be relied upon as accurate to four decimal places under most conditions encountered by systems operating in open atmosphere.

To express this in words rather than equations, light traveling along paths with an *OPL* less than the average tends to race ahead of the rest of the wavefront while light traversing longer optical paths will lag behind, and the amount of lead or lag in position

corresponds to the amount by which the paths are longer or shorter. As an increase in position along the propagation vector is linked to a decrease in path length, the signs are reversed. Because of this, *OPD* serves as the conjugate of the wavefront, which is often what is desired for producing a correction of the wavefront. Despite being conjugates of each other, the terms wavefront and *OPD* are often used interchangeably, especially in applications where the magnitude of the distortions in the wavefront is of primary interest.

2.2.2. Propagation over Intermediate Distances

In Fig. 2.1, the rays are depicted as continuing in straight and parallel lines, despite the curvature present in the wavefront. This is a simplification of things, as Huygens' Principle indicates the rays should be deflected in such a manner as to be perpendicular to the wavefront. In the experimental studies described in chapters 5 and 7 of this dissertation, the actual induced wavefront variations were normally measured in fractions of microns, and the actual curvature and angles of deflection induced were likewise very small. As a result, over the few centimeters of propagation through the flow, the deviation of the traveling rays from the paths they would have followed without those sources of aberration was so small as to be inconsequential. In such cases, the simplification of assuming a straight path through the flow is reasonable, and only the differences in the direction of propagation need to be considered.

Over longer distances, this assumption can break down. Any non-zero deflection angle will eventually produce noticeable shifts in lateral position with a great enough propagation length. Thus, a different model of propagation is needed over those longer distances. Understanding this model requires a better understanding of Huygens' principle.

As alluded to earlier, in constructing his principle, Huygens assumed that each point on a wavefront defined at time *t* could be thought of as a point-source emitting a spherical wavefront that propagates with some velocity *v*, as shown in Fig. 2.2. After a short time Δt , the wavefront will have propagated a distance of $v\Delta t$, corresponding to the radius that these spherical wavelets attain in that time. The shape of the wavefront at *t* + Δt corresponds to the envelope formed by all of the spherical wavelets.³ At each point, this envelope is tangential to one of these wavelets. As mentioned earlier, line drawn from this point on the wavefront at *t* + Δt to the origin point of the wavelet on the wavefront at *t* will be the shortest distance between the two surfaces at that point, and will be perpendicular to the surface defining the wavefront at *t*.



Figure 2.2: The construction of Huygens' principle.

It should be noted that this construction as presented so far is intended to describe a continuous wavefront of uniform intensity. Variations in intensity over a wavefront can be accounted for with point sources of differing strength. As shown in Fig. 2.2, it is also
possible and even quite common for a wavefront to be generated over a limited area, or to be reduced to a limited area by passing through an aperture. The region inside the boundaries of such wavefronts, or within the projection of these boundaries, is known as the region of geometric brightness. In applying Huygens' construction, a common simplification is to ignore portions of the spherical wavefronts that would extend outside of this region. As only a small number of "wavelets" extend outside of this region, this is a generally acceptable approximation for propagation over short distances. However, over longer distances, a significant portion of light may "leak out" into the region of geometric shadow, and yet another model of propagation is needed.

A more rigorous model for plotting and calculating the propagation of light is the use of a complex transmission function, derived again from Maxwell's equations, of the form

$$U(x, y) = A(x, y)e^{i\varphi(x, y)}.$$
 (2.12)

In this expression, the function U(x,y) is defined over a plane in x and y, located at $z = z_0$. As in the construction by Huygens, the light passing through the plane is treated as if each point on this plane were a point source emitting a spherical wave. The A(x,y)portion of Eq. 2.12 is an amplitude component, based on the component of the light at (x,y) that is normal to the z_0 plane. This usually corresponds to the peak electrical field strength of the electro-magnetic waves that make up light, but other measures of amplitude can be used. While a wavefront is defined as a locus of points corresponding to a single phase, the points on this plane may vary in phase, hence the phase component $e^{i\varphi(x,y)}$. If two planes are defined, at $z = z_1$ and $z = z_2$ respectively, and $U(x_1,y_1)$ is known for all points at z_1 , then U at a point (x_2,y_2) at z_2 can be found by the summation of all the spherical waves from z_1 as they reach the point at z_2 . As shown in Fig. 2.3, the distance from a point on z_1 to a point (x_2,y_2) will vary with x_1 and y_1 . This difference in path lengths will produce differences in phase, in addition to whatever variations in phase may already exist in U on z_1 . Thus, some elements of this sum may add their amplitudes together, and some may cancel each other out, depending on their relative phase. This is known as constructive and destructive interference.



Figure 2.3: Propagation by complex transmission functions.

The distance (z') from (x_1,y_1) to (x_2,y_2) is

$$z' = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} .$$
(2.13)

For light with a wavelength of λ , the phase of the wave at (x_2, y_2) will be

$$\varphi = \varphi(x_1, y_1) + 2\pi \frac{z'}{\lambda}. \qquad (2.14)$$

The amplitude of a spherical wave falls off with 1/z' so the optical wave coming from (x_1,y_1) can be expressed at (x_2,y_2) as

$$\frac{A(x_1, y_1)}{z'} e^{i\varphi(x_1, y_1) + i2\pi \frac{z'}{\lambda}} = \frac{U(x_1, y_1)}{z'} e^{i2\pi \frac{z'}{\lambda}}.$$
(2.15)

The wave function $U(x_2, y_2)$ at $z = z_2$ is then the sum of all the arriving waves from the point sources at z_1 . In addition to phase considerations, this summation must also take into account that the function U and its amplitude component, A, are based on the component of the waves normal to the plane over which U and A are defined. Including this consideration, and taking the limit as the plane at z_1 is divided into smaller sources, this sum becomes the integral expression

$$U(x_{2}, y_{2}) = \frac{1}{2\pi} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} U(x_{1}, y_{1}) \cos(\theta_{z'}) \left(\frac{1}{z'} - i\frac{2\pi}{\lambda}\right) \frac{e^{i2\pi\frac{z}{\lambda}}}{z'} dx_{1} dy_{1}, \qquad (2.16)$$

7'

in which $\theta_{z'}$ is the angle between the path z' and a vector normal to the z_2 plane. This approach is more complete than tracing individual rays or wavefront propagation by Huygens' principle. However, solving Eq. 2.16 can be computationally intensive. Solving it at multiple points on z_2 for a complete rendering of the complex transmission function $U(x_2,y_2)$ will be even more so. Thus, this approach is rarely used unless conditions allow for approximations that simplify Eq. 2.16. It should be kept in mind that the derivation of Eq. 2.16 assumes that *n* is a constant over the region between the planes at z_1 and z_2 . Accounting for variations in *n* would require replacing the physical distance *z*' with an *OPL* as defined in Eq. 2.3.

It is possible to perform propagation calculations of this form based on an initial wavefront rather than a complex transmission function, as illustrated in Fig. 2.4. In this

case, Eq. 2.13 will be modified as z_1 is replaced with the *z* position of the wavefront. The phase on a wavefront is constant, but the local vector of propagation is not. The result at z_2 will be a complex transmission function rather than a wavefront, unless one takes the time to adjust z_2 for each (x_2 , y_2) as to find a surface of constant phase from the result. Doing this further complicates matters, as the local vector normal to this surface changes direction.



Figure 2.4: Propagation by a complex transmission function, as found from a near-field wavefront.

Most often, an assumption is made that the variations in z across the wavefront are relatively small, on the order of a wavelength or less. A second assumption is that the angle between local propagation vectors on the wavefront and a vector normal to a nearby z_1 plane is likewise small. If these assumptions hold true, then the z location of a wavefront of constant phase φ_0 can be converted to the phase component of a complex transmission function at z_1 by

$$\varphi(x_1, y_1) = \varphi_0 + \frac{2\pi}{\lambda} (z(x_1, y_1) - z_1).$$
(2.17)

The amplitude component of the transmission function must be found or provided separately for this conversion.

The propagation models shown in Figs. 2.3 and 2.4 trace the development of disturbances within a wavefront, but not the introduction of those disturbances. Figure 2.5 shows a sketch of a more rigorous approach, combining these models with the one shown in Fig. 2.1. If variations in the local index of refraction are taken into account, then the phase contribution from (x_1,y_1) to (x_2,y_2) may be adjusted by the *OPL* rather than the physical path length. This may be achieved by multiplying the propagation length (z') in Eq. 2.13 by the average value for *n* along the path from (x_1,y_1) to (x_2,y_2) . This would account for both the development of existing disturbances and the introduction of disturbances by propagation through the medium.



Figure 2.5: Propagation by a complex transmission function, through a medium with a variable index of refraction.

However, keeping track of the index field, n(x,y,z), and finding the average over the optical path for each (x_1,y_1) to (x_2,y_2) pair may be computationally intensive, and this degree of rigor is not required for all applications. It should be noted that this assumes the propagation lengths involved are short enough that deflection angles induced upon these optical paths do not have significant effect upon those paths.

As a final note, optical propagation of this sort is reversible. Tracing individual rays, as shown in Fig. 2.1, is reversible in that if a ray is traced through a field of variable n to some point at the exit pupil with an exit vector, a ray introduced at that point with a vector in the opposite direction will follow the same path to the starting point of the first ray. Likewise, the complex transmission model of Eq. 2.16 is reversible.

2.2.3. Far-Field Propagation

The term far field has already been introduced without formally defining it. Here, "far field" is used to indicate that the distance traveled (*z*') is much greater than both the wavelength of the light ($z' \gg \lambda$) and relevant spatial length scale perpendicular to this path of travel. ($z' \gg x, z' \gg y$) Far-field conditions can also be achieved by passing the light through a lens. The image resolved at the focal length of the lens is equivalent to what is seen at a far-field distance of infinity.³ Meeting the spatial scale criteria usually indicates that the light in question is a directed beam with a finite cross-sectional width.

As noted in the previous section, using Huygens' principle to calculate the propagation of a wavefront of finite area neglects effects seen at the boundaries of that area. Over short distances of propagation, these effects can be safely disregarded, but they may become significant over distances that could be considered far-field cases. Thus, the complex transmission function is often the best approach for dealing with longer propagation distances. Fortunately, the aforementioned far-field conditions serve to simplify Eq. 2.16. If $z' >> \lambda$, then that equation can be approximated by

$$U(x_{2}, y_{2}) = \frac{i}{\lambda} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} U(x_{1}, y_{1}) \cos(\theta_{z'}) \frac{e^{i2\pi \frac{z}{\lambda}}}{r'} dx_{1} dy_{1}.$$
(2.18)

~ '

If z' >> x and z' >> y for any values of x and y of interest in the plane of origin or the plane at the far-field distance, then $\theta_{z'}$ will be very small and $\cos(\theta_{z'}) \cong 1$, further reducing the integral to

$$U(x_2, y_2) = \frac{i}{\lambda} \int_{-\infty-\infty}^{\infty} U(x_1, y_1) \frac{e^{i2\pi \frac{z}{\lambda}}}{z'} dx_1 dy_1.$$
 (2.19)

If we define *Z* as the distance from the z_1 plane to the z_2 plane, or $Z = z_2 - z_1$, then Eq. 2.13 can be rewritten as

$$z' = Z_{\sqrt{1 + \left(\frac{x_2 - x_1}{Z}\right)^2 + \left(\frac{y_2 - y_1}{Z}\right)^2}} .$$
(2.20)

To second order, $\sqrt{1+\varepsilon} \cong 1 + \frac{1}{2}\varepsilon - \frac{1}{8}\varepsilon^2$, provided $\varepsilon \ll 1$, which has already been established in the far-field conditions of spatial scale. Thus,

$$z' \cong Z + \frac{1}{2} \frac{(x_2 - x_1)^2}{Z} + \frac{1}{2} \frac{(y_2 - y_1)^2}{Z} - \frac{1}{8} \frac{((x_2 - x_1)^2 + (y_2 - y_1)^2)^2}{Z^3}.$$
 (2.21)

This can be further simplified if more stringent definitions of "far field" are applied. The Fraunhofer limit is one in which

$$Z >> \frac{(x_1^2 + y_1^2)}{\lambda}$$
(2.22)

for all values of x_1 and y_1 of consequence in the plane of origin. This is sufficient to set $z' \cong Z$ in the portions of Eq. 2.19 where z' is not an exponent. The term $e^{i2\pi z'\lambda}$ requires a

bit more finesse. Dividing Eq. 2.21 by λ and applying the conditions of 2.22 allows for several terms to be discarded, yielding

$$\frac{z'}{\lambda} \cong \frac{Z}{\lambda} + \frac{1}{2} \frac{x_2^2 + y_2^2 - 2x_1 x_2 - 2y_1 y_2}{Z\lambda}.$$
(2.23)

Inserting these approximations into the integral and moving terms that do not contain x_1 or y_1 outside of the integral leads to the following form:

$$U(x_{2}, y_{2}) = \frac{i}{Z\lambda} e^{i2\pi\frac{Z}{\lambda}} e^{i\pi\frac{x_{2}^{2}+y_{2}^{2}}{Z\lambda}} \int_{-\infty-\infty}^{\infty} U(x_{1}, y_{1}) e^{-i2\pi\frac{x_{1}x_{2}+y_{1}y_{2}}{Z\lambda}} dx_{1} dy_{1}.$$
(2.24)

Equation 2.24 is much more manageable than Eq. 2.16 and can be used to compute farfield propagation. However, it is useful to perform one last alteration, based on an exchange of variables as follows: $f_x = x_2/\lambda Z$, $f_y = y_2/\lambda Z$. By applying these, Eq. 2.24 becomes

$$U(f_{x},f_{y}) = \frac{i}{Z\lambda} e^{i2\pi \frac{Z}{\lambda}} e^{i\pi Z\lambda (f_{x}^{2}+f_{y}^{2})} \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} U(x_{1},y_{1}) e^{-i2\pi (f_{x}x_{1}+f_{y}y_{1})} dx_{1} dy_{1}.$$
(2.25)

This is significant, because the integral expression in the right-hand side of Eq. 2.25 is the form of the two-dimensional Fourier transform of $U(x_1,y_1)$.^{3,9} The widespread availability of tables of Fourier transforms for various functions as well as Fast-Fourier-Transform subroutines in many software packages can be used to greatly speed up these calculations. The use of these and related methods forms the basis for Fourier Optics.¹⁰

It should be noted that the propagation models presented in this section so far are based on near-field models that only account for the effects of distortions already present in a wavefront. They do not deal with the mechanisms that induce distortions. Introducing these mechanisms into these models was addressed in the text associated with Fig. 2.5 in section 2.2.2. However, over longer propagation lengths, it is unusual to have the detailed knowledge of the index field, n(x,y,z), necessary to implement the measures described in that section.

Even if that knowledge were at hand, making use of it would most likely have prohibitive computational costs when dealing with propagation lengths on a large scale. Additionally, the model of the complex transmission function of Eqs. 2.12 and 2.16, on which these far-field models are based, assumes that all point-to-point propagation paths are straight lines. Long propagation distances make it more likely that the deflection of rays may have a significant role

One of the most common approaches to modeling propagation over long distances in a medium involves the periodic use of phase screens. A phase screen is an expression of the variations in phase ($\Delta \varphi$) induced upon a wavefront by index variations in the medium or other effects. The phase component of a complex transmission function is then modified as

$$U'(x, y) = U(x, y)e^{i\Delta\phi(x, y)} = A(x, y)e^{i\phi(x, y)}e^{i\Delta\phi(x, y)} = A(x, y)e^{i[\phi(x, y) + \Delta\phi(x, y)]}.$$
 (2.26)

As noted, details of n(x,y,z) over a long propagation path are unlikely to be known, but some measure or characterization of the phase variations likely to be induced may be found without knowing those details. The characteristics of phase distortions induced by the free atmosphere have been the subject of study for some time.¹ Details of propagation through free atmosphere are presented in chapter 4, but for now it is sufficient to state that the probabilistic distribution of characteristics for atmospheric phase distortions are known and random phase screens with these characteristics can be generated for the purpose of propagation modeling. For other forms of flow, a different model or screens based on experimental data would be needed. The extended propagation model of this form breaks the long optical path into a series of shorter optical paths. Over each segment, Eq. 2.24 or Eq. 2.25 is applied to trace the development of the wavefront over that distance. At the end of each segment, a phase screen with characteristics appropriate to the conditions expected over that segment is applied and the process is repeated for the next segment. The number of segments required depends on the fidelity desired, the frequency of changes in conditions encountered along the path, and the distance over which acquired deflections of individual rays might become too great to ignore. As always, if wavefronts are the preferred form of expressing this form for analysis or graphical display, Eq. 2.17 can be used to convert from wavefronts to complex transmission functions and back.

It should be noted that the goal of this type of propagation model is not to simulate what actually happens as the light propagates, but to produce an end result with the same characteristics as those produced under a given set of physical conditions.

2.2.4. Irradiance Patterns

In many applications, the primary interest in the light reaching a point is the timeaveraged intensity of the light falling on an object. In this case, "time averaged" refers to the time scale of the electromagnetic waves as they travel, not the time scale of changing disturbances to the wavefronts due to changes in the flow they pass through. The proper optical terminology to refer to the flux density of incoming energy is "irradiance", which is measured in units of energy flow per unit time per unit area.³

Irradiance (*I*) can be found quite easily from a complex transmission function, being proportional to $|U(x,y)|^2$. Most often, the amplitude term A(x,y) within a complex transmission function U(x,y) is defined as the peak electric field strength at that point over a full cycle from 0 to 2π in phase. If this is so, then the proportionality relationship may be expressed as³

$$I(x, y)\frac{watt}{m^{2}} = \frac{1}{240\pi} \frac{watt}{volt^{2}} \left| U(x, y)\frac{volt}{m} \right|^{2}.$$
 (2.27)

Using the expression for U from Eq. 2.25, this becomes

$$I(f_{x}, f_{y})\frac{watt}{m^{2}} = \frac{1}{240\pi} \frac{watt}{volt^{2}} \left(\frac{1}{Z\lambda}\right)^{2} \left(\hat{U}(x_{1}, y_{1})\right)^{2} \frac{volt^{2}}{m^{2}},$$
(2.28)

where \hat{U} indicates the Fourier transform of U.

There are three specific forms of irradiance pattern that occur frequently enough to be mentioned here. Those resulting from rectangular apertures with uniform amplitude, those resulting from circular apertures with uniform amplitude, and Gaussian beams.

In using complex transmission functions, an aperture is defined by setting the amplitude outside of the aperture's open region to zero. Thus a wave of uniform amplitude and phase passing through a rectangular aperture of dimensions a and b is expressed as follows:

$$U(x, y) = \begin{cases} A_0 e^{i\varphi_0} & \text{if } |x| < \frac{a}{2} \text{ and } |y| < \frac{b}{2}, \\ 0 & \text{otherwise} \end{cases}$$
(2.29)

with A_0 and φ_0 set to constant values. With the above definition, there exist propagation distances that fulfill conditions such as Z >> x mentioned in the previous section for far-field propagation. In this particular case for a rectangular aperture, a separation of variables between x and y (or f_x and f_y) can be performed and Eq. 2.25 becomes:

$$U(f_x, f_y) = \frac{i}{Z\lambda} e^{i2\pi \frac{Z}{\lambda}} e^{i\pi Z\lambda (f_x^2 + f_y^2)} A_0 e^{i\varphi_0} \int_{-a/2}^{a/2} e^{-i2\pi f_x x_1} dx_1 \int_{-b/2}^{b/2} e^{-i2\pi f_y y_1} dy_1 .$$
(2.30)

The solution to the integrals in Eq. 2.30 is of the form sin(x)/x, commonly defined as the sinc function:

$$U(f_{x},f_{y}) = \frac{i}{Z\lambda} e^{i2\pi\frac{Z}{\lambda}} e^{i\pi Z\lambda(f_{x}^{2}+f_{y}^{2})} A_{0} e^{i\varphi_{0}} \frac{a}{2} \frac{\sin(\pi f_{x} \frac{a}{2})}{\pi f_{x} \frac{a}{2}} \frac{b}{2} \frac{\sin(\pi f_{y} \frac{b}{2})}{\pi f_{y} \frac{b}{2}}$$
(2.31)

and

$$I(f_{x}, f_{y}) \propto \left(\frac{1}{Z\lambda}\right)^{2} \frac{A_{0}^{2} a^{2} b^{2}}{4} \left(\frac{\sin(\pi f_{x} \frac{a}{2})}{\pi f_{x} \frac{a}{2}}\right)^{2} \left(\frac{\sin(\pi f_{y} \frac{b}{2})}{\pi f_{y} \frac{b}{2}}\right)^{2}.$$
 (2.32)

The form of $sinc(\pi x)$ and the square of this function are shown in Fig. 2.6. The primary characteristic of these curves are the large central lobe and the smaller, periodic lobes that decrease in magnitude with increasing distance from the origin. Eq. 2.32 indicates that the far-field irradiance pattern produced by a rectangular aperture will have the form of a sinc-squared function along one axis, multiplied by another sinc-squared function in the other axis, often referred to as a double-sinc pattern.



Figure 2.6: Sinc and squared sinc functions.

Figure 2.7 shows plots of a rectangular aperture as defined in Eq. 2.29 and the double-sinc irradiance pattern produced by an aperture of these dimensions according to Eq. 2.32. An interesting trait found in far-field patterns of this sort is that the central bright spot is wider in the direction of the aperture's narrower dimension and vice versa. More importantly, a portion of the energy projected through the aperture is projected into lobes off to the sides, rather than the central spot in the far field. While the aperture producing this irradiance pattern is finite, the far-field pattern is not, with faint sidelobes extending outward into infinity, even if the magnitude of those lobes drops below any significant threshold. Therefore, the spot size of this pattern is defined by the first set of points at which the irradiance drops to zero in the double-sinc pattern, at $f_x \cdot a/2 = 1$ and $f_y \cdot b/2 = 1$ for an aperture of dimensions $a \ge b$.



Figure 2.7: (a) A rectangular aperture (b) The associated far-field pattern of irradiance.

A circular aperture is defined in much the same way as the rectangular aperture, with U as a constant within some radius of $(x_1, y_1) = (0, 0)$ and U = 0 beyond that radius. It becomes more convenient to work in polar coordinates for this case. It is also convenient to define a scaled far-field variable $f_r = r_2/\lambda Z$, much like f_x and f_y were defined for use in Cartesian coordinates. In polar coordinates, with a circular aperture of radius r_0 , a constant transmission function $U = A_0$ across the aperture, and the aforementioned change of variables Eq. 2.24 becomes:

$$U(f_{r},\theta_{2}) = \frac{i}{Z\lambda} e^{i2\pi\frac{Z}{\lambda}} e^{i\pi Z\lambda f_{r}^{2}} A_{0} e^{i\varphi_{0}} \int_{0}^{r_{0}} r_{1} dr_{1} \int_{0}^{2\pi} e^{-i2\pi f_{r}r_{1}\cos\theta_{1}} d\theta_{1} .$$
(2.33)

As neither U at the aperture nor the aperture itself vary in θ , the solution to Eq. 2.33 does not vary in θ and can be considered solely a function of f_r :

$$U(f_r) = \frac{2\pi r_0^2 i}{Z\lambda} e^{i2\pi \frac{Z}{\lambda}} e^{i\pi Z\lambda f_r^2} A_0 e^{i\varphi_0} \frac{J_1(2\pi r_0 f_r)}{2\pi r_0 f_r}$$
(2.34)

and

$$I(f_r) \propto \frac{4\pi^2 A_0^2 r_0^4}{Z^2 \lambda^2} \left(\frac{J_1(2\pi r_0 f_r)}{2\pi r_0 f_r} \right)^2.$$
(2.35)

where J_1 is a first-order Bessel Function of the first kind. Fig. 2.8 shows the form of $J_1(r)/r$ and of the square of this function.



Figure 2.8: Bessel function-based curves.

As with the sinc function, there is a central lobe of large amplitude, and smaller lobes of decreasing size with increasing distance. In polar coordinates, the central lobe will be circular, and the smaller lobes will form rings around the central lobe. This irradiance pattern is known as the Airy disk, and it is shown in Fig. 2.9 (b). The spot size for this far-field pattern is defined as the radius or diameter of the first dark ring where the irradiance goes to zero. This ring contains 84% of the total energy flux in this pattern.¹¹



Figure 2.9: (a) A circular aperture (b) The associated far-field pattern of irradiance.

The previous two cases for rectangular and circular apertures assume that the light across the aperture is of uniform amplitude. This is known as a "top hat" beam. However, many systems that output directed beams produce beams that vary in irradiance over the cross-section of the beam. The most common form of this is the Gaussian beam, with a profile that corresponds to, or can be approximated by, a Gaussian curve of the form

$$I(r) = I_0 e^{-\left(\frac{x^2 + y^2}{\omega^2}\right)} = I_0 e^{-\left(\frac{r}{\omega}\right)^2}.$$
 (2.36)

Fig. 2.10 below shows the intensity of a Gaussian beam in comparison to a top hat beam.



Figure 2.10: Near-field irradiance for a top hat (a) and a Gaussian (b) beam, both of radius 0.5.

The Gaussian profile lacks the distinct edges of the top hat profile, and so defining the size of the beam is open to interpretation. The most common definition is for a beam "diameter" of 2ω . A circle of radius ω around the central axis of a Gaussian beam contains 85% of the beam power,¹² which makes it a near equivalent to the 84% in the central lobe of an Airy Disk.

As in the other cases, Eq. 2.24 or 2.33 can be used to find the far-field irradiance pattern. However, unlike the previous two examples, there is no finite aperture, and so the area of integration extends out to infinity. This would seem to invalidate many of the assumptions used to arrive at Eq. 2.24, as even an infinite propagation distance cannot be significantly larger than the dimensions across an infinite near-field expanse. However, the Gaussian curve falls off so that portions of the curve more that 2ω from the centerline of the beam have very little contribution to the integral, and propagation distances such that $Z >> 4\omega^2/\lambda$ can be considered to fulfill the Fraunhofer limit presented in Eq. 2.22.

Another change from the previous examples is that the amplitude function $U(x_1,x_2)$ is not a constant, but is instead proportional to the square root of the Gaussian

curve in Eq. 2.36, since irradiance is proportional to the square of the magnitude of this function. That is:

$$U(r_1) \propto \sqrt{I_0} e^{\frac{-r_1^2}{2\omega^2}}$$
 (2.37)

which is also a Gaussian curve. The Fourier transform of a Gaussian is also a Gaussian. The transmission function in the far field is then proportional to this new Gaussian

$$U(f_r) \propto \omega \sqrt{I_0 2\pi} e^{-2\pi^2 f_r^2 \omega^2}$$
(2.38)

and the irradiance is proportional to the square of the magnitude of this transmission function,

$$I(f_r) \propto \omega^2 I_0 2\pi e^{-\pi^2 f_r^2 \omega^2},$$
 (2.39)

which is yet another Gaussian.

An important aspect of Gaussian beams is that they can be focused only to some minimum spot size, and no further. This spot size is then inversely proportional to the divergence of the beam. As noted, the value ω in Eqs. 2.36 through 2.39 is considered to be the radius of the beam. If the wavefront of the beam is truly planar at a given point, then that radius is also the smallest possible radius for that beam, known as the "waist radius" or ω_0 . From that point onward the beam will expand with distance, as indicated by the term $f_r = r_2/\lambda Z$ in Eqs. 2.38 and 2.39. Reducing the waist radius of such a beam requires an increase in the expansion angle. In fact, this phenomenon is seen in beams passing through apertures as well, as shown in Fig. 2.7 where the narrower dimension of the rectangular aperture produces a wider dispersion in the far field. The scaled far-field position variables of f_r , f_x , and f_y in the preceding derivations are based on this expansion of the beam, and the fact that for relatively small angles of expansion, $\theta \cong \sin(\theta) \cong r/Z$. They are also based on the fact that the expansion angle is proportional to the wavelength of the light, so that

$$\theta = \frac{\lambda}{\pi \omega_0}, \qquad (2.40)$$

measured relative to the center-line of the beam.

As a beam expands, the wavefronts go from being planar to spherical, with a radius of curvature centered on the point at which the beam was at its narrowest radius. This phenomenon reflects the fact that a narrow beam or small aperture comes to resemble a point source with increasing distance.

In all of the cases above, the beam is assumed to be initially of uniform phase, and the only effect on the far-field patterns is the interference based on the near-field amplitude pattern, whether that pattern is defined by a finite aperture or a Gaussian curve. This is known as the diffraction limited case, and is viewed as the best possible result in transmission or imaging of light.

2.3. Fluid-Optic Interaction

In many materials, including the collection of gasses known as air, *n* varies primarily with density (ρ). This is often a linear relationship, at least over limited ranges of density variations, and is expressed in the Gladstone-Dale¹³ relationship

$$n = 1 + K_{GD} \cdot \rho , \qquad (2.41)$$

where K_{GD} is a constant for constant wavelength. Expressions for the refractive index of air are continuously under refinement, but one of the currently accepted expressions for dry air under standard conditions (288.15 K, 101,325 Pa) and light of wavelength λ is¹⁴

$$n = 1 + \left(\frac{5792105\mu m^{-2}}{238.0185\mu m^{-2} - \frac{1}{\lambda^2}} + \frac{167917\mu m^{-2}}{57.362\mu m^{-2} - \frac{1}{\lambda^2}}\right) 10^{-8}.$$
 (2.42)

In the case of the previously mentioned heat-shimmer, the rising hot air is subject to the ideal gas law,

$$P = R\rho T . (2.43)$$

In free atmosphere, pressure (*P*) will be very nearly the same for both the hotter and cooler air, but of course the temperature (*T*) is higher for the hotter air. Since the ideal gas constant (*R*) is a constant for a given medium and the medium is air (R = 287.056 m³Pa/(kg K)) throughout, the density of the warmer air must be lower than that of the cooler air. Thus, the warm air rises through the less dense air around it, and the changes in density "bend" the light passing through this flow. According to Eq. 2.43, the density of air under the previously mentioned standard conditions (ρ_0) is 1.2250 kg/m³.

Combining Eqs. 2.41, 2.42, and 2.43 the index of refraction of air can be written as

,

$$n = 1 + \frac{10^{-8}}{287.056 \cdot 1.2250} \left(\frac{5792105}{238.0185 - \frac{1}{\lambda^2}} + \frac{167917}{57.362 - \frac{1}{\lambda^2}} \right) \frac{P}{T}.$$
 (2.44)

for values of λ in µm, *P* in Pascals, and *T* in Kelvin. Refinements to this model can be made for the presence of water vapor and for variations in the amount of CO₂,^{14,15} however, such measures are beyond the scope of this work. In or near the visible range of light, Eq. 2.44 can be approximated by

$$n = 1 + 7.77 \cdot 10^{-7} \frac{P}{T}.$$
 (2.45)

From this expression, variations in the index of refraction can be written as¹⁶

$$dn = 7.77 \cdot 10^{-7} \frac{P}{T} \left(\frac{dP}{P} - \frac{dT}{T} \right).$$
(2.46)

Thus, variations in either pressure or temperature can produce variations in the index of refraction.

2.3.1. Near-Field Effects

As with propagation, the approach taken in calculating and describing optical effects may differ between the near field and the far field. In the near field, these effects are often expressed in terms of wavefronts, deflection angles, and complex transmission functions. As noted in the previous sections on propagation, the best case scenario is based on having no phase variation in the near field, which requires setting these values to zero, or to some constant value.

The fundamentals of near-field propagation through a medium with variations in n were covered in the first part of section 2.2.1, particularly the text associated with Fig. 2.1 and Eqs. 2.1 through 2.3. As was addressed in those passages, light travels faster in regions with lower values of n, and as a result portions of a wavefront may lead or lag other portions of the wavefront. These changes in n are likely to correspond to regions of different density as indicated by Eq. 2.41. The variations in the wavefront can also be expressed in terms of *OPD*.

As the wavefront may have a different value of displacement or *OPD* at each point, it is often more convenient to describe the overall magnitude of these deviations from the mean through the root-mean-squared (rms) value of the individual local variations. While OPD_{rms} is a good indicator of the degree of optical distortion produced by a given flow or set of conditions, it is variations in phase that determine the effects in

the far field. Equation 2.14 can be used to convert from wavefront displacement (Δz) to phase, and since *OPD* is the conjugate of this displacement:

$$\varphi(x, y) = \varphi_0 + \frac{2\pi}{\lambda} \left(\Delta z(x, y) - \overline{z} \right) = \varphi_0 - \frac{2\pi}{\lambda} OPD. \qquad (2.47)$$

Likewise, the root-mean-square of phase variations can be found by scaling OPD_{rms} by $2\pi/\lambda$. This is often called the phase variance and is written as σ_{φ} or σ_{θ} depending on what symbol one is using to indicate phase.

For some applications, deflection angles matter as well as variations in wavefront location. Deflection angles refer to a single ray traced through the flow. Earlier discussions indicated that angular deflections were small enough that actual displacement of the angled beam compared to a straight beam parallel to the overall direction of propagation was negligible. That is true, but only over relatively short propagation distances, as the actual side-to-side deflection of a ray or beam is such that $\Delta x = \tan(\theta_x)L$ for propagation over some distance *L*. As the ray encounters variations in the medium, it deviates from its original direction of travel, albeit ever so slightly. After experiencing many incremental deviations, it finally emerges from the region of aberrating effects with some net deflection, labeled as $-\theta_x$ in Fig. 2.11. These deflection angles can be measured, even when the propagation lengths are short enough that said deflection has little direct effect on the form of the wavefront.



Figure 2.11: An optical ray as part of a wavefront passing through a distorting flow field.

The refractive bending equation¹⁷ indicates that the bending of a ray in a medium with variations in the index should be such that the radius of curvature of the ray's path (R_C) should be

$$R_{C} = \frac{1}{\left|\nabla \ln(n)\right|} = \frac{n}{\left|\nabla n\right|},\tag{2.48}$$

The incremental change in deflection angle $(\Delta \theta)$ associated with an incremental progression along the path of the ray (Δs) is then described by

$$\tan(\Delta\theta) = \frac{\Delta s}{R_c} \cong \Delta\theta , \qquad (2.49)$$

with the approximation applying for small changes in the deflection angle. The gradient in Eq. 2.48 reflects the change in the index perpendicular to the path. Taking these incremental steps to their limit,

$$\theta_x = \int_{s_1}^{s_2} \frac{1}{n} \frac{\partial n}{\partial s_\perp} ds \tag{2.50}$$

to first approximation, where s_{\perp} is the direction perpendicular to *s* at each point along *s* within the plane of the page in Fig. 2.11.

As was addressed in section 2.2.1, it may be acceptable to treat this curved path as effectively a straight line under some circumstances. Those circumstances would be if the magnitude of variations of the path from side to side in the flow were insignificant compared to the size of features in the flow causing these variations. In the arrangement shown in Fig. 2.11, variations in x should be equal to the distance traveled in y, multiplied by the tangent of the deflection angle. If the deflection angle is small, then the tangent of that angle can be approximated by the angle itself.

$$\Delta x = \Delta y \tan(\theta_x) \cong \Delta y \theta_x. \tag{2.51}$$

Thus, if this product, $\Delta y \theta_x$, is small relative to any length scale of significance for variations of *n* in the flow, then the path can be reasonably approximated as a straight line. If this is true and the condition $\Delta y \gg \Delta x$ is also true, then the approximation made in Fig. 2.1 may be used in calculating *OPD* and similar properties. With these approximations, *s* can be assumed to be effectively in the direction of the *y* axis and Eq. 2.50 becomes

$$\theta_x = \int_{y_1}^{y_2} \frac{1}{n} \frac{\partial n}{\partial x} dy.$$
 (2.52)

A wavefront corresponds to the position of all rays generated from the same source at the same time, at some point in time later on. By Huygens' Principle a wavefront always travels perpendicular to itself. Thus each point on the wavefront corresponds to a ray that is traveling perpendicular to the surface defined by the wavefront at that point. From this principle of perpendicularity, the deflection angle of this ray will also be the angle between a line tangent to the wavefront at that point and a line defined by the slope the wavefront had before aberrations were produced.

In the circumstances shown in Fig. 2.11, this means θ_x is not only the angle between a ray and the y-axis, but also the angle between a line tangent to the wavefront at that point and the x-axis. This tangent line is defined by the slope of the wavefront at that point, which means $\tan(\theta_x)$ at a point equals the derivative of the wavefront in x at that point. The approximation of $\tan(\theta_x) \cong \theta_x$ for small values of θ_x has already been mentioned, which leads to the following relationships:

$$\theta_x \cong \tan(\theta_x) = \frac{\partial}{\partial x} y_{WF}(x).$$
 (2.53)

where $y_{WF}(x)$ is the locus of points defining the wavefront and

$$y_{WF}(x) \cong \int \theta_x dx \,. \tag{2.54}$$

The measurement and reconstruction techniques used in the experimental work described in chapters 5 and 7 is based upon the relationship of Eq. 2.54, and will be described in more detail in chapter 3.

2.3.2. Far-field Effects

Section 2.2.3 described how the description of a near-field wavefront may be converted into a far-field wavefront or irradiance pattern. Optical correction is often performed in the near field, but the motivation for dealing with these quantities is to counteract effects in the far field. The forms of aberration seen in the far field tend to fall into various categories, often determined by the size of the beam compared to the size of the structures in the flow causing the aberration. If the material that a wavefront propagates through has a higher average index of refraction than what was assumed or expected, then that wavefront will lag behind the expected position and the phase of the light arriving in the far field will likewise lag in phase. A lower average value of *n* will produce wavefronts and phases that lead the expected results. This sort of mean lag or lead is called piston. Piston has no effect on the irradiance pattern in the far field, and so it is often ignored in applications where the far-field irradiance is the only matter of concern. On the other hand, some communications or sensing applications are dependent on relative phase.

Tip and tilt (T/T) is a net deflection in the beam. Just as single rays can be deflected, so can larger beams acquire some net deflection. In the near field, this is seen as a linear change in phase or a slope on a wavefront or *OPD*. In the far field, this has little effect on the overall shape of the irradiance pattern, but causes a shift in the location of that pattern, as seen in Fig. 2.12. As shown in the figure, three wavelengths worth of tilt across an aperture can move the irradiance pattern so as to miss the target point entirely. Time-varying net deflection of a beam as shown here is commonly referred to as drift if it occurs slowly or jitter if it happens at higher frequencies. As shown in Fig. 2.12, the simulated aberration producing tilt is defined by a linear function over the aperture, and lacks any higher-order or higher spatial frequency components. Thus, T/T aberrations are primarily associated with variations in the material that have a length scale much larger than the beam diameter.



Figure 2.12: Phase varying linearly over an aperture (a) and a displaced far-field pattern (b).

If the wavefront across the aperture is a convex curved surface, rather than a flat surface as shown in Fig. 2.12, then the far-field pattern will be blurred or spread, as shown in Fig. 2.13. A concave wavefront can focus the beam, producing a far-field pattern at some nearer distance, but such wavefronts become convex and defocused at distances past this focal length. As with T/T, defocus is associated with variations in n on a scale larger than the beam diameter. However, there is a limit to how large these variations can be relative to the beam and still produce noticeable curvature.



Figure 2.13: A spherical phase aberration over an aperture (a) and an unfocused far-field pattern (b).

If aberrations are produced on a length scale close to or smaller than that of the beam size, referred to as higher-order aberrations, then it may cause the far-field pattern to break up into more than one pattern. Since the far-field irradiance pattern is proportional to a Fourier transform of the near-field wavefront, aberrations with higher-order spatial frequencies divert energy into side lobes in the far field. Unlike the net deflection of the entire beam caused by T/T, this splits energy off from the main beam rather than diverting the entire beam. Aberrations with a higher spatial frequency will cause a wider divergence. In extreme cases, most of the energy may be diverted into the side lobes and the main beam may effectively disappear.

In imaging applications this can produce double or "ghost" images. In applications of directing a beam at a target or receiver, this both spreads the energy and makes aiming somewhat problematic. As an example, Fig. 2.14 shows the results of an aberration defined by two full cycles of a sine wave across the aperture. The resulting far-field pattern is centered on the origin, but most of the energy in the pattern passes to one side or the other of this target point.



Figure 2.14: Phase variations over an aperture (a) and a divided far-field pattern (b).

Aberrations produced by fluid-optic interaction are almost never of a single type or spatial frequency. Fig. 2.15 shows a simulated wavefront aberration with random finescale aberrations, which also includes some measure of T/T and defocusing curvature. The resulting distortions in the far field are more characteristic of what might actually be observed in light passing through a severely aberrating turbulent flow.



Figure 2.15: Phase variations over an aperture (a) and a blurred and fragmented far-field pattern (b).

As with near-field effects, it is often convenient to have a way of expressing the severity of the far-field aberrations as a single numerical value. The most commonly used indicator of this sort is the Strehl ratio. In the unaberrated patterns shown in Figs. 2.7 and 2.9, the highest level of irradiance (I_0) is found at the center of the pattern. This point, also referred to as the "on-axis" point, represents the point towards which the light is aimed or directed. As shown in Figs. 2.12, 2.14, and 2.15, aberrations divert energy from this point by shifting or splitting the pattern. The Strehl ratio is the ratio of the flux density at the on-axis point to the flux density that would exist in the ideal, diffraction-limited case. (I / I_0) If phase variations are not too large then the Strehl ratio can be estimated by¹⁸

$$SR \cong e^{-\sigma_{\varphi}^2} = e^{-\left(\frac{2\pi}{\lambda}OPD_{rms}\right)^2}.$$
(2.55)

While this approximation was originally developed for relatively small perturbations in the wavefront ($\sigma_{\varphi} < 2$ radians) it has been found to be accurate within 10% for values of the Strehl ratio down to St = 0.3.

Since aberrations tend divert energy from this center point, the Strehl ratio is almost always less than or equal to one. It is possible for some strange form of phase aberration to divert energy to the center point, rather than away from it, but this is almost never seen in naturally occurring aberrations. Another quantity, beam quality (BQ) has been defined in a number of ways, but the most common definition is based on the Strehl ratio¹¹ as

$$BQ = \sqrt{\frac{I_{O}}{SR}} = \sqrt{\frac{I_{O}}{I}}$$
 (2.56)

Just as the Strehl ratio is normally less than or equal to one, with SR = 1 being the ideal, BQ is normally greater than one, with BQ = 1 being the ideal.

The Strehl ratio is based on the irradiance at a single point, but many applications have a somewhat greater tolerance for spreading or divergence of the irradiance pattern. In communications applications, the receiving sensor will usually have some finite area, and the total energy falling on that region is what matters, not the distribution of energy within this region. In other applications, energy delivered to one point spreads to the surrounding region. Thus, for some applications it makes more sense to define a "bucket" as the area within some radius of the on-axis point, and then to judge results by the ratio of total energy flux into this bucket between the aberrated and unaberrated conditions, rather than just at the on-axis point. However, what constitutes a reasonable size for this bucket varies with the application in question, and often depends more on personal judgment than on any widely recognized standard.

- ¹ Tatarski, V.I., *Wave Propagation in a Turbulent Medium*, McGraw-Hill, New York, 1961.
- ² Teare, S.W., Restaino, S.R., Introduction to Image Stabilization, SPIE Press, 2006
- ³ M Klein. Optics. Wiley, New York, 1970.
- ⁴ Mason, J.O., Aguirre, R.C., and Catrakis, H.J., "Computational aero-optics and electromagnetics: compressible vortices and laser beam propagation," Int. Assoc. Mech. Eng. Trans. 42, pp 1973-1981 (2005).
- ⁵ Truman, C.R. and Lee, M.J., "Effects of organized turbulence structures on the phase distortion in a coherent optical beam propagating through a turbulent shear flow," Physics of Fluids A, 2, pp 851–857, 1990
- ⁶ Tsai, Y.P. and Christiansen. W.H., "Two-dimensional numerical simulation of shear layer optics," AIAA J., 28:2092, 1990.
- ⁷ Zubair, F.R. and Catrakis, H.J., "Aero-Optical Resolution Robustness in Turbulent Separated Shear layers at Large Reynolds Numbers," American Institute of Aeronautics and Astronautics Journal, Vol. 45, No. 11, pp. 2721-2728, 2007.
- ⁸ Mani, A., Wang, M., Moin, P., "Resolution requirements for aero-optical simulations," J. Comp. Phys., Vol 227, No. 21.
- ⁹ R.K. de Jonckheere, E.J. Jumper, et. al., "An Engineering Approach to Optical Phase Distortion Through Turbulent Thermal Boundary Layers", AIAA 20th Fluid Dynamics, Plasma Dynamics, and Lasers Conference, Buffalo, New York, June, 1989,
- ¹⁰ Goodman, J.W., *Introduction to Fourier Optics*, 2004 Roberts and Company.
- ¹¹ Tyson, R.K., *Principles of Adaptive Optics*, Academic Press, San Diego, 1991.
- ¹² Kasap, S.O., Optoelectronics and Photonics: Principles and Practices, Prentice Hall, Upper Saddle River, New Jersy, 2001.
- ¹³ de Jonckheere, R.K., Jumper, E.J., et. al., "An Engineering Approach to Optical Phase Distortion Through Turbulent Thermal Boundary Layers", AIAA 20th Fluid Dynamics, Plasma Dynamics, and Lasers Conference, Buffalo, New York, June, 1989
- ¹⁴ Ciddor, P.E., "Refractive Index of Air: New Equations for the Visible and Near Infrared", Applied Optics, Vol. 35, No. 9, 1996, pp. 1566-1573
- ¹⁵ Ciddor, P.E., "Refractive Index of Air: 3 The Roles of CO₂, H₂O and Refractivity Virals", *Applied Optics*, Vol. 41, No. 12, 2002, pp. 2292-2298
- ¹⁶ Jumper, G.Y., Beland, R.R., "Progress in the Understanding and Modeling of Atmospheric Optical Turbulence", 31st AIAA Plasmadynamics and Lasers Conference, Denver, Colorado, June, 2000
- ¹⁷ Born, M. and Wolf, E., *Principles of Optics*, 6th ed., 1980, Pergamon Press, Oxford, UK.
- ¹⁸ Mahajan, V.N., "Shrehl ratio for aberration in terms of their aberration variance," J. Opt. Soc. Am. 73, 860-861, 1983Find reference for Strehl ratio estimate

CHAPTER 3:

METHODS OF OPTICAL MEASUREMENT AND CORRECTION

3.1. Overview

The concept of an adaptive-optic system, or AO system, was briefly introduced in chapter 1 as a system that can be used to measure and then compensate for optical aberrations. This may be done at the receiving end of an optical system, to correct aberrations that have occurred as the beam propagated, or this may also be done in a transmitting system, if the aberrating effects of the flow in the transmission path are known. In this latter case, a conjugate wavefront is imposed on the beam prior to projecting it through the flow, so that the beam emerges from the aberrating medium restored to or near diffraction-limited performance.¹

One realization of an of the adaptive-optic beam train is the Notre Dame adaptiveoptic system designed by Xinetics in cooperation with the Boeing, SVS. A schematic of the system is shown in Fig. 3.1. What can be noted in Fig. 3.1(a) is that the correction is applied in two stages. The first element the incoming beam encounters is a tip-tilt (T/T) mirror; an enlargement of the components of the T/T mirror is shown in Fig. 3.1(b). The T/T mirror is re-imaged on the deformable mirror (DM) which, in effect, adds an ability of the DM to tip and tilt. Note also that the T/T mirror is controlled by a separate, in this case analog, stand-alone processor. It should be noted that the system also incorporates near-stationary correction components for some experiments to remove well-defined, mean aberrations such as spherical aberrations.



Figure 3.1: Notre Dame Adaptive-Optic system. (a) Full system. (b) Detail of the Tip-Tilt compensation portion of the system.

Many AO systems use a similar two-stage approach to correction, and the Notre-Dame system will be referenced as an example throughout this chapter. This chapter will address the two types of aberrations addressed in these two stages, how such aberrations may be measured, and how corrections may be applied. It will also address some of the concerns and limitations of measurement and corrective systems.

3.2. Tip/Tilt Aberrations

As noted in chapter 2, a ray passing through an aberrating medium may be deflected at some angle α from its original course. Likewise, an entire beam as a whole can be deflected, causing the far-field pattern to wander. An example of this was shown in Fig. 2.12 on page 41.

3.2.1. Definitions of Tilt

As noted, the purpose of tilt correction is to remove net tilt from a wavefront and center the far-field pattern. However, "tilt" and "center" are not as clearly defined as one might think. Huygens' Principle defines tilt and a vector of propagation locally at each point in a wavefront. For tilt of a wavefront as a whole, there are two prevalent definitions.

Gradient-tilt, or G-tilt, is an average of all the local gradients on a wavefront, based on the assumption that an average of all the local propagation vectors on a wavefront should yield an average vector of propagation for the entire wavefront.² Zernike-tilt, or Z-tilt, comes from the set of Zernike polynomials that are often used to construct approximations of wavefronts and other surfaces. The Z_1^1 and Z_1^{-1} Zernike polynomials are of the form $z = r \cos(\theta)$ and $z = r \sin(\theta)$ respectively, or z = x and z = yin Cartesian coordinates. The polynomial Z_0 is of the form z = 1. Therefore, using the first three Zernike polynomials to approximate a wavefront takes the form of a leastsquares fit to a flat line (A + Bx) in two-dimensional constructions or a plane (A + Bx +Cy) in three-dimensional approximations.³ As the desired end result of AO correction is a flat line or planar surface that is perpendicular to the desired vector of propagation, one might expect that applying correction to remove the slope of this approximation to be an effective first step in achieving this.

However, as can be seen in Fig. 3.2, these two forms of tilt do not necessarily agree. Using a full cycle of a sine wave as a simulated wavefront, the average of the local propagation vectors is a vector perpendicular to the x-axis, which indicates no net angle of deflection and no net tilt. On the other hand, a least-squares fit with a function

of the form $y = A + B \cdot x$ shows a significant slope across the extent of this wavefront, indicating tilt.



Figure 3.2: G-tilt defined by averaging local tilt, Vs Z-tilt defined by a linear fit.

Either definition of tilt may be considered to be correct, depending on the intended application.

G-tilt is the average of the deflection angles or slopes across a wavefront. If the wavefront is of uniform irradiance and is defined by a surface z(x) across a finite aperture extending from x_1 to x_2 , then the x-component of G-tilt is defined by

$$G - Tilt_{x} = \int_{x_{1}}^{x_{2}} \frac{dz}{dx} dx \bigg/ \int_{x_{1}}^{x_{2}} dx = \frac{z(x_{2}) - z(x_{1})}{x_{2} - x_{1}}.$$
(3.1)

Thus, for a wavefront of uniform irradiance over a one-dimensional aperture, G-tilt can be found by drawing a line between the endpoints of the wavefront at the edges of an aperture. In the example of Fig. 3.2, these endpoints are on a sine wave, separated by one full cycle, so $z(x_2) = z(x_1)$ and the G-tilt = 0. This property could be used as a means of detecting tilt across an aperture by measuring phase or OPD around the edges of an aperture.
Z-tilt is found by selecting constants A and B to minimize the expression

$$\int_{x_1}^{x_2} (z(x) - A - B \cdot x)^2 dx .$$
 (3.2)

The value of *B* then corresponds to the overall tilt of the wavefront by this definition. If the point of reference is shifted so that the aperture of width *d* is centered around x = 0,

$$A = \frac{\int_{-d/2}^{d/2} z(x)dx}{d}$$
(3.3)

and

$$B = \frac{12 \int_{-d/2}^{d/2} x \cdot z(x) dx}{d^3}.$$
 (3.4)

In the example of Fig. 3.2, if the sine wave used to represent a wavefront has an amplitude of *a* and a period of Λ , then

$$B = \frac{12 \int_{-d/2}^{d/2} x \cdot a \sin(2\pi/\Lambda x) dx}{d^3} = \frac{6a \left(2 \sin \frac{d\pi}{\Lambda} - d \cos \frac{d\pi}{\Lambda}\right)}{d^3 \pi}.$$
 (3.5)

These expressions ignore variations in *y*, or cases in which irradiance is not constant in the near field. In cases with varying irradiance, such as that in a Gaussian beam, the wavefront displacement, z(x), must be weighted by the irradiance in the wavefront at that point.⁴

3.2.2. Measuring Tilt

While tilt can be inferred from the near-field wavefront as indicated in the preceding paragraphs, the most common means of measuring tilt is to find the center of

the irradiance pattern in the far field. The point of T/T correction is to align the beam onto a target, thus the angle between the intended axis of propagation and the vector pointing to the center of the far-field pattern is a very practical definition of tilt.

This angle is most often found with some form of position sensing device, which receives an optical intensity pattern, and returns a value based on the position of the center of this pattern. If the irradiance pattern of a beam is found to be centered at (x_c , y_c) while the desired on-axis location would be (0, 0), and it is known that the deflection occurred at some distance *L* from the sensor, then the angle of deflection (θ_x) can be found to be

$$\theta_x = \arctan \begin{pmatrix} x_c \\ L \end{pmatrix} \cong \frac{x_c}{L}, \quad \theta_y = \arctan \begin{pmatrix} y_c \\ L \end{pmatrix} \cong \frac{y_c}{L}.$$
 (3.6)

The approximations $\theta_x \cong x_c / L$ and $\theta_y \cong y_c / L$ hold if x_c and $y_c \ll L$.

However, just as there is more than one definition of tilt for a wavefront, there is more than one way of defining the center of an irradiance pattern. This is illustrated by the two most common types of position sensors, quad cells and centroiding devices.

The four-element quadrant detector, or quad cell, is a set of four irradiance sensors, often some form of photodiode.¹ Arranged into a four-quadrant pattern, each cell produces a signal proportional to the total energy flux due to light falling on that cell. A representation of this arrangement is shown in Fig. 3.3. If we let *A*, *B*, *C*, and *D* represent the signals from these cells, then for small variations in from a centered position, the x-position of the irradiance pattern will be proportional to the quantity ((A + B) - (C + D))/(A + B + C + D). That is, the pattern is held to be centered at x = 0 when the total energy falling on the left half of the quad cell equals the total energy falling on the right half. Scaling the difference in energy falling on the two halves by the total energy falling on the entire sensor produces a result that will be proportional to the lateral shift in the pattern, with a constant of proportionality that will remain roughly the same even if there are changes in the pattern or overall irradiance. A similar relation holds for the y-position.



Figure 3.3: Sensor regions of a quad cell.

As noted, the proportionality for shifts in the location of the irradiance pattern only applies if that location is near to the center of the sensor, compared to the size of the pattern or the size of the central spot of the pattern if it has one. If a significant majority of the light falling on the sensor falls into one cell, the sensor has no way of telling where in the cell this concentration of incoming energy may be. A quad cell works best when the size of the sensor is only slightly larger than the size of the pattern for which a location is to be found. If the pattern is larger than the sensor, then significant portions of the light in the pattern may miss the sensor and will not be accounted for. If the pattern is small enough to fit into one cell, then the dynamic range of the sensor will be limited. Despite these limitations, quad cells are commonly used because their relative simplicity lends itself to implementations that are reliable and durable when used in the field.

Another type of position sensor is the lateral effect detector, which is also known as a position sensing device. This sensor consists of a sheet of photoelectric material with electrodes along the four sides of the sheet,⁵ as shown in Fig. 3.4. When photons strike the photoelectric material, free electrons are produced that flow into the electrodes. The number of electrons produced at a point on the sensor is determined by the energy flux density of the light falling on that point. Some of the electrons produced then become current flowing into the four electrodes. These free electrons are more likely to flow into the electrode closest to their point of generation. If the light were focused to an infinitesimal spot, then the position of that spot between electrodes A and B in Fig. 3.4 would be proportional to the difference in the current flowing into A and the current flowing into B, divided by the sum of the two currents. That is, $x \propto (A - B)/(A + B)$ where A and B are the currents flowing into the respective electrodes. A similar relation holds for the vertical position between electrodes C and D.



Figure 3.4: Components and arrangement of a Lateral Effect Detector sensor.

At first glance, this appears to be equivalent to the operation of the quad cell. However, when the energy within the intensity pattern is not concentrated on a single point, then a sensor of this sort will produce a result based on a weighted-average centroid of the form,

$$x_c = \frac{\iint x \cdot I(x, y) \, dx \, dy}{\iint I(x, y) \, dx \, dy}.$$
(3.7)

with a similar relationship for position in *y*. Since each point on the sensing area of the detector effectively acts as a separate sensor, this type of sensor is not as susceptible to considerations of pattern size or larger pattern shifts as the quad cell.

These different definitions of the center of an irradiance pattern are all quite valid, just as they different definitions of tilt in section 3.2.1 are valid. However, they are different, and if one does not keep those differences in mind, then that can lead to problems in trying to deal with tilt.

As tilt is primarily of interest in aligning the far-field intensity pattern on a target, a look at the far-field pattern is instructive. Figure 3.5 shows far-field intensity patterns for the sinusoidal wavefront in Fig. 3.2, with a peak-to-valley phase variance of 0.8π radians. The solid vertical line in Fig. 3.5 (a) indicates the location of the far-field intensity pattern's centroid, as found with a weighted average of the form in Eq. 3.7. The dashed vertical line in Fig. 3.5 (a) indicates the point corresponding to a quad cell definition of center, at which the total intensity to the left of that point (ΣI_x -) equals the total intensity to the right of that point (ΣI_x +).



Figure 3.5: A far-field irradiance pattern based on a sinusoidal wavefront (a) and this pattern if the wavefront is corrected according to Z-tilt.

As can be seen in Fig. 3.5, the locations of these two definitions of the center are not equivalent. The centroid definition of centering the pattern is at x = 0, which indicates there was no tilt in the original wavefront and agrees with earlier evaluation of G-tilt for the wavefront in Fig. 3.2. On the other hand, Fig. 3.5 (b) shows the far field for a wavefront with tilt calculated according to the Z-tilt definition and removed. Performing T/T correction based on this definition of tilt shifts the far-field pattern so that the center point according to the quad cell definition of center is placed near x = 0. Interestingly, this also places the point of highest intensity in the far field closer to x = 0.

Increasingly, arrays of charge-coupled device (CCD) photosensors, of the sort that serve as the basis for most digital cameras, have been used as position sensors.¹ Each pixel in the array produces a signal or digital value proportional to the irradiance falling on that sensor. If a far-field pattern falls across several pixels of the array, then a weighted average centroid can be found through numerical integration approximating Eq. 3.7. If the size of the pattern is on the order of one or two cells across, then a set of four pixels in a square pattern can be used as a quad cell. Larger blocks of pixels can be used as the equivalent of larger quad cells, but it is rare to do so if enough pixels are involved to make centroiding a viable option.

A limitation of CCD array centroiding is that spatial and temporal resolutions become factors in this form of sensor. Properly approximating Eq. 3.7 requires pixels small enough to resolve relevant features in the far-field pattern and enough pixels to encompass the pattern, including spreading and wandering of the pattern due to aberrations. As noted, four pixels that are larger than or on the order of the pattern size can be used in the manner of a quad-cell, but as it has also been noted, quad-cells and centroiding sensors have different definitions of tilt. The use of pixels of intermediate size will produce measured values of tilt that do not properly correspond to either Z-tilt or G-tilt, but are likely to lie somewhere between the two.

Quad cells and lateral effect detectors output only four signals. Those four signals are reduced to two values for tilt in x and y through operations of addition, subtraction, and division that are simple enough to be carried out by analog circuitry. Each pixel in a CCD outputs a separate signal that must be read and recorded in order to perform the calculations. The computation to convert these values into T/T is also more involved than that for quad cells and lateral effect detectors. The time required to read in all values and perform the computation can limit the sampling rate for a CCD-based sensor.

Before leaving this discussion, it should be noted that recognition of the difference in G and Z-tilt and their ramifications to tilt correction were derived independently during work that will be described in chapter 8. As presented here, the implications of tilt measurement and correction at first glance appear as only subtleties,

but in practice they have very real unintended consequences. Through further study, it became apparent that G and Z-tilt are well-established phenomena and, in that sense, this discovery amounted to rediscovering the wheel. However, in subsequent presentations of results at several national meetings, it became apparent that while the mathematical subtleties of the difference between G and Z-tilt were well established, the implications were not fully appreciated by many who work in this field.

3.3. Wavefront Aberrations

Tip-tilt is only one of the various types of aberration that were discussed and shown in chapter 2, section 2.3.2. Tip-Tilt correction is based on an approximation of a wavefront as a flat surface, though the definitions for determining the slope of that surface may vary. Other forms of aberration are associated with curves and inflections in a wavefront, and require higher-order polynomials or sums of trigonometric functions in order to properly express or approximate the form of that wavefront.

The need for higher-order correction can be seen in Fig. 3.5 of the previous section. Applying tilt correction in the near field, based on G-tilt or Z-tilt, may shift the location of the irradiance pattern in the far field, but it does not change the shape of that pattern. In that example, Z-tilt correction does increase the intensity seen at the target or reception point of x = 0, but that intensity in that example is approximately 70% of what would be achieved in the diffraction-limited case, and that is close to the maximum intensity that may be found anywhere in this intensity pattern.

As a further example, Fig. 3.6 (a) shows a simulated, one-dimensional wavefront in the form of part of a sine wave across an aperture. This wavefront has a noticeable degree of both G-tilt and Z-tilt, along with curves and inflection points. Results of correction of these forms of tilt are also shown in this figure. The far-field irradiance pattern for the uncorrected wavefront is shown in Fig. 3.6 (b), compared to the diffraction-limited ideal for an aperture of this size.



Figure 3.6: (a) A simulated wavefront, with corrections. (b) Uncorrected and diffraction limited far-field patterns. (c) Far-field patterns with various types of tilt correction.

Figure 3.6 (c), shows the far-field patterns associated with the T/T corrected wavefronts. As was the case in Fig. 3.5, T/T correction may shift the position of the

irradiance pattern, but it can not change the shape of the far-field pattern. G-tilt correction improves the Strehl ratio from 0.56 to 0.88, while Z-tilt correction raises it to 0.96. However, neither form of correction can raise it to the ideal of one, because they can not recover energy scattered or diverted into side-lobes. Only higher-order correction to remove the residual distortions that remain after T/T can do this.

3.3.1. Higher-Order Wavefront Measurements

There does not seem to be the same conflict in defining higher-order wavefront distortions that exits for G-tilt and Z-tilt in T/T correction, though the magnitude of higher-order distortions can be expressed in a number of different ways, such as *OPD* or phase variance. However, there are multiple approaches to measuring the form of a wavefront. The choice of approach and design of these sensors is usually driven by considerations of cost, reliably, ease of use, dynamic range, and resolution.

Resolution, in this case, applies not only to spatial resolution for resolving the finer-scale variations, but also to temporal resolution in the form of sampling rates and processing time. Real-time AO correction requires that the measured and reconstructed form of a wavefront be available before changing conditions or evolving flow renders the data obsolete. Wavefront sensors tend to have lower frequency limits than T/T sensors, because there is more information to deal with and that data may require substantial processing to produce a map of a wavefront.

Among the methods for wavefront measurement are: curvature sensors, doubleslit sensors, and various forms of interferometry. Information on these techniques may be found in other sources,^{6,7,8} but the work described in this dissertation was performed primarily with Malley probes, which are in turn based on Shack-Hartman sensors. Both are based on Huygens' Principle, which was as addressed in section 2.2.2. Specifically, they are based on the idea behind this principle that a wavefront can be replaced by a set of point sources along its surface.⁹ As previously mentioned, this leads to the result that wavefronts propagate normal to themselves and that the angle of a ray that is part of a wavefront will correspond to the slope of the wavefront at that point.

The difference between the two is that a Shack-Hartman sensor can be considered a wavefront sensor in that it takes a "picture" of a wavefront at a given time by sampling that wavefront over several locations at the same moment in time. The Malley probe is a technique for reconstructing a wavefront from only two, closely-spaced measurement locations aligned in the direction in which the aberrating flow structures are moving. Based on a set of assumptions about the behavior and development of the flow producing the wavefront aberrations, it extrapolates information about the wavefront at a given time to locations other than the points of measurement, based on data from earlier or later points in time. Another way of saying it is that it measures the slope of the wavefront as it convects with the flow structures past the measurement location.

3.3.2. Shack-Hartman Sensors

Hartmann was the first to realize that Huygens' Principle could be used to measure the figure of wavefronts.¹⁰ He placed an opaque, perforated plate in front of the aberrated wavefront with a photographic plate at a known distance from the perforated plate. By exposing the photographic plate first to an unaberrated beam and then to the aberrated beam he was able to measure the off-axis displacement of beams emerging

from the perforations. Knowing the distance between the perforated and photographic plates he could determine the deflection angles and thus the wavefront slopes at each perforation. With a spatially sampled measurement of the slopes, an approximation of the wavefront can be reconstructed by integration of these slopes.

A Shack-Hartman sensor, as shown in Fig. 3.7, uses lenses rather than pinholes, focusing the light to a point, or at least to a relatively small representation of the far-field irradiance pattern. The displacement of this point or of the centroid of the pattern is an indicator of the average slope of the wavefront over the area of the lens. Many wavefront sensors of this sort use CCD arrays and centroiding routines of the area under the lenslets or a set of position sensors (of the sort addressed in section 3.2.2), rather than the photographic plate used in the original Hartman sensor.



Figure 3.7: A Shack-Hartman sensor.

The deflection angle (θ) can be found from the displacement (d) of the bright spot or centroid by Eq. 3.6 on page 52, with L in that equation corresponding to the distance from the perforated plate to the photographic plate in the case of the Hartman sensor, or the focal length of the lenses in the case of a Shack-Hartman sensor. The approximations in Eq. 3.6 only apply if α is relatively small, but that is frequently the case. As previously noted, the deflection angle of a ray or narrow beam sampled from a wavefront corresponds to the slope at that point of the wavefront, and to the negative of the slope of the *OPD* as the conjugate of the wavefront.

$$\theta(x,t) = -\frac{dOPD(x,t)}{dx}.$$
(3.8)

From this it would seem that the form of the wavefront could be reconstructed by

$$OPD(x,t) = \int -\theta(x,t)dx. \qquad (3.9)$$

However, $\theta(x,t)$ can not be measured for every point. With a spacing between measurement points of Δx , Eq. 3.9 may be approximated in the form

$$OPD(x_{n+1},t) = OPD(x_n,t) - \theta(x_n,t)\Delta x = OPD(x_0,t) - \sum_{1}^{n} \theta(x_n,t)\Delta x.$$
(3.10)

As *OPD* is, by definition, a quantity with zero mean, that property determines the initial value of $OPD(x_0,t)$ in the summation above. Rather than performing the summation as an infinite value problem which can accumulate errors, many two-dimensional Hartman-type sensors a surface is fitted to the array of slopes using some form of least-squares sense approach.

3.3.3. Malley Probes

A Malley probe is based upon the same principles as a Hartman sensor, but also takes advantage of some aspects of a flow in motion. Malley et. al. were the first to recognize that when the aberrating medium is a turbulent flow, the aberrations caused by the convecting flow structures will convect as well.¹¹ Thus, a single beam propagated through the flow could be used to measure a continuous time series of wavefront slopes as the wavefront convects by the measurement location.

The Malley principle has been used at Notre Dame to develop a series of wavefront-measurement devices.^{12,13,14} To the extent that the flow can be treated as slowly varying, the Taylor frozen flow assumption can be used to compute wavefronts up and downstream of the measurement location, which are reasonably accurate for some distance up and downstream. By propagating two small-diameter, closely-spaced, beams through the flow, both the wavefronts slope and its convection speed can be determined.

If the velocity at which the sources of aberration are convecting (U_C) is known, then at some later time, $(t + \Delta t)$ that portion of the flow will be some distance downstream, $(x + \Delta x)$ such that

$$U_c = \frac{\Delta x}{\Delta t}.$$
(3.11)

If the times and distances of convection are sufficiently small that the structures in the flow do not change significantly over that period, then it may be assumed that an optical deflection angle measured at position *x* and time *t* will be close to the deflection angle that would be measured at position $x + \Delta x$ and time $t + \Delta t$. Therefore,

$$\theta(x,t) \cong \theta(x + \Delta x, t + \Delta t) = \theta(x + U_C \Delta t, t + \Delta t)$$
(3.12)

and

$$\theta(x + \Delta x, t) = \theta(x + U_C \Delta t, t) \cong \theta(x, t - \Delta t) = \theta\left(x, t - \frac{\Delta x}{U_C}\right).$$
(3.13)

By this assumption, a time-series of deflection angle measurements can be used to reconstruct the wavefront upstream and downstream of a single measurement point. In this reconstruction around a measurement point x_0 , Eq. 3.9 becomes

$$OPD(x,t) = \int -\theta(x,t)U_C dt. \qquad (3.14)$$

and Eq. 3.10 becomes

$$OPD(x_{n+1}, t_m) = OPD(x_0, t) - \sum_{-n}^{n} \theta(x_0, t_{m-n}) U_C \Delta t$$
(3.15)

The advantage of a Malley probe is that it turns temporal resolution into spatial resolution, allowing for the resolution of fine detail through a high sampling rate rather than by adding more sensors. Malley probes have additional advantages in some cases of implementation and accessing a flow. This technique is originally based on the principles behind a Hartman sensor, in which a localized portion of continuous wavefront is sampled in the form of a narrow beam. However, only the deflection of the narrow beam actually matters in performing the measurement and reconstruction. It is possible to do away with the full wavefront from which the beams would be sampled, and simply generate a set of narrow beams that can be directed through the flow. The deflection of a larger wavefront or not. Therefore, they can be used to reconstruct the aberrations that would be imposed on a full wavefront passing through that flow, whether or not a continuous wavefront was actually involved in making the measurement.

Producing a full wavefront over an area of measurement often requires expanding and collimating a beam without inducing any aberrations into the beam in the process. It also requires that the original beam have sufficient power so that every point in the expanded beam and wavefront that might be sampled will have sufficient irradiance to perform the measurement. Producing a set of narrow, isolated beams requires less specialized equipment and less overall power. Additionally, some experimental conditions, such as the ones that will be described in chapters 5 and 7, do not have full optical access for passing a beam of the desired dimensions through the flow, but the Malley probe technique can be used to extrapolate upstream and downstream into regions that can not be measured directly.

If U_C is not known from the characteristics of the flow, then it can be determined by a comparison of the data from two closely spaced beams, one upstream of the other.¹⁵ As illustrated in Fig. 3.8, the source of aberration encountering the first beam at time *t* should encounter the second beam at time $t + \Delta t$. A cross-correlation of the deflections measured at each beam should provide an indication of the time required for a disturbance to travel from the location of the upstream beam to that of the second beam. The ratio of this travel time and the spacing between the beams indicates U_C by Eq. 3.11, which can then be used to perform the wavefront reconstruction of Eq. 3.15.



Figure 3.8: Indications of convective velocity from Malley probe measurements.

The use of multiple measurement locations along a streamline also allows for another approach in increasing the spatial resolution. Eq. 3.15 is intended for the reconstruction of a wavefront from the deflection data of a single beam, but the data from multiple beams can be combined. If there are two beams at positions x_1 and x_2 , then at a point x_3 , Eq. 3.11 indicates that the deflection angle that would be measured at that point at time *t* should correspond to the deflection angle measured at x_1 at time $t - U_C (x_3 - x_1)$. It also indicates that it should correspond to the deflection angle measured at x_2 at time $t - U_C (x_3 - x_2)$. If x_3 is located between x_1 and x_2 , then it is reasonable to employ an average of these two values. It is even more effective to employ a weighted average to reflect the position between the other two points. Thus, if the indicated deflection angles by extrapolation from x_1 and x_2 are θ_1 and θ_2 respectively. Then the value of θ_3 at x_3 can be extrapolated as

$$\alpha_3 = \frac{\alpha_1(x_2 - x_3) + \alpha_2(x_3 - x_1)}{x_2 - x_1}.$$
(3.16)

These averaged, extrapolated values can be produced for any and all positions between x_1 and x_2 , and then used in the wavefront reconstruction of Eq. 3.15. The use of multiple beams and this averaging scheme is known as the Small Aperture Beam Technique (SABT).¹⁶ Both the Malley probe and the SABT can be expanded to three or more beams.

3.4. Applying Correction

Goal of this work is to advance not only the theoretical understanding of optical aberrations, but the engineering capability to deal with those aberrations. Thus, a review of methods for counteracting these optical distortions is necessary for proper understanding of the problem.

3.4.1. Tip-Tilt Correction

As discussed in section 3.2 Tip-Tilt (T/T) is a net deflection of a beam. Therefore, in order to counteract unwanted deflection, an opposing deflection of equal magnitude and opposing direction must be imposed on the beam. The most common approach is the use of a steering mirror more often referred to as a fast steering mirror or FSM. FSM's are generally made by attaching a flat mirror to actuators of some sort that can change the angle of the mirror. The beam to be corrected is directed onto the mirror, and the mirror is angled so that the reflection is directed onto the desired axis of propagation.

There are other possible approaches, such as the Risley prism pair, which uses a set of prisms to deflect the beam. There are potential advantages and drawbacks to many of these approaches, but the FSM approach is probably more common simply because tipping and tilting a mirror to counter T/T in a beam is a more intuitively obvious solution.

The Notre Dame system shown in Fig. 3.1 uses a quad cell and FSM for correction. An enlargement of this portion of the system is shown again in Fig. 3.9. The control loop between the FSM and the quad cell is an analog system that stands entirely separate from other corrective measures. The incoming beam is focused onto the quad cell with a lens, which produces an irradiance pattern on the sensor in the form of a small far-field spot. If the pattern is not centered on the quad cell, then the signals to the FSM are adjusted accordingly. Provided changes in the T/T component of the aberrations do not exceed the bandwidth for the control loop, the beam passed on for higher-order correction will be centered according to the quad cell definition of centered. As was

68

shown in Fig. 3.5 on page 56, this should correspond reasonably closely to the wavefront being corrected according to the Z-tilt definition of tilt.



Figure 3.9: A schematic of the T/T corrective system within the ND AO system.

3.4.2. Higher-Order Correction

Most forms of higher-order correction rely on some form of phase conjugation. Something is used to impose aberrations on a wavefront that are equal but opposite to the aberrations from other sources. This can be done before or after the other aberrations occur. In either case performing phase conjugation requires some means of detecting those aberrations. Providing this correction beforehand requires foreknowledge of the type and form of aberration that will be encountered. If the conjugation takes place after propagation through the aberrating medium then the distorted beam itself can be used to measure the aberrations. If the conjugation is performed before propagation and incoming beam is required which can be produced from target glint or guidestars,^{17,18} but this is beyond the scope of this dissertation.

Focus and defocus aberrations are something of a special case, in that they can be corrected with lenses. A defocus aberration indicates that the wavefront has some net spherical curvature, causing the rays making up a beam to spread or converge. If a lens is placed in the path of a beam so that the spreading rays can be traced back to a point at the focal length of the lens, then the beam will be collimated thereafter. Alternately if the goal is to focus the wavefront at a finite distance, rather than at infinity as for a collimated beam, this too can be accomplished with one or more lenses. Focused (converging) beams can be dealt with similar manner with a concave lens, or by placing a convex lens at some point after the converging rays have come to a focus, and the focus aberration becomes a defocus aberration. The act of focusing a camera or telescope can be considered a sort of AO system in operation. Orientation of the camera counteracts tip-tilt while adjusting the distance of one or more lenses deals with focus. However, tilt and focus are the simplest forms of aberration seen in wavefronts. For more complex shapes, especially in time-varying systems, a more adaptable means of correction is required. One of the most commonly used forms of higher-order correction is a deformable mirror (DM) with a surface that can be altered into a variety of contours.

Figure. 3.10 shows three sketched illustrations of an aberrated wavefront reflecting from mirrors with different surface contours. In 3.10 (a) the mirror is flat and the phase variation in the direction of propagation remains unchanged, though the left-right orientations relative to this vector have been swapped, as is standard for reflection.



Figure 3.10: (a) flat mirror reflections, (b) shaped mirror conjugate reflection, (c) half-magnitude shaped mirror correction.

In 3.10 (b) the mirror has been deformed to match the shape as of the incoming wavefront and the reflected wavefront now leads where the incoming wavefront lagged and vice-versa. In terms of OPD(x) or other mean-removed indicators of the variations the outgoing wavefront is the conjugate of the incoming wavefront. $(-OPD(x)_{Incoming} = OPD(x)_{Outgoing})$.

This result is very useful if the outgoing light is to transverse the same path as the incoming light. As noted in chapter 2, *OPL* corresponds to an integral of n along a path. (Eq. 2.3) This integral is along the length of the path, and the direction of the path does not matter. Thus, the resulting *OPL* is the same for either direction along the same path. Sending a conjugated wavefront of the sort seen in Fig. 3.10 (b) back through the path that produced the original variations will produce a cancellation of these variations, resulting in a planar wavefront as the light returns to its point of origin.

Fig. 3.10 (c) shows a case halfway between (a) and (b). The mirror is deformed to the same shape as the incoming wavefront, but the variations in the mirror are half the

magnitude of those seen in the wavefront. This half-magnitude deformation produces a planar wavefront as a reflection. This is the correction sought after in the Notre Dame AO system shown in Fig. 3.1.

This sort of deformation and reflection can also be done in reverse, imposing aberrations on an incoming planar wave that will be twice the magnitude of the mirror deformations. This is useful in predictive correction, in which the form of aberration that will be acquired along the beam path is measured by use of return glint, a guide star, or some other means. If a conjugate of these expected deformations is placed on a wavefront before it is sent along the path, then those deformations will be canceled out as those sources of aberration are encountered.

This explanation of higher-order correction by a reflective surface that changes shape is the approach used in the ND AO system and in many other corrective systems.¹ Some other AO systems use special properties of materials in which *n* can be made to vary.¹⁹ A commonly used form of this involves electro-optic materials, in which *n* is a function of the local electrical field; this is known as the Pockels effect if Δn is proportional to the magnitude of the electrical field, or the Kerr effect for variations proportional to the square of the field strength.²⁰ Controlled variations in *n* across a wavefront can be achieved by passing the light through a window comprised of several individually controllable pixels of such material. By this means, *n* can be increased for leading portions of a wavefront, slowing them relative to other sections of a wavefront and to even out variations. Whether it is possible to decrease *n* to speed up lagging portions of the wavefront depends on whether the material is governed by the Pockels or Kerr effect.

72

The ND AO system uses a deformable mirror (DM) produced by Xinetics in the form of a reflective membrane (referred to as a phase sheet) mounted on an array of piezoelectric posts. As different levels of voltage are applied to these posts, they expand or contract, pushing the surface of the mirror outward or pulling it inward. A computer-generated representation of this DM is shown in Fig. 3.11. It should be noted that the dimensions of the posts and magnitude of the deformations in the mirror are not to scale in this representation, but it is presented here to convey the general idea of how this device works; in fact, the displacements of the actual mirror are limited to approximately $\pm 4 \,\mu\text{m}$.



Figure 3.11: A model of the Xinetics deformable mirror.

The operating principles of this system prompted the choice of Z-tilt over G-tilt for T/T correction. Removing Z-tilt, as defined in Eqs. 3.2 and 3.4, minimizes the rms magnitude of higher-order variations within the wavefront (see Figs. 3.2 and 3.6) which minimizes the degree of motion required from the piezoelectric actuators driving the mirror. Any physical AO system will have actuators with a limited degree of motion and some time required to reach a desired position or configuration. This is also true of systems based on electro-optic materials which do not move but will still have a limited range of values of *n* that can be reached and some delay in reaching them. Thus, applying Z-tilt correction before higher-order correction is generally the best choice for keeping the remaining distortions within the achievable range of the higher-order system and minimizing the time required for higher-order correction.

This is also a motivation for performing T/T correction separately. The Xinetics DM and other higher-order correctors can provide T/T correction at the same time as they perform higher-order correction; however, doing so would require a larger range of motion in the actuators, and might make greater demands on the speed of those actuators in changing position. Implementing a separate, simpler system for T/T correction also saves cost, as a greater range of motion or speed in a higher-order corrective system generally makes for a more expensive system.

3.5. Concerns and Limitations of Correction

In practice, perfect correction is not possible because that correction must be carried out by physical and electronic systems with limitations. Some of these concerns were mentioned in passing in previous sections of this chapter, but a more in-depth exploration is worthwhile. These limitations will be of particular relevance in chapters 8 and 9 which address the results and applications of the work described in this dissertation.

3.5.1. Spatial Resolution

Wavefronts are continuous sheets, containing an infinite number of definable locations within a finite area. Each and every point of a wavefront will have some value of phase and *OPD*. Systems to measure and correct wavefronts are made up of a finite number of components. Therefore, it is not possible to achieve perfect resolution of every point in measuring a wavefront, nor can correction be perfectly applied to every point. To illustrate this, a simplified example in one dimension will be used, involving a Shack-Hartman sensor and deformable mirror.

Figure 3.12 shows the simulated wavefront to be used in this example, which is produced by the summation of three sine waves of random phase with periods equal to the aperture length, one-half the aperture length, and one third of the aperture length. The actual magnitudes of the aperture and wavefront aberrations are irrelevant for this example so the units are left undefined.



Figure 3.12: A simulated wavefront produced by summation of sine waves.

Figure 3.13 shows the results of a simulated detection and reconstruction of this wavefront by a Shack-Hartman sensor as described in section 3.3.2. Specifically, the aperture was divided into sections, the slope in each section was found by the G-tilt definition of slope, and then reconstruction by Eq. 3.10 from page 63. The reconstruction with four sensors over the aperture captures some aspects of the overall shape of this wavefront, but completely misses certain prominent features of the wavefront. The reconstruction with six sensors at least suggests the existence of these features, even if it is not a particularly good fit to those features and still misses some of the minor features.



Figure 3.13: Simulated Shack-Hartman reconstruction.

These results are not surprising. It has long been established that resolving a signal or waveform via periodic sampling requires a minimum of two sampled data points per period of the waveform or features to be resolved, i.e., the Nyquist frequency. The simulated wavefront of this example contains a sine wave with three periods within the aperture shown. Therefore, a minimum of six points or regions of measurement is

necessary to meet this requirement. Using fewer data points than this minimum tends to produce aliasing, as the higher-frequency components of the waveform are either lost entirely, or misinterpreted as lower-frequency aspects.

Figure 3.14 shows the results of using an increasing number of sensors to divide the aperture into a greater number of smaller sections for this measurement and reconstruction. With nine or twelve sensors, the major features of the wavefront become more recognizable and the lesser features are at least hinted at if not fully resolved. As Figs. 3.13 and 3.14, show, meeting the minimum of the Nyquist frequency does not necessarily constitute a "good fit" depending on the requirements of a given application; however, the Nyquist frequency does serve as an indicator for an absolute minimum that will be needed.



Figure 3.14: Simulated Shack-Hartman reconstruction.

Just as a finite number of sensors limits the ability to accurately measure and reconstruct a wavefront, so does a limited number of degrees of freedom in the corrective

mechanism limit the ability to apply correction. Figure 3.15 (a) shows the simulated wavefront from Fig. 3.12 with approximations of this wavefront by least-squares fitting of various polynomials. The order of the polynomial in the fit corresponds to the number of degrees of freedom in the corrective mechanism. For a deformable mirror as described in section 3.4.2, this would be determined by the number of actuators driving the mirror. Polynomials of order 4, 6, 9, and 12 have been used to reflect the number of sensors used in the preceding example and the curves necessary to match those reconstructions.



Figure 3.15: (a) 4th, 6th, 9th, and 12th order polynomial fits to the simulated wavefront. (b) Residual wavefront after correction according to the polynomial fits of (a).

The 12th order polynomial matches the wavefront to the degree that it overlies the curve of the wavefront in the figure and can not be seen. However, the polynomials of lesser order have a visible degree of fitting error because they do not have sufficient degrees of freedom to match the form of the simulated wavefront. The results of correction based on these fits are shown in 3.15 (b), with decreasing amplitude of the residual wavefront with increasing order of the fit, and correction by a 12th order fit leaves almost no deviations at all remaining within the corrected wavefront.

It should be noted that these fits are performed to the wavefront itself, with perfect knowledge of that wavefront at every point. Figure 3.16 (a) shows the results of a 12th order polynomial fit to the 12-sensor reconstruction of Fig. 3.14. The errors inherent in the reconstruction from limited data are passed on to the polynomial fit. The deviations seen in the residual wavefront after correction based on this reconstruction, shown in Fig. 3.16 (b), are on the order of those seen for the 9th order fit that was performed with perfect knowledge of the wavefront.



Figure 3.16: (a) A 12th order polynomial fit to the 12-sensor reconstruction of the wavefront. (b) Residual wavefront variations after correction by the reconstruction in (a).

3.5.2. Temporal Resolution

Just as there are limits to measuring a wavefront or applying perfect correction at every point in space, there are limits to doing so at every moment in time. This is especially true for digital systems, which are discrete in time. Thus, rather than a continuous correction, many systems provide correction that is updated periodically. During the interval between corrections, structures in a flow that produce aberrations will move and evolve, and a correction provided for conditions at one moment in time will become less accurate as time goes by. Analog systems, despite having continuous signals, are subject to a similar problem in the form of latency, which will be addressed in section 3.5.3.

If the interval between periodic corrections is long enough relative to the rate of change in the aberrations observed, then those aberrations may change to the point that the corrective measures add to the distortion, rather than alleviating it. For example, this dissertation has made extensive use of sine waves as aberrations in simulated wavefronts. If such an aberration were to be perfectly corrected at a point in time, but that correction were to remain frozen as the aberration convected across an aperture, then eventually the aberration would reach a point of being π radians out of phase with the correction. At that point, instead of canceling out the aberration, the correction would add to it, doubling its amplitude.

Figure 3.17 (a) shows a plot of a simulated corrected wavefront over an aperture, as it develops in time. The simulated wavefront is a sine wave with a period equal to twice the length of the aperture. This aberration moves across the aperture with time, completing one cycle of motion over the period of time shown. This aberration is corrected four times over this cycle, visible as the points at which the aberration suddenly disappears and the wavefront becomes a flat line. In generating these figures, a perfect correction for the wavefront at the moment of correction was assumed, without any of the uncertainty or fitting error described in the previous section.

80



Figure 3.17: Time evolution of periodic correction, (a) near-field phase, (b) far-field irradiance.

Figure 3.17 (b) shows the time-varying far-field pattern produced by these nearfield wavefronts. As can be seen, the far-field pattern achieves the diffraction-limited ideal at the times of correction but the high center lobe of the pattern degrades and shifts from side to side over the interval between these corrections.

Figure 3.18 repeats this process, but for an aberration of the same magnitude but half the length scale. This aberration convects at the same speed as in the previous example, but completes two cycles instead of one over the same interval of time. For the same frequency of correction, the aberration and far-field distortions have time to grow noticeably worse than they did in the previous example. This is because the same interval in time corresponds to a greater degree of motion relative to the shorter length scale of the features in this simulated wavefront.



Figure 3.18: Time evolution of periodic correction, (a) near-field phase, (b) far-field irradiance.

Figure 3.19 shows the Strehl ratio of Figs 3.17 (b) and 3.18 (b) as functions of time. As is to be expected from the previous figures, the Strehl ratio is lower on average for the aberration with the shorter period and reaches lower minimums. In communications applications, these minimums may be a more accurate gauge of system performance than the average Strehl ratio, as even a temporary loss of signal is often unacceptable.



Figure 3.19: Time varying Strehl ratios for periodic correction.

3.5.3. Latency

Latency refers to a delay in applying the correction. In feedback loops, some degree of latency is almost unavoidable, as signals within the system do not reach their destinations instantaneously. Likewise, actuators can not traverse with infinite velocity and require some time to reach a specified position. In analog systems, there may be transient effects associated with some of the components, which take time to settle. Digital systems must interface with such analog components, and computing the proper response to the measured conditions requires some processing time.

The manner in which latency can limit and impair a corrective system is similar to the way in which the periodic corrections addressed in the previous section do so. A perfect correction for a wavefront at time *t* is not so perfect at time $t+\Delta t$ and becomes increasingly less perfect with larger values of Δt . The difference is that instead of starting out with a good or perfect correction at $\Delta t = 0$ that degrades over time in the interval between corrections, the delayed correction associated with latency is always late by some fixed interval of Δt relative to the conditions the correction is meant for. Thus, the degree of error associated with latency tends to be a relatively steady value, rather than the cycles shown in Figs. 3.17, 3.18, and 3.19.

If the interval Δt is of significant duration relative to the rate of change for features in the distortions to be corrected, then applying this correction may increase the overall distortion instead of lessening it. Using the sine-shaped simulated wavefronts of the previous section as examples, the worst-case scenario would be a time delay in applying the correction corresponding to the time associated with the distortions in the wavefront moving a distance equal to half their period. Under those conditions the correction would be 180° out of phase with the distortions to be corrected, which would double the magnitude of those distortions rather than removing them.

This effect of latency is well known in the field of feedback control. A common technique for design of such systems is Bode analysis, which plots the response of a system in terms of relative magnitude and phase delay of the output as functions of the frequency of the input to the system. In this analysis, systems are considered to be unstable if they amplify signals of frequencies associated with phase delays of 180° or more.

3.5.4. Combined Effects

It is very rare for any system to have only one of the limitations above. Any physical system will have some limitations in the resolution of measuring a wavefront, limits in matching the form of the wavefront in applying correction, and some non-zero time delay in applying that correction. As such the net phase variance of a wavefront remaining after correction (σ_r^2) will be comprised of errors produced in measuring the wavefront ($\sigma_{r(meas)}^2$), fitting the wavefront ($\sigma_{r(fit)}^2$), and temporal effects such as periodic correction and latency ($\sigma_{r(menp)}^2$), so that

$$\sigma_r^2 = \sigma_{r(meas)}^2 + \sigma_{r(fit)}^2 + \sigma_{r(temp)}^2.$$
(3.17)

Other sources of error may also be included in this equation, though it should be noted that this assumes the sources of error in this summation are independent of each other. The question is whether the magnitude of these errors and delays is significant relative to the needs of the application and the conditions encountered. ¹ Tyson, R.K., *Principles of Adaptive Optics*, Academic Press, San Diego, 1991

- ² Tyson, R.K. (editor), *Adaptive Optics Engineering Handbook*, Ch. 3, R.J. Sasiela and J.D. Shelton, *Guide Star System Considerations*, Marcel Dekker, New York, 2000.
- ³ Sasiela, R.J, *Electromagnetic Wave Propagation in Turbulence: Evaluation and Application of Mellin Transforms*, SPIE Press, 2007.
- ⁴ Barchers, J.D., Fried, D.L., Link, D.J., Evaluation of the performance of Hartmann sensors in strong scintillation, *Applied Optics*, Vol. 41, No. 6, 20 February 2002, pp. 1012-1021.
- ⁵ Kelly, B.O., "Lateral-Effect Photodiodes," Laser Focus, Mar. 1976, pp. 38-40
- ⁶ Roddier, F., "Curvature sensing and compensation: a new concept in adaptive optics," Appl. Opt., Vol. 27, pp 1223-1225 (1988)
- ⁷ van Workum, J., Plascyk, J.A., Skolnick, M.L., "Laser wavefront analyzer for diagnosing high-energy lasers," Adaptive optical components; Proceedings of the Seminar, Washington, D.C., March 30, 31, 1978. (A79-17170 05-35) Bellingham, Wash., Society of Photo-Optical Instrumentation Engineers, 1978, p. 58-66
- ⁸ Koliopoulos, C., "Avoiding Phase-Measuring Interferometry's Pitfalls," Photonics Spectra, C. Koliopoulos, "Avoiding Phase-Measuring Interferometry's Pitfalls," Photonics Spectra, Oct. 1988, pp 169–176.
- ⁹ M. Klein. Optics. Wiley, New York, 1970
- ¹⁰ Malacara, D., *Optical shop testing*, Wiley, New York, 1978.
- ¹¹ Malley, M., Sutton, G.W., and Kincheloe, N., "Beam-Jitter Measurements for Turbulent Aero-Optical Path Differences," Applied Optics, 31, 1992, pp. 4440-4443.
- ¹² Jumper, E. J., and Hugo, R. J., "Quantification of Aero-Optical Phase Distortion Using the Small-Aperture Beam Technique." AIAA Journal, 33(11), 1995, pp. 2151-2157.
- ¹³ Hugo, R.J., and Jumper, E.J., "Experimental Measurement of a Time -Varying Optical Path Difference by the Small-Aperture Beam Technique". Applied Optics, Vol. 35(22), 1996, pp. 4436-4447.
- ¹⁴ Fitzgerald, E.J., and Jumper, E.J., "Two-Dimensional Optical Wavefront Measurements Using as Small-Aperture Beam Technique Derivative Instrument," Optical Engineering, Vol. 39, No. 12, Dec. 2000, pp. 3285-3293.
- ¹⁵ Gordeyev, S., Jumper, E. J., Ng, T., and Cain, A., "Aero-Optical Characteristics of Compressible, Subsonic Turbulent Boundary Layer." AIAA Paper 2002-3606, Orlando, Florida, 23-26 June, 2003
- ¹⁶ Jumper, E.J., and Hugo R.J., "Quantification of aero-optical phase distortion using the small-aperture beam technique," AIAA J 1995, 33(11) pp 2151–7.
- ¹⁷ Gurski, G.F., Nomiyama, N.T., Radley, R.J., and Wilson, J., "An initial evaluation of performance of adaptive optical systems with extended dynamic targets," J. Opt. Soc. Am. 67, 345-350 (1977)
- ¹⁸ Whiteley, M.R., Welsh, B.M., and Roggemann, M.C., "Optimal modal wave-front compensation for anisoplanatism in adaptive optics," J. Opt. Soc. Am. A, 15, no.8, Aug. 1998, pp 2097-2106.
- ¹⁹ Love, G.D., 2000, Liquid crystal adaptive optics, in Tyson, R.K. (ed.) Adaptive Optics Engineering Handbook, Marcel Dekker, New York, pp 273-285.
- ²⁰ S.O. Kasap, *Optoelectronics and Photonics: Principles and Practices*, Prentice Hall, Upper Saddle River, New Jersy, 2001

CHAPTER 4:

ATMOSPHERIC PROPAGATION

4.1. Overview

In communicating with the community of people concerned with turbulenceinduced aberrations on the figure of a laser-beam's wavefront made to propagate through the turbulence, it has become clear that most fall into one of two camps; they are familiar with the physics and effects of propagation through either the atmosphere (atmospheric propagation) or with the equivalent properties of high-speed turbulent flows in the vicinity of the beam-director's exit pupil (aero-optics). While the two propagation scenarios share the fact that the aberrations are imposed by index-of-refraction variations within the turbulence, little else about the characteristics and effects of these two types of fluid-optic interaction are the same.

At the heart of the field of atmospheric propagation is a single parameter, C_n^2 , which has come to be the accepted standard in characterizing optical turbulence. As will be addressed in this chapter, C_n^2 depends on the turbulence being in the form described by Andrei Kolmogorov. If the turbulence can be characterized as Kolmogorov, then this single parameter, C_n^2 , describes not only the scale size of the aberrating index-ofrefraction fluctuations, but also the magnitude of the aberration associated with them. Coupled with the velocity of the air normal to the optical path, all of the relevant optical parameters of the aberrating turbulence can be deterministically derived, including the
requirements for adaptive-optic mitigation systems. Because of this, the focus of requests for research in aero-optics in the late 1980's asked for determination of C_n^2 , for aerooptic flows. Unfortunately, pursuit of C_n^2 as a characterizing parameter can easily become a goal in and of itself, without regard for whether the underlying assumptions that serve to define C_n^2 are applicable to the conditions under consideration. Because of this, a review of these assumptions and the fundamental definitions of C_n^2 and other commonly used parameters will be beneficial to understanding of the aero-optical problem, and how it differs from the work already established for atmospheric propagation.

4.2. Atmospheric Models

To fully describe and predict the behavior, development, and variations of a flow, the Navier-Stokes equations are required. In a three-dimensional flow, these consist of three momentum equations

$$\frac{\partial(\rho U_1)}{\partial t} + \frac{\partial(\rho U_1^2)}{\partial x_1} + \frac{\partial(\rho U_1 U_2)}{\partial x_2} + \frac{\partial(\rho U_1 U_3)}{\partial x_3} = -\frac{\partial P}{\partial x_1} + \left(\frac{\partial \tau_{11}}{\partial x} + \frac{\partial \tau_{12}}{\partial x_2} + \frac{\partial \tau_{13}}{\partial x_3}\right) - \rho g_1 \quad (4.1)$$

$$\frac{\partial(\rho U_2)}{\partial t} + \frac{\partial(\rho U_2 U_1)}{\partial x_1} + \frac{\partial(\rho U_2^2)}{\partial x_2} + \frac{\partial(\rho U_2 U_3)}{\partial x_3} = -\frac{\partial P}{\partial x_2} + \left(\frac{\partial \tau_{21}}{\partial x_1} + \frac{\partial \tau_{22}}{\partial x_2} + \frac{\partial \tau_{23}}{\partial x_3}\right) - \rho g_2 \quad (4.2)$$

$$\frac{\partial(\rho U_3)}{\partial t} + \frac{\partial(\rho U_3 U_1)}{\partial x_1} + \frac{\partial(\rho U_3 U_2)}{\partial x_2} + \frac{\partial(\rho U_3^2)}{\partial x_3} = -\frac{\partial P}{\partial x_3} + \left(\frac{\partial \tau_{31}}{\partial x_1} + \frac{\partial \tau_{32}}{\partial x_2} + \frac{\partial \tau_{33}}{\partial x_3}\right) - \rho g_3 \quad (4.3)$$

where g_i represents a body force, such as gravity, in the x_i direction and τ_{ij} represents the viscous stresses, which depend upon the nature of the fluid. Air is a Newtonian fluid, for which

$$\tau_{ij} = \mu \left[\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_1} - \frac{2}{3} \left(\frac{\partial U_1}{\partial x_1} + \frac{\partial U_2}{\partial x_2} + \frac{\partial U_3}{\partial x_3} \right) \delta_{ij} \right].$$
(4.4)

A fourth equation governs continuity of mass:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho U_1)}{\partial x_1} + \frac{\partial (\rho U_2)}{\partial x_2} + \frac{\partial (\rho U_3)}{\partial x_3} = 0.$$
(4.5)

Dealing with and keeping track of variations in temperature requires an additional equation for continuity of energy:

$$\frac{\partial \rho E_T}{\partial t} + \frac{\partial \rho U_1 E_T}{\partial x_1} + \frac{\partial \rho U_2 E_T}{\partial x_2} + \frac{\partial \rho U_3 E_T}{\partial x_3} = \frac{\partial}{\partial x_1} \left(\kappa \rho \frac{\partial T}{\partial x_1} \right) + \frac{\partial}{\partial x_2} \left(\kappa \rho \frac{\partial T}{\partial x_2} \right) + \frac{\partial}{\partial x_3} \left(\kappa \rho \frac{\partial T}{\partial x_3} \right) + \frac{\partial}{\partial x_3} \left(\kappa \rho \frac{\partial T}{\partial x_3} \right) + \frac{\partial}{\partial x_3} \left(\kappa \rho \frac{\partial T}{\partial x_3} \right) + \frac{\partial}{\partial x_3} \left(U_1 \tau_{11} + U_2 \tau_{12} + U_3 \tau_{13} \right) + \frac{\partial}{\partial x_2} \left(U_1 \tau_{21} + U_2 \tau_{22} + U_3 \tau_{23} \right) + \frac{\partial}{\partial x_3} \left(U_1 \tau_{31} + U_2 \tau_{32} + U_3 \tau_{33} \right) + \rho \left(U_1 g_1 + U_2 g_2 + U_3 g_3 \right).$$

$$(4.6)$$

where E_T is the total energy defined by internal energy plus kinetic energy per unit mass and κ is the thermal diffusivity of the fluid. For an ideal gas, the internal energy (e) is defined by the specific heat at constant volume ($e = C_v T$), while the kinetic energy per unit mass is $\frac{1}{2}(U_1^2 + U_2^2 + U_3^2)$ for all gasses and fluids, ideal or otherwise.

Pressure, temperature, density, and three components of velocity represent six unknowns in these five equations; therefore an additional equation is required for solutions, usually called the equation of state. It is common to draw this equation from some variation of the ideal gas law or calorically perfect gas relations.

For obvious reasons, these equations are usually written in a more condensed form of notation. However, they are presented in full here to show just how complicated they can be. Very little work is done utilizing the full equations as presented above. Instead, most models, theories, and specific solutions are based on some simplification of these equations, which is often achieved by ignoring one or more factors. A common approach is to treat the flow as incompressible, setting ρ to a constant. Without density variations, buoyancy ceases to be a factor, and body forces may be discarded. This is often accompanied by ignoring temperature variations as well, eliminating the energy equation and equation of state from consideration, leaving a system of four equations and four unknowns:

$$\rho\left(\frac{\partial U_i}{\partial t} + \frac{\partial (U_i U_1)}{\partial x_1} + \frac{\partial (U_i U_2)}{\partial x_2} + \frac{\partial (U_i U_2)}{\partial x_3}\right) = -\frac{\partial P}{\partial x_i} + \left(\frac{\partial \tau_{i1}}{\partial x_1} + \frac{\partial \tau_{i2}}{\partial x_2} + \frac{\partial \tau_{i3}}{\partial x_3}\right) \quad (4.7)$$

for each axis, i = 1, 2, or 3, and

$$\frac{\partial U_1}{\partial x_1} + \frac{\partial U_2}{\partial x_2} + \frac{\partial U_3}{\partial x_3} = 0.$$
(4.8)

This approximation has proven to be reasonably accurate for many engineering purposes. Solutions for the incompressible Navier-Stokes equations have been developed for many commonly encountered situations, along with various approaches to finding solutions.

As was addressed in section 2.3, variations in the index of refraction (*n*) are directly tied to variations in density. At first glance, this would appear to make the incompressible Navier-Stokes equations useless for optical problems. However, it is possible to recover some relevant features of the flow by treating them as passive scalars. A passive scalar is a quality with a scalar value that may vary from point to point in the flow, is carried by the flow, and may spread by diffusion, but does not affect the behavior of the flow. For example, when smoke or dye is injected into a flow for visualization, ideally it should be carried with the fluid without changing the dynamics of the fluid, and

the concentration of this marking material could then be considered a passive scalar. The change over time of a passive scalar *A* at a fixed point is governed by the advection-diffusion equation:

$$\frac{\partial A}{\partial t} + u_1 \frac{\partial A}{dx_1} + u_2 \frac{\partial A}{dx_2} + u_3 \frac{\partial A}{dx_3} = \kappa_A \left(\frac{\partial^2 A}{dx_1^2} + \frac{\partial^2 A}{dx_2^2} + \frac{\partial^2 A}{dx_3^2} \right).$$
(4.9)

where κ_A is the diffusivity of the property *A*. The diffusivity may be a constant or a function of some other property of the flow.

In section 2.3, the relation between variations in n and variations in P and T was found to be

$$dn = 7.77 \cdot 10^{-7} \frac{P}{T} \left(\frac{dP}{P} - \frac{dT}{T} \right)$$
(4.10)

for conditions of standard atmosphere and light in the visible or near-visible range. While free atmosphere can often be modeled adequately as being only driven by pressure variations, as in the incompressible Navier-Stokes equations, steep pressure gradients dissipate at sonic speeds. On the length scale associated with the diameter of a beam or viewing path in most applications, the dT term in Eq. 4.10 will tend to dominate the dP term. The small pressure fluctuations that are coupled to velocity variations in atmospheric turbulence have a low order impact on optical propagation compared to the effect of temperature fluctuations,¹ even if those temperature fluctuations are not of sufficient magnitude to significantly affect the flow. Thus, it is reasonable to assume that tracking and modeling *T* and variations in *T* as a passive scalar is an adequate means of tracking *n* and variations of *n* in free atmosphere flow. This has proven to be the case in many studies, models, and applications.^{2,3}

4.2.1. Turbulence as a Separate Quality

In cases of atmospheric propagation, the length of the optical path may be on the order of kilometers or more. On the other hand, the aspects of the flow that produce significant optical aberrations are often on the order of the beam or aperture diameter or smaller, and these diameters are often measured in meters or centimeters. Finding an analytical or numerical solution that resolves the necessary small scale detail within a domain on the scale of the path length is rarely feasible. Instead, it may be more practical to resolve the time-averaged qualities of the flow.

A common approach of this type is to separate each property of the flow into mean and varying portions, such that $U_i = \langle U_i \rangle + u_i$, $P = \langle P \rangle + p$, and so on. It should be noted that the averaging in this case, denoted by $\langle \rangle$, is not a time average, but an ensemble average, as if running the same experiment multiple times. Thus, $\langle U \rangle$ represents the average of the realized flows at a point in space and can change with time while *u* represents the deviation from this mean in a particular case. Inserting these modifications into the incompressible Navier-Stokes equations (Eqs. 4.7 and 4.8) and taking another ensemble average of the result leads to what are known as the Reynolds-Averaged Navier-Stokes (RANS) Equations

$$\rho \left(\frac{\partial \langle U_i \rangle}{\partial t} + \frac{\partial (\langle U_i \rangle \langle U_1 \rangle)}{\partial x_1} + \frac{\partial (\langle U_i \rangle \langle U_2 \rangle)}{\partial x_2} + \frac{\partial (\langle U_i \rangle \langle U_3 \rangle)}{\partial x_3} \right) = \frac{\partial \partial \langle U_i \rangle}{\partial x_1} + \frac{\partial \partial \langle U_i \rangle}{\partial x_2} + \frac{\partial \partial \langle U_i \rangle}{\partial x_2} + \frac{\partial \partial \langle U_i \rangle}{\partial x_3} + \frac{\partial \langle U_i \rangle}{\partial x_3} + \frac{\partial \langle U_i u_1 \rangle}{\partial x_3} + \frac{\partial \langle u_i u_2 \rangle}{\partial x_2} + \frac{\partial \langle u_i u_3 \rangle}{\partial x_3} \right)$$

$$(4.11)$$

and

$$\frac{\partial \langle U_1 \rangle}{\partial x_1} + \frac{\partial \langle U_2 \rangle}{\partial x_2} + \frac{\partial \langle U_3 \rangle}{\partial x_3} = 0.$$
(4.12)

The terms dealing with the varying components in Eq. 4.11 are known as the Reynolds stresses, and may be written as $\rho \langle u_i u_j \rangle = R_{ij}$. Despite the name, these are not actually stress terms but instead represent the average momentum flux due to turbulence, though they do have a similar effect on a flow as a stress tensor. When i = j, the Reynolds stress corresponds to twice the average kinetic energy associated with the turbulent velocity in the x_i direction per unit volume:

$$R_{ii} = \rho \langle u_i u_i \rangle = 2 \cdot \frac{1}{2} \rho \langle u_i^2 \rangle.$$
(4.13)

The sum of the squared velocity components along all three axes is often used as an indicator of the average turbulence kinetic energy (*TKE*) and used as an indicator of the strength of the turbulence:

$$\langle q^2 \rangle = \langle u_1^2 \rangle + \langle u_2^2 \rangle + \langle u_3^2 \rangle = \frac{1}{\rho} (R_{11} + R_{22} + R_{33}) = \frac{2}{\rho} TKE.$$
 (4.14)

The incompressible Navier-Stokes equations (Eqs. 4.7 and 4.8) can be reapplied to the turbulent velocities to find an equation for the change and motion of this kinetic energy:

$$\rho \frac{\partial \frac{1}{2} \langle q^2 \rangle}{\partial t}$$
(4.15)(a)

$$+\rho\left(\langle U_1\rangle\frac{\partial \frac{1}{2}\langle q^2\rangle}{\partial x_1}+\langle U_2\rangle\frac{\partial \frac{1}{2}\langle q^2\rangle}{\partial x_2}+\langle U_3\rangle\frac{\partial \frac{1}{2}\langle q^2\rangle}{\partial x_3}\right)=$$
(4.15)(b)

$$-\left(\begin{array}{c}R_{11}\frac{\partial\langle U_{1}\rangle}{\partial x_{1}}+R_{12}\frac{\partial\langle U_{1}\rangle}{\partial x_{2}}+R_{13}\frac{\partial\langle U_{1}\rangle}{\partial x_{3}}+R_{21}\frac{\partial\langle U_{2}\rangle}{\partial x_{1}}+R_{22}\frac{\partial\langle U_{2}\rangle}{\partial x_{2}}\\+R_{23}\frac{\partial\langle U_{2}\rangle}{\partial x_{3}}+R_{31}\frac{\partial\langle U_{3}\rangle}{\partial x_{1}}+R_{32}\frac{\partial\langle U_{3}\rangle}{\partial x_{2}}+R_{33}\frac{\partial\langle U_{3}\rangle}{\partial x_{3}}\end{array}\right)$$
(4.15)(c)
$$-\rho\left(\frac{\partial\frac{1/2}{\langle q^{2}u_{1}\rangle}}{\partial x_{1}}+\frac{\partial\frac{1/2}{\langle q^{2}u_{2}\rangle}}{\partial x_{2}}+\frac{\partial\frac{1/2}{\langle q^{2}u_{3}\rangle}}{\partial x_{3}}\right)$$
(4.15)(d)
$$-\left(\left\langle u_{1}\frac{\partial p}{\partial x_{1}}\right\rangle+\left\langle u_{2}\frac{\partial p}{\partial x_{2}}\right\rangle+\left\langle u_{3}\frac{\partial p}{\partial x_{3}}\right\rangle\right)$$
(4.15)(e)

$$-\mu \left(\left\langle u_{1} \frac{\partial^{2} u_{1}}{\partial x_{1}^{2}} \right\rangle + \left\langle u_{1} \frac{\partial^{2} u_{1}}{\partial x_{2}^{2}} \right\rangle + \left\langle u_{1} \frac{\partial^{2} u_{1}}{\partial x_{3}^{2}} \right\rangle + \left\langle u_{2} \frac{\partial^{2} u_{2}}{\partial x_{1}^{2}} \right\rangle + \left\langle u_{2} \frac{\partial^{2} u_{2}}{\partial x_{2}^{2}} \right\rangle + \left\langle u_{2} \frac{\partial^{2} u_{2}}{\partial x_{2}^{2}} \right\rangle + \left\langle u_{3} \frac{\partial^{2} u_{3}}{\partial x_{1}^{2}} \right\rangle + \left\langle u_{3} \frac{\partial^{2} u_{3}}{\partial x_{2}^{2}} \right\rangle + \left\langle u_{3} \frac{\partial^{2} u_{3}}{\partial x_{3}^{2}} \right\rangle + \left\langle u_{3} \frac{\partial^{2} u_{3}$$

Equation 4.15 is broken into several parts not only because of its length, but because each of the labeled sections corresponds to a different mechanism or aspect of the turbulent activity. Part (a) is, of course, the change in turbulent energy with respect to time at a fixed point. Part (b) is the convection of that energy with the mean flow. Part (c) is the production of turbulent energy through interaction with the mean flow. Part (d) is the advection of turbulent energy through turbulent mixing. Part (e) is the transfer of turbulence through pressure fluctuations, while part (f) contains viscous effects of dissipation and diffusion.

A problem with dealing with these equations of turbulence is one of closure. Between the variables for the mean flow and those for the turbulence, there are not enough equations to find solutions for all the variables. Additionally, introducing the mean and turbulent properties as separate quantities is a significant step backwards in simplifying the equations. In order to reduce the process of solving the equations and of performing numerical simulations to a manageable level, further simplifying assumptions are needed.

A common set of simplifying assumption applied to turbulence are that it is homogeneous and isotropic. Homogeneity indicates the average magnitude and behavior of the turbulent velocities are the same everywhere in space, so that u_i may be a function of position in a particular realization of the flow, but the ensemble average, $\langle u_i \rangle$, is not. This also applies to R_{ij} , and $\langle q^2 \rangle$. Isotropy includes the property of homogeneity and indicates the properties of the turbulence do not change with orientation, so that $R_{11} = R_{22}$ $= R_{33}$ for any set of axes with equal scales and $R_{ij} = R_{kl}$ for all $i \neq j$ and $k \neq l$. This assumption has some justification from the equations alone, as parts (d) and (e) in Eq. 4.15 serve to spread turbulence. Part (f) also does this, as well as transferring turbulent energy between axes of orientation.

Applying this assumption causes most terms containing a derivative in a spatial dimension of a turbulent quantity to go to zero. In Eq. 4.11, the mean flow becomes uncoupled from the turbulent flow. The uniform-intensity turbulent flow has no effect on the mean flow, though the mean flow can still play a role in production of turbulence. Equation 4.15 is greatly simplified by this assumption as all the terms involving convection, advection, or diffusion of the turbulent energy go to zero, leaving:

$$\rho \frac{\partial \frac{1}{2} \langle q^{2} \rangle}{\partial t} = -\left(R_{11} \frac{\partial \langle U_{1} \rangle}{\partial x_{1}} + R_{12} \frac{\partial \langle U_{1} \rangle}{\partial x_{2}} + R_{13} \frac{\partial \langle U_{1} \rangle}{\partial x_{3}} + R_{21} \frac{\partial \langle U_{2} \rangle}{\partial x_{1}} + R_{22} \frac{\partial \langle U_{2} \rangle}{\partial x_{2}} \right) \\ + R_{23} \frac{\partial \langle U_{2} \rangle}{\partial x_{3}} + R_{31} \frac{\partial \langle U_{3} \rangle}{\partial x_{1}} + R_{32} \frac{\partial \langle U_{3} \rangle}{\partial x_{2}} + R_{33} \frac{\partial \langle U_{3} \rangle}{\partial x_{3}} \right) \\ - \mu \frac{1}{2} \left[\left\langle \left(\frac{\partial u_{1}}{\partial x_{1}} + \frac{\partial u_{1}}{\partial x_{1}} \right)^{2} \right\rangle + \left\langle \left(\frac{\partial u_{2}}{\partial x_{2}} + \frac{\partial u_{2}}{\partial x_{2}} \right)^{2} \right\rangle + \left\langle \left(\frac{\partial u_{3}}{\partial x_{3}} + \frac{\partial u_{3}}{\partial x_{3}} \right)^{2} \right\rangle \\ + 2 \left\langle \left(\frac{\partial u_{1}}{\partial x_{2}} + \frac{\partial u_{2}}{\partial x_{1}} \right)^{2} \right\rangle + 2 \left\langle \left(\frac{\partial u_{1}}{\partial x_{3}} + \frac{\partial u_{3}}{\partial x_{1}} \right)^{2} \right\rangle + 2 \left\langle \left(\frac{\partial u_{2}}{\partial x_{3}} + \frac{\partial u_{3}}{\partial x_{2}} \right)^{2} \right\rangle \right].$$
(4.16)

The right-hand side of Eq. 4.16 consists of two terms, one relating to the production of turbulent energy through interaction with the main flow, and one for the viscous dissipation of that energy. The viscous dissipation term is sometimes called $\langle \varepsilon \rangle$.

4.2.2. Spectral Analysis of Turbulence

Central to many discussions of turbulence are correlation functions. For two points, $\vec{x} = (x_1, x_2, x_3)$ and $\vec{x'} = (x_1, x_2, x_3)$, and a property *A*, this is defined as $B(\vec{x}, \vec{x'}) = \langle A(\vec{x})A(\vec{x'}) \rangle$. The most common correlation to focus on is a correlation of the turbulent velocities, which may include velocities in different directions and also may be a function of time,

$$B_{ij}\left(\vec{x}, \vec{x'}, t\right) = \left\langle u_i\left(\vec{x}, t\right) u_j\left(\vec{x'}, t\right) \right\rangle.$$
(4.17)

In the case of homogeneous turbulence, the location of the points \vec{x} and $\vec{x'}$ becomes irrelevant, only the vector from one to the other matters:

$$\vec{r} = \vec{x'} - \vec{x}, \tag{4.18}$$

$$B_{ij}(\vec{r},t) = \left\langle u_i(\vec{x},t)u_j(\vec{x}+\vec{r},t) \right\rangle.$$
(4.19)

In isotropic turbulence, the orientation of the vector \vec{r} has no bearing on the value of B_{ij} , only the magnitude, $(r = |\vec{r}|)$ which reflects the distance between two arbitrary points.

Correlation functions tend to have their maximum value when the distance of separation is zero, the two points are the same, and

$$B_{ij}(0,t) = \left\langle u_i(\vec{x},t)u_j(\vec{x},t) \right\rangle.$$
(4.20)

For small magnitudes of \vec{r} , the flow at one point will be linked to the flow observed at the other, but there are likely to be differences in behavior as well. As the separation distance increases, the link between one point and the other becomes more tenuous and is increasingly overwhelmed by the effects from the flow at points other than the two in question. Thus, correlation functions approach zero as distance approaches infinity, often swiftly enough that they become effectively zero within a finite distance. It is common practice to assume correlations functions are Gaussian in shape, as this produces results in accordance with the observed power-law behavior described later in section 4.2.3.

Spectral functions are then defined as the Fourier transform of the correlation functions. In non-isotropic flows, this should be a three-dimensional transform of the form:

$$\Phi_{ij}(\vec{k},t) = \frac{1}{8\pi^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} B_{ij}(\vec{r},t) e^{-i\vec{k}\cdot\vec{r}} d^3\vec{r}.$$
(4.21)

Under isotropic conditions, a spectral function may be defined threedimensionally as above, or one-dimensionally. The magnitude of the wavenumber vector, $k = |\vec{k}|$, has units of $(\text{length})^{-1}$, therefore, $(k)^{-1}$ can be considered a length scale. Likewise, the spectral function is a function of k rather than of \vec{k} under isotropic conditions:

$$\hat{B}_{ij}(k,t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} B_{ij}(r,t) e^{-ikr} dr \,.$$
(4.22)

The inverse of the one-dimensional transform is:

$$B_{ij}(r,t) = \int_{-\infty}^{\infty} \hat{B}_{ij}(k,t) e^{-ikr} dk .$$
 (4.23)

Note that for r = 0, B_{ij} is equal to the Reynolds stress (R_{ij}) as defined in section 4.2.1, divided by the density. Setting r = 0 and i = j, one finds that:

$$\frac{1}{2}B_{ii}(0,t) = \frac{R_{ii}(r,t)}{2\rho} = \frac{1}{\rho} \left(\frac{1}{2}\rho \left\langle u_{i}(t)^{2} \right\rangle \right) = \frac{1}{2} \int_{-\infty}^{\infty} \hat{B}_{ii}(k,t) dk$$
(4.24)

which indicates the turbulent kinetic energy in the x_i direction is an integral of the spectral function over the wavenumbers represented by k. From this, \hat{B}_{ii} /2 can be interpreted as the distribution of this energy per unit mass over the wavenumbers.

The total kinetic energy of a three-dimensional flow is the sum of the kinetic energy in all three axes. Again, isotropic conditions indicate this energy to be the same in all three axes, so that total energy per unit mass is equal to $\frac{3}{2}B(0,t)$. Likewise, the sum of the spectral functions can be considered the full energy spectrum:

$$E(k,t) = \frac{1}{2} \left(\hat{B}_{11}(k,t) + \hat{B}_{22}(k,t) + \hat{B}_{33}(k,t) \right) = \frac{3}{2} \hat{B}(k,t)_{(isotropic)}$$
(4.25)

and

$$\frac{1}{2} \left[\left\langle u_{1}(t)^{2} \right\rangle + \left\langle u_{2}(t)^{2} \right\rangle + \left\langle u_{3}(t)^{2} \right\rangle \right] = \int_{0}^{\infty} E(k,t) dk \quad .$$
(4.26)

It is possible to arrive at the same form of energy spectrum by Fourier transforms of the turbulent velocities rather than the correlation functions, but as $\langle u_i \rangle$ does not go to zero with distance in the manner that the correlation function does, resolving the integrals becomes somewhat more complicated with that approach.

Applying the Fourier transform to the Navier-Stokes equations can also be used in a spectral approach, and can be used to derive a spectral equivalent to Eq. 4.16, producing the wavenumber-dependent change of energy with time:

$$\frac{\partial E(k,t)}{dt} = G(k,t) + F(k,t) - 2\nu k^2 E(k,t) .$$
(4.27)

In this equation, G(k,t) represents the generation of turbulence, adding turbulent energy to the flow through interaction with the mean flow or by other means. This is sometimes referred to as forcing or production. The term $2vk^2E(k,t)$ represents the viscous dissipation of energy. It should be noted that with proportionality to k^2 , dissipation becomes more dominant at higher wavenumbers, which correspond to smaller length scales. F(k,t) is the energy flux term, representing the transfer of energy between length scales. Due to aforementioned closure problems with the RANS equations, this term requires a cubic velocity correlation that is not defined by the RANS, and expressions for this term may vary with different models. For isotropic turbulence and a cubic correlation function $\xi(r,t)$, this term can be found to be⁴

$$F(k,t) = \frac{1}{3} \left(\left\langle u_1(t)^2 \right\rangle + \left\langle u_2(t)^2 \right\rangle + \left\langle u_3(t)^2 \right\rangle \right) \frac{k}{\pi} \int_0^\infty \frac{\sin(kr)}{r} \frac{\partial}{\partial r} \left[\frac{1}{r} \frac{\partial}{\partial r} r^4 \xi(r,t) \right] dr \,. \tag{4.28}$$

In many derivations and analyses, it is useful to note that F(k,t) involves only the transfer of existing energy, not the addition or dissipation of energy. Therefore,

$$\int_{0}^{\infty} F(k,t)dk = 0.$$
 (4.29)

4.2.3. The Kolmogorov Spectrum

This model of turbulence was put forth by Andrei Kolmogorov in 1941.⁵ The events of World War II prevented his work from becoming known in the west for some years, and during that time others arrived at the same derivations independently.^{6,7} Kolmogorov published corrections to his theory in 1962,⁸ but many works continue to refer back to his original paper.

The Kolmogorov model makes many assumptions about turbulence. As was often done in the preceding sections, it assumes that incompressibility and an isotropic distribution of ensemble-averaged turbulence are good descriptors of the flow. In the formulation of this model, it is also assumed that the turbulence is quasi-stationary, with either no mean flow or a mean flow that is largely irrelevant to turbulent activity on the length scales of interest. It is then assumed that the turbulence is fully developed and has reached a form of equilibrium, such that the ensemble-averaged properties do not change with time.

If there is no change in time, then Eq. 4.27 becomes

$$2\nu k^{2} E(k) = G(k) + F(k).$$
(4.30)

Integrating over all values for *k* causes the flux term to drop out, as it deals only with transfer of existing energy, leaving

$$\langle \varepsilon \rangle = \int_{0}^{\infty} G(k) dk$$
 (4.31)

That is, the net rate of turbulent energy entering the flow must equal the net rate of viscous dissipation of energy from the flow, which is to be expected for equilibrium conditions of this sort.

A further assumption of the Kolmogorov model is that the production process and the dissipation process occur at widely separated length scales. The large scale is dominated by large persistent eddies, sustained by the production of turbulent energy. In the atmosphere, this may correspond to large-scale wind shear, plumes due to solar heating, or other phenomena that tend to exist on the larger scales. As noted previously, k^2 is a component of the viscous dissipation term in spectral form. Thus, it reaches greatest prominence at large values of *k*, corresponding to small length scales.

The range of scales between these two is known as the inertial range. In this range, both the addition and dissipation of energy are negligible. By the assumptions regarding the large and small scales, as well as Eq. 4.31, this indicates that energy enters the flow at the larger length scales, is transferred to the smaller length scales at a constant rate over this range, and that the rate of energy entering or leaving this range is equal to the net dissipation rate, $\langle \varepsilon \rangle$. From this, it is assumed that the energy spectrum in this region is of the form

$$E(k) = C_{\kappa} \langle \varepsilon \rangle^{n} k^{m}.$$
(4.32)

E(k) has units of velocity squared per wavenumber, which can also be written as length cubed over time squared, or l^3/t^2 . The units for ε and k are l^2/t^3 and 1/l, respectively. To match units of time, n must equal 2/3, which requires m to equal -5/3 to match units of length, leading to the result:

$$E(k) = C_{K} \left\langle \varepsilon \right\rangle^{2/3} k^{-5/3}.$$
(4.33)

Sometimes C_K and $\langle \varepsilon \rangle^{2/3}$ are combined into one constant, particularly if the magnitude of $\langle \varepsilon \rangle$ is unknown or irrelevant to the application in question. Under the approximation of Eq. 4.33, $|E(k)| = |C_K \langle \varepsilon \rangle^{2/3}|$ when k = 1. Therefore, this quantity corresponds to the energy associated with the 1-m scale in the flow. However, C_K has no units and the units of $\langle \varepsilon \rangle^{2/3}$ are $l^{4/3}/l^2$, which are not correct for an expression of energy. There is also a tendency to confuse this constant with C_v^2 , which is associated with the structure function described later in this section.

Figure 4.1 shows a sketch of an idealized energy spectrum for atmospheric turbulence. The peak of this curve corresponds to the length scales where turbulent production or forcing is most dominant in the flow. Energy transfers to both higher and lower wavenumbers, but preferentially transfers to larger values of k. This produces the inertial range which follows the $k^{-5/3}$ trend. With increasing k, viscous forces eventually overtake the inertia of the fluid, and the spectrum plot drops rapidly as the energy is dissipated.



Figure 4.1: Turbulent Spectra Approximation.

Kolmogorov's work also describes turbulence in terms of a structure function, defined as the average magnitude-squared difference in a property over a given separation

$$D_{v}(\vec{r}) = \left\langle \left(u(\vec{x} + \vec{r}) - u(\vec{x}) \right)^{2} \right\rangle.$$
(4.34)

This is related to the previously-defined correlation function $\langle u(\vec{x} + \vec{r})u(\vec{x}) \rangle$ by

$$\left\langle \left(u(\vec{x} + \vec{r}) - u(\vec{x}) \right)^2 \right\rangle = \left\langle u(\vec{x} + \vec{r})^2 \right\rangle + \left\langle u(\vec{x})^2 \right\rangle - 2\left\langle u(\vec{x} + \vec{r})u(\vec{x}) \right\rangle.$$
(4.35)

In the isotropic case, the relation with the correlation function $(B_\nu(r))$ can also be written as

$$D_{v}(r) = 2(B_{v}(0) - B_{v}(r)).$$
(4.36)

The subscript v in these expressions is used to indicate a velocity correlation or structure function.

It was found by Kolmogorov⁹ and Obukhov¹⁰ that this structure function has the form:

$$D_{\nu}(r) = 2\int_{0}^{\infty} (1 - \cos(kr)) \frac{Ce^{-k^{2}l^{2}}}{(k^{2} + 4\pi^{2}/L^{2})^{5/6}} dk .$$
(4.37)

As with the energy spectrum, there is a large scale, a small scale, and an inertial scale between the two. The variables L and l in Eq. 4.37 represent the length scales which can be considered "large scale" or "small scale", respectively.

Figure 4.2 shows the structure function for an example in which L=10 m and l=3 mm. It should be kept in mind that the horizontal axis in this figure corresponds to the separation distance r, not the wavenumber k, so large scales are towards the right on this figure, as opposed to being on the left for the Kolmogorov spectrum in Fig. 4.1. In the small-scale range, the structure function acts as a quadratic function. In the inertial range, it increases according to a 2/3 power law, and it remains constant in the large scales. Over the inertial range, the structure function can be approximated with the simple function $D_v(r) = C_v^2 r^{2/3}$. Note that *C* in Eq. 4.37 is not necessarily the same as C_v^2 in the $r^{2/3}$ fit. In Figure 4.2, *C* was set to 1 in generating the plot, but $C_v^2 = 3.5$ for the power-law fit to this plot.



Figure 4.2: Structure function with $r^{2/3}$ fit over the inertial range.

To review and clarify the difference between these two functions, the energy spectrum (E(k)) is based on the correlation function and represents the kinetic energy contained in eddies of a size 1/k. The structure function ($D_v(r)$) represents the average difference in velocity, squared, seen at two points separated by some distance r. By the relation between the structure function and the correlation function, the structure function and energy spectrum are interlinked in such a way that if the structure function can be approximated by

$$D_{v}(r) = C_{v}^{2} r^{n}. ag{4.38}$$

over the inertial range, then the energy spectrum can be approximated by

$$E(k) = \frac{\Gamma(n+1)}{2\pi} \sin\left(\frac{\pi n}{2}\right) C_{\nu}^{2} k^{-(n+1)}$$
(4.39)

over the range of wavenumbers corresponding to the length-scales of *r* for the inertial range.¹¹ In the case of n = 2/3, this comes to

$$E(k) \cong \frac{1}{8} C_{\nu}^{2} k^{-5/3}.$$
(4.40)

which matches the power law found in Eq. 4.33.

It should be noted that the units of C_v^2 should be $L^{4/3}/t^2$ for n = 2/3 in Eq. 4.38, which will also fulfill the units for Eq. 4.40. However, failure to remember that the expression for the energy spectrum in terms of C_v^2 contains an additional constant of proportionality, approximately equal to 1/8 in the case of the one-dimensional spectrum, is sometimes a source of confusion in discussions on this subject.

4.3. Optical Turbulence

Obukhov¹² applied Kolmogorov's ideas to a passive scalar in a turbulent flow. The kinetic energy spectrum associated with turbulence has units of l^3/t^2 because it corresponds to velocity (*l*/*t*) squared per wavenumber (1/*l*), and (*l*/*t*) (*l*/*t*) / (1/*l*) = l^3/t^2 . The "energy" spectrum for concentrations or intensity of a quality *A* is then defined as having units of A^2 per wavenumber, or A^2/l . If the turbulence behaves according to the Kolmogorov model, then $\langle \varepsilon \rangle$ and *k* will play roles in the transport and mixing of *A*, but there will also be diffusion of this quality, which can be represented by a second dissipation term, ε_A . This term will have units of A^2/t , and by dimensional analysis,

$$E_{A}(k) = K_{A} \langle \varepsilon \rangle^{m} \varepsilon_{A}^{n} k^{p} = K_{A} \langle \varepsilon \rangle^{-1/3} \varepsilon_{A} k^{-5/3}$$
(4.41)

for some value of the constant K_A . Thus, this scalar spectrum also follows the -5/3 rule found for the energy spectrum. In fact, a scalar spectrum for a flow will tend to have the same overall shape as the turbulent energy spectrum, though not necessarily exactly the same shape due to slightly different governing dynamics.

While the energy spectrum indicates energy content for the flow, scalar spectra such as this are more indicators of spatial frequencies in distribution of whatever property *A* represents. In some cases and applications, it is more physically meaningful to look at a structure function of the form

$$D_A(\vec{r}) = \left\langle \left(A(\vec{x} + \vec{r}) - A(\vec{x}) \right)^2 \right\rangle$$
(4.42)

as an indicator of the average variation in *A* over a separation distance of $|\vec{r}|$ between two points. As with the scalar spectrum, scalar structure functions tend to follow the 2/3 power law presented for velocity structure functions towards the end of section 4.2.3.

As was addressed in section 4.2, temperature is often treated as a passive scalar when dealing with incompressible flows. Therefore, there will be some temperature structure function, $D_T(\vec{r}) = \left\langle \left(T(\vec{x} + \vec{r}) - T(\vec{x})\right)^2 \right\rangle$, which can be approximated by

 $D_T(r) = C_T^2 r^{2/3}$ over the inertial range. As was noted in section 2.3, index of refraction in dry air varies with pressure and temperature in such a way that

$$n \cong 1 + 7.77 \cdot 10^{-7} \,\frac{P}{T} \tag{4.43}$$

and

$$dn \cong 7.77 \cdot 10^{-7} \frac{P}{T} \left(\frac{dP}{P} - \frac{dT}{T} \right)$$
(4.44)

for wavelengths near the visible range, pressure in Pascals, and temperature in Kelvin. These variations in *n* can then be described in terms of an index structure function $D_n(r)$ that can be approximated by $D_n(r) = C_n^2 r^{2/3}$ over the inertial range. If temperature fluctuations are the only significant contributor to variations in index, then

$$C_n^2 \cong \left(\frac{7.77}{10^7} \frac{P}{T^2}\right)^2 C_T^2.$$
 (4.45)

Since *n* is a ratio and without units, D_n is likewise unitless, and C_n^2 has units of length to the -2/3 power. The optical index spectrum can then be approximated by

$$E_n(k) \cong \frac{1}{8} C_n^2 k^{-5/3} \tag{4.46}$$

over the inertial range.

In many cases, one will see the refractive index power spectrum written as

$$\Phi_n(k) = 0.033 C_n^2 k^{-11/3}.$$
(4.47)

This is based on a three-dimensional Fourier transform, as seen in Eq. 4.21 rather than the one-dimensional form of Eq. 4.22 used throughout the preceding derivations. This approach is more physically meaningful for some optical applications, as it reflects an optical environment as a set of overlapping three-dimensional volumes of various sizes with variations in the average value of n for each region. Again, failure to remember the constants of 1/8 and 0.033 in Eqs. 4.46 and 4.47 is sometimes a source of confusion in discussions of optical turbulence.

4.3.1. Coherence Length

Since the presumption is that the single parameter, C_n^2 , fully characterizes the strength of the index of refraction in both scale size and amplitude, it follows that it should be possible to determine the smallest scale size of "significant" disturbances induced on a wavefront passing through the flow. Fried defined a coherence length, r_0 , now referred to as the Fried parameter that is associated with this presumption.

As addressed in previous sections, the phase variation on the wavefront of the beam propagating through the atmosphere is a result of integrating the effects of indexof-refraction variations along the propagation path. If those index variations are distributed according to a structure function following a 2/3 power law, then the phase structure function over the aperture can be described by a 5/3 power law ($D_{\varphi}(r) \cong C_{\varphi}^2 r^{5/3}$), the change in power from 2/3 to 5/3 being a consequence of the integration. The strength of an electromagnetic wave is usually defined by its amplitude (*A*); however, in many optical propagation problems, intensity (*I=A*²) indicates the energy falling on the target. A log-amplitude structure function can be defined in terms of the unaberrated amplitude, A_0 :

$$D_{A}(r) = \overline{\left(\ln(A(r)) - \ln(A_{0})\right)^{2}} = \left(\ln\frac{A(r)}{A_{0}}\right)^{2}.$$
 (4.48)

This is useful as it can be combined with the phase structure function as an overall wave structure function, D(r), such that $D(r) = D_{\varphi}(r) + D_A(r)$. Tatarski¹¹ also dealt with this definition of structure functions and found that over a propagation path from point s_1 to point s_2 :

$$D_{\varphi}(r) + D_{A}(r) = 2.91 \left(\frac{2\pi}{\lambda}\right)^{2} r^{5/3} \int_{s_{1}}^{s_{2}} C_{n}^{2}(s) ds \,.$$
(4.49)

Fried¹³ assumed that if one was attempting to recover information from an optical or electromagnetic signal received through an aperture, then there would be some signal of amplitude A_S , modulated onto a carrier wave of amplitude A_0 , such that $A_0 >> A_S$. This modulation might be one of amplitude, phase, or frequency. Based on this assumption, a circular aperture of diameter *d*, and a combined structure function of $D_{\varphi}(r) + D_A(r)$, a normalized signal to noise ratio can be written in terms of *d*, normalized by some length scale r_0 :

$$\Psi\left(\frac{d}{r_0}\right) = \frac{32}{r_0^2 \pi} \int_0^d \frac{1}{2} \left(\cos^{-1}\frac{r}{d} - \frac{r}{d} \left(1 - \left(\frac{r}{d}\right)^2\right)^{1/2}\right) e^{-\frac{6.88}{2} \left(\frac{r}{r_0}\right)^{5/3}} r dr \,. \tag{4.50}$$

This function, shown in Fig 4.3, is asymptotic to a constant value for large diameters, and proportional to the square of the diameter for small apertures. This may be explained as increasing the aperture gathers more light which increases signal strength, with the area being proportional to the square of the diameter. However, a larger diameter also leads to larger degrees of phase variation across the aperture and increasing noise. Eventually, the growing noise due to phase variation overtakes the greater light intake, and the signal to noise ratio levels off to a constant value.



Figure 4.3: Fried's normalized signal-to-noise ratio.¹³

Setting r_0 such that

$$D_{\varphi}(r) + D_{A}(r) = 6.88(r/r_{0})^{5/3}$$
(4.51)

normalizes this curve to a value of 1 over the larger apertures and places the intersection of the two asymptotes to the point where $d/r_0 = 1$. This also indicates that having an aperture larger than this value of r_0 will not significantly improve the signal to noise ratio of a signal or resolution of an image.

Fried combined Eqs. 4.49 and 4.51 to find an expression for r_0 in terms of C_n^2 :

$$r_0 = \left(0.423 \ \frac{4\pi^2}{\lambda^2} \int_{s_1}^{s_2} C_n^2(s) ds\right)^{-3/5}.$$
 (4.52)

As r_0 also represents a limit in improving the resolution of an image, it is often identified as the size of the largest aperture that can be considered diffraction limited. However, this is an inherently imprecise definition, as any variation from a planar or spherical wave over the aperture will result in a deviation from the diffraction-limited case, even if that deviation is vanishingly small.

It so happens that the value of r_0 arrived at as above is such that the mean-squared value of phase variations within a circular aperture of diameter r is

$$\sigma_{\varphi}^{2} = 1.04 \left(\frac{r}{r_{0}}\right)^{5/3}.$$
(4.53)

Thus, a more quantitative physical interpretation often given for r_0 is that it is the size of an aperture over which the rms of phase variations across the aperture is approximately 1 radian. This commonly given definition of r_0 is an incidental byproduct that is a few steps removed from the basis of its definition.

However, coupled with Eq. 4.49, it is clear that r_0 not only defines a length scale, but also defines the amplitude of the phase variation over that length scale, and r_0 can be used as a basis for designing an adaptive-optic system. As addressed in section 3.5.1, corrective systems are limited by the available degrees of freedom which are usually connected to the number of actuators that can be placed in a finite area, and so are often unable to completely match the form wavefront in performing a fit for correction. As a rule of thumb,¹⁴ if the points of actuation in a corrective system have a distance of separation, r_s , then the residual phase variance after correction by this system will be

$$\sigma_{r(fit)}^{2} = \kappa \left(\frac{r_{s}}{r_{0}}\right)^{5/3}.$$
(4.54)

The fitting error constant, κ , depends on the type and shape of correction associated with each actuator. If the actuators only correct for piston in the associated area, then $\kappa \approx 1.26$. For a segmented mirror, $\kappa \approx 0.28$ and for a truly deformable mirror $\kappa \approx 0.23$. These types of correction will be explained and explored in greater detail in chapter 8.

It should be remembered that all of these results are dependent on Kolmogorov turbulence being the source of these optical effects and on the length scales involved being in the range for which power law approximations apply. It should also be kept in mind that these results reflect averages and that the degree of optical aberration observed in a particular instantaneous case may be quite different.

4.3.2. The Greenwood Frequency

The analysis of turbulence and optical turbulence up to this point has been for quasi-stationary turbulence, ignoring any net motion of turbulent features in the flow, as well as most other time-varying qualities. However, this work is directed towards realtime corrections, in which case the temporal scale necessary for correction is of equal importance with the spatial scale.

While features seen in turbulent atmospheric flows can and do change with time, the optical effect of these changes is usually small compared to the motion of these sources of aberration relative to a beam or viewing path. Thus, the Taylor "frozen flow" hypothesis can often applied. The motion may be due to wind, or the optical path itself may be in motion due to movement of objects defining the endpoints of this path. If a source of aberration of length scale r has a relative velocity of V perpendicular to the path, then that suggests a time scale of r/V. If r_0 from the preceding section were taken as the characteristic length scale of the flow, then one might define a characteristic time scale of $\tau_0 = r_0/V$ or a characteristic frequency of V/r_0 . On the other hand, r_0 is associated with a circular aperture, not a one-dimensional length scale that can be oriented in the direction of the flow. Additionally, as was noted in the previous section, the qualities of an optical environment may change over the path length. Likewise, the relative velocity of the path may change with distance along the path. This is seen not only in the case of different wind speeds at different altitudes, but also in cases where a system and an object being tracked by the system are in motion relative to each other.

Fried and Greenwood coauthored a paper examining the power spectra of phase variations seen at an aperture, with regards to the bandwidth required from actuators driving a segmented mirror to correct these aberrations¹⁵. They found power spectra associated with a deformable mirror for correcting piston alone, power spectra for tilt correction, and differences in these spectra for segments on the edges of the segments surrounded by and influenced by other segments of the mirror. While this level of detail in their analysis is laudable, it is also a bit much for providing guidance in engineering applications. With this in mind, Greenwood simplified their work by first looking at the case of piston-only correction, and then taking the limit as the segment size shrunk to zero¹⁶. From this, the power spectral density for phase variation at an infinitesimal point in the aperture is found to be

$$PSD_{\varphi}(f) = 0.0326 \left(\frac{2\pi}{\lambda}\right)^2 f^{-8/3} \int_{s_1}^{s_2} C_n^2(s) |V(s)|^{5/3} ds$$
(4.55)

for an optical path from s_1 to s_2 . This expression allows for possible variations along the path in C_n^2 and the mean velocity (*V*) perpendicular to the path of propagation.

An actuator driving a point on a deformable mirror, or any other form of optical correction, is likely to have some frequency response, which can be expressed as a complex function, $H(f,f_c)$, where f_c is a characteristic frequency of the system. This correction will effectively filter out part of the aberrations. The power spectral function of the residual, uncorrected phase error (PSD_r) will then be

$$PSD_{r}(f) = |1 - H(f, f_{c})|^{2} PSD_{\varphi}(f)$$
 (4.56)

An interesting and useful effect of power spectra of this sort is that the total power of the spectrum, found by integrating over all frequencies, is equal to the average phase variance of the associated wavefronts. Therefore the residual phase variance due to the frequency response of the AO system, $\sigma_{r(f)}^2$, will be

$$\sigma_{r(f)}^{2} = \int_{0}^{\infty} \left| 1 - H(f, f_{c}) \right|^{2} PSD_{\varphi}(f) df .$$
(4.57)

If the frequency response is assumed to have a slow roll-off of the sort associated with RC-filters,

$$H(f, f_c) = (1 + i \cdot f / f_c)^{-1}$$
(4.58)

then for a desired value of σ_r one should design a system with a characteristic frequency such that

$$f_{c} = \left[0.102 \left(\frac{2\pi}{\lambda \sigma_{r(f)}} \right)^{2} \int_{s_{1}}^{s_{2}} C_{n}^{2}(s) |V(s)|^{5/3} ds \right]^{3/5}.$$
 (4.59)

As noted, the Fried parameter, which is considered to be "maximum aperture size for a diffraction-limited image," is a length scale that corresponds to a phase variance of approximately 1 radian. Thus, phase variance of 1 radian is often seen as the limit of an acceptable degree of phase variance and it is common practice to select a value of 1 for σ_r in the equation above. The resulting value of f_c is called the Greenwood frequency, denoted by f_G . This frequency is commonly used as a guideline for the bandwidth required of a corrective system intended to deal with optical turbulence. While the proper calculation of $\sigma_{r(f)}^2$ should use Eq. 4.57 to reflect the characteristics of a specific system, it is common¹⁴ to use the rule of thumb

$$\sigma_{r(f)}^2 = \left(\frac{f_G}{f_{3dB}}\right)^{5/3}.$$
(4.60)

in which f_{3dB} is the frequency associated with -3dB gain (output power = ½ input power) for the system, often referred to as the bandwidth of the system. Latency and time delays in the corrective system are normally included in the overall frequency response of the system, but if one wishes to consider the effects of latency alone, then a time delay of Δt produces a residual phase variance of¹⁷

$$\sigma_{r(\Delta t)}^2 = 28.4 (f_G \Delta t)^{5/3}.$$
(4.61)

4.4. The Limitations of C_n^2

As was noted in at the beginning of this chapter, C_n^2 and parameters based on C_n^2 , such as r_0 and f_G , have become the commonly accepted standard indicators in characterizing the optical effects of turbulence. However, they have become so commonly used that they are sometimes sought after or applied without consideration of whether or not they are applicable to the conditions encountered. As was addressed in preceding portions of this chapter, C_n^2 is defined by being the parameter in a 2/3-power-law curve-fit to the structure function $D_n(r)$, characterizing average magnitude of variations in *n* over separation distances. The existence of this structure function and the applicability of a 2/3-power-law fit to this function are in turn based on conditions of Kolmogorov turbulence, which has its own set of assumptions, including a requirement for the air encountered to be in a state of homogeneous, isotropic, semi-equilibrium with respect to turbulence. Furthermore, in finding D_n , it is common to assume that temperature fluctuations are the dominant contributor to optical effects, and that D_n is proportional to D_T , the structure function for variations in temperature. Finally the curve-fit of which C_n^2 is a parameter is intended only for the inertial range of turbulence, and so only describes effects associated with that range. If any of these conditions or assumptions are not applicable to the conditions encountered, then C_n^2 and parameters based on it become ill-defined or meaningless.

¹ Friehe, C.A., et. al., "Effects of temperature and humidity fluctuations on the optical refractive index in the marine boundary layer", *Journal of the Optical Society of America*, Vol. 65, No. 12, pp. 1502-1511, Dec 1975.

² Andrews, L.C., Phillips, R.L., Shivamoggi, B.K., Beck, J.K., and Joshi, M., "A statistical theory for the distribution of energy dissipation in intermittent turbulence," Phys. of Fluids A 1, pp 999-1006 (1989).

³ Hill, R J. and Clifford, S.F., "Modified spectrum of atmospheric temperature fluctuations and its application to optical propagation," Optical Society of America, Journal, vol. 68, July 1978, pp. 892-899.

⁴ Batchelor, G.K., 1953. Homogeneous Turbulence, Cambridge University Press, Cambridge.

⁵ Kolmogorov, A.N., On degeneration of turbulence in incompressible viscous fluid, C.R. Akad. Sci. SSR (Dokl.) Vol 31(6), 1941, pp 538-541.

⁶ Onsager, L., "The distribution of energy in turbulence," (Abstract) Phys. Rev. 68, (1945) page 286.

⁷ Von Weizsäcker, C.F., "Das Spektrum der Turbulenz bei großen Reynoldsshen Zahlen," Z. Physik. 124, (1948) pp. 614-627.

- ⁸ Kolmogorov, A.N. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. Journal of Fluid Mechanics, 13: 82–85, 1962.
- ⁹ Kolmogorov, A.N., "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. Dokl. Akad. Nauk SSSR 30, pp 301-305 (1941).
- ¹⁰ Obukhov, A.M., "Spectral energy distribution in turbulent flow," Izv. Akad. Nauk SSSR 5, pp 453-566 (1941).
- ¹¹ Tatarski, V.I., Wave Propagation in a Turbulent Medium, McGraw-Hill, New York, 1961.
- ¹² Obukhov, A. M.: Temperature field structure in a turbulent flow, Izv. AN SSSR (Geogr. and Geophys. series), 13, 58–69 1949.
- ¹³ Fried, D.L., "Optical Heterodyne Detection of an Atmospherically Distorted Signal Wave Front," Proceedings of the IEEE, Vol. 55, Jan 1967, pp 57-67.
- ¹⁴ Tyson, R.K., *Principles of Adaptive Optics*, Academic Press, San Diego, 1991.
- ¹⁵ Greenwood, D.P., and Freid, D.L., Power Spectra Requirements for Wave-Front-Compensative Systems, J. Opt. Soc. Am., Vol. 66, March 1976, pp 193-206.
- ¹⁶ Greenwood, D.P., Bandwidth Specification for Adaptive Optics Systems, J. Opt. Soc. Am., Vol. 67, March 1977, pp 390-393.
- ¹⁷ Hardy, J, Adaptive Optics for Astronomical Telescopes, Equation 9.57, Oxford Press, 1998.

CHAPTER 5:

AERO-OPTIC SHEAR LAYER EXPERIMENTS AT AEDC

5.1. Overview

In 1994, Ron Hugo and Eric Jumper of Notre Dame were offered access to facilities at the Arnold Engineering Development Center (AEDC). More detail concerning these experiments can be found in papers by Hugo^{1,2} and Fitzgerald.^{3,4} At the time of these experiments, the facility in question was uniquely suited for the study of high-speed shear layers and their optical effects. The results of these experiments differed from those predicted by the prevailing theories and schools of thought at the time, and a search for and explanation of these results prompted the line of inquiry leading to the work presented in this dissertation.

A shear layer is produced when two flows with parallel but unequal vectors of convective flow come into contact and interact. The layer of interaction between the two flows is subject to shear stresses due to the difference in velocity from one side to the other, hence the term, "shear layer." Such flows are also known as "mixing layers" especially if the gas or fluids on each side of the layer are of different compositions or otherwise have different characteristics. Among the commonly-observed characteristics of shear layers is the tendency for relatively small initial disturbances to form into larger coherent vortical structures that "roll" along between the two streams of differing velocities.⁵ Both the thickness of the shear layer and the size of the structures in the shear layer tend to grow over time.

A more detailed description of shear layers and their driving mechanics will be presented in chapter 6, along with theoretical explanations for the optical effects observed in the AEDC experiments. This chapter will concern itself the AEDC experiment, its results, and some of the initial implications of those results.

5.2. The Acoustic Research Tunnel

A schematic of the Acoustic Research Tunnel (ART) at AEDC is shown in Fig. 5.1. Using the AEDC Propulsion Wind Tunnel complex as a source of high-pressure air, this facility produced a shear layer comprised of a stream of approximately 0.8 Mach running parallel with a stream that had been slowed to ≈ 0.1 Mach by passing through an expansion and a high-density screen. The unit Reynolds numbers for these streams during these experiments were 1.4 and 12.7 x 10^6 m⁻¹ respectively. As both streams came from the same source, the total temperature of the two streams was equal and was measured to be $\approx 27^{\circ}$ C. Static pressure within the ART was a constant ≈ 0.6 Atm across the shear layer.⁶ The contraction of the test section was required to maintain constant mean static pressure along the length of the test section as low-speed air was entrained into the shear layer and accelerated. These conditions produced a convective velocity for structures and disturbances within the shear layer of ≈ 160 m/s.



Figure 5.1: AEDC ART facility.(Havener)⁶

Three observation stations of Schlieren-grade windows for optical measurement were built into the AEDC facility. These stations included not only the 20.32-cmdiameter side windows shown in Fig. 5.1, but also windows of the same shape and size, located on the top and bottom of the tunnel at the same streamwise locations as the side windows. The upper and lower windows were used to pass a set of narrow beams through the flow for measurements using the Small Aperture Beam Technique (SABT) described in section 3.3.3.

5.2.1. Initial Results

The ART facility proved to be subject to vibration in the range of 8 Hz to 2 kHz, largely due to the facility being located near and dependent upon the larger 16T/16S wind tunnel complex at AEDC as a source of high-speed air. This vibration was measured by attaching accelerometers to the ART and by measuring SABT beam jitter when running the 16T facility without directing any air from that facility through the ART.⁷ Hugo chose to deal with this by post-processing the data with a high-pass Butterworth filter

with a cutoff frequency of 2.5 kHz.⁸ Examples of time series of wavefronts reconstructed by Hugo from this data are shown in Figs. 5.2 and 5.3.



Figure 5.2: Station 1 wavefront reconstruction, from data filtered at 2.5 kHz. $(Hugo)^8$



Figure 5.3: Station 2 wavefront reconstruction, from data filtered at 2.5 kHz. $(Hugo)^8$

Average OPD_{rms} of the wavefronts at station 2 was on the order of 0.06-0.160 μ m with periods of greater and lesser distortion. The wavefronts at station 1 were of lesser amplitude, on the order of 0.04 μ m.

This work was the first time that sets of continuous time series of spatially resolved wavefronts had been measured under such conditions. Earlier work was capable of taking individual snapshots of wavefronts at isolated moments to resolve the features of wavefronts from such flows, but was not able to trace the evolution of those features with time. Previous estimates of performance degradation for optical systems dealing with flows of this nature were forced to rely on a statistical approach that treated the turbulence as random, much like the approach used in atmospheric propagation described in chapter 4.

5.2.2. Re-reduction of the AEDC Data

The experiment carried out by Hugo at the AEDC was largely intended to show that such measurements were possible. A more thorough analysis and explanation of what the measured wavefronts and their characteristics meant was carried out by Edward Fitzgerald at Notre Dame. In analyzing the AEDC data, he found that for measurements taken through the shear layer, with air was passing through the ART facility, vibration seemed to be a significant contributor to the measured beam jitter only for frequencies less than 500-700 Hz.⁹ Therefore, it seemed reasonable that the cutoff frequency for filtering out vibration from the measurements might be lowered from the 2.5 kHz used by Hugo. Figure 5.4 shows the results of repeating wavefront reconstruction for station 1, as performed by Hugo from the same data, but with different cutoff frequencies of filtration.


Figure 5.4: Wavefront reconstruction from AEDC data for station 1 with varying cutoff frequencies in high-pass filtering: (a) 2000 Hz, (b) 1200 Hz, (c) 750 Hz, (d) 0 Hz (unfiltered). (Fitzgerald, 2002.)⁹

Of note in parts (a) through (c) of Fig. 5.4 is that in changing the cutoff frequency from 2000 Hz to 750 Hz, the overall appearance of results from station 1 does not change significantly, although OPD_{rms} does increase by about 27% from 0.045 µm to 0.057 µm. When filtration is removed entirely in part (d), there is significant tilt off to one side. As was alluded to earlier and will be explained more fully in chapter 6, structures and features of shear layers normally begin as small-scale disturbances which grow and combine over time into larger structures. Station 1 of the ART was located at a point in the shear layer just after the two streams came into contact. In this region, the only features expected to be present in the flow are disturbances fed into the flow by from the boundary layer on the splitter plate. These disturbances eventually grow and develop into larger coherent structures, but this would not happen within the region of measurement at station 1 for the conditions within the ART facility. Therefore, features on the scale of the steady tilt seen over most of the wavefront series shown in Fig. 5.4 (d) are most likely a result of a corruption of the data, such as by low-frequency vibration, rather than a product of features in the flow.

Figure 5.5 repeats the reconstruction process of Fig. 5.4 with the data from station 2. In this case, changing the cutoff frequency from 2000 Hz to 750 Hz raises the OPD_{rms} by 37% to 0.274 µm from 0.167 µm. Beyond that, the overall character of the disturbances seen in these reconstructed wavefronts changes. The wavefronts become increasingly dominated by back-and-forth cycles of tip-tilt, with the average period of that cycle growing longer with decreasing cutoff frequency.



Figure 5.5: Wavefront reconstruction from AEDC data for station 2 with varying cutoff frequencies in high-pass filtering: (a) 2000 Hz, (b) 1200 Hz, (c) 750 Hz, (d) 0 Hz (unfiltered). (Fitzgerald, 2002.)⁹

The oscillating tilt of Fig. 5.5 (c) is indicative of structures larger than the 5 cm region of reconstruction passing by. The relations between structure size, aperture size, and T/T will be explored more thoroughly in chapter 8, but in general, flow structures and associated wavefront features of a length scale larger than the region of interest tend to produce T/T while shorter scales are associated with wavefront distortions requiring higher-order correction.

For a cutoff frequency of 750 Hz, the typical period of the T/T in the station 2 reconstructions is about 0.72 ms, which can also be expressed as a frequency of \cong 1400 Hz. This is why these features were not visible in the reconstruction of Fig. 5.5 (a) based on data filtered at 2000 Hz or the original reconstruction that was filtered at 2500 Hz. This frequency should be determined by the average size of the structures (Λ) and the convection velocity (U_c) at which they pass through the region so that

$$f = \frac{U_C}{\Lambda} \,. \tag{5.1}$$

With $f \cong 1400$ Hz and $U_C \cong 160$ m/s for the flow in the ART facility, $\Lambda \cong 11.4$ cm. As predicted, this is larger than the 5 cm reconstruction region and so manifests primarily as T/T effects in the wavefronts. Frequencies of 2500 Hz and higher, as were used in the original reconstruction, would correspond to structure sizes of $\Lambda \cong 6.4$ cm and smaller, which produce the higher-order disturbances with only a moderate amount of tilt seen in Fig. 5.3.

5.2.3. Severity of the Distortion

As previously noted, the goal at the time of the AEDC experiment was to show that measurements of this sort were possible. Explorations of what the results meant waited until afterwards. The first implication of this data is that sheer layers involving sub-sonic, compressible flows present a significant problem for optical applications Figure 5.6 shows time-varying near field wavefronts from the 750 Hz filtration cases in Figs. 5.4 and 5.5 as well as the associated far-field intensity patterns based on those wavefronts for light with a wavelength of $1.0 \,\mu\text{m}$. These patterns include the effects of tilt correction, but still the central lobe of this intensity pattern wavers and splits into lobes. Hints of an Airy Disk (in one dimension) can be seen now and again, in the results from station one. In the results for station two the irradiance pattern has splintered into a jumbled mess with peak irradiance being, at best, roughly half that seen from station one, and the location of that peak wanders over a greater range of locations.



Figure 5.6: One-dimensional wavefronts and far field patterns as a function of time for station 1 (left) and 2 (right). (Cicchiello, 2001)¹⁰

These far field patterns produce time varying Strehl ratios shown in Fig. 5.6. The average value for this ratio over this period of time at station one is 0.41, indicating that almost three-fifths of the potential on-target intensity may be lost in transmission due to these aberrations. If the application were for a high-bandwidth communications system, then the requirement to avoid losing bits in the data stream would most likely be some required minimum in the instantaneous Strehl ratio rather than its average value. By this standard, the performance is even worse, as the on-axis intensity effectively drops to zero at one point.¹¹ At station two, the Strehl ratio has an average value of 0.07, and over much of the recorded time period is very close to zero. From these results, it would appear that optical applications in environments where flows of this sort may be encountered will face serious problems if some form of correction is not provided.



Figure 5.7: Strehl ratios produced by the irradiance patterns in Fig 5.6. $(Cicchiello, 2001)^{11}$

5.2.4. Predictions and Prevailing Theory

A second implication of these results is that the commonly used assumptions and models at the time do not predict the severity of the measured optical distortions. Much of the work concerning optical effects of shear layers prior to this analysis focused on two-index mixing, in which the two streams making up the shear layer are comprised of different gasses. In cases where the two streams are of the same gaseous composition, this may still be applicable if one stream has been heated or cooled to a different stagnation temperature than the other. However, this is not the case in the shear layer produced in the ART facility, as both streams are air and have the same total temperature as they drawn from the same source.

Attempts at refining the two-index model to conditions like those of the ART facility have been made based on the idea that if the two streams are at different Mach numbers, then they will have different static temperatures, pressures, and densities, even if their stagnation properties are the same.¹² Another approach is to map out the velocity fields in the flow and then assume that localized temperature is solely a function of local velocity or Mach number by the relation:

$$T = T_0 - \frac{\left|\vec{V}\right|^2}{2C_p} = T_0 / \left[1 + \frac{(\gamma - 1)}{2}M^2\right].$$
 (5.2)

This is known as, "adiabatic heating."

These models all have two things in common. First, they all assume that all changes in index of refraction in the air are primarily temperature-based, much as was done for atmospheric propagation in chapter 4. Secondly, they all predict a lower degree of optical distortion for the flow conditions in the ART facility than what was actually measured. Table 5.1 lists maximum differences in *OPL* as predicted by the various models, as well as the "worst case" observed in reconstructions of the AEDC experimental data with a 750 Hz filter.

	2-index mixing	Modified 2-	Adiabatic	AEDC
		index	Heating	Experimental
				Results
OPD peak-to-	0.339	0.241	0.310	1.340
peak (µm)				

Table 5.1: Comparison of maximum peak-to-peak *OPD* for predictive models and experimental results.¹³

Even looking at average conditions instead of worst-case incidents, average OPD_{rms} as predicted by these models is on the order of $\cong 0.12 \,\mu\text{m}$ while the experimental results shown in Fig. 5.5 (c) have an average value of $\cong 0.27 \,\mu\text{m}$. For a comparison in visual form, Fig. 5.8 shows results of a simulation wavefronts produced by ART flow conditions, based on the adiabatic heating model side by side with another example of results from the AEDC experiment for the same flow conditions and location.



Figure 5.8: (left) Wavefronts generated by computational simulation and an adiabatic heating model. (right) AEDC results for flow conditions similar to those in the computational simulation.(Fitzgerald)¹³

All in all, it seems that these models predict optical distortions half the magnitude or less of those measured. Therefore, it would seem that a new look at the fundamentals of shear layers, their development and behavior, and the validity of these models and their assumptions is in order. New models may be required to adequately deal with optically-aberrating flows of this nature. This will be the subject of the next chapter.

- ¹ Hugo, R.J., Jumper, E.J., Havener, G. & Stepanek, C. "Time-resolved aero-optical measurements of a wavefront aberrated by a compressible free shear layer," AIAA Paper 95-1979.
- ² Hugo, R.J., Jumper, E.J., Havener, G. & Stepanek, C. "Time-resolved wave front measurements through a compressible free shear layer," AIAA J. Vol. 35, 671–677.
- ³ Fitzgerald, E.J. and Jumper, E.J., "Aperture Effects on the Aerooptical Distortions Produced by a Compressible Shear Layer," AIAA J., Vol. 40, No. 2, February 2002
- ⁴ Fitzgerald, E.J. and Jumper, E.J., "The optical distortion mechanism in a nearly incompressible free shear layer," J. Fluid Mech. (2004), vol. 512, pp. 153–189.
- ⁵ Papamoschou D., Roshko A., "The compressible turbulent layer: an experimental study," J Fluid Mech, vol 197, 1988, pp 453–77.
- ⁶ Havener, G. and Heltsey, F. "Design aspects and preliminary holographic-PIV measurements for a subsonic free shear layer flow channel" AIAA paper 94-2550, Jun 1994.
- ⁷ Hugo, R.J., Jumper, E.J., Havener, G., and Stepanek, S.A., "Time Resolved, Aero-Optical Measurements of a Wavefront Aberrated by a Compressible Shear Layer," AIAA Paper 95-1975, June 1995.
- ⁸ Hugo, R.J., Jumper, E.J., Havener, G., and Stepanek, S.A., "Time Resolved, Aero-Optical Measurements of a Wavefront Aberrated by a Compressible Shear Layer," AIAA J., Vol. 35, No. 4, pp 671-677.
- ⁹ Fitzgerald, E.J. and Jumper, E.J., "Aperture Effects on the Aerooptical Distortions Produced by a Compressible Shear Layer," AIAA J., Vol. 40, No. 2, February 2002.
- ¹⁰ Cicchiello, J.M, "Low-Order Representation of Dynamic Aero-Optic Distortions", PhD Dissertation, University of Notre Dame, April 2001
- ¹¹ J.M Cicchiello, "Low-Order Representation of Dynamic Aero-Optic Distortions", PhD Dissertation, University of Notre Dame, April 2001
- ¹² Dimotakis, P.E., Catrakis, H.J., and Fourguette, D.C., "Flow Structure and Optical Beam Propogation in High-Reynolds-Number Gas-Phase Shear Layers and Jets," J. Fluid Mech., No. 433, 2001, pp 105-134.
- ¹³ Fitzgerald, E.J. and Jumper, E.J., "The Optical Distortion Mechanism in a Nearly Incompressible Free Shear Layer," J. Fluid Mech. (2004), vol. 512, pp. 153–189.

CHAPTER 6:

THE PHYSICAL AND OPTICAL MECHANISMS OF SHEAR LAYERS

6.1. Overview

Atmospheric optical propagation, which was explored and explained in chapter 4, takes a spectral or stochastic approach to characterizing the distortions produced by the free atmosphere. The formulas and parameters used to characterize such flows assume that variations in temperature are the primary source of optical aberration, and that these variations encountered in an optical path through the atmosphere are essentially random. Experimental results from a sub-sonic but compressible shear layer, presented in chapter 5, indicate that such flows are dominated by semi-regular structures of a particular size, and that temperature fluctuations are insufficient to explain the severity of the optical distortions observed. Therefore, an exploration of what does take place in a flow of this sort would be beneficial in understanding how to characterize the optical effects of such flows and to move towards some guidelines for applying correction as exist for atmospheric propagation.

Shear layers are of interest in airborne applications as they can be found in the boundaries of separated flows on aircraft or turrets, as shown in Fig. 6.1. Also shown in this figure are a number of other flows that might share some of the characteristics of a compressible shear layer and might act as additional sources of aberration.



Separation and Reallachment Line

Figure 6.1: Separation, shear layer, and other potentially-aberrating flows for a possible turret configuration.

The shear layer is of particular interest, as it occupies a position directly in the optical path when that path is directed downstream. Figure 6.2 shows results found by Demos Kryazis¹ for a spherical turret under flight conditions. It reveals a sharp drop in delivered intensity at angles greater than 90° from forward. The reasons for this will be shown in the following sections.



Figure 6.2: Strehl ratio as a function of azimuth look angle. (Demos Kryazis)¹

6.2. Shear Layer Mechanics

The basic form of a shear layer is shown in Fig. 6.3, consisting of two regions of fluid flowing in parallel, but with different velocities. In fact, U_1 and U_2 may be in opposite directions. Shear layers are also known as mixing layers, though this term is usually applied when there are distinct differences in density, temperature, or other properties of the fluids in the two regions making up the shear layer.



Figure 6.3: The basic flow of a shear layer.

In the physical world, the sharp discontinuity in velocity indicated in Fig. 6.3 can not persist. Momentum is transferred from faster fluids near the boundary to slower fluids near the boundary, creating a layer of transition between the two main flows. Michalke² noted that the profiles of average velocity observed in many real shear layers could be approximated by a hyperbolic-tangent function,

$$U(y) = \frac{1}{2} [1 + \tanh(y)].$$
 (6.1)

which may be shifted and scaled in y and U to fit a particular case. This profile is shown in Fig. 6.4.



Figure 6.4: Hyperbolic-tangent velocity profile approximation for shear layers.

6.2.1. Kelvin-Helmholtz Instability

In 1868, Hermann von Helmholtz noted that surfaces marking some sort of discontinuity of velocity, such as those seen around regions of separated flow, were

inherently unstable.^{3,4} Lord Kelvin proved this instability mathematically three years later.⁵ In honor of their contributions in this area, this sort of behavior is known as Kelvin-Helmholtz instability.

Batchelor⁶ explained the physical mechanism behind this instability in 1967, using vorticity dynamics. In this approach, the fluid is considered to be inviscid, and the boundary between the two regions of flow is considered to be a line or sheet of vortices. If the line is perfectly flat, as shown in Fig. 6.5 (a), then the forces and induced velocities on a portion of the line from vortices comprising other portions of the vortex line sum to zero. However, if a sinusoidal disturbance is imposed on the vortex line, as shown in Fig. 6.5 (b), then the portions of the line that are displaced downward will induce motion on the portions displaced upwards, and vice versa. These forces cause the elevated and depressed portions of the line to move in opposite directions along the boundary, as well as forcing the displaced portions of the vortex line to move further from the initial positioning of the boundary.



Figure 6.5: (a) Vortex line, (b) perturbed vortex line with non-zero net forces and torques.

Michalke² performed a stability analysis of this flow assuming a mean flow with a hyperbolic-tangent profile as shown in Fig. 6.4 and perturbations of this flow of the form:

$$c = c_r + i \cdot c_i , \qquad (6.2)$$

$$\Psi(x, y, t) = \Re\left[\phi(y)e^{i\alpha(x-ct)}\right],\tag{6.3}$$

$$u' = \frac{\partial \psi}{\partial y} = \phi'(y)e^{i\alpha(x-ct)}, \qquad (6.4)$$

$$v' = \frac{-\partial \psi}{\partial x} = -i\alpha\phi(y)e^{i\alpha(x-ct)}.$$
(6.5)

In these equations and their analysis, u' and v' are the perturbation velocities, ψ is a defined stream function, ϕ is the amplitude of a disturbance, α is the wavenumber of this disturbance, and the complex number $c = c_r + i \cdot c_i$ contains the phase velocity (c_r) and an indicator of the growth rate for the disturbances. (c_i) These perturbation velocities are then inserted into Euler's equation of motion for inviscid flow,

$$\rho \left(\frac{\partial \vec{U}}{\partial t} + \vec{U} \cdot \nabla \vec{U} \right) = -\nabla P .$$
(6.6)

Michalke's results for a stability analysis of a linearized approximation of this flow are shown in Fig. 6.6. He found that disturbances with wavenumbers in the range of $0 < \alpha$ <1 are amplified by the governing mechanics of the flow, and grow with time. The most amplified case is for a wavenumber of $\alpha = 0.4446$.



Figure 6.6: Stability curve for a hyperbolic-tangent velocity profile. (Michalke)

The wave number, α , from Michalke's analysis corresponds to the streamwise dimension of a disturbance (Λ) by the relationship

$$\alpha = \frac{2\pi}{\Lambda}.$$
(6.7)

Thus, the most amplified disturbances are those with a streamwise length of $2\pi / 0.4446$, or 14.13, relative to the vertical length scale for the profile shown in Fig. 6.4. These disturbances will move with the convection velocity (U_C) such that

$$U_{c} = \frac{U_{1} + U_{2}}{2}.$$
 (6.8)

This motion will then produce a frequency in the fluctuations observed at a fixed point, such that

$$f = \frac{U_c}{\Lambda}.$$
(6.9)

It is common practice to characterize flows such as this by a dimensionless ratio known as the Strouhal number, which is defined as

$$St = \frac{f \cdot L}{V}.$$
(6.10)

The convection velocity provides a characteristic velocity (V), and a characteristic frequency is identified in Eq. 6.9. The characteristic length scale (L) normally used is the thickness of the shear layer. However, there is more than one definition of thickness for flows of this sort.

The simplest definition of thickness may be the visual thickness (δ_{vis}) which is an estimate based on how thick the shear layer appears to be when made visible by Schlieren photography or other flow-visualization techniques. However, this estimate is inherently imprecise and subjective.

Vorticity thickness (δ_{ω}) reflects a linear fit of profiles along the lines of the hyperbolic-tangent approximation shown in Fig. 6.4. In this fit, shown in Fig. 6.7, there are three regions: two where the velocity is approximated by a constant equal to the free-stream velocities ($u = U_1$ or $u = U_2$) and one in which the velocity gradient, du/dy, is a constant equal to the highest value for this gradient seen in the actual profile. The thickness of this linear-fit profile then becomes the difference in free stream velocities divided by the maximum velocity gradient:

$$\delta_{\omega} = \frac{\left| U_{2} - U_{1} \right|}{\left[\frac{\partial u}{\partial y} \right]_{\max}}.$$
(6.11)

It is customary⁷ to assume that $\delta_{\omega} \cong 0.5 \ \delta_{vis}$.



Figure 6.7: Linear-fit basis of vorticity thickness.

Momentum thickness (δ_{θ}) is a concept that was originally developed for boundary layers by Von Karman.⁸ Viscous effects cause the fluid in contact with a solid surface to have a tangential velocity of zero with respect to that surface, as well as reducing the parallel velocity of fluid near this surface. Momentum thickness represents this loss of velocity and momentum in terms of a layer of fluid moving at the free-stream velocity. Specifically, a layer of fluid with velocity U_{∞} and thickness δ_{θ} would have a net momentum equal to that which is missing from the flow due to the velocity deficits found in the boundary layer.



Figure 6.8: The basis of momentum thickness for a boundary layer.

A flow of thickness δ_{θ} , a uniform velocity U_{∞} , and a uniform density ρ_{∞} , would have a momentum flux past a fixed point of

$$(\rho_{\omega}U_{\omega}\delta_{\theta})U_{\omega} = \rho_{\omega}U_{\omega}^{2}\delta_{\theta}$$
(6.12)

per unit length in the cross-stream direction. If the dimension of thickness, δ_{θ} , is to be matched to the reduction in momentum flux due to changes in density and velocity from the free-stream values, then

$$\rho_{\infty}U_{\infty}^{2}\delta_{\theta} = \int_{0}^{\infty} u(y)\rho(y) (U_{\infty} - u(y)) dy.$$
(6.13)

Thus, δ_{θ} can be found by the following integral:

$$\delta_{\theta} = \frac{1}{\rho_{\infty}U_{\infty}^2} \int_0^{\infty} u(y)\rho(y) (U_{\infty} - u(y)) dy = \int_0^{\infty} \frac{u(y)\rho(y)}{U_{\infty}\rho_{\infty}} \left(1 - \frac{u(y)}{U_{\infty}}\right) dy.$$
(6.14)

Unlike boundary layers, shear layers are not bounded by a solid surface and the mean of u may or may not go to zero at any point. However, the mean velocity does

approach some value U_2 for increasingly positive y and some value U_1 for increasingly negative y. This can be used to find and equivalent δ_{θ} for shear layers in the form:

$$\delta_{\theta} = \int_{-\infty}^{\infty} \frac{\rho(y)}{\rho_2} \frac{u(y) - U_1}{U_2 - U_1} \left(1 - \frac{u(y) - U_1}{U_2 - U_1} \right) dy.$$
(6.15)

The ratio between δ_{θ} and δ_{ω} is not fixed and depends on flow conditions.⁹

This dissertation will primarily use δ_{θ} , δ_{ω} , and occasionally δ_{vis} as measures of thickness, but these are not the only such definitions in use. Some define shear layer thickness in terms of endpoints where the average velocity is within some percentage of the velocities at $\pm \infty$, relative to $\Delta U = |U_1 - U_2|$. Displacement thickness is defined as the distance by which a streamline that would run parallel to a surface is displaced by the boundary layer near that surface. Like momentum thickness, this can be adapted to a shear layer, but it is used less frequently than momentum thickness in these types of flow.

The hyperbolic-tangent profile for a specific case can be written in terms of the two stream velocities and the momentum thickness,

$$U(y) = \frac{U_2 - U_1}{2} \left[1 + \tanh\left(\frac{y}{2\delta_{\theta}}\right) \right] + U_1.$$
(6.16)

From this, $2\delta_{\theta}$ would be the proper length scale for scaling the results from Michalke's stability analysis. Thus, from Eq. 6.7, the streamwise length scale of the most amplified disturbance should be

$$\Lambda = \frac{2\pi}{0.4446} 2\delta_{\theta} = \frac{4\pi\delta_{\theta}}{0.4446}.$$
(6.17)

Combining this with Eqs. 6.9 and 6.10, the most amplified Strouhal number with regards to momentum thickness (St_{θ}) should be

$$St_{\theta} = \frac{f\delta_{\theta}}{U_{c}} = \frac{1}{\Lambda}\delta_{\theta} = \frac{0.4446}{4\pi\delta_{\theta}}\delta_{\theta} = 0.0354.$$
(6.18)

More recent and refined analysis by Mokewitz, Heuerre, and Ho indicates a value of 0.032 for the most amplified Strouhal number.^{10,11} Despite differences in the analysis, these numbers are fairly close to each other, and numerous experiments confirm that the most amplified Strouhal number is not far removed from these values.^{12,13,14,15}

6.2.2. Initial Development

As previously noted, the infinitely thin dividing line and sharp jump in velocity between the two flows indicated by Fig. 6.3 is something that exists only in a theoretical case. Likewise, it takes some finite time for the hyperbolic tangent profile of Fig. 6.4 to develop. In the real world, the two flows will have existed and traveled from some previous point, developing boundary layers as they progressed. Figure 6.9 shows a shear layer formed by two flows initially separated by a thin wedge or plate. Each flow has a boundary layer of flow retarded by contact and interaction with the plate. Once the plate has been left behind, the two boundary layers come into contact with each other, producing a wake in which the flow is less than the mean velocity of either of the two flows. This wake persists for a short period, before approaching a time-averaged velocity profile that can be approximated with the hyperbolic-tangent function presented in section 6.2.1 and Fig. 6.4. While this model is for two flows coming into contact, setting the lower flow, U_1 , to zero can be used to represent a flow over a solid object with a separation region behind the object.



Figure 6.9: Sketch of spatially developing profiles and momentum thickness of a shear layer.

These initial profiles have a momentum thickness, which is in turn associated with preferential amplification of certain frequencies and wavelengths of disturbance, as described in section 6.2.1. The initial existence of the wake appears to have little effect on which wavelengths are most amplified.¹⁶ Noting the development and changes in the flow downstream is then best expressed in terms of $x/\delta_{\theta 0}$, where x is the downstream distance and $\delta_{\theta 0}$ is the initial momentum thickness. Ho and Huang¹⁴ indicate that scaling results in relation to the momentum thickness of the boundary layer of the higher-speed flow produces more consistent results than using the momentum thickness of the entire wake. This is attributed to the high-speed boundary layer containing most of the initial vorticity in the flow.

As was noted by Helmholtz, discontinuities in a flow are inherently unstable, which includes the velocity deficit in the initial wake. The amplification of disturbance waves in accordance with Michalke's stability analysis has been observed in the region $x/\delta_{\theta 0} \cong 1$. On the other hand, as indicated by the sketch in Fig 6.9, the flow does not immediately achieve something comparable to the hyperbolic-tangent profile. This can lead to the development of an initial sub-shear layer, with the expected characteristics of a shear layer, but bracketed on one or both sides by regions that retain characteristics more indicative of boundary layers.¹⁷ Signs of this may be seen out to a distance of $x/\delta_{\theta 0} \cong 100$. This can affect the most amplified Strouhal number in this region, as that number is usually based on the initial thickness of the boundary layers, but the sub-shear layer may not partake of the entirely of these layers. This is particularly true if one or both of the incoming boundary layers are turbulent, in which case the initial disturbances may owe more to the length scale and velocities of the inner boundary layer, rather than those of the whole boundary layer.¹⁷

It was once thought that shear layers were examples of stochastic turbulence, with random variations in velocity imposed on top of the mean velocity profile. This is the basis of Michalke's analysis, which assumes that disturbances of all frequencies are initially present in the flow with roughly equal strengths, but that some range of frequencies is preferentially amplified while disturbances outside that range decay. However, it was eventually found that large, coherent structures are in fact dominant in a developed shear layer.¹⁸ These structures initially form as the growing disturbances in the shear layer "roll up" into vortices.

As a sinusoidal disturbance grows, it extends further in to the flow on each side. Portions extending into the higher-speed flow tend to be pushed forward, relative to the convective velocity carrying the disturbances in the flow. Similarly, portions extending into the low-speed flow are slowed, as indicated in Fig. 6.10



Figure 6.10: Effects of disturbances extending into the flows of a shear layer.

Eventually, the portions extending into the high-speed side catch up with the portions in the low-speed side, and the vortex line or sheet representing the boundary between the flows rolls up in a manner somewhat similar to a cresting wave. In this manner, a sinusoidal disturbance becomes a series of vortices, as shown in Fig. 6.11.



Figure 6.11: Vortex rows resulting from Kelvin-Helmholtz instability. (Regularized with forcing.) (Roberts,¹⁹ 1982)

The point at which this first roll-up occurs depends on the nature of the incoming boundary layer.¹⁸ If this boundary layer is laminar, then the vortices may be expected to have fully formed by the time the flow reaches the location $x/\delta_{\theta 0} \cong 1000$. With a turbulent boundary layer, the process is accelerated, with fully developed vortices at $x/\delta_{\theta 0} \cong 300$.

Once these large-scale vortical structures are established, all important parameters of the flow become independent of Reynolds number and become similar when nondimensionalized by a single velocity and length scale. This condition is known as selfpreservation.²⁰ Non-dimensionalization by δ_{θ} and U_C renders the layer self-similar once it has reached this point.¹⁸ Within this region, non-linear dynamics of vortical structures growing and interacting with each other comes to dominate the flow. Despite this transition, many aspects of Michalke's linear stability analysis still seem to apply.¹⁴ As part of the non-linear process, the amplification of particular wavelengths is seen to saturate. This saturation is evidence of a shift in the dominant fluid dynamics, as it is not something that can be predicted by linear stability analysis techniques. However, the frequency associated with the neutral stability of this state is predicted by Michalke's analysis, and can be seen in Fig. 6.6, where $\alpha = 1$. The associated Strouhal number for this state of neutral stability would be:

$$St_n = \frac{f_n \delta_\theta}{U_c} = 0.0796.$$
(6.19)

Values close to this have been repeatedly observed experimentally for the appropriate conditions and regions.¹⁴ This self-similar region can extend to values of $x/\delta_{\theta 0}$ in the thousands, often past the point where most experimentalists stop taking data.

6.2.3. Growth

In all stages of development, shear layers grow as they progress. However, the rate of this growth, $d\delta_{\theta}/dx$, can change as different regions are governed by different fluid dynamics and mechanisms of growth. The most commonly cited rule is for the growth of shear layers to be linear, with $d\delta_{\theta}/dx$ equal to a constant. If non-dimensional values such as the Strouhal number are to be constant, as it is in the self-similar region, then for a constant U_C the frequency must vary inversely, so that frequency is proportional to x^{-1} just as the thickness, δ_{θ} , is proportional to x.

This linear growth rule of thumb is based on the behavior observed in the selfsimilar region. In the initial stages of development, the growth rate is closer to being proportional to the square root of x and may not exhibit an invariant Strouhal number.¹¹ As noted, the amplification indicated in Michalke's linear analysis shifts towards neutral stability in this region, and so the mechanism for growth must come from some other aspect of the flow. Despite this shift, there remain parallels between the nature of the self-similar shear layer and its form during initial development. One important parallel can be seen in comparing Figs. 6.5 and 6.11. In the former, a simplified shear layer in its initial stages is modeled as a line of small, overlapping vortices. In the latter figure developed shear layers, having been regularized with forcing, exist as lines of relatively large and separate vortices. As with the vortex line model, if one of the large vortices in the self-similar shear layer becomes displaced vertically relative to its neighbor, then the vortices will tend to influence each other, each drawing the other closer in the streamwise direction while pushing the other further from their common center in the cross-stream direction.

148

This effect can be seen in the right half of the upper portion of Fig. 6.11. It is shown more thoroughly in Fig. 6.12, as well as what follows. The two interacting vortices begin to rotate around each other and combine into one vortex in a process called pairing. Individual vortices can and do grow with time, via turbulent entrainment of irrotational fluid around them. However, as two adjacent vortices grow in this manner, they come to exert greater influence on each other, which eventually triggers the pairing process of those two vortices.



Figure 6.12: Vortex pairing. (Winant, 1974)²¹

With each pairing, the length scales double. These length scales include both the shear layer thickness, and the streamwise spacing of the vortices. As the frequency is based on these structures being carried past a point ($f = U_C / \Lambda$) with each doubling in the structure sizes of δ_{θ} and Λ the frequency is halved. This fits with the necessary relationship to maintain a constant Strouhal number addressed above.

This pairing process indicates that the instantaneous thickness of a shear layer may be subject to sharp jumps as each pairing doubles the thickness of the layer at that point. However, these pairing events occur at random points unless some form of regularizing forcing is in effect, so the mean growth rate appears linear. In this respect, shear layers are unlike jets, in which the growth rate of δ_{θ} is truly a linear and relatively smooth process.²²

Growth rates tend to be dependent on the ratios of properties of the two flows, such as U_2/U_1 and ρ_2/ρ_1 . The ratio of the difference in velocities to the convective velocity is also commonly used, but can be rewritten in terms of the velocity ratio.

$$\frac{\Delta U}{U_c} = \frac{U_2 - U_1}{(U_2 + U_1)/2} = \frac{2\binom{U_2}{U_1} - 1}{\binom{U_2}{U_1} + 1}.$$
(6.20)

If the two streams are comprised of different gasses, then the ratio γ_2/γ_1 may also play a role, but this dissertation is focused primarily on conditions in which both streams are standard air. While the nature of the growth can be explored analytically, establishing exact values has lain in the realm of experimental work and numerical simulation. As such, there are various estimates for this growth rate from various sources based on experimental or numerical-model results. Figure 6.13 was complied by Ho and Huerre¹¹

using data from multiple sources. Their ratio of choice (*R*) is equivalent to 0.5 $\Delta U/U_C$, and this figure seems to indicate a growth rate of



$$\frac{d\delta_{\theta}}{dx} \approx 0.047R = 0.0235 \frac{\Delta U}{U_c}.$$
(6.21)

Figure 6.13: Momentum thickness growth rate for incompressible flows. (Ho and Huerre, 1984)¹¹

Some of the data points in Fig. 6.13 come from a paper²³ by Brown and Roshko. Latter work by Papamoschou and Roshko⁷ builds on this paper and indicates the growth rate for the visual thickness to be

$$\frac{\partial \delta_{vis}}{\partial x} = 0.17 \frac{\left[1 - \frac{U_2}{U_1}\right] \left[1 + \left(\frac{\rho_2}{\rho_1}\right)^{1/2}\right]}{1 + \frac{U_2}{U_1} \left(\frac{\rho_2}{\rho_1}\right)^{1/2}}.$$
(6.22)

As noted earlier, they also estimate that $\delta_{\omega} \cong 0.5 \ \delta_{vis}$. This equation for a growth rate promises to be more useful for aero-optic flows, as the optical effects of interest are generally due to compressibility effects, which means it is likely that the two streams will be of different densities as well as velocities.

6.3. Weakly-Compressible Modeling

The fluid dynamic theory and empirical laws in the first portion of this chapter include differences in velocity and density, but not temperature or pressure. Pressure differences between the two streams do not appear in these equations because if a shear layer propagates in a straight line, then the average static pressure (\overline{P}) on the two sides of the shear layer must be the same. If there were a pressure differential ($\overline{P_1} \neq \overline{P_2}$) then the flow would shift towards the side with the lower pressure until these pressures were equalized, or curve around the lower pressure region if equalization of the pressure was not possible. Temperatures may well be different in the two streams, but if pressure, density, and velocity or Mach number are known, then temperature can be calculated from those values.

At the time of the AEDC shear layer experiment described in chapter 5, it was also common practice to assume that fluctuations in static pressure $(P = \overline{P} + p')$ were negligible $(p' \cong 0)$ even in the case of shear layers with stream velocities and relative differences in velocity high enough to be considered compressible flows. This assumption grew out of work with compressible boundary layers. In the earliest applications, this was merely a convenient assumption for extracting temperature data from hot-wire measurements.^{24,25} Later work with supersonic boundary layers^{26,27} seemed to indicate that the strong Reynolds analogy (SRA),²⁸ which indicates that fluctuating pressures are insignificant compared to the overall pressure ($p'/\overline{P} \ll 1$), was an accurate description of such flows.^{29,30,31,32} However, this work dealt with boundary layers along a surface, not shear layers in free space which contain the large vortical "rollers" mentioned earlier in this chapter.

A relatively simple calculation can be performed to check this. Continued experimental work and simulations associated with shear layers indicate that the rolling structures within them have overall characteristics of a rigid rotator with angular velocity

$$V_{\theta}(r) = \omega r \,. \tag{6.23}$$

This approximation also fits with a linear approximation of the hyperbolic-tangent velocity profile of shear layers, which is associated with the definition of vorticity thickness, as shown in Fig. 6.7. As part of this assumption, Eq.6.23 should describe the velocity field of the structure out to some radius R, at which point the vortical velocities match the free-stream velocities of the two streams making up the shear layer. (U_1 , U_2) Assuming the vortex is convicting at a the average convective speed of

$$U_c = \frac{U_1 + U_2}{2}, \qquad (6.24)$$

and that $U_1 > U_2$ then in a frame of reference moving with this roller, the angular velocity at radius *R* in the direction of stream 1 will be

$$V_{\theta}(R) = \omega R = U_1 - U_C = U_1 - \frac{U_1 + U_2}{2} = \frac{U_1 - U_2}{2}.$$
 (6.25)

On the side towards stream 2 this will be

$$V_{\theta}(R) = \omega R = U_{c} - U_{2} = \frac{U_{1} + U_{2}}{2} - U_{2} = \frac{U_{1} - U_{2}}{2}, \qquad (6.26)$$

which indicates that

$$\omega = \frac{\Delta U}{2R}.$$
(6.27)

The radial pressure gradient for a vortex of this sort is³³

$$\frac{dp}{dr} = \rho r \omega^2 = \rho r \frac{\Delta U^2}{4R^2}.$$
(6.28)

As was shown in the experimental results of chapter 5, the vortical structures in the shear layer under study were a significant source of optical aberration, which indicates that density was not a constant within those structures. However, for the sake of simplicity, density will be treated as a constant while performing this simplified estimate. Integrating this from the center outward produces the following expression for pressure within the roller

$$p(r) = p_{\min} + \rho r^2 \frac{\Delta U^2}{8R^2},$$
 (6.29)

where p_{min} is the pressure at the center of the roller. From this, the pressure differential from the center to the edge should be

$$\Delta p = p(R) - p(0) = p_{\min} + \rho R^2 \frac{\Delta U^2}{8R^2} - p_{\min} = \frac{\rho \Delta U^2}{8}.$$
 (6.30)

It is interesting to note that the size of the structure, R, drops out of this equation, leaving only the average density and velocity differential as factors.

The conditions of the AEDC experiment were such that $\Delta U \cong 260 \text{ m/s} - 35 \text{ m/s} = 225 \text{ m/s}$ and static pressure was approximately 0.6 Atm.³⁴ This is in line with standard compressible-flow tables which indicate a Mach 0.8 flow would be expected to have a

static pressure that is 65% that of the stagnation pressure. Going back to those tables, the expected density associated with these Mach and pressure conditions would be 70-75% of the stagnation value, which would presumably be standard atmospheric density. 70% of standard atmospheric density would be about 0.90 kg/m³. Based on these estimates, one of the rolling structures in this shear layer should have a pressure differential from center to edge of

$$\Delta p = \frac{0.90 kg / m^3 (225m/s)^2}{8} = 5700 \frac{kg}{ms^2} = 5700 Pa \,. \tag{6.31}$$

This indicates a ratio of pressure fluctuations to average pressure of

$$\frac{\Delta p}{\overline{P}} \cong \frac{5700 Pa}{101325 Pa \cdot 0.6} = 0.094.$$
(6.32)

In other words, the pressure fluctuations associated with the vortical structures in this shear layer may be almost 10% of the overall static pressure in the flow. This is a very crude estimate of this pressure difference, but it is sufficient to indicate that neglecting p' may not be a justifiable assumption in dealing with these flows.

6.3.1. Discrete Vortex Modeling

These studies led to the development of the Weakly Compressible Model (WCM) for computational simulation. The basis for this approach to simulation is the Discrete Vortex Method (DVM), which models a shear layer as a line of point vortices, as shown in Fig. 6.5 on page 136 and Fig. 6.14 below. The original code for these simulations was produced by Hugo³⁵ for the modeling of low-speed planar jets, and then modified by Fitzgerald³⁶ for use with compressible flows.



Figure 6.14: Representation of a two-dimensional shear layer by discrete vortices.

A point along the x-axis of this simulation is designated as the end of the modeled splitter-plate. Upstream of this point, the discrete vortices move with the convective velocity imposed upon the model, but are not free to deviate in their spacing or position in the *y* direction. Downstream of this point, the vortices are allowed to move in accordance with the influence of other vortices in the model along with the convection velocity influence. The original simulation did not extend downstream infinitely, thereby producing odd behavior. The end of the vortex sheet tended to roll up faster than would be the case for a sheet extending out to infinity, and if more vortices accumulated on one side of the centerline than the other, it caused the vortex sheet to veer off in that direction. This has been corrected by extending the simulation to $x = \pm \infty$, imposing an analytic solution of the velocity field for a shear layer on points outside the region of computational simulation and providing a correction for the influences of these regions.

Computational round-off error in releasing new vortices from the splitter plate and effects from correction for the semi-infinite domain produce perturbations in the computed flow sufficient to induce the kind of amplification and roll-up observed in physical flows. If the distance between two consecutive vortices increases past some critical distance, r_{crit} , the strength of those two vortices is decreased by one-third, and a new vortex is inserted between the two original vortices, such that the total strength for all three vortices is equal to that of the two original vortices. This conserves overall circulation and spatial resolution along the simulated vortex sheet.

The velocity (in polar coordinates) induced by each vortex in this model is

$$U_{\theta}(r) = \frac{\Gamma}{2\pi r} \,. \tag{6.33}$$

where Γ is the strength of the vortex. To avoid difficulties with singularities at the center of these vortices, the core of each vortex where *r* is less than some *r*_{core}, is simulated by solid body rotation of the form

$$U_{\theta}(r) = \frac{\Gamma}{2\pi r_{core}^2} r.$$
(6.34)

This model is inherently inviscid, but Eqs. 6.33 and 6.34 together produce a piecewise linear approximation of the velocity field for a decaying vortex in a viscous fluid, which is

$$U_{\theta}(r,t) = \frac{\Gamma}{2\pi r} \left[1 - \exp\left(\frac{-r^2}{4\upsilon t}\right) \right].$$
(6.35)

The time-varying component of Eq. 6.35 is caused by momentum transfer from inner regions of the vortex to outer regions, which in turn is due to viscosity. This temporal behavior can be approximated in the inviscid model by increasing r_{core} with time for each vortex. This "diffusion rate" can be characterized by the diffusion of vorticity between two parallel laminar streams, which can be found in a similarity solution by Lock.³⁷ Lock's solution is specifically intended for high-speed flows, and involves the similarity variable

$$\eta = y_{\sqrt{\frac{U_1}{v_x}}},\tag{6.36}$$

in which U_1 is the velocity of the higher-speed stream. The position in the *x* direction can then be substituted for a temporal relation via the convection velocity:

$$\eta = y \sqrt{\frac{U_1}{v U_C t}} = y \sqrt{\frac{2U_1}{(U_1 + U_2)vt}}.$$
(6.37)

Solving for *y* and using the velocity ratio, $r_u = U_2/U_1$,

$$y = \eta \sqrt{\frac{(1+r_u)vt}{2}}$$
 (6.38)

From this, the core of a discrete vortex in this model should have the same growth rate as seen in this similarity solution for a laminar flow. Using the specific cases presented in Lock's work and assuming a linear variation with r_u , r_{core} for each vortex should be:

$$y = \sqrt{\frac{(1+r_u)(6-4r_u)^2 vt}{2}} + r_{core,t=0}.$$
 (6.39)

where $r_{core,t=0}$ is the initial radius of that core when the vortex is introduced into the model.

As noted in section 6.2.2, the initial development of the shear layer is dependent on the initial thickness of the boundary layers feeding into the shear layer. Matching the DVM model to a physical example then requires setting the upper and lower velocities, the initial boundary layer thickness, and viscosity of the fluid. Other values in the model, such as Γ , r_{core} , and vortex spacing are determined by these conditions.

This model, as described so far, is based on dynamics of incompressible flow. The original version was used by Hugo to model the mixing of two incompressible fluids with different indexes of refraction. However, it has been found that even for transonic
speeds, such as those produced in the ART facility, compressibility effects have very little effect on the growth and development of a shear layer.^{38,39,40,41} Thus, the velocity fields and position of the vortex sheet can be modeled in an incompressible manner while still producing a shear layer with relative velocities, growth rates, and length scales that are roughly in accordance with those seen in the AEDC facility, and with predictions of growth rate and velocity profiles for shear layers based in the empirical work of others.⁴²

6.3.2. Optical Modeling

The code used in this model was initially developed for a case of two-index mixing, in which the fluid in the high-speed portion of the shear layer had a different index of refraction than the fluid in the low-speed flow. The vortex sheet represents the boundary between the two flows and the fluids comprising those flows. Therefore, keeping track of the location of this sheet defines which fluid occupies each point in the flow field. As pointed out in section 2.3.1, effective length of an optical path through a flow can be found by integrating the index along that path:

$$OPL = \int_{s_1}^{s_2} n(s) ds \,. \tag{6.40}$$

If the initial optical path is perpendicular to the shear layer, and the deflection angles acquired in the flow field are relatively small, then changes in *x*-position of this path will be negligible, and the path can be approximated as a straight line parallel to the initial vector of the beam. If this vector is aligned with the *y*-axis, then:

$$OPL(x) \cong \int_{y_1}^{y_2} n(x, y) dy$$
. (6.41)

In the two-index case, n(x,y) is determined by which side of the vortex sheet the point (x,y) lies. This becomes slightly more complicated as the shear layer rolls up and the vortex sheet coils around itself, but the principle remains unchanged. In the ART facility at AEDC, the fluid in each portion of the flow is air, with the same total temperature in the high and low speed portions of the flow. Thus, the mechanisms producing the optical distortions observed must be different than those in the incompressible two-index case.

As alluded to earlier, it was originally thought that the SRA assumption of negligible pressure fluctuations was essentially correct. Therefore, the optical effects observed in the AEDC experiments were originally attributed to adiabatic heating.⁴³ This was addressed briefly towards the end of chapter 5, but will be repeated here. The adiabatic heating model assumes that the SRA applies and that variations in index of refraction are due primarily to temperature fluctuations and that local temperature is a function of local velocity or Mach number in the form

$$T = T_0 - \frac{|V|^2}{2C_p} = \frac{T_0}{1 + \frac{\gamma - 1}{2}M^2}.$$
 (6.42)

From this temperature, the Gladstone-Dale relationship, and the ideal gas law, the local value of n can be found as

$$n = 1 + \rho K_{GD} = 1 + \frac{p K_{GD}}{RT}.$$
(6.43)

Determining T and n in this form of the model then requires determining the local velocity at each point along the optical path from the contributions of all vortices in the

vortex sheet, which in turn increases the computational and storage requirements of the model.

Implementation of this adiabatic-heating model with conditions matching those of the AEDC experiments failed to produce optical aberrations on the scale observed in the experiment, as shown in Fig. 5.8 on page 130, indicating that the model was incomplete. Examination of the velocity fields produced by this model led to questions about whether the SRA assumptions were correct in this case, and initial calculations of the sort performed in section 6.3 led to the conclusion that the SRA assumption was likely to be incorrect. Fig. 6.15 shows one such velocity field in a frame of reference moving with the convection velocity. This flow field was taken from a position in the model corresponding to station 2 in the ART facility. As alluded to earlier, it appears to have a form roughly similar to that of a rigid rotator as was used in the estimate of Δp in section 6.3, or at least it has that profile in the y-direction through the center of this structure. In the x-direction, this structure interacts with similar structures before and after, producing what are effectively stagnation points relative to the convection velocity.



Figure 6.15: Velocity field produced by the DVM model for AEDC conditions, centered at a position of x=48.3 cm.(Fitzgerald)⁴²

Pressure differences can be computed from the velocity field via the unsteady Euler equations:

$$\frac{\partial p}{\partial x} = -\rho \left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} \right). \tag{6.44}$$

and

$$\frac{\partial p}{\partial y} = -\rho \left(\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} \right).$$
(6.45)
162

Temperature is a function of pressure as well as velocity, and so the temperature field given by adiabatic heating must then be adjusted for pressure changes. This is done via the Hilsh approximation,⁴⁴

$$\frac{T(x,y)}{T_{ad}} = \left[\frac{p(x,y)}{p_{\infty}}\right]^{\frac{(\gamma-1)}{\gamma}}.$$
(6.46)

Inclusion of these pressure differentials into the model produces what is called the Weakly Compressible Model (WCM). It is considered "weakly" compressible as the velocity field is produced by a model without compressibility effects, but the optical results are based off of compressibility effects of that velocity field. Actual implementation of the model uses the adiabatic-heating model as an initial guess for the density field, and then refines the density, pressure, and temperature distribution iteratively, using the Euler equations and Hilsh approximation.

6.3.3. Model Results

The rough estimate of pressure fluctuations performed near the beginning of section 6.3 indicated that the magnitude of those fluctuations was a function only of the difference in velocity of the two streams making up the shear layer. However, as was pointed out regarding the velocity field in Fig. 6.15, these rollers are not precisely rigid rotators, in part due to the interaction with similar structures preceding and following them in the flow. Figures 6.16 and 6.17 show the simulated vortex line and associated pressure field for two different times at the same location in the simulation, corresponding to station 2 in the AEDC facility.



Figure 6.16: (left) Perturbed and rolled-up vortex line from WCM simulation. (right) Pressure field associated with the simulated conditions to the left.⁴²

As was addressed in section 6.2.3, shear layer growth is associated with vortex pairing in which the observed vortical structures combine into larger structures. Figure 6.16 shows the vortex line has rolled up into two vortices that are close together and about to combine. In the associated pressure field, the low pressure at the center of each vortex does not quite reach the level of 5.7 kPa below the average that was indicated in the earlier rough estimate. However, the points between these rollers, where they interact with each other, form points of semi-stagnation at which the pressure is actually higher than that found in the free stream to either side of the shear layer. The difference in pressure between these highs and lows is on the order of that predicted in the simple rigid rotator estimate.



Figure 6.17: (left) Perturbed and rolled-up vortex line from WCM simulation showing one large roller. (right) Pressure field associated with the simulated conditions to the left.⁴²

Figure 6.17 shows simulated results for the same positions as was shown in Fig. 6.16, but at a different time. Unless some form of forcing is applied to regularize the behavior of the shear layer, vortex pairing can and will occur at random intervals. In this case, the structures of the size shown in Fig. 6.16 have paired before reaching this location, producing one large vortex. This is also the simulated case shown in Fig. 6.15, but without the velocity indicators in that figure. In this case, the pressure drop at the center of the structure is almost twice that indicated in the rough estimate, while including the high pressure of the stagnation points between raises this to almost three

times the result of that estimate, with net Δp equal to about 25% of the mean static pressure in the flow.

As previously noted, the adiabatic heating model produced wavefronts that were significantly less severe than those observed experimentally. Fig. 6.18, previously presented in chapter 5, shows the results of this simulation for the conditions of the ART facility at station 2, alongside the measured results for this position. As can be seen, the aberrations as predicted by this model are relatively benign compared to the actual results obtained in the ART facility.



Figure 6.18: (a) Wavefronts generated by computational simulation and an adiabatic heating model. (b) AEDC results for flow conditions similar to those in the computational simulation. (Fitzgerald)⁴²

In light of the simulation results for pressure shown in Figs. 6.16 and 6.17, the inability of constant-pressure models based on adiabatic heating to reproduce the optical distortions observed in the AEDC ART facility is to be expected. Figure 6.19 shows wavefront results from the WCM for the same position and aperture size as was used in Figs. 6.15 through 6.18. As can be seen, the wavefronts generated by the simulation do

seem to capture the overall magnitude and length scales of the measured distortions shown in Fig. 6.18 (b), but the simulated wavefronts are smooth, lacking the roughness and fine detail seen in the experimental results.



Figure 6.19: WCM wavefronts, corresponding to AEDC ART station 2. $(Fitzgerald)^{42}$

Experimental results for ART station 1, shown in Fig. 6.20 also show this fine structure, which can not be accounted for by the WCM as the shear layer at this point is still in the early stages of perturbation and has not yet rolled up into the vortical structures that are the primary source of optical distortion in this model. It is believed that these distortions in the early stage of the shear layer are due to structures produced in the boundary layer on the splitter plate feeding into the shear layer.^{45,46}



Figure 6.20: Reconstructed wavefronts from measurements at station 1 of the AEDC ART facility. $(Fitzgerald)^{42}$

Figure. 6.21 shows the results of superimposing the experimental results of Fig. 6.20 onto the simulation results of Fig. 6.19. The combined wavefronts do look remarkably like the experimental results for station 2, shown in Fig. 6.18 (b), although it is not a perfect match as the small-scale detail overlying the dominant larger-scale aberrations in Fig. 6.18 seems to be of greater relative amplitude. Some additional small-scale distortions in the experimental results may be due to a growing boundary layer along the upper wall of the ART facility.



Figure 6.21: Superimposing station 1 wavefronts onto WCM simulation for station 2. $(Fitzgerald)^{42}$

In more concrete numbers, table 6.1 shows the prediction of the WCM for compressible model for maximum peak-to-peak OPD as compared to the experimental results from the AEDC experiments as well as predictions of models based on the previously prevailing assumptions presented in chapter 5. The WCM prediction is within 2% of the experimental result, while the predictions of the other models are off by a factor of four or more.

	2-index	Modified 2-	Adiabatic	WCM	AEDC
	mixing	index	Heating		Experimental
					Results
OPD peak-	0.339	0.241	0.310	1.323	1.340
to-peak					
(µm)					

Table 6.1: Comparison of maximum peak-to-peak *OPD* for predictive models and experimental results.⁴²

6.4. Conclusions

This review of shear-layer dynamics and theory led to a new computational model that seems to predict the overall form and magnitude of the optical effects observed in the experiment at AEDC. However, this model also predicts significant localized changes in pressure as the primary source of these optical distortions, which is contrary to the school of thought that was commonly accepted at the time of the AEDC experiment and before the development of this model. Additionally, the flow depicted in this model, with pressure driven changes in the index of refraction and somewhat coherent and organized structures in the flow has little similarity to the atmospheric model upon which the commonly-used parameters for characterizing fluid-optic effects are based. Such a radical shift from accepted theory demands empirical confirmation of the model, which will be the subject of the next chapter.

¹ Demos Kryazis, personal communication.

² Michalke, A., "On the inviscid instability of the hyperbolic-tangent velocity profile", Journal of Fluid Mechanics, Vol 19, 1964, pp. 543-556.

³ Helmholtz, H. von, "Uber discontnuirliche Flussigkeitsbewegungen", Monats. Konigl. Preuss. Akad. Wiss, Berlin, Vol. 23, 1868, pp. 858-872.

⁴ Helmholtz, H. von, translated as "On discontinuous movements of fluids", Phil. Mag., Vol 36, No. 4, 1986, pp. 337-346.

⁵ Lord Kelvin, "Hydrokinetic solutions and observations", Phil. Mag, Vol. 42, No. 4, 1871, pp. 362-377.

⁶ Batchelor, G. K., An Introduction to Fluid Dynamics, 1967, Cambridge University Press.

⁷ Papamoschou, D., and Roshko, A., "The Compressible Turbulent Shear Layer: An Experimental Study," *Journal of fluid Mechanics*, Vol. 197, 1988, pp. 453-477.

⁸ Von Karman, T., "On Laminar and Turbulent Friction", Z. Angew. Math. Mech, Vol 1, 1921, pp. 235-236.

⁹ Qibing Li, Haixin Chen, and Song Fua, "Large-scale vortices in high-speed mixing layers," Physics of Fluids, Vol. 15, No. 10, 2003, pp 3240-3243.

¹⁰ Mokewitz, P. A., and Heuerre, P., "The influence of the velocity ration on the spatial instability of mixing layers", Phys. Fluids., Vol. 25, 1982, pp. 1137-1143.

¹¹ Ho, C. M., and Huerre, P. "Perturbed free shear layers", Annual Review of Fluid Mechanics, Vol. 16, 1984, pp. 365-424.

- ¹² Sato, H., "The stability and transition of a two-dimensional jet," *Journal of Fluid Mechanics*, Vol. 7, 1960. pp. 53-80.
- ¹³ Fremuth, P., "On transition in a separated laminar boundary layer", *Journal of Fluid Mechancis*, Vol. 25, 1966, pp. 683-704.
- ¹⁴ Ho, C. M., and Hauang, L. S., "Subharmonics and vortex merging in mixing layers", *Journal of Fluid Mechanics*, Vol. 119, 182, pp. 443.473.
- ¹⁵ Hussain, A. K. M. F., and Zaman, K. B. M. Q., "An experimental study of organized motions in the turbulent phale mixing layer", *Journal of Fluid Mechanics*, Vol. 159, 1985, pp. 85-104.
- ¹⁶ Miskand, R. W., "Experiments on the nonlinear stages of free-shear-layer transition", *Journal of Fluid Mechanics*, Vol. 56, 1972, pp. 695-719.
- ¹⁷ Morris, S.C. and Foss, J.F., "Turbulent boundary layer to single-stream shear layer: the transition region", *Journal of Fluid Mechanics*, Vol. 494, 2003, pp. 187-221.
- ¹⁸ Dimotakis, P. E., and Brown, G. L., "The mixing layer at high Reynolds number: large-structure dynamics and entrainment", *Journal of Fluid Mechanics*, Vol. 78, 1976, pp. 535-560.
- ¹⁹ Roberts, F. A., Dimotakis, P. E., and Roshko, A., "Kelvin-Helmholz instability of superposed streams", In *Album of Fluid Motion*, ed. M. Van Dyke, p. 85, Stanford, California, Parabolic, 1982.
- ²⁰ Liepmann, H.W., and Laufer, J., "Invenstigation of free turbulent mixing", NACA Technical Note, No. 1257, 1947.
- ²¹ Winant, C.D., and Browand, F.K., "Vortex pairing: the mechanism of turbulent mixing-layer growth at moderate Reynolds number", Journal of Fluid Mechanics, Vol 63, 1974, pp. 237-255.
- ²² Thomas, F.O., "Structure of mixing layers and jets", Appl. Mech. Rev., Vol. 44. 1991, pp. 119-153.
- ²³ Brown, G.L. and Roshko, A. "On density effects and large structures in turbulent mixing layers", J. Fluid Mech. Vol. 64, pp. 775-781.
- ²⁴ Kistler, A. L. 1959 Fluctuation measurements in a supersonic turbulent boundary layer. Phys. Fluids 2, 290.
- ²⁵ Kovasnay, L. S. G. 1950 Hot wire anemometer in supersonic flow. J. Aeronaut. Sci. 17, 565–572.
- ²⁶ Kistler, A. L. & Chen, W. S. 1963 The fluctuating pressure field in a supersonic turbulent boundary layer. J. Fluid Mech. 16, 41–64.
- ²⁷ Gross, K. P., McKenzie, R. L. & Logan, P. 1987 Measurements of temperature, density, pressure, and their fluctuations in supersonic turbulence using laser-induced fluorescence. Exps. Fluids 5, 372– 380.
- ²⁸ Morkovin, M. V. 1962 Effects of compressibility on turbulent flows. In M´ecanique de la Turbulence (ed. A. Favre), pp. 367–380. Paris: CNRS.
- ²⁹ Lutz, S. A. 1989 Modeling of density fluctuations in supersonic turbulent boundary layers. AIAA J. 27, 822–823.
- ³⁰ Smith, D. R. & Smits, A. J. 1993 Simultaneous measurement of velocity and temperature fluctuations in the boundary layer of a supersonic flow. Exp. Thermal Fluid Sci. 7, 221–229.
- ³¹ Lele, S. K. 1994 Compressibility effects on turbulence. In Annu. Rev. Fluid Mech. 26, 211–254.
- ³² Smits, A. J. & Dussauge, J.-P. 1996 Turbulent Shear Layers in Supersonic Flow. American Institute of Physics.

- ³³ Vanyo, J.P., "Rotating Fluids in Engineering and Science," 2001, Courier Dover Publications, pp. 170-172.
- ³⁴ Havener, G. & Heltsley, F. 1994 Design aspects and preliminary holographic-PIV measurements for a subsonic free shear layer flow channel. AIAA Paper 94-2550.
- ³⁵ Hugo, R.J., "Quantifying the Spatio-Temporal Effects of Optically-Active Turbulent Flowfields on a Coherent Optical Wave", PhD thesis, University of Notre Dame, Notre Dame, Indiana, Apr 1995.
- ³⁶ Fitzgerald, E.J., and Jumper, E.J., "The optical distortion mechanism in a nearly incompressible free shear layer," J. Fluid Mech. (2004), vol. 512, pp. 153–189.
- ³⁷ Lock, R. C., "The velocity distribution in the laminar boundary layer between parallel streams." *Quarterly Journal of Mechanics and Applied Math*, IV: 42-57, 1951.
- ³⁸ Sandham, N.D. and Reynolds, W.C., "Compressible mixing layer: Linear theory and direct simulation", AIAA Journal, 28(4): 618-624.
- ³⁹ Sandham, N.D. and Reynolds, W.C, "Three-dimensional simulations of large eddies in the compressible mixing layer", Journal of Fluid Mechanics, 224:133-158, 1991.
- ⁴⁰ Hall, J.L, Dimotakis, P.E., and Roseman, H., "Experiments in nonreacting compressible shear layers", AIAA Journal, 31(12):2247-2254, 1993.
- ⁴¹ Leep, L.J., Dutton, J.C., and Burr, R.F., "Three-dimensional simulations of compressible mixing layers: Visualizations and statistical analysis", AIAA Journal, 31(11):2039-2046, 1993.
- ⁴² Fitzgerald, E.J., and Jumper, E.J., "The optical distortion mechanism in a nearly incompressible free shear layer," J. Fluid Mech. (2004), vol. 512, pp. 153–189.
- ⁴³ Fitzgerald, E. J. & Jumper, E. J., "Shear layer optical distortions due to compressibility vs. passive scalars," AIAA Paper 98-2834.
- ⁴⁴ Hilsh, R., "The Use of the Expansion of Gases in a Centrifugal Field as Cooling Process," The Review of Scientific Instruments, 18(2), 1946, pp. 108-113.
- ⁴⁵ Fitzgerald, E. J. and Jumper, E. J., "Aperture effects on the aerooptical distortions produced by a compressible shear layer, " (2002), AIAA J. 40, 267–275.
- ⁴⁶ Gordeyev, S., Jumper, E. J., Ng, T. T. & Cain, A. B., "Aero-optical characteristics of compressible, subsonic turbulent boundary layers," AIAA Paper 2003-3606.

CHAPTER 7:

OPTICAL CHARACTERIZATION OF A SHEAR LAYER

7.1. Overview

The experimental, theoretical, and computational work performed by others and presented in chapters 5 and 6 indicated that the existing theory and commonly used approximations for shear layers were not sufficient to explain the optical distortions observed at AEDC. The weakly compressible model (WCM), presented in chapter 6, produced simulated results more in line with those observed experimentally. However, any hypothesis must be verified by observation of physical phenomena in nature or the laboratory, and a good hypothesis will make predictions that can be checked in this manner. The WCM predicts localized regions of significant low pressure. Furthermore, it predicts that the structures containing these pressure wells will have a size and spacing determined by the point at which they are observed and the growth rate of the shear layer.

Additionally, verification of the WCM is a step toward the goal of characterizing and dealing with the optical distortions produced by such flows. In pursuit of this goal, an optical survey of a transonic shear layer was performed, leading to further insights as to the nature of the distortions produced by such flows, and eventually to potentially useful scaling laws and guidelines for corrective systems presented in chapter 8. 7.2. Notre Dame Weakly-Compressible Shear Layer Facility

Use of the AEDC facility was a unique opportunity. Gaining access to this facility in the future was not something that could be scheduled for or relied upon. Therefore, in order to study aspects of flows of the sort produced in the AEDC facility in more detail, a Weakly-Compressible Shear Layer (WCSL) facility was constructed at Notre Dame's Hessert Laboratory for Aerospace Research.

7.2.1. Facility Design

A schematic of the Notre Dame WCSL facility, in its earliest form, is shown in Fig. 7.1. It consisted of an inlet nozzle and test section mated with one of Notre Dame's three transonic in-draft, wind-tunnel diffusers. The diffuser section was attached to a large, gated plenum. The plenum was, in turn, connected to three Allis Chalmer 3,310 cubic feet per minute (CFM) vacuum pumps.^{*} Depending on the gate-valve arrangements, each of these pumps could be used to power separate diffusers, or they could be used in combination to power a single diffuser.

^{*} These pumps have since been replaced by two Dekker DEH5000P vacuum pumps; however, the Allis Chalmer pumps were used for the experiments and research presented in this dissertation.



Figure 7.1: Schematic of the Notre Dame Weakly-Compressible Shear Layer Facility. (First configuration.)

Being an in-draft tunnel, the feeding source was air from the room with total pressure and temperature determined by atmospheric conditions at the time of the experiment. The test section was fed from a 104-to-1 inlet nozzle directly from room total pressure on the high-speed side. On the low-speed side, room-total-pressure air was first passed through a settling tank with a ball valve and butterfly valve on the inlet. Passing through these valves at high speed produced a loss in total pressure. The purpose of the settling tank was to slow down and quiet the flow. Air drawn from the settling tank was then passed through an expansion section that was part of and partitions from the high-speed flow in the same physical nozzle. (See Figs. 7.1 and 7.2) The two flows met at the end of a splitter plate, located at the beginning of the test section.

The WCSL facility was constructed not only to have such a facility readily at hand, but to overcome some limitations of the ART facility at AEDC. Figuring prominently in these limitations were the limited optical access points in the ART. For this reason the test section, shown in Fig. 7.2, was constructed entirely of ³/₄ inch clear

Plexiglas. This material proved to be of sufficient optical quality for taking Schlieren images and Malley probe measurements, while the upper and lower walls of the test section could be fitted with optical flats when necessary or desirable. The vertical dimensions and contraction ratio of this test section were sized to match those of the ART. However, the horizontal width of this version of the WCSL facility was only 3.81 cm, compared to a width of about 33 cm in the ART facility at AEDC.



Figure 7.2: Schematic of WCSL test section.

The Plexiglas construction of this test section not only provided optical access to the flow, alviet of poor optical quality, but made it quite easy to drill static pressure taps at desired locations and to create ports for inserting other sensors and probes into the flow. Fig. 7.2 shows a line of such pressure taps running the length of the test section. The vertical bar at the midpoint in this schematic represents a sliding panel in the side wall that could be adjusted up or down and then clamped into place. This panel incorporated a set of pressure taps arranged in a vertical line, and was located 0.5 m downstream of the origin point of the shear layer. This placement was chosen to correspond to the position at measurement station 2 in the ART facility, which was 0.483 m from the end of the splitter plate in that facility.

7.2.2. Modifying the Facility

Over the course of performing experiments with the WCSL facility, it was modified to remove or reduce disturbances that might influence the development of the shear layer. According to the stability analysis done in chapter 6 and experimental work by others, shear layers are sensitive to disturbances introduced at the point of initial contact between the two flows.^{1,2,3,4,5,6} One of these modifications was a throat section added between the test section and the diffuser. The purpose of this section was to create a point of contraction at which the flow would reach Mach 1.0, which in turn would prevent disturbances from the pumps or exhaust system from propagating upstream into the test section. A photo of this throat section is shown in Fig. 7.3. The shape of the contraction and expansion of this throat was determined with 4th-order splines to match inlet slope, outlet slope, and desired minimum area with a smooth curve. While most of this section was made from Plexiglas, The lower block providing the shape of this throat was cut from a block of wood, made from wooden sheets laminated together, with the upper surface sanded and given a coating of wax for smoothness.



Figure 7.3: Sonic throat between test section and diffuser. Flow is from left to right.

The settling tank was modified by adding an internal baffle, constructed from wooden two-by-fours and quarter-inch plywood as shown in the sketch of Fig.7.4. The baffle and other interior surfaces of the tank were covered with sound-dampening foam. The purpose of this modification was to prevent turbulence and acoustic effects produced as air passed through the valves into the tank, from propagating downstream and influencing the shear layer at its origin. The baffles and foam dampen the sound, and force the air to take a longer circuitous path through the tank.



Figure 7.4: Settling tank interior, with baffles and sound-dampening foam.

Afterwards, the valves themselves were replaced with "quiet valves" made up of bundles of tubes. A drawing of one such device is shown in Fig. 7.5. The purpose of these devices was to produce the same drop in total pressure that was produced by the ball and butterfly valves in the original configuration. By using the wall friction, entrance, and exit effects of a relatively long tube, this pressure drop could be achieved with less noise than was produced with the valves. The quiet valve segment in Fig. 7.5 was constructed from seven segments of half-inch PVC pipe, surrounded by a sheath of 4-inch PVC pipe and capped with a circular block of wood at each end. These end caps were shaped with an outer diameter to match that of the sheath, holes to match the inner diameter of the half-inch PVC tubes, and recessed areas to receive those tubes and hold them in place. Wood glue and silicone calk provided sufficient bonding and sealing to hold the device together and prevent leakage.



Figure 7.5: Schematic of a "quiet valve" segment.

In measurements of flow conditions with the sonic throat in place an atmospheric pressure of about 98 kPa was associated with a static pressure within the test section of about 65 kPa within the test section and a low-speed flow of about 0.08 M. With these values and an assumption that the total temperature in both streams was 296 K, that of the free atmosphere, density and total pressure for the low-speed flow can be calculated from standard relationships for compressible flow. From these, a mass-flow rate in the low-speed portion of the WCSL flow of approximately 0.20 kg/s was indicated, as well as a drop in total pressure from the atmosphere to the settling tank interior of about 33 kPa.

In calculating the loss of total pressure through a tube, there are two factors: the friction from the tube walls and the effects of the entrance and exit. Friction losses are found via the Darcy-Weisbach equation,⁷

$$\Delta p = f \frac{L}{D} \rho \frac{V^2}{2}, \qquad (7.1)$$

in which D and L are the inner diameter and length of tube respectively. In this equation, f represents a friction factor, which may be found by from Moody diagrams or iteratively from the Colebrook-White equation,⁸

$$\frac{1}{\sqrt{f}} = -2\log_{10}\left(\frac{\varepsilon}{3.7D} + \frac{2.51}{\operatorname{Re}\sqrt{f}}\right),\tag{7.2}$$

in which ε is the roughness of the interior pipe surface and the Reynolds number is based on the diameter of the tube. The total pressure loss associated with an inlet with sharp corners is $\Delta p = 0.25\rho V^2$ and the exit loss is $\Delta p = 0.5\rho V^2$, regardless of the shape of the exit.

Despite the name, half-inch schedule-40 PVC pipe actually has an internal diameter of about 0.62 inches. Seven such pipes have a net intake area of about 13.7 cm². If the air in the room has a density of 1.16 kg/m^3 , then a velocity of about 125 m/s is needed to move the desired 0.20 kg/s through that area. Estimates for roughness of PVC vary, but a value of 5×10^{-5} mm was used in these calculations. From these values and the equations of the preceding paragraph, a quiet valve of the form shown in Fig. 7.5 with a length of 30.5 cm (1 ft.) should produce a drop in total pressure of about 16.5 kPa, which is half of the desired 33 kPa pressure drop produced with a ball valve in the earlier configuration.

This was a first-order estimate to guide the design; it neglects compressibility effects and the different material of the end caps. However, when a pair of one-foot quiet-valve sections were constructed, attached end-to-end with a standard 4-inch PVC connector, and similarly attached to the settling tank in place of the ball valve, the resulting pressure drop was close to the desired value, and could be adjusted to the desired value by rotating one device relative to the other, moving the outer ring of tubes in one segment out of alignment with those in the other.

The Plexiglas used to construct the test section of the facility is a fairly poor material from an optical standpoint. It is fine for observing the overall structure of the flow. Even in making optical measurements involving the deflection of narrow beams, this material has proven adequate. However, the distortions this Plexiglas will induce on a wavefront of larger dimensions are of an inconvenient and unacceptable degree in making measurements where accuracy to fractions of microns is desired. The original test section was replaced with one in which portions of the upper and lower walls could be removed and replaced with glass windows of higher optical quality. Photos of this test section and one of the removable portions are shown in Fig. 7.6.



Figure 7.6: Test section with removable Plexiglas segments.

A sketch of the modified tunnel, with sonic throat and quiet valves in place, is shown in Fig. 7.7. This is the form of the facility used in performing the optical measurements presented in section 7.4.



Figure 7.7: Schematic of the Notre Dame Weakly-Compressible Shear Layer Facility. (Modified)

Many of these modifications to the tunnel that existed at the outset of my work were directly incorporated into the design of a second-generation WCSL facility, with a larger cross-section. This wider version of the facility was not used in the course of the research described in this dissertation, but has been used extensively by other researchers at Notre Dame, who are carrying on with research in this field. As the construction and design of the second facility was an outgrowth of the work described in this dissertation, details on the second-generation WCSL and its design may be found in Appendix A.

7.2.3. Flow Conditions

The study of shear layers described in this dissertation was carried out in two stages, with somewhat different flow conditions for each stage. The first stage of experiments verified the weakly-compressible model (WCM), described in section 6.3, by flow visualization and by measuring pressure fluctuations. Results of these experiments are described in section 7.3 below. This verification was carried out before the modifications described in the previous section were made to the WCSL facility. The original configuration of the facility was adequate for this set of experiments, as the presence or absence of the large pressure wells indicated by the WCM was unlikely to be affected by small disturbances.

Measurements of static pressure from the ports in the side-walls of the test section, compared to the stagnation pressure measured by a probe in the low-speed flow and in the room from which the high-speed flow drew air indicated flow conditions of Mach 1.09 in the upper flow and Mach 0.17 in the lower flow. At the pressure and temperature conditions measured and calculated, this indicates speeds of about 60 m/s on the low-speed side, and 340 on the high-speed side, with an average convective velocity of about 200 m/s in the shear layer between the two.

In addition to calculating these general values, a probe to measure stagnation pressure was inserted through a sealable port in the upper surface of the test section. This instrument was positioned so that the end of the probe was located 1 mm downstream of the end of the splitter plate, and was so constructed that it could be raised or lowered to various positions above and below the splitter plate. Average static pressure is constant across a shear layer and was measured via the aforementioned pressure ports in the side of the test section. Stagnation temperature of the incoming air in the high-speed flow was that of the room the WCSL facility was housed in, while the low-speed air was assumed

184

to be matched to this stagnation temperature. Static and total pressure and temperature can be used to derive Mach number and velocity by the following relationships:

$$M = \left\{ \frac{2}{\gamma - 1} \left[\left(\frac{P_0}{P} \right)^{\frac{\gamma - 1}{\gamma}} - 1 \right] \right\}^{\frac{1}{2}},$$
(7.3)

$$T = \frac{T_0}{1 + \frac{(\gamma - 1)}{2}M^2},$$
(7.4)

$$a = \sqrt{\gamma RT} , \qquad (7.5)$$

$$V = M \cdot a \,. \tag{7.6}$$

Fig. 7.8 shows the profile of this shear layer in terms of Mach number and Fig. 7.9 shows the same profile in terms of velocity. These values come from the mean rms of the static pressure at that position, and from total pressure data from pitot probe data at various vertical positions 1 mm downstream of the splitter plate. Eqs. 7.3 through 7.6 were then used to derive Mach number and velocity from this data. The horizontal error bars on the figures indicate the error in the derived quantity based on the standard deviation of 2048 values recorded for each data point, while the points themselves are based on the rms of those values, along with the resolution error of the data acquisition board. The magnitude of this uncertainty in pressure was essentially the same at every measurement location; however, this results in smaller uncertainties for larger values of P_0 when Eqs. 7.3 is applied.



Figure 7.8: Mach number profile at the splitter plate.



Figure 7.9: Streamwise velocity profile at the splitter plate.

It can be noted that there was a small deficit in the velocity field 1 mm downstream from the splitter plate, consistent with a wake produced by the boundary layers feeding the shear layer as addressed in section 6.2.2. Studies by Ho and Huang⁹ indicate that the thickness of the high-speed boundary layer influences the development of a shear layer more than the thickness of the entire wake. Using the data shown in Fig. 7.9 and Eq. 6.14, the momentum thickness of the high-speed-side boundary layer was approximately 0.38 mm. Initial measurements and analyses seem to indicate that the high-speed boundary layer was turbulent, while the low-speed boundary layer was laminar.

As previously noted, both the WCSL facility, and the ART facility that it was based on, had cross sections with a steady contraction in the vertical dimension in order to maintain a steady static pressure in the streamwise direction. Low-speed air entrained into a shear layer experiences a loss in static pressure as it accelerates, and entrained high-speed air increases in pressure as it decelerates. Under the conditions produced in these facilities, the pressure loss in the low-speed air is greater than the pressure gain in the high-speed air, requiring a counteracting decrease in volume in order to maintain a steady static pressure. The WCSL test section has a contraction ratio of approximately 10 to 7 along its 1 m length.

Figure 7.10 shows static pressure recorded from the pressure taps along the length of the test section. The values shown here are the rms of 2048 values recorded at each point. The uncertainties shown reflect the standard deviation in those values and the resolution limitations of a 12-bit data acquisition card. As can be seen in the figure, with Mach numbers of 1.09 and 0.17, there is a slight favorable pressure gradient. The

contraction of the test section was designed to maintain a constant pressure for flows with Mach numbers of 0.8 and 0.1, but from these results it appears to do a reasonable job of maintaining pressure under other flow conditions.



Figure 7.10: Static pressure along the WCSL test section.

The second stage of experiments recorded optical effects of the shear layer by passing the beams of a Malley probe through the flow. These experiments, described in section 7.4, were conducted after the modifications to the WCSL facility described in section 7.2.2 were put in place. These modifications produced changes in the flow conditions. In particular, the sonic throat restricted the air flow through the facility, limiting the Mach numbers in the flow during this set of experiments to 0.88 and 0.06 for the high-speed and low-speed flows, respectively. At the pressure and temperature conditions in the test section this translates to velocities of approximately 285 m/s and 22

m/s, with an expected convection velocity for the shear layer between these flows around 153 m/s. Exact speeds for any given experiment varied somewhat with valve settings and ambient conditions of the day, but they tended to be close to these values when all the modifications described in section 7.2.2 were in place.

Figure 7.11 shows results for static pressure measurements along the test section, as was done in Fig. 7.10, but for the flow conditions in the modified tunnel. The mean static pressure within the test section is significantly higher than for the unmodified tunnel, reflecting the lower velocities and Mach numbers in the flow with the sonic throat in place, but it still shows a similar slightly favorable pressure gradient.



Figure 7.11: Test section static pressures under optical-survey run conditions.

7.3. Verification of the Weakly Compressible Model

The initial use of the WCSL facility was to provide experimental verification of the WCM simulations. Specifically, verification that the shear layer was rolling up into coherent vortical structures and evidence of significant pressure wells in those structures. The original plan included the capture of Schlieren images to visualize the density variations in the flow; however, the humidity on the Notre Dame campus during the days of these experiments proved sufficient to produce condensation mist in the low-pressure portions of the shear layer, which could be illuminated with a strobe lamp. Fig. 7.12 shows two examples with concentrations of condensed vapor to the right of center in the images, revealing the location of low pressure regions. In each example there are also regions free of such vapor to the left of center, indicating higher pressure. The form of these visualized structures bears a striking resemblance to the predictions of the WCM in chapter 6. In particular, the simulated velocity field shown in Fig. 6.15 on page 162, as well as the associated pressure distribution in Fig. 6.17 on page 165 bear a striking resemblance to the condensation-visualized structures in Fig. 7.12.



Figure 7.12: Condensation-based flow visualizations of a structure in the WCSL facility.

While the condensation images gave direct evidence of the formation of coherent structures, the condensation itself is only an indirect indication of low-pressure cells. Quantitative evidence was acquired through measurement of the unsteady pressures associated with these structures passing by. The test section schematic in Fig. 7.2 shows a rake of pressure taps in a vertically sliding panel set into one of the side walls at the 0.5 m *x*-position. For these verification experiments, four of these taps, spaced 1.9 cm apart, were fitted with surface-mounted 5-psi Kulite ultra-miniature pressure transducers. As the pressure inside the test section was already lower than the static pressure in the room, and the point of this experiment was to measure the existence of regions with even lower pressure in the flow, the difference between the static pressure of the room and the lowest pressure to be measured in the flow would have exceeded the dynamic range of these

transducers. Thus, it was necessary to provide a vacuum pressure as a reference pressure. The existence and location of this additional vacuum system is noted in the tunnel schematic shown in Fig. 7.1.

As the coherent structures passed by the 0.5 m *x*-location, the transducer array measured the unsteady pressure; however, these structures were not perfectly uniform in frequency, stage of formation, spacing, or vertical position at this point. Thus, a Kulite at a fixed *y* location was unlikely to capture a time-position cut through the same portion of each successive structure as it passed through the region. To cope with this and facilitate the comparison and averaging of these data, recording of these data were triggered by the Kulite closest to the centerline of the shear layer reaching a threshold of low pressure. This threshold was set low enough that the trigger fired only when the pressure reached an absolute minimum.

Because of this threshold triggering, the data recorded reflected cases with particularly severe pressure fluctuations at that location, rather than what might be considered average cases. Referring back to the predictions of the WCM, Figs. 6.16 and 6.17 show predictions for flow at the same location, but in one case the flow exhibits two smaller structures while in another the smaller structures combined into a larger structure with a greater magnitude of pressure variations. As this method of triggering seeks instance of particularly low pressure, it tends to capture the cases corresponding to the paired example of Fig. 6.17 while passing over cases corresponding to Fig. 6.16.

Once triggered, 2048 measurements were collected at each transducer location at 50 kHz. Two hundred such sets of data were collected at each transducer location and then ensemble averaged. By adjusting the *y* location of the four pressure ports, and

triggering off the most shear-layer-central transducer, six y-location, ensemble pressure traces were captured. For comparison, Fig. 7.13 shows an ensemble average of 200 phase-locked data series overlaid with a single data series. It can be seen that after the first cycle, the averaging reduces the amplitude of oscillation of the static pressure; whereas, the single trace continues to have large excursions in amplitude compared to the ensemble averaged trace. What does seem to be captured is the essential character of the first cycle, which is triggering the data collection. The ensemble appears to capture the average frequency of subsequent cycles, despite the variations in amplitude and period seen in the individual trace. It is interesting, however, that the time between the lowpressure trigger event that begins each data series and the next valley in the ensemble average indicates a frequency of approximately 1 kHz, while the following peaks and valleys of the ensemble average have a frequency of approximately 1.3 kHz. This is consistent with noting that the data acquisition was triggered by the pressure reaching low points that were lower than what was seen in most structures passing that point, and so the first cycle of the pressure fluctuations in the ensemble tends to be slightly larger than the subsequent cycles.



Figure 7.13: Phase-Locked Average of 200 Trigger Traces Overlaid by a Single Trace.

Figure 7.14 shows the ensembled pressure and location-averaged results for only the first cycle. In this instance, location averaged refers to averaging two or more sets of 200 traces from different Kulite transducers for different positions of the array of sensors, with sets being combined based on their position relative to the transducer being used to trigger the data collection. Thus, the averaged traces associated with positions furthest from the y = 0 position are based on only a single set of 200 traces while four such sets averaged together make up the y = 0 position. To allow for further variations in measured results due to differences in position of the rake, each set of traces recorded at the same time was scaled by the value of the first cycle peak-to-peak average for the largest trigger of the four triggered traces. It should be noted that even this method, because the peak-to-peak values are averaged, slightly underestimates the actual static-
pressure well depth and clearly underestimates the pressure peaks found at the saddle point in the braid between successive structures. It should also be noted that the horizontal axis in Fig. 7.14 is plotted as position rather than time. This position was derived from the time history by multiplying the time by the convective velocity in the previously cited frozen-flow relationship of $\Delta x = U_C \Delta t$. In the present case, referring back to Fig. 7.9, the convective velocity was 198.4 m/s.



Figure 7.14: Average pressures associated with a coherent structure at x = 0.5 m.

These data only include points after the trigger event, which corresponds to points in the flow upstream of the pressure well that serves as the trigger. The original intent in devising this experiment was to use a central trigger for these data, so that data preceding the trigger event for some period would be recorded as well as data following the instant of this trigger. However, this proved infeasible with the data acquisition system in use at that time. As may be noted in Figs. 6.16 and 6.17, the structures and static pressure wells predicted by the WCM are not symmetric in the up- and downstream directions; however, in order to get a sense of the well shape for comparison with predictions of the WCM, the data from upstream direction was reflected into the downstream direction and plotted as a surface shown in Fig. 7.15; the pressure well from Fig. 6.17 is replotted at the same scale in Fig. 7.15



Figure 7.15: Pseudo-reconstruction of an average static pressure well (left) compared to WCM simulation. (right)

This pseudo-reconstruction, when compared to the WCM prediction, not only indicates that the predicted pressure well exists, but it shows unmistakable similarities in relative location, spatial scale, and magnitude these of variations in pressure. For reasons already addressed, this is far from a complete picture of the flow and the structures within it, but it is sufficient to show that the WCM captures aspects of the flow not present in the previously-accepted thinking and modeling prior to these experiments.

Before leaving this section, it is worth mentioning why the pressures measured on the side walls of the test section represent the static pressure in the coherent structures as they pass by. While the pressure taps clearly represent the locations relevant to the structures, a boundary layer on the side wall separates the measurement location in the wall-normal direction from the pressure away from the wall in cells as they pass by. Because the structures are well-defined and effectively two-dimensional, they have twodimensional pressure fields which are identical in any wall-parallel plane cut at any wallnormal location in the test section up to the boundary layers on the side-walls of the test section. Thus, the pressures at the edge of the boundary layers are the pressures these experiments attempted to capture. One of the basic premises of boundary layers is that, at least on average, boundary layers cannot support pressure gradients in the wall-normal direction. In this case "on-average" refers to unsteady fine scales within the laminar or turbulent boundary layer itself.¹⁰ These fine scales are at much higher frequency than the passage frequency of the coherent structures in the shear layer. In the present case the boundary layer at the 0.5 m location was turbulent.

A number of researchers have studied the frequencies and amplitudes of pressure fluctuations in the boundary layer.^{11,12,13} Buckner et. al.¹² found that on average these frequencies are associated with structures that are approximately equal in size to the boundary layer thickness and bass by at approximately 0.8 the outer velocity of the boundary layer. In the data in Fig. 7.13, it is probable that the highest frequency fluctuations on the single trace an be associated with the boundary layer through which the pressures are being measured.

7.4. Optical Measurements

With this verification of the WCM, attention was again turned to the optical effects of the flow, using a Malley probe to study these effects. Details on principles and operation of Malley probes are described in section 3.3.3. The basic principle is that beams with a diameter that is small relative to the optical distortions being induced are directed through the flow, and the deflection of these beams corresponds to the local slope that would be induced on a planar wavefront passing through the flow. Use of convection velocities, frozen flow assumptions, and combining of data from different beams can then be used to reconstruct wavefronts over larger streamwise lengths up and downstream of the measurement location. A sketch of the approximate placement and orientation of these beams in the flow and test section during these studies is shown in Fig. 7.16.



Figure 7.16: Malley probe beams directed through the flow and WCSL test section.

7.4.1. Measurement System

A Malley probe makes use of the average convective velocity in the flow to extrapolate measurements to points upstream and downstream of a measurement point. If the flow at a given location in space and time induces a beam deflection of $\theta(x,t)$, and the velocity at which the sources of aberration are convecting (U_c) is known, then at some later time, ($t + \Delta t$) that portion of the flow will be some distance downstream, ($x + \Delta x$) such that

$$U_c = \frac{\Delta x}{\Delta t}.$$
(7.7)

If the times and distances of convection are sufficiently small that the structures in the flow do not change significantly, then it may be assumed that an optical deflection angle (θ) measured at position *x* and time *t* will be close to the deflection angle that would be measured at position $x + \Delta x$ and time $t + \Delta t$. Therefore,

$$\theta(x,t) \cong \theta(x + \Delta x, t + \Delta t) = \theta(x + U_C \Delta t, t + \Delta t)$$
(7.8)

and

$$\theta(x + \Delta x, t) = \theta(x + U_C \Delta t, t) \cong \theta(x, t - \Delta t) = \theta\left(x, t - \frac{\Delta x}{U_C}\right).$$
(7.9)

By this assumption, a time-series of deflection-angle measurements can be used to reconstruct the wavefront upstream and downstream of a single measurement point. Of course, there are limits to how far one can go from the measurement point and maintain an acceptable degree of accuracy as the flow does change as it progresses towards the measurement location from upstream and as it continues downstream. To capture the convection velocity necessary for this extrapolation, a closelyspaced pair of beams is used, one directly upstream of the other. Per Eq. 7.8, sources of distortion passing through the location of the first beam at some time *t* reach the second beam some distance Δx downstream at time $t + \Delta t$. A cross-correlation between the deflection angles recorded for the two beams reveals the average Δt and with a known spacing of Δx between the beams, U_C can be deduced from Eq. 7.7.

The probe used in this study was constructed using two beams from a 1 mW helium-neon laser. The single beam from the laser was first passed through a telescope with a pinhole at the narrowest focal point as a spatial filter to remove side lobes and aberrations from the beam. A beam splitter was used to generate two parallel beams from this single beam, which were then directed upward through the flow. After passing through the flow, the beams were brought back down to the optical bench, where they were passed through polarizing filters to adjust for intensity before being focused on to position sensing devices. A schematic of this arrangement is shown in Fig. 7.17 and a photo is shown in Fig. 7.18.



Figure 7.17: Plan-view schematic of optical bench and sensors.



Figure 7.18: WCSL facility with Malley probe setup on the optical bench.

Principles of lateral effect detector sensors are given in section 3.2.2. The sensors were of model SC-10D from OSI Optoelectronics, formerly UDT. Experimental experience with this facility has shown that the range of displacement of the beam on the sensor during an experiment in the WCSL facility tends to be less than 1 mm when the optical path and arrangement shown in Figs. 7.17 and 7.18 is used. As such, it is only necessary to establish a calibration curve over the linear range, which may be achieved by taking data for a few points in this range and performing a linear fit. Each data point on the example calibration curves of Fig. 7.19 is an average of hundreds of readings taken over the course of a few seconds.



Figure 7.19: Calibration curves for sensors used in study of the ND WCSL facility.

The use of 1-meter lenses to focus the beams onto the sensors simplifies the

conversion of displacements measured by the lateral effect detector to beam deflection

angles. Provided the detector is positioned at the nominal focal point of the lens, the position of the point at which the focused beam falls on the detector will be dependent only on the angle at which the beam enters the lens, not the location on the lens where this occurs. Therefore, the deflection angle associated with a measured displacement of d will be

$$\theta = \arctan\left(\frac{d}{1m}\right) \cong \frac{d}{1m}$$
 (7.10)

The approximation in equation 7.10 applies if $d \ll 1$ m, which is always the case.

In recording these signals, a low-pass filter with a cutoff frequency of 18 kHz was used to prevent aliasing. In post processing, a high-pass filter with a cutoff of 500 Hz was used to remove tunnel vibration. Figure 7.20 (a) shows a long series of beam deflection data, gathered at a position of 20.4 cm downstream of the splitter plate. Figure 7.20 (b) shows the OPD reconstructed from these deflections with a convective velocity of 153 m/s indicated by the flow conditions and the equation

$$OPD(x_{n+1}, t_m) = OPD(x_0, t) - \sum_{-n}^{n} \theta(x_0, t_{m-n}) U_C \Delta t$$
(7.11)



Figure 7.20: Deflection data (a) and OPD reconstruction (b) from a position of 20.4 cm.

As noted earlier, there are limits to how far upstream and downstream the wavefront can be accurately extrapolated. Additionally any beam or field of view directed through the flow will also have finite limits. Thus, it is both necessary and practical to place an aperture on the reconstructed wavefront, so that for any given time, only those points falling within some finite distance around the measurement point at that time will be considered. Figure 7.21 (a) denotes a portion of the reconstructed wavefront, from Fig. 7.20, centered at a point corresponding to a time of 3.3 ms into the sampling period of the measurement, and extending 5 cm to either side of that point.



Figure 7.21: Partitioning of reconstructed OPD with an aperture (a) and the partitioned section (b).

The *x*-position value of 0.5 m in this figure comes from the time of 3.3 ms multiplied by the convection velocity of 153 m/s. Realistically, the frozen-flow approximation does not hold over the full 1.5-m range of the x-axis in Fig. 7.21 (a), but a smaller portion of the longer OPD series can be taken as a reasonable reconstruction of the wavefront seen over that smaller area at the associated sampling time. Thus, the smaller portion appearing in Fig. 7.21 (b) can then be taken as an approximate reconstruction of the wavefront that would have been produced at a time of 3.3 ms, in a beam with a 10 cm diameter, centered at the 20.4 cm location in the flow.

The OPD series of Fig. 7.21 (a) is plotted vs. a position in x covering a range of almost 1.5 m, but the reconstruction is based on data recorded at is actually arbitrary, a time of 3.4 ms corresponds to a position of

In examining these wavefronts, it was made standard practice to remove the mean and net slope, also called piston and tip-tilt, from the wavefront within the defined aperture at each time step. The initial purpose of this line of research was to examine the requirements of a corrective system, particularly for higher-order correction. In the Notre Dame Adaptive Optics system, described in chapter 3, piston is considered to be irrelevant to the performance of the system, and a fast steering mirror serves to remove tip-tilt. Higher-order distortions that remain after correction by the steering mirror are then left to a deformable mirror for correction, and it was these distortions that were originally of interest. Figure 7.22 shows the process of removing piston and tilt from the portion of the reconstruction in the aperture. This is done by performing a least-squares fit of the function $y = a \cdot x + b$ to the values of OPD(x) in the aperture, as shown in Fig. 7.22 (a), and then subtracting the resulting equation from OPD(x). The result of doing this to the example wavefront is shown in Fig. 7.22 (b).



Figure 7.22: Applying a linear fit to the reconstructed wavefront within an aperture (a) and the wavefront after T/T correction (b).

As noted earlier, using a Malley probe with two beams, one upstream of the other, can reveal things about general qualities and particular components of the flow. Fig. 7.23 shows the power spectra of beam deflection for the two beams as recorded at a location of 3.9 cm. Notable features of these spectra are the sharp spike at a frequency of 16.7 kHz and a broad-band hump stretching from 5 kHz to 14 kHz, with a peak value near 8 kHz.



Figure 7.23: Beam-deflection power spectra: upstream beam (a), downstream beam (b).

Power spectral densities such as these are produced by performing a Fourier transform of the deflection data and multiplying the resulting complex values by their own conjugates, so that for a set of deflection values, $\theta(t)$, comprised of *N* points,

$$PSD_{\alpha}(f) = \frac{\hat{\theta}(t)\hat{\theta}(t)^{*}}{N}$$
(7.12)

with $\hat{\theta}(t)$ representing the fast-Fourier transform (FFT) of $\theta(t)$ and $\hat{\theta}(t)^*$ representing the complex conjugate of $\hat{\theta}(t)$. The spectra shown in Fig. 7.23 and in the following sections of this chapter are averages of the spectra produced from 200 or more sets of data. The averaging is used to produce cleaner, smoother spectra.

With two beams, a spectral cross-correlation can be performed in the form of

$$PSD_{CC}(f) = \frac{\hat{\theta}_{1}(t)\hat{\theta}_{2}(t)^{*}}{N}.$$
(7.13)

The PSD produced by Eq. 7.12 has real values, due to multiplying the complex values produced by the Fourier transform or FFT by their own conjugates. However, the cross-correlation spectrum of Eq. 7.13 will have complex values, which can be expressed as magnitude and phase, of the form

$$PSD_{CC}(f) = \frac{\hat{\theta}_1(t)\hat{\theta}_2(t)^*}{N} = A(f)e^{iB(f)}.$$
(7.14)

With an assumption of frozen flow over the time required for sources of aberration to travel from the location of the first beam to a second beam some distance Δx downstream, the phase of this cross-correlation PSD will be such that¹⁴

$$B(f) = 2\pi f \frac{\Delta x}{U_c}.$$
(7.15)

Figure 7.24 shows phase of the cross-correlation produced from the deflection data used in Fig. 7.23. Over the range of frequencies corresponding to the broad-band hump seen in Fig. 7.23, the phase as a function of frequency does indeed take a form that can be approximated by a straight line, and the slope of this line indicates a convective velocity of 260 m/s for a measurement taken with beams 4 mm apart. This indicates that the sources of distortion at this point are associated more with the high-speed stream than

with a fully-developed shear layer, whose aberrating structures travel at the slower convection velocity.



Figure 7.24: Phase of beam-deflection cross-correlation.

Also of interest is that the sharp peak in magnitude at 16.7 kHz has a corresponding phase which deviates significantly from that associated with near-by frequencies. The sharpness of this peak is suspicious, compared to the broad-band nature of other features, but the phase difference is further indication that this peak has a source other than the shear layer, and should be ignored.

As noted, the disturbances in the early stages of development for the shear layer are often more associated with the high-speed flow than with a convective velocity that reflects and average between the two flows. For comparison, a deflection spectrum and phase plot from a measurement location of 12 cm is shown in Fig. 7.25. A fit to phase for the frequencies associated with the hump of greatest activity indicates a velocity of approximately 140 m/s, which is more in line with the prediction of velocity in the shear layer based on the flow conditions. For frequencies higher than this range the phase plot shows a shallower slope, which is in line with the higher velocity of disturbances associated with the high-speed flow, possibly from the boundary layer on the upper surface of the test section.



Figure 7.25: Phase of beam-deflection cross-correlation.

7.4.2. Results

A series of Malley probe measurements were made at various locations starting at 0.7 cm downstream from the edge of the splitter plate, where the high- and low-speed flows first come into contact, out to 48.2 cm from the splitter plate, with a spacing of approximately 0.5 cm between measurement locations. The displacement data from the position sensing devices in the Malley probe was sampled at a rate of 40 kHz, with 25

sets of 16384 points per set being recorded at each measurement location. The two beams of the Malley probe were separated by a distance of 3.8 mm. 4th-order Butterworth filters were applied to remove signals below 500 Hz, which were dominated by physical vibration of the facility and optical bench, and signals above 18 kHz to prevent aliasing.

A serious concern in taking these measurements was avoiding contamination of the optical data by other sources. As noted in the previous section, performing a spectral cross-correlation and plotting the phase can identify features that are not associated with the shear layer. Additionally, at some points during the collection of Malley probe data, the optical bench on which the optics comprising the Malley probe were fastened was briefly pulled sideways so that the beams of the Malley probe passed to one side of the test section, rather than through the test section and the flow it contained. Despite passing the beams through still air, the sensors still recorded disturbances, most likely due to vibration of the equipment and electronic noise. The PSDs associated with these signals are shown in Fig. 7.26



Figure 7.26: PSD of vibration and electronic noise data from the Malley probe.

The highest noise peak seen in both beams is seen at 860 Hz, and is most likely a remnant of physical vibration outside the range of the 500 Hz high-pass filter intended to deal with vibration effects. The narrow-frequency nature of the peaks seen at higher frequencies is suggestive of electronic noise, though it is not truly possible to separate physical effects from electronic effects in this noise data. Comparing this to the spectra used as an example in Figs. 7.23 and 7.24, the large vibration peak at 860 Hz is only 1/3 the magnitude seen in the broad-band hump associated with the effects of the shear layer, and lies outside the range of frequencies associated with that hump. As will be shown

later, the optical-distortion effects produced by the shear layer shift to lower frequencies as the measurement location is moved downstream, but the magnitude associated with those distortions also grows larger, rendering this vibration noise insignificant by the time frequencies of that range become important to the measurements.

There is a sharp spike at 17 kHz in the noise of the upstream beam that may correspond to the 16.7 kHz spike seen in Fig. 7.23, and the ability to identify such spikes as non-physical noise was demonstrated in the example using that figure. Thus, the data from these measurements can be used with some confidence that noise corrupting the data can either be identified as such, or will have minimal impact on the results.

In processing these data, a measurement location of 1.8 cm was the first point at which the PSDs show a clear hump of activity and the cross-correlation of the two beams showed a clear convection velocity associated with this frequency range. Figure 7.27 shows PSDs for deflection of the downstream beam, plotted to an arbitrary but proportional scale and shifted vertically to reflect the position of that beam for each measurement. The downstream beam was chosen for this figure as less noise seems to be present in the data recorded for this beam, as seen in the spectra of Fig. 7.26.



Figure 7.27: PSD of beam deflection, shifted by position.

The broad-band hump of activity in the spectra of Fig. 7.26 can be seen to shift to lower frequencies and increase in magnitude as measurement position is shifted downstream. This trend becomes clearer with more measurement points further downstream added, as shown in Fig. 7.28.



Figure 7.28: PSD of beam deflection, shifted by position.

This trend is to be expected if the disturbances are produced by the coherent structures associated with the shear layer, discussed earlier in chapter 6. As noted in section 6.2.3, these structures grow and combine such that the streamwise length associated with one of these structures, Λ , is proportional to the distance downstream from the origin point of the shear layer. Although the effective origin point for these structure may be at a slightly different position than the physical origin of the shear layer, due to changing dynamics at various stages in development. Frequencies associated with these structures are determined by how swiftly the convection velocity carries them past the point of observation. Therefore, if

$$\Lambda \propto x \tag{7.16}$$

and

$$f = \frac{U_c}{\Lambda},\tag{7.17}$$

then

$$f \propto \frac{1}{x}$$
 (7.18)

In the figures above, a number of local peaks can be seen at each measurement location, but these appear on top of a broad hump that not only shifts to lower frequencies with increasing distance from the splitter plate but grows in peak intensity with position. Again, if the pressure wells predicted by the WCM are the primary contributor to optical distortions then one would expect those distortions to get worse as the structures not only become larger but the pressure differences become more severe. Some of the local peaks might be associated with noise of the sort shown in Fig. 7.26, but as addressed earlier, the magnitude of such noise will not be large enough to obscure the broad-band hump representing the effects of the shear layer.

The fact that the shear layer manifests as a broad hump rather than a sharp spike in these spectra is due to the shear-layer structures not being truly periodic, at least not without applied forcing to make them periodic. As shown in Figs. 6.16 and 6.17, predictions from the WCM indicate that structures passing a given point at different times may be at different stages in their development. In capturing the strobe-illuminated images in Fig. 7.12, it was not possible to "freeze" the illuminated structures with the strobe. To the eye, the illuminated structures would move about jerkily and change shape in the same jerky fashion. However, by matching the passage frequency or a subharmonic of that frequency, these images would appear to settle down somewhat, relative to other strobe frequencies. Fig. 7.13 shows the variations in the pattern of pressure changes seen at a single point over the course of a few cycles, but also uncovers a fundamental frequency in the ensemble average of several individual traces. Likewise, the OPD reconstruction in Fig. 7.20 shows activity that is near-periodic over short periods, but tends to shift around some characteristic frequency over longer periods.

There are different approaches to finding this characteristic frequency of the shear layer. One approach is to use a weighted average of these frequencies, of the form

$$f_C(x) = \frac{\int PSD(f, x) f df}{PSD(f, x) df}.$$
(7.19)

However, while this weighted average has been used in characterizing results of WCM simulations,¹⁵ it was found to be a poor choice in characterizing the data gathered from the physical shear layer in the WCSL facility. As shown in Fig. 7.25, not all of the

disturbances recorded by the Malley probe are from the shear layer. The flows in the test section also produce boundary layers on the upper and lower surfaces of the test section. Phase plots of cross-correlations indicate that the optical deflection and distortion effects associated with frequencies higher than those associated with the visible hump in the spectrum tend to have velocities closer to that of the high-speed flow than that of the shear layer. Thus, an estimate of characteristic frequency based on the frequency with the greatest magnitude seen in the PSD for beam deflection was judged to be a more practical approach, although this required the elimination of noise-related peaks from consideration in some cases. Figure 7.29 plots the peak frequency of these jitter-signal spectra as a function of position downstream from the origin of the shear layer. The trend in frequency matches a 1/x pattern, which is consistent with expected behavior of this flow.



Figure 7.29: Frequency with position and a/x + b fit.

The fits are actually performed only to the points more than 5 cm downstream from the origin. As was noted in sections 6.2.2 and 6.2.3, the model of linear growth does not apply to the initial stages of development for a shear layer. As was shown in the cross-correlation of Fig. 7.25, and has been confirmed by other researchers,¹⁶ the initial development of a shear layer often begins in the boundary layer associated with the highspeed flow, as it takes time for the flow to assume the standard shear-layer profile often represented by a hyperbolic tangent. A fit based on growth proportional to \sqrt{x} has been performed on the points for which x < 5 cm, in accordance with the growth rate expected for that region.¹⁶

Figure 7.30 shows the structure length estimated from the frequencies in Fig. 7.29 and the convection velocity from Eq. 7.17. Equation 6.22 in section 6.2.3 gave an expected growth rate for shear layers in terms of the visual thickness, δ_{vis} . This equation requires the ratio of densities in the flows, as well as the ratio of the velocities of the flow. For Mach numbers of 0.88 and 0.06, the static pressure in the test section, and a total temperature in each flow equal to that in the room the air is drawn from, the density of each flow can be calculated to be 1.01 kg/m³ in the high speed flow and 0.70 kg/m³ in the low speed flow. From this and the velocities of 285 m/s and 22 m/s, Eq. 6.22 predicts a growth rate of $d(\delta_{vis})/dx = 0.27$. Sources^{17,18} on this subject do not agree on the ratio between this thickness and the streamwise spacing between vortical structures, giving ratios varying from 1.5 to 2, which in turn indicate the growth rate for $d\Lambda/dx$ could range from 0.40 to 0.54. The linear fit in Fig. 7.30 has a slope of 0.54, corresponding to a ratio close to 2 between Λ and δ_{vis} .



Figure 7.30: Characteristic Λ and linear growth rate.

One of the purposes of this study was to examine the requirements for optical correction of compressible shear layers, with the intent of using the adaptive optic system shown in Fig. 3.1 for this correction. Part of the design of this system, and of many corrective systems, is that the correction takes place in two stages. First, there is tip-tilt (T/T) correction with a fast steering mirror to center the beam on target, and then a deformable mirror is used to deal with wavefront distortions that remain after the T/T correction. Therefore, it was useful to examine the results with regard to this residual, higher-order distortion.

Fig. 7.31 shows the time-averaged rms OPD with T/T removed, as reconstructed from the data taken at various streamwise positions. Error bars are not shown as the uncertainty in these measurements as shown on this figure would be smaller than the size

of the symbols used to graph the data on this figure. There are five separate curves in this figure, but all of them are based on the same data recorded at those positions. The difference in the curves is the size to which the aperture was set in reconstructing the OPD from the recorded beam deflections. An aperture of 300 cm can effectively be considered an infinite aperture, relative to the structures in the flow, and for that aperture size OPDrms grows almost linearly with position. However, applying a smaller aperture to the data reduces the magnitude of the OPDrms seen within that aperture. Additionally, the shape of the curves produced changes from something close to a straight line to a curve of decreasing slope that levels off at some point. After observing this effect in the reconstructed data, an exploration of the reasons behind this was undertaken, which is described in the next chapter.



Figure 7.31: Optical aberration from a shear layer, varying with position and aperture size.

- ¹ Roberts, F. A., Dimotakis, P. E., and Roshko, A., "Kelvin-Helmholz instability of superposed streams", In *Album of Fluid Motion*, ed. M. Van Dyke, p. 85, Stanford, California, Parabolic, 1982.
- ² Winant, C.D., and Browand, F.K., "Vortex pairing: the mechanism of turbulent mixing-layer growth at moderate Reynolds number", Journal of Fluid Mechanics, Vol 63, 1974, pp. 237-255.
- ³ Ho, C-M, and Huerre, P., "Perturbed Free Shear Layers," Annual Reviews of Fluid Mechanics, 16, 1984, pp. 365-424.
- ⁴ Duffin, D., "Feed-Forward Adaptive-Optics Correction of Aero-Optical Aberrations Caused by a Two-Dimensional Heated Jet," 36th AIAA Plasmadynamics and Lasers Conference, AIAA-2005-4776, June 2005.
- ⁵ Rennie, M., Duffin, D., and Jumper, E., "Characterization of a Forced Two-Dimensional, Weakly-Compressible Shear Layer," 38th AIAA Plasmadynamics and Lasers Conference, AIAA-2007-4007, June 2007.
- ⁶ Nightingale, A., Mitchell, B., Goodwine, B., and Jumper, E.J., "Feedforward' Adaptive-Optic Mitigation of Aero-Optic Disturbances," 39th AIAA Plasmadynamics and Lasers Conference, AIAA-2008-4211, June 2008.
- ⁷ Darcy, H.,. "Recherches Experimentales Relatives au Mouvement de L'Eau dans les Tuyaux," ("Experimental Research Relating to the Movement of Water in Pipes"), Mallet-Bachelier, Paris, 1857
- ⁸ Colebrook, C.F. "Turbulent flow in pipes, with particular reference to the transition region between smooth and through pipe laws," Jour. Ist. Civil Engrs., London (Feb. 1929)
- ⁹ Ho, C. M., and Hauang, L. S., "Subharmonics and vortex merging in mixing layers", *Journal of Fluid Mechanics*, Vol. 119, 182, pp. 443.473.
- ¹⁰ Schlichting, H., Gersten, K., Boundary-layer theory, Springer, New York, 2000
- ¹¹ Thomas, F.O., Putnam, C.M., Chu, H.C, "On the mechanism of unsteady shock oscillation in shock wave/turbulent boundary layer interactions," Experiments in Fluids, Volume 18, Issue 1-2, pp. 69-81, (1994).
- ¹² Buckner, A., Gordeyev, S., and Jumper, E.J., "Conditional Measurements of Optically-Aberrating Structures in Transonic Attached Boundary Layers," 44th AIAA Aerospace Sciences Meeting and Exhibit, AIAA-2006-0079
- ¹³ Cress, J.A. and Gordeyev, S., "Aero-Optical Measurements in a Turbulent, Subsonic Boundary Layer at Different Elevation Angles," 39th Plasmadynamics and Lasers Conference, Seatle, AIAA 2008-4214
- ¹⁴ Gordeyev, S. and Jumper, E.J., "Aero-Optical Characteristics of Compressible, Subsonic Turbulent Boundary Layers," 34th AIAA Plasmadynamics and Lasers Conference, AIAA-2003-3606
- ¹⁵ Nightingale, A., Duffin, D., Lemmon, M., Goodwine, B., and Jumper, E., "Adaptive-Optic Correction of a Regularized Compressible Shear Layer," 37th AIAA Plasmadynamics and Lasers Conference, AIAA-2006-3072
- ¹⁶ Morris, S.C. and Foss, J.F., "Turbulent boundary layer to single-stream shear layer: the transition region", *Journal of Fluid Mechanics*, Vol. 494, 2003, pp. 187-221.
- ¹⁷ Brown, G.L., and Roshko, A., "On Density Effects and Large Structure in Turbulent Mixing layers," Journal of Fluid Mechanics, Vol 64, No. 4, 1974, pp. 775-816.
- ¹⁸ Dimotakis, P.E., "Two-Dimensional Shear Layer Entrainment," AIAA Journal, Vol. 24, No.11, 1986, pp. 1791-196.

CHAPTER 8:

EFFECTS AND REQUIREMENTS OF CORRECTION

8.1. Overview

A number of guidelines to predict the effectiveness of systems for optical correction were given in chapter 4, particularly section 4.3. However, these were specifically based upon the characteristics of free atmospheric turbulence in a state of effective equilibrium. Chapters 5, 6, and 7, explored the optical effects and some aspects of the underlying fluid dynamics for compressible flows that might be found in the vicinity of an aircraft, particularly shear layers. In the course of this studying these flows, termed aero-optic flows to distinguish them from atmospheric turbulence, it was made clear that many of the underlying assumptions used in characterizing atmospheric turbulence do not apply to aero-optic flows. Likewise, the guidelines for designing and predicting the performance of corrective systems presented in chapter 4 are not useful in dealing with aero-optic flows, as they are based on parameters such as C_n^2 or the Fried parameter (r_0) that are not meaningfully defined for cases other than atmospheric turbulence.

To find guidelines for the design of optically corrective systems for dealing with the effects of non-Kolmogorov flows, it has proven useful to go back to basics in some regards. This begins with a simplifying assumption that the optical disturbances can be represented as a sinusoidal function. This is quite true for some forms of aero-optic flows, particularly ones that have been intentionally regularized with artificial forcing. Admittedly, this is a very simplistic approximation, but as will be shown in chapter 9, the results produced with this approximation have proven to be applicable even to conditions where the assumption itself does not apply.

With this assumption, modeling correction for tip-tilt (T/T) or other forms of correction becomes relatively easy. In particular, finding the average effect of correction for an average level of distortion can be found and expressed as a magnitude gain of a transfer function for the corrective system, dependent on the frequency of the disturbances to be corrected.

8.2. Assumption of a Sinusoidal Wavefront

As noted in chapter 6, the character of a shear-layer flow tends to be dominated by large vortical structures, and at a given streamwise location these structures tend towards a certain size and spacing, based on the thickness of the shear layer at that point. This aspect of the theory presented in chapter 6 was verified in the experimental results of chapter 7, such as the ensemble-averaged pressure measurements at a fixed position in Fig. 7.13 and the frequency hump that moves with position in the spectra of Figs. 7.27 and 7.28. The peak values of the data used in Figs. 7.27 and 7.28 were in turn used to produce the estimates of overall structure size in Fig. 7.30, which fall within the predicted range for shear layer growth rates.

From these results, it might seem plausible to approximate the optical disturbances of shear-layer flows by using a sine wave of the appropriate frequency and

amplitude. However, as can be seen in Fig. 7.13, the pressure fluctuations associated with a single trace have some quasi-periodic characteristics. Although the nature of a single trace of recorded pressure is quite different from the steady frequency and smaller amplitude of the ensemble averages, it is still somewhat sinusoidal, at least over a time period of one or two cycles of that sinusoid. The reconstruction of OPD shown in Fig. 7.20 has similar characteristics. The variations in frequency and amplitude of a trace of pressure or OPD are consistent with the broadness of the peaks in Fig. 7.28 indicating a range of wavelengths around the most probable frequency being averaged together.

Then again, periodic signals and even finite intervals of non-periodic signals can be approximated by a summation of sine and cosine functions. Additionally, shear layers are susceptible to forcing. Oster and Wygnanski¹ found that applying forcing, in the form of a moving flap at the initial point of a shear layer where the two streams first come into contact, will force a shear layer to grow more quickly than it would without such intentional perturbations. They also found that this higher growth rate persists only until it reaches a certain thickness, determined by the frequency of the forcing, and that thickness will remain nearly-constant for some distance down stream until the flow arrives at the location at which that thickness would be reached by the normal, unforced, growth rate, as shown in Fig. 8.1.



Figure 8.1: Momentum thickness of forced shear layers. (Oster and Wygnanski, 1982)¹

This phenomenon has been further explored and expanded upon,^{2,3,4} indicating that the period of steady layer thickness also represents a regularization of the rolling vortices in the shear layer, producing a coherent train of structures with greater uniformity in size and spacing. Reference [4] indicates that the initial roll-up point described in section 6.2.2 is unaffected by forcing of this nature, but thereafter the rollers tend to combine in groups of three or more, rather than the usual pairing, producing the increased growth rate and the train of relatively large and regular structures in the following region.

Work by others at Notre Dame has supported this view of shear layer behavior, both computationally,⁵ and experimentally.^{6,7} Simulations using the weaklycompressible model indicate the existence of clear, regular structures and greatly regularized optical distortions associated with those structures. Figure 8.2 below shows examples of computational results from the weakly-compressible model (WCM) of an unforced and a forced shear layer. Figure 8.3 shows beam-jitter at a position of 0.415 m in the simulation for unforced and forced cases, indicating a very close match in form between the sinusoidal forcing and the narrow beam deflection during forcing.



Figure 8.2: WCM realizations of a shear layer without forcing (left) and with forcing (right). (Nightingale, 2005)⁵



Figure 8.3: Normalized beam jitter according to WCM realization of a shear layer without forcing (left) and with forcing (right). (Nightingale, 2005)⁵

Figure 8.4 shows a reconstructed wavefront from optical experimental data taken through a M 0.78 shear layer being forced at the trailing edge of the splitter plate. While the reconstructed *OPL* from the experiment is not as close to a pure sine wave as the simulated beam jitter in Fig. 8.3, it does show a fairly steady base frequency.



Figure 8.4: Segment of OPL history for 800 Hz, forced shear layer at $x = 570 \text{ mm.} (\text{Rennie}, 2006)^6$

Additionally, research by others at Notre Dame has lead to the successful use of a sinusoidal approximation for real-time optical correction of a forced low-speed heated jet.⁸ Plans for automated application of the same basic corrective system to a high-speed shear layer are also under development, with encouraging results in simulation.^{9,10}

In light of these results, a sinusoidal approximation of wavefronts is certainly justifiable for shear layers under forcing. Additionally, with Fourier series and power spectra representing optical phenomena as sets of sinusoidal components, there is reason to hope that some of the results found by using this approximation may be applicable to non-regularized flows as well.

8.3. Tip-Tilt Correction within an Aperture

Tit-Tilt (T/T) and correction for T/T were addressed in sections 3.2 and 3.4.1. To reiterate here, tilt may be defined as Z-tilt, which is a linear fit to the wavefront itself, or as G-tilt, which is an average of the local gradient at each point on the wavefront. Correcting for G-tilt is best for putting the centroid of the associated far-field pattern on a target point. Correcting for Z-tilt is best for minimizing overall actuator stroke if the corrected wavefront is to be passed on to a higher-order corrective system, as is the case in the ND AO system shown in Fig. 3.1 on page 48. For this reason, the work performed at Notre Dame has focused on Z-tilt correction.

In applying T/T correction to a sinusoidal simulation of a wavefront, an interesting aspect of such correction becomes apparent. Figure 8.5 shows an example of an aperture in one dimension admitting a sine shaped wavefront. In this instance, the length scale of the variations on the wavefront is larger than the aperture. The portion of the wavefront over this aperture has relatively little curvature, and performing T/T correction based on Z-tilt leaves a nearly-flat remnant in the aperture, and the resulting far field pattern is close to the diffraction-limited ideal.



Figure 8.5: T/T correction for distortions with a long length scale.

Figure 8.6 repeats this example with sinusoidal wavefront aberrations of a length scale smaller than the aperture.



Figure 8.6: T/T correction for distortions with a short length scale.

Both wavefronts have the same magnitude of aberration, and will have the same average OPD_{rms} over time if the source of that aberration is in motion relative to the optical path. The uncorrected far field in both cases has Strehl ratio (on axis intensity relative to the diffraction limited case) of about 0.8, which fits with the prediction of the large-aperture approximation,
$$SR \cong e^{-\sigma_{\varphi}^2} = e^{-\left(\frac{2\pi}{\lambda}OPD_{rms}\right)^2}.$$
(8.1)

However, this may be a fluke as the term "large-aperture" in the name of the approximation indicates that the aperture should be significantly larger than the length scale of the major optical distortions, which is clearly not the casein Fig. 8.5.

However, effects of T/T correction are quite different for these two cases. In the case with a longer length scale for the disturbance, the overall amplitude of the wavefront aberrations in the aperture is greatly reduced, and the far field intensity pattern associated with the corrected near field wavefront shifts to bring the point of highest intensity back in line with where the center of the diffraction limited pattern would be. It should be noted that the form of the far field intensity pattern is not actually changed, but merely shifted to a new location. This is an inherent trait of T/T correction that will be explored further in latter sections. For now, it is sufficient to note that T/T correction tends to have a greater impact for disturbances with a length scale larger than the aperture over which the correction is being performed.

8.3.1. Simple Simulated Correction

Figure 8.7 was generated by using a pure sine function of extended coherence length as the beam deflection, θ . Each point is the time-averaged OPD_{rms} , after T/T removal for a fixed period (Λ) of the sinusoidal jitter and a fixed size of aperture (A_p). Each set of points shares a common value for A_p while Λ varies.



Figure 8.7: T/T corrected OPD for simulated sinusoidal disturbances.

For an infinite aperture, the average OPD_{rms} grows linearly with Λ , much as the OPD_{rms} of the experimental results with a large aperture grew linearly with position in Fig. 7.31. This is a natural consequence of the method of wavefront reconstruction. As was addressed in section 3.3.3, the deflection angle of the narrow beams used in this study correspond to the local slope that would be seen in a continuous wavefront,

$$\theta_x(t) = -\frac{dOPL(t)}{dx}.$$
(8.2)

Reconstructing the wavefront is then a matter of integrating this derivative. The data represented by θ is a time series, but can be translated into a set of points in space via the convection velocity,

$$OPL(t) = \int \frac{dOPL(t)}{dx} dx = \int \frac{dOPL(t)}{dx} \frac{dx}{dt} dt = \int \frac{dOPL(t)}{dx} U_C dt = \int -\theta_x U_C dt .$$
(8.3)

If θ is a sine wave, as it was in this very simple simulation, then

$$OPL_x(t) = \int -K\sin(2\pi f\tau)U_C d\tau = U_C \frac{K}{2\pi f}\cos(2\pi ft).$$
(8.4)

As frequency corresponds to structure size by $f = U_C / \Lambda$,

$$OPL_{x}(t) = U_{C} \frac{K}{2\pi} \frac{\Lambda}{U_{C}} \cos\left(2\pi \frac{U_{C}}{\Lambda}t\right) = \Lambda \frac{K}{2\pi} \cos\left(2\pi \frac{U_{C}}{\Lambda}t\right).$$
(8.5)

The rms value of this is then

$$OPL_x(t) = \Lambda \frac{K}{2\sqrt{2}\pi}.$$
(8.6)

which is proportional to Λ as seen in Fig. 8.7.

However, when a finite aperture with T/T correction is imposed on the reconstructed wavefront in this simulation, then the OPD_{rms} follows the linear grown trend only for apertures smaller than the structure size. As Λ approaches A_p , OPD_{rms} levels off, reaches a maximum around $\Lambda = A_p$, and then tapers off asymptotically as Λ increases. As was shown in Figs. 8.5 and 8.6, T/T correction is more effective for wavefront distortions with a longer length scale, but does little for distortions with a length scale such that $\Lambda < A_p$.

One striking feature of this behavior is that the curves indicated on Fig. 8.7 for finite apertures seem to have self-similar shapes. The scaling of these shapes is directly tied to A_p in the axis for Λ , and the effect shown in Eq. 8.2 through 8.6 produces a similar scaling of OPD_{rms} with respect to Λ and thence to A_p . Both Λ and OPD_{rms} have units of length, as does A_p . If the values of Λ and OPD_{rms} in Fig. 8.7 are non-dimensionalized by A_p , then the simulated results collapse onto one curve, shown in Fig. 8.8.



Figure 8.8: Non-dimensionalized results for sinusoidal disturbances.

This result is promising, as scaling laws are valuable tools in applying laboratory results to field applications, but the use of a sine wave as an approximation for the disturbances is only an approximation. This needs to be linked more directly to the flow with theory and backed up with experimental data.

8.3.2. Optical Scaling for Shear Layers

The preceding section defined *OPL* in terms of the deflection angles often measured and used to reconstruct *OPL*. As addressed in section 2.3.1, a more proper definition of *OPL* is the integral of the index of refraction along a beam's path, which for air is a function of density.

$$OPL(x) = \int (\overline{n} + \Delta n(s)) ds \cong \int (1 + \Delta n(s)) ds = \int (1 + K_{GD}(\overline{\rho} + \Delta \rho(s))) ds$$
(8.7)

From there it follows that the magnitude of variation in the *OPL*, which is *OPD*, will be proportional to the integral of density variations through the flow field.

$$OPD(x) = OPL(x) - \overline{OPL} \cong \int (1 + K_{GD}(\rho + \Delta \rho(s)))ds - \int (1 + K_{GD}(\rho))ds$$
$$= K_{GD} \int (\Delta \rho(s))ds$$
(8.8)

For isentropic flows, changes in density are proportional to changes in pressure and inversely proportional to changes in temperature. From work with the WCM, and even the preliminary calculations starting on page 153 in chapter 6, the pressure drop inside a vortical structure of a shear layer is approximately proportional to the square of the characteristic velocity. Thus, one can find an extending chain of equalities and proportionality,

$$\frac{\Delta\rho}{\overline{\rho}} = \frac{1}{\gamma} \frac{\Delta p}{\overline{p}} \propto \frac{1}{\gamma} \frac{\rho}{\overline{p}} U_c^2 = \frac{1}{\gamma} \frac{U_c^2}{R\overline{T}} = \frac{U_c^2}{a^2} = M_c^2, \qquad (8.9)$$

where $U_C = (U_2-U_1)/2$ and is the characteristic velocity of the structures, *a* is the local speed of sound, and M_C is a convective Mach number. *OPD*, as noted above, is proportional to an integral of $\Delta \rho$ over some distance, which in turn will be proportional to the magnitude of the changes in density and the length over which those variations are integrated. As a definition of thickness, δ_{ω} or δ_{vis} should be proportional to the size of the rollers and to each other and so are certainly proportional to the significant integration length for this case:

$$OPD_{rms} \propto \int K_{GD} \Delta \rho \, dy \propto K_{GD} \Delta \rho \delta_{vis} \propto K_{GD} \overline{\rho} M_C^2 \delta_{vis} \,. \tag{8.10}$$

A more thorough exploration of the density and Mach number aspects of this relationship can be found in a paper by Fitzgerald.¹¹ From the exploration of shear-layer

dynamics in chapter 6, particularly Eq. 6.22 on page 151, δ_{vis} is proportional to x and by the definition of δ_{vis} as the apparent size of the visualized shear layer; Λ is likely to be proportional to and on the order of δ_{vis} :

$$OPD_{rms} \propto \rho M_{C}^{2} \Lambda \propto \rho M_{C}^{2} x. \qquad (8.11)$$

Thus, if OPD_{rms} does indeed scale with Λ , as it did in the model associated with Figs. 8.7 and 8.8, then it should scale with position in a similar manner, especially as the average density and convective Mach number should be constants in the flow produced in the ND WCSL facility. If the finite aperture and T/T removal apply over a mix of frequencies as well as a distortion in the form of a sine wave, then all the salient points in the sine function test that make the scaling work should be present in a physical shear layer and optical system. Thus, for a given set of flow conditions, there is reason to hope that there is some function, g, such that

$$\frac{OPD_{rms}(x,A_p)}{A_p} = g\left(\frac{x}{A_p}\right).$$
(8.12)

It is left now to examine some actual optical data to see if this does apply to a real system. The results for OPD_{rms} with T/T correction and varying aperture size were shown in Fig. 7.31, and are shown again in Fig. 8.9 below. There are some similarities in the behavior of data points shown in Fig. 8.9, and the simulation results in Fig. 8.7. An aperture of 300 cm can effectively be considered an infinite aperture, and for that aperture size, OPD_{rms} reconstructed from the data grows almost linearly with position, as it did in the simplified simulation. Also note that from Fig. 7.30 on page 220, Λ reaches 5 cm in length at a position somewhere around 11 cm, and in Fig. 8.9, the data points for OPD_{rms} values generated with a 5 cm aperture begin to level off at about that point.



Figure 8.9: Optical aberration from a shear layer, varying with position and aperture size.

From the results of the simulation in Fig. 8.8, one might expect the data points to fall off at downstream positions where the characteristic structure length is larger than the aperture. However, as can be seen in the power density spectrums for the beam deflection data in Fig. 7.27 and 7.28, on page 215, the data includes high frequency effects as well as the lower frequencies associated with the pressure wells in the rollers. This may be caused by smaller irregularities and vortices that roll up into the larger structures; however, the boundary layer forming and growing along the upper surface of the test section with the high speed flow is also a likely contributor in this regard.¹² The effect of the aperture with T/T correction does not remove these smaller scale optical distortions. Despite these differences, applying the same practice of non-

dimensionalization to the data produces a similar collapse onto a single curve. This is very encouraging as the data used in this figure is taken from an unforced shear layer for which the sinusoidal approximation is questionable, but the results arrived at using that approximation still apply.



Figure 8.10: Non-dimensionalized Shear Layer Results.

8.4. The Aperture Filter

The phenomenon of a spatial filter producing a clean image or far-field pattern has been observed experimentally for some time, and can be traced back thousands of years in the writings of various natural philosophers concerning variations on the pinhole camera. As was shown in chapter 2 and in section 8.3, higher-frequency features of a wavefront tend to produce sidelobes in a far-field pattern, which is also seen when the light is brought to a focal point by a lens or other means. A small aperture placed at such a focal point will physically block these sidelobes, and the wavefront propagating away from this aperture will likewise lack the higher-frequency features that produced those sidelobes.

The examples of T/T correction for sinusoidal wavefronts in sections 8.3 and 8.3.1 demonstrated that passing a wavefront through an aperture can also remove features, or at least change how those features are seen over the limited area of the aperture. Because of this, T/T correction within an aperture tends to have different effects for different relative sizes of the disturbance to be corrected, Λ , and the aperture over which the correction is to be performed, A_p . The aperture serves as a spatial filter, separating the effects of disturbances into those that are primarily T/T when $\Lambda > A_p$, and those that would require higher-order correction effects when $\Lambda < A_p$. If the sources of these distortions are in motion relative to the optical path then this spatial length scale will be associated with a temporal frequency by the relationship

$$f = \frac{U_C}{\Lambda},\tag{8.13}$$

in which $U_{\rm C}$ indicates the convective velocity perpendicular to the optical path.

8.4.1. Filter Gain

Filters are commonly characterized by their frequency response. For physical systems, this function can be found experimentally by feeding signals of known frequency into the system and measuring the result. The gain of the frequency response is the ratio between the input signal and the resulting output. This approach follows in the footsteps of Greenwood^{13,14} in dealing with a corrective system in terms of how it deals with different frequencies of disturbance, as was described in section 4.3.2. It also

has parallels with an approach used and expanded upon by Robert Tyson.^{15,16} However, both the original analysis by Greenwood and the ongoing work by Tyson have been oriented primarily towards the atmospheric turbulence, and once a set of guidelines for that case has been developed, the analysis and principles used in developing those guidelines tend to be forgotten. Additionally, Greenwood and Fried treated the problem stochastically, based on structure functions, while Tyson's work is based on Fourier transforms. Although these approaches have some benefits in performing the associated mathematical operations, I prefer to work in the domains of standard time and space as an aid to understanding what is happening physically.

The phenomenon of a spatial filter cleaning up an image or far field pattern has been observed experimentally for some time, and can be traced back thousands of years in the writing of various natural philosophers concerning variants of the pinhole camera. On the other hand, while ; a more precise definition of the can for this filter can be found analytically. For a one-dimensional wavefront in the form of OPD(x,t), observed over an aperture of size A_p , the OPD_{rms} is:

$$OPD_{rms}(A_p, t) = \sqrt{\frac{1}{A_p} \int_{-A_p/2}^{A_p/2} [OPD(x, t)]^2 dx}, \qquad (8.14)$$

The residual OPD_{rms} over the aperture after T/T removal is

$$OPD_{rms}(A_p, t)_{T/T} = \sqrt{\frac{1}{A_p} \int_{-A_p/2}^{A_p/2} [OPD(x, t) - (A(t) + xB(t))]^2} dx , \qquad (8.15)$$

where *A* and *B* are coefficients defining the tilt and piston being removed. Z-tilt is defined as a linear fit to the wavefront and corresponds to values of *A* and *B* that minimize Eq. 8.15. These values can be found by taking the derivative of Eq. 8.15 with

respect to these coefficients, which will identify potential minimums at locations where these derivatives equal zero. For this purpose is sufficient to do this for just the integral within Eq. 8.15:

$$\frac{\partial}{\partial A} \int_{-A_p/2}^{A_p/2} [OPD(x,t) - (A(t) + xB(t))]^2 dx = \int_{-A_p/2}^{A_p/2} 2[-OPD(x,t) + (A(t) + xB(t))] dx = 0, \quad (8.16)$$

$$\frac{\partial}{\partial B} \int_{-A_p/2}^{A_p/2} [OPD(x,t) - (A(t) + xB(t))]^2 dx = \int_{-A_p/2}^{A_p/2} [-OPD(x,t) + (A(t) + xB(t))] x dx = 0.$$
(8.17)

For a given wavefront, OPD(x,t), this is a set of two equations and two unknowns with the solution:

$$A = \frac{\int_{-A_p/2}^{A_p/2} OPD(x,t)}{A_p},$$
(8.18)

$$B = \frac{12 \int_{-A_p/2}^{A_p/2} OPD(x,t)}{A_p^3}.$$
(8.19)

As addressed in section 8.2, the expected wavefront produced by propagation through a shear layer may be approximated with a sine function. As a frequency response is sought for this effect, it makes sense to do so again;

$$OPD(x,t) = K \sin\left[2\pi \left(f \cdot t - \frac{x}{\Lambda}\right)\right] = K \sin\left[2\pi \left(\frac{U_C}{\Lambda} \cdot t - \frac{x}{\Lambda}\right)\right].$$
(8.20)

With this substitution, the minimizing values for *A* and *B* are found to be:

$$A = \frac{1}{2} \frac{K\Lambda \left[\cos \left(\frac{\pi \left(-A_p + 2U_C t \right)}{\Lambda} \right) - \cos \left(\frac{\pi \left(-A_p + 2U_C t \right)}{\Lambda} \right) \right]}{\pi A_p}, \qquad (8.21)$$

$$B = -\frac{1}{\pi^2 A_p^{-3}} \left\{ 3K\Lambda \left[sin\left(\frac{\pi \left(-A_p + 2U_C t\right)}{\Lambda}\right)\Lambda + \pi cos\left(\frac{\pi \left(-A_p + 2U_C t\right)}{\Lambda}\right)A_p \right] - sin\left(\frac{\pi \left(A_p + 2U_C t\right)}{\Lambda}\right)\Lambda + \pi cos\left(\frac{\pi \left(A_p + 2U_C t\right)}{\Lambda}\right)A_p \right] \right\}.$$
 (8.22)

With Eqs. 8.20, 8.21, and 8.22, expressions for the uncorrected and corrected OPD_{rms} of Eqs. 8.14 and 8.15 can be found. However, a more meaningful quantity for comparison is $(OPD_{rms})^2$. As was addressed in chapter 3, if a wavefront is represented as a one-dimensional waveform, as has been done in this analysis, then the net disturbances in the wavefront and the power spectrum of that waveform (PSD_W) are related by

$$OPD_{rms}^{2} = \int_{-\infty}^{\infty} PSD_{W}(f) df . \qquad (8.23)$$

As noted previously, the approach by Greenwood¹⁴ in defining the frequency that bears his name was to treat a corrective system as having a transfer function. If that transfer function is defined as G(f), then the disturbance remaining after correction is applied will be

$$OPD_{rms}^{2}(A_{p})_{T/T} = \int_{-\infty}^{\infty} G(A_{p}, f) PSD_{W}(f) df .$$
(8.24)

The power spectrum of a sinusoid with amplitude *K* and frequency U_C/Λ , such as the one defined in Eq. 8.20 and used for the wavefront, would be:

$$P_{W}(f) = \frac{K^{2}}{4} \left[\delta \left(f - \frac{U_{C}}{\Lambda} \right) + \delta \left(f + \frac{U_{C}}{\Lambda} \right) \right].$$
(8.25)

One of the defining characteristics of the Dirac delta function (δ) is that an integration of a product of this function with another function equals the value of the second function at the "trigger point" of the delta function. Therefore, the expected uncorrected OPD_{rms} would be

$$OPD_{rms}^{2} = \int_{-\infty}^{\infty} \frac{K^{2}}{4} \left[\delta \left(f - \frac{U_{c}}{\Lambda} \right) + \delta \left(f + \frac{U_{c}}{\Lambda} \right) \right] df = \frac{1}{2} K^{2}.$$
(8.26)

The power spectrum of Eq. 8.25 assumes an effectively infinite aperture, but this is also the result found in averaging the OPD_{rms} across a finite aperture over one or more full cycles in time.

$$\overline{OPD}_{rms}^{2}\left(A_{p}\right) = f \int_{0}^{\frac{1}{f}} \frac{1}{A_{p}} \int_{\frac{Ap}{2}}^{\frac{Ap}{2}} K^{2} \sin^{2}\left[2\pi\left(f \cdot t - \frac{x}{\Lambda}\right)\right] dx dt = \frac{1}{2}K^{2}.$$
(8.27)

This is in line with Parseval's identity:

$$\int_{-\infty}^{\infty} \left| h(t) \right|^2 dt = \int_{-\infty}^{\infty} \left| \hat{h}(f) \right|^2 df , \qquad (8.28)$$

for which $\hat{h}(f)$ is the Fourier transform of h(t).

A transfer function, such as $G(A_p, f)$ is commonly defined as a ratio of an output from the system to the input to the system that produced that output. In terms of timeaveraged values, this would be

$$G(A_p, f) = \frac{\overline{OPD_{rmw}^2}(A_p, f)_{T/T}}{\overline{OPD_{rmw}^2}(A_p, f)} = \frac{2 \cdot \overline{OPD_{rmw}^2}(A_p, f)_{T/T}}{K^2}.$$
(8.29)

This also fits with the spectral approach,

$$OPD_{rms}^{2}\left(A_{p}, f = \frac{U_{C}}{\Lambda}\right)_{T/T} = \int_{-\infty}^{\infty} G(A_{p}, f) \frac{K^{2}}{4} \left[\delta\left(f - \frac{U_{C}}{\Lambda}\right) + \delta\left(f + \frac{U_{C}}{\Lambda}\right)\right] df$$
$$= \frac{K^{2}}{4} \left[G\left(A_{p}, -\frac{U_{C}}{\Lambda}\right) + G\left(A_{p}, \frac{U_{C}}{\Lambda}\right)\right].$$
(8.30)

Assuming that $G(A_p, f) = G(A_p, f)$, which is true for most frequency-dependent systems,

$$OPD_{rms}^{2}\left(A_{p}, f = \frac{U_{C}}{\Lambda}\right)_{T/T} = \frac{1}{2}K^{2}G\left(A_{p}, f = \frac{U_{C}}{\Lambda}\right).$$

$$(8.31)$$

Which indicates that

$$G\left(A_{p}, f = \frac{U_{C}}{\Lambda}\right) = \frac{2 \cdot OPD_{rms}^{2} \left(A_{p}, f = \frac{U_{C}}{\Lambda}\right)_{T/T}}{K^{2}},$$
(8.32)

which is the same result as was found by the use of time-averaged values in Eq. 8.29.

Combining equations 8.15 through 8.22 and averaging over time produces the result of

$$\overline{OPD_{rms}^{2}}(A_{p})_{T/T} = K^{2} \left[\frac{1}{2} - \frac{\frac{1}{2} + \cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)}{\left(\frac{\pi A_{p}}{\Lambda}\right)^{2}} + \frac{3\sin\left(\frac{\pi A_{p}}{\Lambda}\right)\cos\left(\frac{\pi A_{p}}{\Lambda}\right)}{\left(\frac{\pi A_{p}}{\Lambda}\right)^{3}} - \frac{3}{2} \frac{\sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)}{\left(\frac{\pi A_{p}}{\Lambda}\right)^{4}} \right].(8.33)$$

The non-denominational quantity A_p/Λ appearing in Eq. 8.33 is useful for envisioning what happens in the aperture during correction, with values less than one indicating circumstances like those shown in Fig. 8.5 while a ratio greater than one would correspond to the shorter distortion length scales shown in Fig. 8.6. Additionally, by the relationship $f = U_C/\Lambda$, it acts as a non-denominational indicator of frequencies of the disturbance in the form of a Strouhal number based on the length scale of the aperture.

$$\frac{A_p}{\Lambda} = \frac{A_p}{U_c / f} = \frac{A_p f}{U_c} = St_{Ap}.$$
(8.34)

Applying the definitions for the transfer function from Eqs. 8.29 and 8.32 leads to the following expression for the effective gain of T/T correction over a one-dimensional aperture:

$$G(St_{Ap}) = 1 - \frac{1 + 2\cos^2(\pi St_{Ap})}{(\pi St_{Ap})^2} + \frac{6\sin(\pi St_{Ap})\cos(\pi St_{Ap})}{(\pi St_{Ap})^3} - \frac{3\sin^2(\pi St_{Ap})}{(\pi St_{Ap})^4}.$$
 (8.35)

This gain function is plotted in Fig. 8.11, along with markers to show that the 50% (3-dB) cutoff of this system is found around $St_{Ap} \cong 0.85$.



Figure 8.11: System gain for T/T correction over a one-dimensional aperture.

A common practice in dealing with a filter frequency response or other form of system transfer function is to plot it in log-log scale or decibels against a logarithmic scale, as had been done for Eq. 8.35 in Fig. 8.12. In this form, it is often easy to see how

an otherwise complicated function can be approximated by simpler functions of the form $a(x^b)$ over certain ranges of x. In Fig. 8.11, it is fairly obvious that for values of St_{Ap} greater than the cutoff point, $G(St_{Ap})$ approaches a value of one, or 0 dB, as T/T correction becomes ineffective in dealing with disturbances of a shorter length scale. In Fig. 8.12, it can be seen that for values of St_{Ap} smaller than the cutoff point, the function has a logarithmic "slope" of 40 dB per decade, indicating that it can be approximated as a function of St_{Ap} to the 4th power.



Figure 8.12: System gain for T/T correction over a one-dimensional aperture.

More specifically, a Taylor series expansion of Eq. 8.35, or of the trigonometric functions within that equation, reveals that

$$G(St_{Ap}) \cong \frac{\pi^4 St_{Ap}^4}{45} \cong 2.16St_{Ap}^4$$
(8.36)

for small values of St_{Ap} .

8.4.2. Two-Dimensional Circular Aperture Filter

The shear layer experiments that were performed at AEDC and the ND WCSL facility described in this dissertation have dealt with one-dimensional wavefronts aligned in the direction of the flow. However, the majority of applications requiring T/T correction involve a beam or optical viewing path with a two-dimensional cross-section, which is often circular. The analysis of the proceeding section for a one-dimensional aperture can be repeated for T/T corrections over a two-dimensional circular aperture, replacing Eqs. 8.14 and 8.15 with

$$OPD_{rms}(A_p,t) = \sqrt{\frac{4}{\pi A_p^2} \int_{-\sqrt{(A_p/2) - x^2}}^{\sqrt{(A_p/2) - x^2}} \int_{-A_p/2}^{A_p/2} [OPD(x, y, t)]^2 dxdy}$$
(8.37)

and

$$OPD_{rms}(A_p,t) = \sqrt{\frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{\sqrt{(A_p/2)-x^2}}^{\sqrt{(A_p/2)-x^2}} OPD(x,y,t) - [A(t) + xB(t) + yC(t)] \right\}^2 dydx} \quad (8.38)$$

As with the one-dimensional aperture, the definition of Z-tilt correction is such that the coefficients A, B, and C are set to values that minimize the integral in Eq. 8.38. The full derivation of the gain function for Z-tilt correction over a circular aperture can be found in section B.4.2 of appendix B. For the purposes of the exploration in this chapter, it is sufficient to provide the resulting gain function found at the end of this derivation, which is:

$$G(St_{Ap}) = 1 - \frac{16J_0^2(\pi St_{Ap}) + 4J_1^2(\pi St_{Ap})}{(\pi St_{Ap})^2} + \frac{64J_0(\pi St_{Ap})J_1(\pi St_{Ap})}{(\pi St_{Ap})^3} - \frac{64J_1^2(\pi St_{Ap})}{(\pi St_{Ap})^4}.$$
 (8.39)

in which J_0 and J_1 indicate Bessel functions of the first kind. The two-dimensional gain is plotted in Fig. 8.13 and has a 3-dB cut-off at $St_{Ap} \cong 0.93$.



Figure 8.13: System gain for T/T correction over a two-dimensional aperture.

Plotting the gain in decibels and a Taylor series expansion indicates that the function $G = 1.52(St_{Ap})^4$ serves as an approximation for Eq. 8.39 below the cutoff frequency.



Figure 8.14: System gain for T/T correction over a two-dimensional aperture.

8.4.3. G-Tilt Aperture Filters

As noted previously, Z-tilt is not the only definition of tilt. G-tilt is defined as the average of the local gradients at each point of the wavefront over the aperture, so that

$$G-tilt(t) = \frac{-\int\limits_{-A_p/2}^{A_p/2} \frac{d}{dx} OPD(x,t) dx}{\frac{A_p}{2} - \left(-\frac{A_p}{2}\right)} = -\left[\frac{OPD\left(\frac{A_p}{2},t\right) - OPD\left(-\frac{A_p}{2},t\right)}{A_p}\right].$$
 (8.40)

The negative sign in front of the integral in Eq. 8.40 reflects the fact that *OPD* is the conjugate of the wavefront, having equal magnitude (for a medium in which $n \cong 1$) but opposite sign.

For correction of G-tilt, one can again define a one dimensional aperture and correction over that aperture such that

$$OPD_{rms}(A_p, t) = \sqrt{\frac{1}{A_p} \int_{-A_p/2}^{A_p/2} [OPD(x, t)]^2} dx , \qquad (8.41)$$

and

$$OPD_{rms}(A_p, t)_{T/T} = \sqrt{\frac{1}{A_p} \int_{-A_p/2}^{A_p/2} [OPD(x, t) - (A(t) + xB(t))]^2} dx .$$
(8.42)

However, for correcting G-tilt as defined in Eq. 8.40, the coefficient B is selected so that

$$_{corrected}G - tilt(t) = -\frac{\left[OPD\left(\frac{A_p}{2}, t\right) - \frac{A_p}{2}B(t)\right] - \left[OPD\left(-\frac{A_p}{2}, t\right) + \frac{A_p}{2}B(t)\right]}{A_p} = 0, \quad (8.43)$$

which leads to

$$B(t) = \frac{OPD\left(\frac{A_p}{2}, t\right) - OPD\left(-\frac{A_p}{2}, t\right)}{A_p}.$$
(8.44)

The coefficient A(t) corresponds to piston and has no effect on G-tilt. G-tilt correction is primarily intended to shift the centroid of a far-field pattern to a target point, and average piston has no effect on the far-field pattern. Technically, piston does not affect Z-tilt either, except that piston corresponds to the Z_0^0 Zernike mode which is normally removed before defining the Z_1^1 and Z_1^{-1} modes which correspond to tilt. For the sake of comparison with Z-tilt, A(t) will be defined as to remove the mean piston from the corrected wavefront. The rest of this derivation can be found in detail in section B.5.1 of appendix B. The end result of this derivation is a gain function of

$$G(St_{Ap}) = \frac{1}{3} \left[4 - \cos^2(\pi St_{Ap}) + 6 \frac{\sin(\pi St_{Ap})\cos(\pi St_{Ap})}{(\pi St_{Ap})} - 9 \frac{(1 - \cos^2(\pi St_{Ap}))}{(\pi St_{Ap})^2} \right].$$
(8.45)

Equation 8.45 is plotted in Fig. 8.15, and has a 50% cutoff of $St_{Ap} \approx 0.77$. Of particular note is that for $St_{Ap} > 1$, the gain becomes larger than one. This means that when dealing with distortions in the wavefront with a length scale smaller than the aperture, not only is G-tilt correction not effective in lowering the average OPD_{rms} within the aperture, but it may actually increase it, which is likely to produce a reduction in Strehl ratio according to the large-aperture approximation of Eq. 8.1.



Figure 8.15: System gain for G-tilt correction over a one-dimensional aperture.

Figure 8.16 shows an illustration of why this is so. By definition, Z-tilt minimizes OPD_{rms} , but this is not the definition of G-tilt. Correcting G-tilt for a sinusoidal waveform with a period shorter than the aperture over which the correction is to take place, as shown in Fig. 8.16, may actually induce a significant degree of Z-tilt that was not present and so increase the overall *OPD*. Note that the gain in Fig. 8.15 goes to unity

for integer values of St_{Ap} and tends to be greater than one for other values where the disturbance length scale does not divide evenly into the aperture length.



Figure 8.16: G-tilt correction for $A_p > \Lambda$.

As G-tilt correction moves the centroid of the far-field pattern to a fixed spot, a view of the effects in the far field may also aid in understanding this phenomenon. Figure 8.17 shows the far-field intensity patterns for the corrected and uncorrected wavefronts of Fig. 8.16. The Fourier-transform aspects of far-field propagation cause energy to be transferred from the central lobe to sidelobes in this pattern. Shorter length scales of wavefront distortions are associated with sidelobes further from the main lobe or intended target point. The way in which the aperture in Fig. 8.16 contains a non-integer number of cycles of the disturbance causes more energy to be transferred to one side than the other. Under such circumstances, the centroid of the pattern is located partway between the peak of the center lobe and the peak of the large sidelobe. Shifting

the pattern to put the centroid on the target point actually shifts the main energy-

containing lobe of the pattern off of the target point, reducing the intensity at that point.



Figure 8.17: Far-field effects of G-tilt correction for short length scales.

Figure 8.18 plots Eq. 8.45 in log-decibel form, and a Taylor series expansion indicates that $G = 2.16(St_{Ap})^4$ serves as an approximation for G-tilt correction below the cutoff frequency.



Figure 8.18: System gain for G-tilt correction over a one-dimensional aperture.

As with Z-tilt, G-tilt correction can also be applied over a circular aperture, and in fact most applications outside the laboratory will involve apertures of this sort. The full derivation is presented in section B.5.2 of appendix B, with the end result of

$$G(St_{Ap}) = 1 + J_1^2 \left(\pi St_{Ap} \right) + \frac{8J_0 \left(\pi St_{Ap} \right) J_1 \left(\pi St_{Ap} \right)}{\left(\pi St_{Ap} \right)} - \frac{20J_1^2 \left(\pi St_{Ap} \right)}{\left(\pi St_{Ap} \right)^2}.$$
 (8.46)

Equation 8.46 is plotted in Fig. 8.19 and shows the same sort of behavior with higher values of St_{Ap} as was seen for the one-dimensional case, in that it reaches values greater than one. On the other hand, the amplitude of the initial overshoot past $G(St_{Ap}) = 1$ is not as large and the function seems to approach a steady value of $G(St_{Ap}) \cong 1$ more quickly than the one-dimensional case. The 50% cutoff for this function is found at $St_{Ap} \cong 0.88$



Figure 8.19: System gain for G-tilt correction over a circular aperture.

A Taylor series approximation and plotting the gain in log-decibel form indicates that $G = 1.52(St_{Ap})^4$ is a good approximation for Eq. 8.46 with small values of St_{Ap} . Interestingly, the same functions found as approximations for Z-tilt correction across onedimensional and circular apertures serve as an adequate approximation for G-tilt over roughly the same range. So it would seem that the tendency to equate G-tilt and Z-tilt in many works and applications is not unreasonable, but only over a range of values in St_{Ap} that corresponds to disturbances that are primarily T/T in nature.



Figure 8.20: System gain for G-tilt correction over a circular aperture.

8.4.4. Piston-Only Correction

Before pressing onward to higher-order forms of correction, it may be useful to take a step back, and look at correction for piston alone. As will be addressed later, some higher-order corrective systems are based on localized piston-only correction.

Additionally, there are some applications that are naturally piston-corrected in that piston has no effect on the aspects of interest. This is notably true when resolution or intensity in the far-field is the goal. The far-field intensity pattern produced by a nearfield wavefront is unaffected by the average phase or *OPL* of that wavefront. A wavefront of constant phase across an aperture will produce a diffraction-limited pattern in the far field. It does not matter what that phase is, provided it is the same value at all points in the aperture. Only variations in phase and deviations from this mean value have any effect on the far-field pattern. This is why the approximation for the Strehl ratio in Eq. 8.1 is known as the large aperture approximation; it assumes that the aperture is larger than the length-scale of the optical distortions. If the length-scale of the distortions is larger than the aperture, then those distortions may manifest partly or even primarily as piston across the aperture.

Therefore, if Strehl ratio or other aspects of the far field intensity pattern are of primary concern, then a wavefront passing through a finite aperture may already be considered piston-corrected, and is subject to the same sort of filtration effect as was seen for T/T correction. The derivation is similar to that for Z-tilt or G-tilt correction but setting the B(t) coefficient for tilt to zero, and can be found in full detail in sections B.3.1 and B.3.2 of appendix B. The end results of these derivations are gain functions of

$$G(St_{Ap}) = 1 + \frac{\cos^2(\pi St_{Ap}) - 1}{(\pi St_{Ap})^2}$$
(8.47)

for a one-dimensional aperture and

$$G(St_{Ap}) = 1 - \frac{4J_1^2(\pi St_{Ap})}{(\pi St_{Ap})^2}$$
(8.48)

for a circular aperture.

Both these gain functions are plotted in Fig. 8.21. The one-dimensional case has a 50% cutoff of $St_{Ap} \cong 0.45$ and the gain for a circular aperture has its cutoff at $St_{Ap} \cong 0.52$. Also of potential interest are the points at which the functions reach 90%, which are $St_{Ap} \cong 0.74$ for the one-dimensional aperture and $St_{Ap} \cong 0.87$ for the circular. These may be used as an indicator of the beginning of the range for which the large-aperture approximation of Eq. 8.1 applies. For that matter, whenever integrating a power spectrum to find the phase variance, as in Eq. 8.23, for the purpose of using the phase variance in the large aperture approximation of Eq. 8.1, the spectrum should be filtered with the appropriate piston-correction gain function. The piston-component of wavefront variations has no effect on the far-field pattern or Strehl ratio.



Figure 8.21: System gain for piston correction over one-dimensional (left) and circular (right) apertures.

The piston-correction gain functions are plotted again in Fig. 8.22 in decibel vs. log-scale form, with curves of $G = 3.29(St_{Ap})^2$ and $G = 2.47(St_{Ap})^2$ serving as adequate approximations for these functions below their cutoff values of St_{Ap} .



Figure 8.22: System gain for piston correction over one-dimensional (left) and circular (right) apertures

8.5. Higher-Order Correction

In some applications, particularly ones intended to track an object or target of some sort, T/T correction may be the only form of correction applied. In other applications, such as the Notre Dame Adaptive Optics System described in Chapter 3, T/T correction may be a single stage in the process. Figures 8.5 and 8.6 show examples of a properly of T/T compensation, in that it does not change the intensity pattern seen in the far field, but can only shift its location. Tip-tilt and tracking serve to center a beam or a sensor's field of view onto a desired target, but improving the overall quality of the beam or received image requires higher order correction of some sort.

In Notre Dame's in-house system, a deformable mirror from Xinetics is controlled by actuators on the non-reflective side. This is a common approach that can be found in other applications.^{17,18} In these applications, the size of the entire aperture becomes less important than the distance between actuators within the aperture. Thus, for the purposes of analysis, it becomes practical to divide the aperture up into a number of sub-apertures, with each sub-division either centered on a single actuator or encompassing the region between actuators.

8.5.1. Basis-Spline Correction Fitting

Guidelines for higher-order correction of atmospheric distortions based on the effects of local actuators already exist and were mentioned in section 4.3.1. In particular there is the equation

$$\sigma_r^2 = \kappa \left(\frac{r_s}{r_0}\right)^{5/3},\tag{8.49}$$

in which r_0 is the Fried parameter for atmospheric propagation defined in section 4.3.1, r_s is the spacing between actuators in this piston-only corrective system, κ is a constant that depends on the type of correction being performed, and σ_r^2 is the residual phase variance after correction.

Equation 8.49 comes from an analysis of fitting error performed by Richard Hudgin¹⁹ in the 1970's. In this analysis, the correction to a wavefront is created by summation of a set of basis functions. This is also known as a basis-spline or B-spline fit, in which the fit to a curve or data set is built up out of a set of basis functions, represented by $R_j(x,y)$. The fit to the curve encountered at a given time is produced by scaling each basis function by some value appropriate to that curve, $S_j(t)$, and the fit to the curve will be

$$Fit(x, y, t) = \sum_{j=1}^{N} \left(S_{j}(t) R_{j}(x, y) \right).$$
(8.50)

An example of a B-spline fit is shown in Fig. 8.23, using a set of basis functions in the form of overlapping triangles to produce a fit in the form of a series of connected line segments.



Figure 8.23: B-spline curve fitting: A set Basis functions (left) are scaled and summed to produce a least-squares fit to a given curve or set of data points. (right)

In the case of wavefront correction, this curve may be OPD(x,y,t) and if so then the residual disturbances in the wavefront after correction with this fit will be described by

$$_{cor} OPD(x, y, t) = OPD(x, y, t) - Fit(x, y, t) = OPD(x, y, t) - \sum_{j=1}^{N} \left(S_{j}(t) R_{j}(x, y) \right)$$
(8.51)

and

$$_{cor} OPD(t)_{rms}^{2} = \frac{1}{A} \int_{Ap} \left[OPD(x, y, t) - \sum_{j=1}^{N} \left(S_{j}(t) R_{j}(x, y) \right) \right]^{2} dx dy$$
(8.52)

where *A* is the area of the aperture.

The basis functions, $R_j(x,y)$, can be of any form, but some work better than others for specific applications. In designing a corrective system with *N* actuators, it is common practice to have *N* basis functions, with each function representing the effect and influence that a particular actuator will have on the system. In performing the corrective fit with a basis set of this form, it becomes a relatively simple matter to map the corrective fit onto the control signals needed to provide that correction, as the proper control for the *j*th actuator will then be a function of the corresponding coefficient of fitting, $S_j(t)$.

An optimal fit is one with values of $S_j(t)$ to minimize Eq. 8.52 for each value of t. As was done with the T/T correction, the optimal values of these coefficients will correspond to values for which the derivative of Eq. 8.52 with respect to that coefficient is zero.

$$\frac{\partial_{cor} OPD(t)_{rms}^2}{\partial S_k} = 0 = \frac{1}{A} \int_{A_p} 2 \left[OPD(x, y, t) - \sum_{j=1}^N \left(S_j(t) R_j(x, y) \right) \right] \left[-R_k(x, y) \right] dxdy.$$
(8.53)

This leads to the equation

$$\frac{1}{A} \int_{A_p} OPD(x, y, t) R_k(x, y) dx dy = \sum_{j=1}^N \left[S_j(t) \frac{1}{A} \int_{A_p} R_j(x, y) R_k(x, y) dx dy \right].$$
(8.54)

Using the following notation:

$$OPD(t)_{rms}^{2} = \frac{1}{A} \int_{Ap} [OPD(x, y, t)]^{2} dx dy, \qquad (8.55)$$

$$B_{j}(t) = \frac{1}{A} \int_{Ap} OPD(x, y, t) R_{j}(x, y) dx dy, \qquad (8.56)$$

$$C_{jk} = \frac{1}{A} \int_{Ap} R_j(x, y) R_k(x, y) dx dy, \qquad (8.57)$$

Eq. 8.52 can be rewritten as

$$_{cor} OPD(t)_{rms}^{2} = OPD(t)_{rms}^{2} - 2\sum_{j=1}^{N} B_{j}(t)S_{j}(t) + \sum_{j=1,k=1}^{N} C_{jk}S_{j}(t)S_{k}(t)$$
(8.58)

and Eq. 8.54 can be written as

$$B_{k} = \sum_{j=1}^{N} \left[S_{j}(t) C_{jk} \right].$$
(8.59)

The indexed sets of B_k , S_j , and C_{jk} can also be treated as vectors and matrices, and the set of equations represented by Eq. 8.59 can be written as

$$\vec{B}(t) = \vec{C}\vec{S}(t)$$
. (8.60)

Technically the matrix \vec{C} in Eq. 8.60 should be \vec{C}^T , the transpose of \vec{C} . However, by Eq. 8.57, $C_{jk} = C_{kj}$ which indicates \vec{C} is a symmetric matrix, for which $\vec{C}^T = \vec{C}$. Likewise, Eq. 8.58 can be written as

$$_{cor} OPD(t)_{rms}^{2} = OPD(t)_{rms}^{2} - 2\vec{B}(t)^{T}\vec{S}(t) + \vec{S}(t)^{T}\vec{C}\vec{S}(t) .$$
(8.61)

The solution to Eq. 8.60 is quite clearly

$$\vec{S}(t) = \vec{C}^{-1}\vec{B}(t)$$
. (8.62)

Substituting this into Eq. 8.61, the minimum achievable wavefront distortion after correction is

$$\begin{bmatrix} _{cor} OPD(t)_{rms}^{2} \end{bmatrix}_{min} = OPD(t)_{rms}^{2} - 2\vec{B}(t)^{T}\vec{C}^{-1}\vec{B}(t) + \vec{B}(t)^{T}\vec{C}^{-1}\vec{C}\vec{C}^{-1}\vec{B}(t)$$
$$= OPD(t)_{rms}^{2} - \vec{B}(t)^{T}\vec{C}^{-1}\vec{B}(t) = OPD(t)_{rms}^{2} - \sum_{j=1,k=1}^{N} C_{jk}^{-1}B_{j}(t)B_{k}(t).$$
(8.63)

The analysis by Hudgin¹⁹ continues from this point with a focus on aberrations produced by the atmosphere, characterized by Kolmogorov turbulence. As turbulence of this type is not periodic or predictable except in the form of ensemble averages, his analysis deals with an ensemble average of the residual wavefront disturbances, in the form

$$\left\langle _{cor} OPD(t)_{rms}^{2} \right\rangle = \left\langle OPD(t)_{rms}^{2} \right\rangle - \sum_{j=1,k=1}^{N} C_{jk}^{-1} \left\langle B_{j}(t)B_{k}(t) \right\rangle.$$
(8.64)

In section 4.2.3, a structure function of the form

$$D_A((x_2, y_2) - (x_1, y_1)) = \left\langle \left[A((x_2, y_2), t) - A((x_1, y_1), t) \right]^2 \right\rangle$$
(8.65)

was defined to describe the average difference in some quantity *A* at points separated by $\vec{r} = (x_2, y_2) - (x_1, y_1)$. In the case of homogeneous and isotropic turbulence, this becomes a function of the separation distance, $r = |\vec{r}|$, as the location of the first point and direction to the second point become irrelevant. In terms of a structure function for the *OPD* (*D*_{*OPD*}), the residual differences can be written as

$$\left\langle _{cor} OPD(t)_{rms}^{2} \right\rangle = \frac{1}{2A^{2}} \int_{Ap} D_{OPD}(r) dx dy$$
$$+ \frac{1}{2A^{2}} \sum_{j=1,k=1}^{N} C_{jk}^{-1} \int_{Ap} R_{j}(x_{1}, y_{1}) R_{k}(x_{1}, y_{1}) D_{OPD}(r) dx dy .$$
(8.66)

As was addressed in section 4.3.1, the structure function for wavefront variations produced by Kolmogorov turbulence, as seen in a receiving aperture, is of the form

$$D_{\varphi}(r) + D_A(r) = 6.88(r/r_0)^{5/3}$$
(8.67)

and the mean-squared phase variation seen over an aperture of size A_p with atmosphereinduced distortions is

$$\sigma_{\varphi}^{2} = 1.04 \left(\frac{A_{p}}{r_{0}}\right)^{5/3}.$$
(8.68)

Therefore, atmosphere-induced OPD_{rms} over an aperture is such that

$$OPD_{rms}^{2} = \left(\frac{\lambda}{2\pi}\right)^{2} \sigma_{\varphi}^{2} = \left(\frac{\lambda}{2\pi}\right)^{2} 1.04 \left(\frac{A_{p}}{r_{0}}\right)^{5/3}.$$
(8.69)

From this, Hudgin arrived at a prediction for residual wavefront distortions of the form

$$\left\langle _{cor} OPD(t)^2_{rms} \right\rangle \left(\frac{2\pi}{\lambda} \right)^2 = \sigma_{\varphi}^2 = \kappa \left(\frac{r_s}{r_0} \right)^{5/3}$$
 (8.70)

where r_s represents the spacing between the actuators producing each basis function of R_j in the corrective system and the constant κ is determined by the basis functions being used in the corrective fit.

8.5.2. Piston-Only Sub-Apertures

A relatively simple form of higher-order correction, and a good starting point for exploration of this subject, is a system that only corrects for piston, but provides this correction on a local basis in many small regions over the aperture. This is easily visualized as a set of independent flat mirrors on pistons, with each mirror being moved up or down to intercept the local portion of the wavefront at the appropriate point. In actual implementation, this may instead take the form of an array of pixels made of liquid crystal or other materials in which the index of refraction can be altered by applying voltage or current. This allows the relative phase of light passing through each pixel to be adjusted on command.²⁰

In dealing with aero-optic disturbances, time averages will serve the same function as ensemble averages did for atmospheric disturbances in Hudgin's analysis.

$$\overline{_{cor} OPD(t)_{rms}^2} = \overline{OPD(t)_{rms}^2} - \overline{\vec{B}(t)}^T \vec{C}^{-1} \vec{B}(t) = \overline{OPD(t)_{rms}^2} - \sum_{j=1,k=1}^N \overline{C_{jk}^{-1} B_j(t) B_k(t)} .$$
(8.71)

In using a sinusoidal wavefront of the form

$$OPD(x, y, t) = K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right),$$
 (8.72)

this time average can be found by integrating a function over the rage $\varphi = -\pi$ to $\varphi = \pi$ and then dividing by 2π .

For correction of piston and piston alone over sub-apertures, the basis function may be written as

$$R_{j}(x) = \begin{cases} for\left(X_{j} - \frac{sub}{2}A_{p}\right) \leq x < \left(X_{j} + \frac{sub}{2}A_{p}\right) \\ and\left(Y_{j} - \frac{sub}{2}A_{p}\right) \leq y < \left(Y_{j} + \frac{sub}{2}A_{p}\right) \\ 0 & otherwise \end{cases}$$
(8.73)

The parameters X_j and Y_j represent the center of this basis function and presumably correspond to the location of an actuator driving the local piston correction in a physical system of this sort. This analysis assumes that the actuators and corrective segments are arranged in a rectangular grid with spacing such that each actuator is a distance of _{sub} A_p from each of its closest neighbors.

Considering a one-dimensional aperture, if the aperture is divided into subapertures of equal size, then the number of sub-apertures will be

$$N = \frac{A_p}{_{sub}A_p}.$$
(8.74)

The center location (X_i) of each sub-aperture will then be

$$X_{j} = -\frac{A_{p}}{2} + \frac{sub}{2} A_{p} (j-1), \qquad (8.75)$$
$$B_{j}(t) = \frac{1}{A_{p}} \int_{-sub}^{sub} \int_{A_{p}/2+X_{j}}^{A_{p}/2+X_{j}} K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right)(1) dx$$

$$=\frac{K}{2}\frac{\Lambda}{\pi A_{p}}\left[\cos\left(\frac{-\pi A_{p}+2\pi_{sub}A_{p}(j-1)+\Lambda\varphi}{\Lambda}\right)-\cos\left(\frac{-\pi A_{p}+2\pi_{sub}A_{p}(j)+\Lambda\varphi}{\Lambda}\right)\right],\quad(8.76)$$
and

$$C_{jk} = \frac{1}{A_p} \int_{-sub}^{sub} A_p / 2 + X_j} \int_{-sub}^{(1)} (1) (1) dx = \frac{\left[sub A_p / 2 + X_j\right] - \left[-sub A_p / 2 + X_j\right]}{A_p} = \frac{sub A_p}{A_p}$$
(8.77)

for j = k, while $C_{jk} = 0$ for $j \neq k$. If the index *j* is assigned so that R_j corresponds to the *j*th sub-aperture from one end of the aperture, then

$$\vec{C} = \frac{{}^{sub}A_p}{A_p}\vec{I}$$
(8.78)

where \vec{l} represents the identity matrix and

$$\vec{C}^{-1} = \frac{A_p}{{}_{sub}A_p}\vec{I}$$
(8.79)

where \vec{I} represents the identity matrix. Because of this, only cases where j = k will contribute to the summation in Eq. 8.63 and so

$$_{cor} OPD(t)_{rms}^{2} = OPD(t)_{rms}^{2} - \sum_{j=1}^{N} C_{jj}^{-1} (B_{j})^{2} .$$
(8.80)

It has been well established in previous sections that the time-averaged

uncorrected $(OPD_{rms})^2$ is $\frac{1}{2}K^2$. The average element for summation in Eq. 8.71 is then:

$$\sum_{j=1}^{N} \overline{C_{jj}^{-1}(B_{j})^{2}} = \sum_{j=1}^{N} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{A_{p}}{sub} \left(\frac{K}{2} \frac{\Lambda}{\pi A_{p}} \right)^{2} \left[\cos\left(\frac{-\pi A_{p} + 2\pi_{sub}A_{p}(j-1) + \Lambda\varphi}{\Lambda}\right) \right]^{2} d\varphi$$
$$- \cos\left(\frac{-\pi A_{p} + 2\pi_{sub}A_{p}(j) + \Lambda\varphi}{\Lambda}\right) \right]^{2} d\varphi$$
$$= \sum_{j=1}^{N} \frac{K^{2} \Lambda^{2} \left(1 - \cos\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right)^{2}\right)}{2\pi^{2}(A_{p})_{sub}A_{p}}.$$
(8.81)

Using the value of N from Eq. 8.74

$$\overline{\int_{cor} OPD(t)_{rms}^2} = \frac{1}{2}K^2 - \frac{K^2\Lambda^2 \left(1 - \cos\left(\frac{\pi_{sub}A_p}{\Lambda}\right)^2\right)}{2\pi^2 subA_p^2}.$$
(8.82)

In previous sections, the ratio A_p/Λ was found to be equivalent to a Strouhal number based on the length scale of the aperture. Likewise,

$$\frac{\sup A_p}{\Lambda} = \frac{\sup A_p}{U_C / f} = \frac{\sup A_p \cdot f}{U_C} = St_{subAp} .$$
(8.83)

Using this relationship and the corrective "gain" as defined in previous sections,

$$G(St_{subAp}) = \frac{\overline{(cor OPD(t))_{rms}^2}}{\overline{OPD(t)}_{rms}^2} = 1 - \frac{\left(1 - \cos\left(\pi St_{subAp}\right)^2\right)}{\left(\pi St_{subAp}\right)^2}.$$
(8.84)

Interestingly, this is the same result arrived at in Eq. 8.47 for one-dimensional piston correction, but scaled to the corrective sub-apertures rather than the full aperture.

8.5.3. Segmented Correction

The next stage in higher-order correction would seem to be adding T/T correction to the sub-apertures. However, letting the correction in each sub-aperture remain independent of its neighbors fails to capture an important aspect of devices like the aforementioned deformable mirror. In such devices, adjacent sub-apertures are frequently not independent. Extending one actuator may not only move that portion of the mirror, but may produce a strain that can shift adjacent actuators. Even if this is not the case, the mirror is a continuous sheet and the shape of any section will be determined not only by the position of the actuator associated with that portion, but by the position of surrounding actuators as well. An approximation of this last property can be made by placing an additional restriction on a set of sub-apertures, requiring the linear fits over these sub-apertures to meet at the edges. With this restriction in place, the envisioned corrective system changes from a set of independent T/T mirrors to a continuous segmented mirror in which each segment is connected to its neighbors by "hinges" of some sort. Figure 8.24 illustrates the difference. As shown in Fig. 8.24, the additional restriction may prevent a perfect fit in each sub-aperture, but it does eliminate discontinuities at the boundaries between sub-apertures.



Figure 8.24: Segmented Mirror Correction.

A set of basis functions that produce this sort of fit were shown in Fig. 8.23 as an example of B-spline curve fitting. A detailed analysis and derivation using this basis function, as was done for piston-only sub-apertures, can be found in section B.7.3 of Appendix B. For now, it is sufficient to note that with this basis function the matrix \vec{C} is defined by values of

$$C_{jj} = \frac{2}{3} \frac{{}_{sub} A_p}{A_p}$$
(8.85)

along the diagonal of the matrix and values of

$$C_{j(j\pm 1)} = \frac{1}{6} \frac{s_{ub} A_p}{A_p}$$
(8.86)

in the positions adjacent to the diagonal. Inversion of this matrix is not as simple as that of the purely diagonal matrix found for local piston correction, but it is possible. Again, the full analysis may be found in section B.7.3, but the average element for summation in Eq. 8.71 may be found to be:

$$\overline{C_{jk}^{-1}B_{j}B_{k}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} C_{jk}^{-1}B_{j}B_{k}d\varphi$$

$$= \frac{3}{2\pi} \frac{K^{2}\Lambda^{4}e^{i\pi(j+k)}}{\pi^{4}A_{p} \left(sub A_{p}\right)^{3} \left[(45+26\sqrt{3})(7+4\sqrt{3})^{N}-3-2\sqrt{3}\right]}$$

$$\bullet \left[(2+\sqrt{3})(2+\sqrt{3})^{j+k} + (26+15\sqrt{3})(7+4\sqrt{3})^{N}(2+\sqrt{3})^{-(j+k)} - (2+\sqrt{3})(2+\sqrt{3})^{|j-k|} - (26+15\sqrt{3})(7+4\sqrt{3})^{N}(2+\sqrt{3})^{-(j+k)} \right]$$

$$\begin{bmatrix}
-4\sin\left(\frac{\pi X_{j}}{\Lambda}\right)\cos\left(\frac{\pi X_{j}}{\Lambda}\right)\sin\left(\frac{\pi X_{k}}{\Lambda}\right)\cos\left(\frac{\pi X_{k}}{\Lambda}\right)\cos^{4}\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right) \\
-4\cos^{2}\left(\frac{\pi X_{j}}{\Lambda}\right)\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right)\cos^{4}\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right) + \cos^{4}\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right) \\
+2\cos^{2}\left(\frac{\pi X_{j}}{\Lambda}\right)\cos^{4}\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right) + 2\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right)\cos^{4}\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right) \\
+8\sin\left(\frac{\pi X_{j}}{\Lambda}\right)\cos\left(\frac{\pi X_{j}}{\Lambda}\right)\sin\left(\frac{\pi X_{k}}{\Lambda}\right)\cos\left(\frac{\pi X_{k}}{\Lambda}\right)\cos^{2}\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right) \\
+8\cos^{2}\left(\frac{\pi X_{j}}{\Lambda}\right)\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right)\cos^{2}\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right) + 2\cos^{2}\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right) \\
-4\cos^{2}\left(\frac{\pi X_{j}}{\Lambda}\right)\cos^{2}\left(\frac{\pi x_{k}}{\Lambda}\right) - 4\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right)\cos^{2}\left(\frac{\pi x_{k}}{\Lambda}\right) \\
-4\sin\left(\frac{\pi X_{j}}{\Lambda}\right)\cos\left(\frac{\pi X_{j}}{\Lambda}\right)\sin\left(\frac{\pi X_{k}}{\Lambda}\right)\cos\left(\frac{\pi X_{k}}{\Lambda}\right) \\
-4\cos^{2}\left(\frac{\pi X_{j}}{\Lambda}\right)\cos^{2}\left(\frac{\pi X_{j}}{\Lambda}\right) + 2\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right) \\
-4\cos^{2}\left(\frac{\pi X_{j}}{\Lambda}\right)\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right) X_{k}}{\Lambda}\right) \\
-4\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right)\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right) \\
-4\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right)\cos^{2}\left(\frac{\pi X_{k}}{\Lambda}\right) \\
-4\cos^{2}$$

with X_j and X_k representing the locations of the *j*th or *k*th actuator. From Eq. 8.71, the corrective gain should be

$$\frac{\overline{COPD(t)_{rms}^2}}{\overline{OPD(t)_{rms}^2}} = 1 - \frac{\overline{\vec{B}(t)^T \vec{C}^{-1} \vec{B}(t)}}{\overline{OPD(t)_{rms}^2}} = 1 - \frac{2}{K^2} \sum_{j=1,k=1}^N \overline{C_{jk}^{-1} B_j B_k} .$$
(8.88)

Finding a general solution to Eqs. 8.87 and 8.88 for any value of *N* from this is left for the future.

8.5.4. Continuous Surfaces

An actual continuous surface does not have "hinges" of the sort seen in the segmented correction of the previous section. To achieve continuity in derivatives of the curve produced well as position at the borders between sub-apertures, a third-order or

higher fit is necessary. Xianetics actually recommends using third-order functions to describe their deformable mirror for control purposes.

A common form of B-spline basis functions are those produced by the Cox-de Boor recursion formula, described in detail in section B.7.3 of appendix B. The basis functions used for piston-only and segmented-mirror correction in the previous sections were examples of the 0-order and first-order basis functions produced by the Cox-de Boor formula. The third-order basis function of this form is shown in Fig. 8.25 below, graphed against intervals of length $subA_p$. It should be noted that the effects of this basis function extend past the locations of the adjacent actuators, to reflect a system in which moving one actuator puts a strain on the adjacent actuators and may cause them to shift as well.



Figure 8.25: Third-order basis function.

The full analysis for this system may be found in section B.7.4 of appendix B.

For now, it is sufficient to note that the matrix defining this system is found to be

$$\vec{C} = \frac{1}{5040} \frac{{}_{sub}A_p}{A_p} \begin{bmatrix} 2416 & 1191 & 120 & 1 & 0 & \cdots & 0\\ 1191 & 2416 & 1191 & 120 & \ddots & \vdots\\ 120 & 1191 & 2416 & \ddots & 0\\ 1 & 120 & \ddots & 120 & 1\\ 0 & \ddots & 2416 & 1191 & 120\\ \vdots & \ddots & 120 & 1191 & 2416 & 1191\\ 0 & \cdots & 0 & 1 & 120 & 1191 & 2416 \end{bmatrix}, \quad (8.89)$$

and that finding a generalized solution for the effective gain of this type of correction is theoretically possible, but is left for the future.

8.5.5. Higher-Order Correction by Simulation

Gain functions of the sort shown in sections 8.4.1 through 8.4.4 can be found numerically by running a simulation of a sine-shaped wavefront in an aperture, computing the correction for multiple cases of relative length scales and relative positions of the aperture and wavefront disturbances, and averaging where appropriate. The onedimensional gain of Fig. 8.11 was originally found in this manner, with the analytical derivation of Eq. 8.35 coming afterwards.

Figure 8.26 shows the gain found in this manner for a one-dimensional aperture that has been divided into one to five sub-apertures, with independent piston correction occurring in each sub-aperture.



Figure 8.26: Numerically-computed piston-only corrective gain with subapertures.

The curves located further to the right on the figure are associated with more subapertures. A greater number of smaller sub-apertures within an aperture of fixed size allows correction to be performed on a finer scale and allows the system to deal effectively with disturbances of smaller scales and higher frequencies. It should also be noted that the shape of the curve changes somewhat with the number of sub-apertures. This deviation seems to occur around a point at which the sub-aperture length scale is about half that of the wavefront disturbance.

Figure 8.27 shows the gain curves of Fig. 8.26, plotted against a Strouhal number based on the length scale of the sub-apertures rather than that of the overall aperture. Curves for 10, 25, 50, and 100 sub-apertures have also been added. The curves fall on top of each other, with only slight variations, and are all very similar to the onedimensional gain curve for piston correction found in Eq. 8.47 and plotted in Fig. 8.21. This agrees with the prediction of Eq. 8.84 in section 8.5.2.



Figure 8.27: Numerically-computed piston-only corrective gain scaled by sub-aperture dimensions.

This can likewise be performed for correction with other forms of basis function. The results of using a first-order basis function, corresponding to the segmented mirror of section 8.5.3 are shown in Fig. 8.28. Fig. 8.28 (a) is similar to Fig. 8.26 in showing results for one to five sub-apertures. Just as was the case for the piston-only correction, more sub-apertures allow the system to deal with smaller length scales and higher frequencies of disturbance, shown by the gain curves moving to the right. However, the shape of the curve seems to change significantly with different numbers of apertures.



Figure 8.28: Effects of segmented correction.

Figure 8.28 (b) scales the results seen in Fig. 8.28 (a) by the dimensions of the subaperture, as was done for piston-only correction in Fig. 8.27. As was also done for piston-only correction, this figure includes results for 10, 25, 50, and 100 sub-apertures. The numerically-calculated gain curves appear to converge onto a smooth curve as the sub-aperture count increases. However, unlike what was seen for local piston correction, the shape of the curve being converged towards is different than that of T/T correction over a single aperture. This is because the correction provided in each sub-aperture of the segmented mirror has a constraint not seen in the single-aperture case, in that the endpoints of the segment in each sub-aperture must match the endpoints of the adjoining sub-apertures. The converging curve appears to have a 3-dB point around $St_{subAp} \cong 0.5$, compared to the previously found 3-dB point of $St_{Ap} \cong 0.83$ for T/T in a single aperture.

As was done for piston and segmented correction, results of one-dimensional simulations for third-order correction are shown in Fig. 8.29. A correction by a single third-order fit across the aperture has a 3-dB point of $St_{Ap} \cong 1.62$. Adding more sub-

apertures causes the system to converge on a gain curve with 3-dB point near $St_{subAp} \cong 0.5$, just as was the case for the segmented correction.



Figure 8.29: Effects of correction with a continuous surface.

Figure 8.30 compares the gain curves for first order (segmented) and third-order (continuous surface) correction with 200 sub-apertures, a degree of resolution in correction that produces fairly smooth curves. The two curves are similar, with 3-dB points at $St_{subAp} \cong 0.5$, though curve for the third-order correction seems to be a bit steeper. Both curves also show a slight deviation occurring at $St_{subAp} \cong 0.5$, which appears to be an artifact of curve-fitting that occurs when the period of the disturbance is twice that of the aperture.



Figure 8.30: Comparing segmented and continuous surface correction for large numbers of sub-apertures.

Finding a simple approximation for these corrective-gain curves is a bit more difficult that was the case for the analytic expressions found in section 8.4. As with any discreteized simulation, there are always limits of resolution both time and space within the simulation. Achieving finer resolution requires more computing time and eventually some practical limit is reached. Additionally, precision and round-off error may become factors when the values or differences in values involved become very small. Simply plotting the analytic solutions found in section 8.4 tended to encounter problems of this sort at values of $St_{subAp} < 0.01$ when using PC-based MatLab routines of the sort that were also used to perform these simulations.

Figure 8.31 shows results for a simulation segmented correction with 1000 subapertures, plotted in log-decibel form. The results level off for values of St_{subAp} below 0.2 to a gain of about 0.002 (-27 dB), which is likely due to the aforementioned limitations of the simulation. The region from $St_{subAp} = 0.1$ to $St_{subAp} = 1.0$ shown here never quite achieves the characteristics that would allow a power-law fit to be carried out with much confidence, but a function of $8(St_{subAp})^4$ seems to serve as a first-order approximation over the range from $St_{subAp} = 0.2$ to $St_{subAp} = 0.5$.



Figure 8.31: Effects of segmented correction.

Figure 8.32 shows results for continuous-surface correction based on a third-order fit with 1000 sub-apertures, plotted in log-decibel form. The results level off to a gain of about 10⁻⁶ (-60 dB) for values of St_{subAp} below 0.15. Again this is more likely due to the limitations of the simulation than a reflection of reality. In the region between $St_{subAp} =$ 0.15 and the 3dB point of $St_{subAp} = 0.5$, a function of $2000(St_{subAp})^{11.7}$ seems to serve as a good approximation. This is not as precise a fit as was achieved in sections 8.4.2, 8.4.3, and 8.4.4 with Taylor series approximations of analytic functions, but it is likely to be a good enough first-order approximation to be useful in engineering and design of optical corrective systems.



Figure 8.32: Effects of segmented correction.

For these forms of higher-order correction, finding better approximations that can be useful in engineering may be possible with greater computer power, or it may be possible to find a generalized analytic solution based on the spline curve-fitting. However, limitations of time and resources place these are beyond the scope of this study. For now, the 50% cutoff point of $St_{subAp} = 0.5$ will suffice as a guideline for minimum spatial resolution required in a corrective system.

8.6. Temporal Effects

Section 3.5 in chapter 3 addressed common limitations of corrective systems. The first such limitation was that of spatial resolution, which was addressed the previous section of this chapter as is the result of having a finite number of sub-apertures with non-

zero width and spacing in the corrective system. Another type of limitation is that of temporal effects, often caused by a physical system having limitations in how quickly it can respond to changes in conditions. This often manifests as the system having some upper bandwidth, being unable to cope with conditions and inputs of a frequency beyond that.

In expressing the corrective system gains developed in this chapter in terms of Strouhal numbers, it would seem that these results include temporal effects, as a Strouhal number contains a frequency along with a velocity and length scale. However, this is merely an artifact of the relationship between the length scales in the flow and the velocity with which they pass through the optical aperture. The analytical and numerical approaches used in arriving at these system gains assumed that the corrective system was able to instantly react to any changes in the wavefront and achieve the best possible correction it could provide for the system without delay.

8.6.1. Periodic Correction

Periodic correction is common in discrete-time systems, in which the correction to be applied is updated at intervals. Section 3.5.2 described the problems and limitations of this, which may occur as the wavefront to be corrected changes during the interval between corrections, while the correction applied remains fixed.

To explore this aspect of corrective systems, the assumption of a sinusoidal wavefront will again be used in the form of

$$OPD(x) = K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right),$$
 (8.90)

where Λ is a length scale associated with the distortions in the wavefront and φ indicates relative position of the distortion, which is assumed to be changing with time. In previous sections of this chapter it was found that average OPD_{rms} of this wavefront over long periods of time and a one-dimensional aperture of length A_p would be

$$\overline{OPD_{rms}^2} = \frac{1}{2}K^2.$$
(8.91)

If correction were applied at the point in time corresponding to φ_0 , in the form of a perfect fit to the wavefront at that moment, then the corrected wavefront will have the form

$$_{cor} OPD(x) = K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right) - K \sin\left(\frac{2\pi}{\Lambda}x + \varphi_0\right).$$
(8.92)

Averaging over the aperture, the period of correction, and all possible starting-points for that correction period leads to:

$$\overline{\left(_{cor} OPD_{rms}\right)^{2}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\Delta \varphi} \int_{\varphi_{0}}^{\varphi_{0} + \Delta \varphi} \frac{1}{A_{p}} \int_{-Ap/2}^{Ap/2} \left[K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right) - K \sin\left(\frac{2\pi}{\Lambda}x + \varphi_{0}\right) \right]^{2} dx d\varphi d\varphi_{0}$$
$$= K^{2} \frac{\Delta \varphi - \sin(\Delta \varphi)}{\Delta \varphi}. \tag{8.93}$$

Intermediate steps of this derivation and integration can be found in section B.6.1 of appendix B. As was done when dealing with spatial resolution, the residual distortions after correction can then be compared to the distortions before correction by taking the ratio of these two results to produce an effective gain function for the corrective system. Derivation of an analytic expression for periodic correction over a two-dimensional circular aperture can also be found in appendix B, but the result of this is the same as for a one-dimensional aperture, given in Eq. 8.94:

$$G(\Delta \varphi) = \frac{\overline{\left(\frac{\cos OPD_{rms}}{\cos \Theta}\right)^2}}{\overline{\left(OPD_{rms}\right)^2}} = 2\frac{\Delta \varphi - \sin(\Delta \varphi)}{\Delta \varphi}.$$
(8.94)

In actually implementing this system, it may be more understandable to express things in terms of the number of corrections per cycle of the disturbances in the wavefront (N_C) or the ratio of the frequency of corrections (f_C) to the frequency associated with the wavefront distortions (f_D). Then again, one may begin with only the length of the correction period (τ_1). The relationships between these terms and the characteristics of the flow (U_C and Λ) are

$$\Delta \varphi = \frac{2\pi}{N_c} = 2\pi \frac{f_D}{f_c} = 2\pi \frac{U_C}{f_c \Lambda} = 2\pi \frac{U_C \tau_1}{\Lambda}.$$
(8.95)

Figure 8.33 plots Eq. 8.94 in terms of the number of corrections per cycle. From this figure, it seems that performing only one or two corrections per cycle does not reduce the averaged OPD_{rms} because performing a perfect correction may reduce the OPD to zero at some points in time, but doing so with a frequency this low also allows the distortions in the wavefront to reach a point at which the correction and the distortion are 180° out of phase with each other, doubling the distortion at those times. As the gain functions are expressed in terms of the square of OPD_{rms} to better relate to power spectra, this doubling becomes a quadrupling of the effect, and more than counterbalances the perfect correction performed at the beginning of the cycle.



Figure 8.33: Reduction in wavefront distortions for different frequencies of periodic correction.

From Eq. 8.94 and Fig. 8.33 a minimum of 3.3 corrections per cycle is necessary simply to avoid making the average distortion worse, and the 3-dB (50%) point is found at 4.9 corrections per cycle. For small values of $\Delta \varphi$ and corresponding large values of N_C , it may be more convenient to use the approximation $sin(x) \cong x - x^3/6$ to arrive at the following expression:

$$G(\Delta\varphi) \cong 2\frac{\Delta\varphi - \Delta\varphi + (\Delta\varphi)^3 / 6}{\Delta\varphi} = \frac{(\Delta\varphi)^2}{3} = \frac{4}{3}\pi^2 \frac{1}{N_c^2}.$$
(8.96)

8.6.2. Latency

As addressed in section 3.5.3, latency refers to a delay in applying the correction. This may be caused by the time needed to calculate the appropriate conjugate of the wavefront to be corrected, the time required for actuators to traverse to the required position, or transient effects and delays in control signals reaching the required points and values.

Again making use of the sinusoidal wavefront approximation, if the wavefront is of the form

$$OPD(x) = K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right),$$
 (8.97)

then the corrected wavefront would be of the form

$$_{cor} OPD(x) = K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right) - K \sin\left(\frac{2\pi}{\Lambda}x + \varphi - \Delta\theta\right), \tag{8.98}$$

with $\Delta \theta$ indicating the delay of the correction in terms of phase relative to the moving distortions in the wavefront. Averaging over the aperture and all points during a cycle of the disturbance leads to:

$$G(\Delta\theta) = \frac{\overline{\left(\frac{\cos OPD_{rms}}{OPD_{rms}}\right)^{2}}}{\left(\overline{OPD_{rms}}\right)^{2}}$$
$$= \frac{2}{K^{2}} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{A_{p}} \int_{-Ap/2}^{Ap/2} \left[K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right) - K \sin\left(\frac{2\pi}{\Lambda}x + \varphi - \Delta\theta\right) \right]^{2} dx d\varphi$$
$$= 2(1 - \cos(\Delta\theta)). \tag{8.99}$$

Details and intermediate steps in this derivation are provided in section B.6.2 of appendix B. As with periodic correction, the results for a circular aperture are the same as for the one-dimensional aperture.

As with periodic correction in the previous section, it is more common to begin with knowledge of the flow and some fixed time delay (τ_2) associated with the system that to know the relative phase. In terms of these factors,

$$\Delta \theta = 2\pi f_D \tau_2 = 2\pi \frac{U_C \tau_2}{\Lambda}.$$
(8.100)

Eq. 8.99, plotted in Fig. 8.34, is periodic, as it is based on an assumption of a periodic disturbance. In the case of a periodic or even quasi-periodic disturbance, if a delayed correction is delayed by enough, then it may regain some of its effectiveness by matching and correcting the following cycle of the disturbance. From Eq. 8.99 and Fig. 8.34, it seems that performing corrections with a latency phase lag of more than 1.05 radians will add to the OPD_{rms} observed, rather than decrease it. This corresponds to a time lag of about 17% of the time associated with one cycle of the disturbance. However, if the disturbances are truly periodic, then a phase delay of more than 5.23 radians may do some good by catching the next cycle as it comes around.



Figure 8.34: Reduction in wavefront distortions for different degrees of latency in the correction.

If one does not wish to rely on the periodicity of the disturbances, then a phase delay of no more than 0.73 radians is recommended, as this represents a 3-dB cutoff for

the system. For latency delays that are relatively small compared to the time scales associated with the wavefront disturbances, it may be more convenient to use the approximation $\cos(x) \cong 1 - x^2/2$ to arrive at the following expression:

$$G(\Delta\theta) \cong 2(1-1+\frac{1}{2}(\Delta\theta)^2) = (\Delta\theta)^2.$$
(8.101)

8.6.3. Periodic Correction with Latency

Most physical systems with periodic correction will also have latency. Again assuming a sinusoidal wavefront, the corrected wavefront would be of the form

$$_{cor} OPD(x) = K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right) - K \sin\left(\frac{2\pi}{\Lambda}x + \varphi_0 - \Delta\theta\right), \qquad (8.102)$$

with φ_0 corresponding to time for which the periodic correction is meant and $\Delta\theta$ corresponding to the delay in applying that correction. Averaging over the aperture, the period of correction, and all possible starting-points for that correction in the manner of Eq. 8.93 involves some mid-stages with excessive numbers of terms, which are shown in section B.6.3 of appendix B. Only the initial integral equation and the end result will be shown here:

$$\overline{\left(_{cor} OPD_{rms}\right)^{2}}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\Delta \varphi} \int_{\varphi_{0}}^{\varphi_{0} + \Delta \varphi} \frac{1}{A_{p}} \int_{-Ap/2}^{Ap/2} \left[K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right) - K \sin\left(\frac{2\pi}{\Lambda}x + \varphi_{0} - \Delta \theta\right) \right]^{2} dx d\varphi d\varphi_{0}$$

$$= \frac{K^{2}}{\Delta \varphi} \left[\Delta \varphi + \sin(\Delta \theta) - \sin(\Delta \theta) \cos(\Delta \varphi) - \cos(\Delta \theta) \sin(\Delta \varphi) \right]. \quad (8.103)$$

The gain associated with this form of correction would then be

$$G(\Delta\varphi,\Delta\theta) = \frac{\left(\frac{\cos OPD_{rms}}{\cos \theta}\right)^2}{\left(\overline{OPD_{rms}}\right)^2} = \frac{2}{\Delta\varphi} \begin{bmatrix} \Delta\varphi + \sin(\Delta\theta) - \sin(\Delta\theta)\cos(\Delta\varphi) \\ -\cos(\Delta\theta)\sin(\Delta\varphi) \end{bmatrix}.$$
 (8.104)

To check this against the results in sections 8.6.1 and 8.6.2, if $\Delta \theta$ is small and significantly smaller than $\Delta \varphi$, then using approximations of $\sin(x) \cong x$ and $\cos(x) \cong 1$:

$$G(\Delta\varphi,\Delta\theta) \cong \frac{2}{\Delta\varphi} \Big[\Delta\varphi + \Delta\theta - \Delta\theta \cos(\Delta\varphi) - \sin(\Delta\varphi) \Big] \cong 2 \frac{\Delta\varphi - \sin(\Delta\varphi)}{\Delta\varphi}, \quad (8.105)$$

which is the result found in Eq. 8.94 for periodic correction without significant latency. Likewise, for $\Delta \varphi$ that is small and significantly smaller than $\Delta \theta$:

$$G(\Delta \varphi, \Delta \theta) \cong \frac{2}{\Delta \varphi} [\Delta \varphi + \sin(\Delta \theta) - \sin(\Delta \theta) - \cos(\Delta \theta) \Delta \varphi] = 2[1 - \cos(\Delta \theta)], \quad (8.106)$$

which is the result found in Eq. 8.99 for continuous correction with latency.

Using these approximations for both $\Delta \varphi$ and $\Delta \theta$ approaching zero leads to a gain of $G(\Delta \varphi, \Delta \theta)=0$. Using the additional terms for an approximation of sine and cosine that were used in sections 8.6.1 and 8.6.2:

$$G(\Delta\varphi,\Delta\theta) \cong \frac{2}{\Delta\varphi} \begin{vmatrix} \Delta\varphi + \Delta\theta - \frac{(\Delta\theta)^3}{6} - \left(\Delta\theta - \frac{(\Delta\theta)^3}{6}\right) \left(1 - \frac{(\Delta\varphi)^2}{2}\right) \\ - \left(\Delta\varphi - \frac{(\Delta\varphi)^3}{6}\right) \left(1 - \frac{(\Delta\theta)^2}{2}\right) \end{vmatrix}$$
$$= \left[\frac{1}{3}(\Delta\varphi)^2 + \Delta\varphi\Delta\theta + (\Delta\theta)^2 - \frac{1}{6}\left(\Delta\varphi(\Delta\theta)^3 + (\Delta\varphi)^2(\Delta\theta)^2\right)\right]. \tag{8.107}$$

Dropping the 4th-order terms from this leads to the expression

_

$$G(\Delta \varphi, \Delta \theta) \cong \frac{1}{3} (\Delta \varphi)^2 + \Delta \varphi \Delta \theta + (\Delta \theta)^2.$$
(8.108)

When actually implementing a system, the data available on the characteristics of the disturbances and system are unlikely to be provided in terms of relative phase, as that can only be determined by examining the disturbances and system together and in relation to each other. In practice, it is more likely that one will be confronted with a disturbance that may be characterized by a period of τ_C or frequency of f_C , such that τ_C =1/ f_C . To deal with this disturbance, it is then necessary to select or construct a system that will have a given update period (τ_1) or frequency (f_1) and latency delay (τ_2). The terms of relative phase in Eqs. 8.104 and 8.108 can then be expressed as

$$\Delta \varphi = 2\pi \frac{\tau_1}{\tau_c} = 2\pi \frac{f_c}{f_1} \tag{8.109}$$

and

$$\Delta \theta = 2\pi \frac{\tau_2}{\tau_c}.\tag{8.110}$$

Using these substitutions, Eq. 8.104 may be written as

$$G(\tau_1, \tau_2, \tau_C) = \frac{\tau_C}{\pi \tau_1} \begin{bmatrix} 2\pi \frac{\tau_1}{\tau_C} + \sin\left(2\pi \frac{\tau_2}{\tau_C}\right) - \sin\left(2\pi \frac{\tau_2}{\tau_C}\right) \cos\left(2\pi \frac{\tau_1}{\tau_C}\right) \\ -\cos\left(2\pi \frac{\tau_2}{\tau_C}\right) \sin\left(2\pi \frac{\tau_1}{\tau_C}\right) \end{bmatrix}$$
(8.111)

and Eq. 8.108 as

$$G(\tau_1, \tau_2, \tau_C) \cong \frac{4\pi^2}{\tau_C^2} \left[\frac{1}{3} \tau_1^2 + \tau_1 \tau_2 + \tau_2^2 \right].$$
(8.112)

8.7. Using the Filter Functions

The preceding sections in this chapter have presented many equations

representing various filter functions as representations of corrective systems. Here is an

illustrative example showing how these y might be applied in predicting the performance of a given adaptive-optic system.

Consider a flow with a characteristic velocity of 80 m/s in which the primary wavefront aberrations have a characteristic frequency of 200 Hz. Correction is to be applied over a circular aperture with a diameter of 25 cm. The first stage in this correction is a T/T mirror applying correction based on the Zernike-tilt definition of tilt. The mirror is controlled by an analog system that tends to have a time delay of 0.25 ms in reaching a position towards which it is corrected. The Strouhal number associated with the disturbance and the aperture is then

$$St_{Ap} = \frac{0.25m \cdot 200 \frac{1}{s}}{80 \frac{m}{s}} = 0.625.$$
 (8.113)

Looking at Fig. 8.13, this is below the 3-dB point associated with correction of this sort; however, Fig. 8.14 indicates this is not quite in the range where the approximation of $G=1.52(St_{Ap})^4$ would apply. Using the full form of Eq. 8.39

$$G = \begin{bmatrix} 1 - \frac{16J_0^2(\pi 0.625) + 4J_1^2(\pi 0.625)}{(\pi 0.625)^2} \\ + \frac{64J_0(\pi 0.625)J_1(\pi 0.625)}{(\pi 0.625)^3} - \frac{64J_1^2(\pi 0.625)}{(\pi 0.625)^4} \end{bmatrix} = 0.16.$$
(8.114)

This indicates that this form of correction cannot reduce the mean-squared OPD or phase variance of this disturbance below 16% of its uncorrected value, due to the inability of a flat T/T mirror to fully match the shape of the wavefront distortions.

For a disturbance of this frequency, the latency in the control system for the mirror represents a phase delay of

$$\Delta\theta = 2\pi \cdot 200 \frac{1}{s} \cdot \frac{0.25}{1000} s = 0.31 \, radians \,. \tag{8.115}$$

By 8.99 there should be an error associated with this, represented by a gain due to temporal effects (G_t), such that

$$G_t = 2(1 - \cos(0.31)) = 0.095$$
. (8.116)

Indicating that the mean-squared inaccuracies due to latency in the applied correction are equivalent to 9.5% of the mean-squared uncorrected disturbance.

As was addressed in section 3.5.4, fitting error and temporal-based error are normally considered to be uncorrelated, which means the net residual phase variance after correction (σ_r^2) can be found by simply adding the mean-squared error from each source. The error in the analysis above has been expressed as something proportional to the uncorrected error (σ^2), so this may be expressed as

$$\sigma_r^2 = \sigma_{r(fit)}^2 + \sigma_{r(temp)}^2 = \sigma^2 (0.16 + 0.095) = 0.26\sigma^2, \qquad (8.117)$$

which indicates this system will reduce the wavefront distortion by approximately a factor of four. Reducing the response time of the system could improve this performance, but a level of 16% residual error remains the absolute limit of T/T correction under these conditions.

It should be kept in mind that the uncorrected error in this analysis includes piston. If piston within an aperture is not normally considered for a given application, which is normally the case when the resolved image or far-field pattern is of primary importance, then a more meaningful comparison might be with that of the pistoncorrected case. According to Eq. 8.48, the expected error after piston correction (σ_{pc}^2) in these circumstances would be 65% of the original error. So the phase variance after T/T correction compared to the wavefront disturbances pertinent to this application before correction would be

$$\frac{\sigma_r^2}{\sigma_{pc}^2} = \frac{0.26\sigma^2}{0.65\sigma^2} = 0.4.$$
(8.118)

Therefore, the T/T correction of this system would produce a reduction in phase variance of about 60% for applications in which net piston across the aperture is not a factor.

8.8. Conclusions

The gain functions developed over the course of this chapter serve as indicators of the effectiveness in applying various types of correction to distorted wavefronts in an aperture, provided those distortions are sinusoidal in nature. However, many forms of analysis express waveforms as sums of sinusoidal functions. A scaling law found by means of a sinusoidal simulation worked when applied to the non-sinusoidal distortions in the experimental work involving a shear layer. As will be shown in the following chapter, these gain functions can be applied with fruitful results to conditions and distortions that can not be approximated with a single sinusoid.

¹ Oster, D., and Wygnanski, I., "The Forced Mixing Layer Between Parallel Streams," Journal of Fluid Mechanics, Vol. 123, 1982, pp 91-130.

² Weisbrot, I., and Wygnanski, I., "On Coherent Structures in a Highly Excited Mixing Layer," Journal of Fluid Mechanics, Vol. 195, 1988, pp. 137-159.

³ Ho, C-M, and Huerre, P. "Perturbed Free Shear Layers," Annual Reviews of Fluid Mechanics, Vol. 16, 1984, pp365-424.

⁴ Ho, C-M., and Huang, L-S., "Subharmonics and vortex merging in mixing layers," Journal of Fluid Mechanics, Vol. 119, 1982, pp. 443-473.

⁵ Nightingale, A.M., Gordeyev, S., Jumper, E.J., Goodwine, B., and Siegenthaler, J.P., "Regularizing Shear Layer for Adaptive Optics Control Applications," AIAA Paper 2005-4774, June, 2005.

- ⁶ Rennie, R.M., Siegenthaler, J.P., and Jumper, E.J., "Forcing of a Two-Dimensional, Weakly-Compressible Subsonic Free Shear Layer," AIAA Paper 2006-561, January, 2006.
- ⁷ Rennie, M.R., Duffin, D.A., 2, and Jumper, E.J., "Characterization of a Forced Two-Dimensional, Weakly-Compressible Shear Layer," AIAA Paper 2007-4007, June, 2007.
- ⁸ Duffin, D.A., "Feed-Forward Adaptive-Optic Correction of Aero-Optical Aberrations Caused by a Two-Dimensional Heated Jet," AIAA Paper 2005-4776, June 2005.
- ⁹ Nightengale, A., et. al., "Adaptive-Optic Correction of a Regularized Compressible Shear Layer," AIAA Paper 2006-3072, June, 2006.
- ¹⁰ Nightengale, A.M., Mitchel, B., Goodwine, B., Jumper, E.J., "Feedforward Adaptive-Optic Mitigation of Aero-Optic Disturbances," 39th Plasmadynamics and Lasers Conference, 23 - 26 June 2008, Seattle, Washington, AIAA paper 2008-4211
- ¹¹ Fitzgeral, E.J and Jumper, E.J., "Scaling Aerooptic Aberrations Produced by High-Subsonic-Mach Shear Layers", AIAA Journal, Vol. 40, No. 7, July 2002, pp. 1373-1381
- ¹² Gordeyev, S., and Jumper, E.J., "Aero-Optical Characteristics of Compressible, Subsonic Turbulent Boundary Layers," AIAA paper 2003-3606.
- ¹³ Greenwood, D.P., and Freid, D.L., "Power Spectra Requirements for Wave-Front-Compensative Systems", J. Opt. Soc. Am., Vol. 66, March 1976, pp 193-206.
- ¹⁴ Greenwood, D.P., Bandwidth Specification for Adaptive Optics Systems, J. Opt. Soc. Am., Vol. 67, March 1977, pp 390-393.
- ¹⁵ Tyson, Robert K., "Using the deformable mirror as a spatial filter: application to circular beams," App. Opt., Vol. 21, No. 5, Mar 1982, pp 787-793.
- ¹⁶ Tyson, Robert K., "Bit-error rate for free-space adaptive optics laser communications," J. Opt. Soc. Am. A., Vol. 19, No. 4, Apr 2002, pp 753-758.
- ¹⁷ van Dam, M.A., Le Mignant, D., and Macintosh, B.A., "Performance of the Keck Observatory Adaptive-Optics System," Appl. Opt., vol 43, pp 5458-5467 (2004)
- ¹⁸ Wilcox, C.C., Andrews, J.R., Restaino, S.R., Teare, S.W., Payne, D.M., and Krishna, S., "Analysis of a combined tip-tilt and deformable mirror," Opt. Lett. 31, pp 679-681 (2006)
- ¹⁹ Hudgin, R., "Wavefront compensation error due to finite corrector-element size," J. Opt. Soc. Am., Vol. 67, No. 3, March 1977.
- ²⁰ Love, G.D., "Liquid Crystal Adaptive Optics", <u>Adaptive Optics Engineering Handbook</u>, Tyson, R.K., Marcel Decker, New York, 2000, pp 273-285.

CHAPTER 9:

APERTURE FILTERS IN APPLICATION

9.1. Overview

The derivations in the previous chapter were performed by approximating the wavefronts to be corrected as sine waves. This, in turn, was based on the expected form of distortions produced by a shear layer. However, this is an extreme simplification. Then again, Fourier series, power spectra, and some other forms of analysis are based on the idea that any waveform can be expressed as a summation of sine functions. Additionally, the scaling law found in section 8.3.2 by use of this approximation did seem to work with the actual data taken from a shear layer, despite the inaccuracies and limitations of this approximation and the presence of a boundary layer as well.

This chapter provides a set of examples in which the corrective-system gain functions found in chapter 8 are applied and compared to experimental data and results. In these examples, the validity of approximating the distortions with a single sinusoid is questionable, or clearly not applicable, yet the gain functions produced with this simple approximation prove to be accurate descriptors of the effects of correction when applied to these cases.

9.2. T/T Correction in a Compressible Shear Layer

The gain function for one-dimensional Z-tilt correction of Eq. 8.35 is intended to be compared to a power spectral density. Producing a power spectrum requires a continuous signal or, in the case of discreet data points, a continuous sequence. Producing a spectrum with a significant degree of frequency resolution requires this to be a relatively long sequence. A power spectrum for the uncorrected case can be produced by reconstructing a long, continuous wavefront, extrapolated upstream and down from the data recorded. These long, continuous wavefronts may be thought of as being viewed through an infinitely large aperture.

However, this is not possible for the T/T corrected wavefronts as the correction applied will change with each time step, and will change the wavefront accordingly. Thus the corrected wavefront within the aperture at each time step must be considered as a separate case, and can not be simply joined to the ends of wavefronts from previous or following points in time. Because of this, a power spectrum of the corrected wavefront would be limited to what could be produced using only the data points within the aperture at a given time.

Averaging several such spectra from different time steps may be useful in some cases, and tends to produce smoother spectra, but this does nothing to improve the frequency resolution. One way to achieve higher frequency resolution is to pack a greater number of sensors into the area of measurement. A major benefit of the Malley probe is that it achieves spatial resolution via temporal resolution, substituting a higher sampling rate for more sensors. However, with either approach, the cost or effort of achieving the required number of data points within a finite aperture for a desired

295

resolution may be prohibitive. This degree of resolution was not available at the time of the experiments described in chapter 7.

On the other hand, it is possible to record the tilt removed at each time step, as represented by the coefficient *B* in Eq. 8.15. A power spectrum can then be generated from this long continuous sequence of tilts. As the gain, *G*, in Eq. 8.35 indicates the degree of aberration remaining after T/T correction, so the function 1 - G should correspond to the power spectrum of the aberrations being removed. Figure 9.1 (a) shows a power spectrum for beam jitter data taken at a position of 13.4 cm in a shear layer comprised of Mach 0.77 and 0.06 flows. Frequencies below 500 Hz have been filtered out to remove vibrations in the experimental equipment. It also shows the power spectrum for tilt removed in applying T/T correction over an aperture of 5 cm, and the function 1 - G scaled in the horizontal axes by the aperture size and the convection velocity of 148 m/s to recover frequencies from the Strouhal number defined in Eq. 8.34. Vertical scales have been normalized by the peak value of the jitter spectrum.



Figure 9.1: Comparing power spectra for beam jitter, tilt removed, and predicted filter function.

For a more direct comparison, Fig. 9.1 (b) shows just the tilt removed spectrum and 1 - G, with the tilt spectrum scaled by its maximum. The degree to which the prediction of 1 - G corresponds with this spectrum can be clearly seen, including the small local maximum seen between 5 and 6 kHz.

Figure 9.2 (a) shows T/T spectra generated from the same data set used in Fig. 9.1 for three different sizes of aperture. The smaller apertures have more energy associated with them as some elements of the wavefront that could only be corrected with higher-order measures for a 20-cm aperture become T/T effects for apertures of 15 or 10 cm.



Figure 9.2: (a) Tilt spectra for different aperture sizes. (b) Scaling tilt spectra by Strouhal number.

Figure 9.2 (b) shows the same spectra as 9.2 (a), but plotted against a Strouhal number based on the aperture size ($St_{Ap} = A_p f / U_C$) rather than raw frequency. Also plotted on this figure is the 1 – *G* function for a one-dimensional aperture. One way to interpret this figure is that as A_p decreases, the spectrum being filtered slides to the left and more of the spectrum enters the region for which T/T correction is effective. As with Fig. 9.1 (b), the way in which the shape of the filter spectrum predicting which frequencies will be removed matches the spectrum for T/T being removed is quite clear.

9.3. Periodic Correction of a Heated Jet

The effects of temporal limitations in applying optical correction were first addressed in section 3.5. To summarize here, any physical system will have some delays in applying a correction. If the disturbances are changing or moving with time, then a correction that is generated for the disturbances seen at time *t* but is not actually applied until time $t + \Delta t$ will become less effective with increasing Δt . In a similar issue, corrective systems designed with current technology tend to rely on computers and microprocessors in processing measurements of the distortions and generating the corrections to deal with them. Such systems are discrete in time, meaning the corrections produced by these systems are applied at updated at specific moments in time, while the aberrations to be corrected change continuously. If the period between corrections is long enough for the aberrations to change significantly before the next correction is applied, then the corrective system becomes less effective.

In section 8.6, the effects of these temporal limitations were analyzed, using a simplified sinusoidal disturbance. In the course of this analysis, it was found that for a disturbance with a period of τ_C , a corrective system that updates with a period of τ_1 , and a delay in applying the correction of τ_2 , the corrective "gain" of the system will be

$$G(\tau_1, \tau_2, \tau_C) = \frac{\tau_C}{\pi \tau_1} \begin{bmatrix} 2\pi \frac{\tau_1}{\tau_C} + \sin\left(2\pi \frac{\tau_2}{\tau_C}\right) - \sin\left(2\pi \frac{\tau_2}{\tau_C}\right) \cos\left(2\pi \frac{\tau_1}{\tau_C}\right) \\ -\cos\left(2\pi \frac{\tau_2}{\tau_C}\right) \sin\left(2\pi \frac{\tau_1}{\tau_C}\right) \end{bmatrix}. \quad (9.1)$$

$$298$$

In 1996, Cicchiello¹ performed a study of the effects of latency and periodic correction in one-dimensional optical distortions produced by a heated jet. This flow has tendencies towards periodic disturbances and dominating structure sizes similar to those seen in a shear layer, although the mechanism behind the optical distortions produced is the temperature difference between the heated air of the jet and the cooler surrounding air. His results indicated that not only was the Strehl ratio achieved by such correction a function of both τ_1 and τ_2 , but the effects of these factors appeared to be decoupled. That is to say, there appeared to exist two independent functions, h(t) and g(t), such that the Strehl ratio (*SR*) achieved after correction could be described by

$$SR(\tau_1, \tau_2) = h(\tau_1)g(\tau_2).$$
 (9.2)

This was based on providing post-processed correction to the wavefronts reconstructed from these measurements with varying correction periods and latency delays, and then constructing the far-field intensity pattern associated with the corrected wavefronts. Figure 9.3 shows curves of Cicchiello's Strehl ratio results associated with particular values of τ_1 which have been normalized by their value when τ_2 was equal to zero.



Figure 9.3: Time-averaged Strehl ratio as a function of time delay between constructing a perfect correction and applying it (τ_2), normalized by the time-averaged Strehl ratio for $\tau_2 = 0.0$ for three values of the period between applying corrections (τ_1). (Cicchiello, 1997)¹

If Eq. 9.2 is correct, then both h(t) and g(t) should equal one when t = 0.

Therefore, the Strehl ratio at $\tau_2 = 0$ should be equal to $h(\tau_1)$ alone. Normalizing all other Strehl ratio results associated with that value of τ_1 by this the Strehl ratio at $\tau_2 = 0$ should produce Strehl ratio values that are functions of τ_2 alone:

$$SR(\tau_1, \tau_2 = 0) = h(\tau_1)g(0) = h(\tau_1), \qquad (9.3)$$

$$\frac{SR(\tau_1, \tau_2)}{SR(\tau_1, \tau_2 = 0)} = \frac{h(\tau_1)g(\tau_2)}{h(\tau_1)} = g(\tau_2).$$
(9.4)

The manner in which the plotted curves of Fig. 9.3 lie close to each other would seem to support this, especially as the results for the most frequent correction rate ($\tau_1 = 1.0 \text{ ms}$) and the least frequent correction rate ($\tau_1 = 1.4 \text{ ms}$) lie practically on top of each other.

The corrective gain found in section 8.6.3 and shown in Eq. 9.1 expresses the ratio of the corrected mean-squared *OPD* or phase variance to the uncorrected case. Cicchiello's results are in terms of Strehl ratio after correction. By the large-aperture approximation,

$$SR \cong e^{-\sigma^2} = e^{-\left(\frac{2\pi}{\lambda}OPD_{rms}\right)^2}, \qquad (9.5)$$

Strehl ratios can be related to values of OPD_{rms} , and also to phase variance, σ , which is equal to OPD_{rms} scaled by $2\pi/\lambda$. Given an average value of $(OPD_{rms})^2$ or σ^2 without correction, the value after correction can be found by multiplying the uncorrected value by the gain function for that form of correction. Therefore, for a disturbance of some period, τ_C , and of amplitude to produce an average uncorrected phase variance of σ_0 , a plausible prediction for the average Strehl ratio after correction would be

$$SR \cong e^{-G(\tau_1, \tau_2, \tau_c)\sigma_0^2}.$$
(9.6)

According to Cicchiello's assertion that the effects of τ_1 and τ_2 are decoupled in the form of Eq. 9.2, it should be possible to separate Eq. 9.1 into independent functions of τ_1 and τ_2 so that

$$SR \cong e^{-G(\tau_1, \tau_2, \tau_C)\sigma_0^2} = e^{-[G_1(\tau_1, \tau_C) + G_2(\tau_2, \tau_C)]\sigma_0^2} = e^{-G_1(\tau_1, \tau_C)\sigma_0^2} e^{-G_2(\tau_2, \tau_C)\sigma_0^2}.$$
 (9.7)

However, Eq. 9.1 does not appear to be separable in this manner. In section 8.6.3, Taylor-series approximations of the trigonometric functions in Eq. 9.1 were used to find an approximation of

$$G(\tau_1, \tau_2, \tau_C) \cong \frac{4\pi^2}{\tau_C^2} \left[\frac{1}{3} \tau_1^2 + \tau_1 \tau_2 + \tau_2^2 \right].$$
(9.8)

However, the term of $\tau_1 \tau_2$ in the midst of this still prevents the separation into $G_1(\tau_1, \tau_C)$ and $G_2(\tau_2, \tau_C)$ such that $G = G_1 + G_2$.

Looking at Cicchiello's results in Fig. 9.3, the smallest value for the correction period used in generating these results is $\tau_1 = 1.0$ ms, while the largest value of correction delay appearing on this figure is half of that. If more terms from the Taylor series are used in approximating the terms associated with τ_1 than those associated with τ_2 , then Eq. 9.1 may be approximated by:

$$G(\tau_{1},\tau_{2},\tau_{C}) \cong \frac{\tau_{C}}{\pi\tau_{1}} \begin{vmatrix} 2\pi \frac{\tau_{1}}{\tau_{C}} + 2\pi \frac{\tau_{2}}{\tau_{C}} - 2\pi \frac{\tau_{2}}{\tau_{C}} \cos\left(1 - \frac{4\pi^{2}}{2} \frac{\tau_{1}^{2}}{\tau_{C}^{2}}\right) \\ -1\left(2\pi \frac{\tau_{1}}{\tau_{C}} - \frac{8\pi^{3}}{6} \frac{\tau_{1}^{3}}{\tau_{C}^{3}}\right) \end{vmatrix}$$
$$= \frac{\tau_{C}}{\pi\tau_{1}} \left[2\pi \frac{\tau_{2}}{\tau_{C}} \frac{4\pi^{2}}{2} \frac{\tau_{1}^{2}}{\tau_{C}^{2}} + \frac{8\pi^{3}}{6} \frac{\tau_{1}^{3}}{\tau_{C}^{3}}\right] = \frac{4\pi^{2}}{\tau_{C}^{2}} \left[\tau_{1}\tau_{2} + \frac{1}{3}\tau_{1}^{2}\right]$$
(9.9)

This also is not separable in the indicated manner. There does not seem to be any selection of the number of terms to be used in this sort of approximation that will produce an equation that can be separated in the manner indicated by Cicchiello's conclusion.

Going back to the paper¹ in which Cicchiello's results were presented, the uncorrected average Strehl ratio associated with his reconstructed wavefronts was 0.265. Using the large-aperture approximation, this indicates an average uncorrected phase variance of $\sigma_0 \cong 1.15$ radians. From time series representations of reconstructed wavefronts associated with this heated jet, such as the one shown in Fig. 9.4, the larger disturbances seem to have a period of $\tau_C \cong 0.005$ seconds. From this, the values of τ_1 used
by Cicchiello correspond to a range of approximately 3.6 to 5 corrections per cycle and the values of τ_2 used correspond to a delay of 0% to 10% of the cycle of the disturbance.



Figure 9.4: Time series of experimental wave fronts for propagation through a heated jet. (Cicchiello, 1997)¹(Hugo, 1996)²

Using these values of σ_0 and τ_C in Eqs. 9.1 and 9.6 over the same range of values of τ_1 and τ_2 used by Cicchiello produces the set of predictions for Strehl ratio shown in Fig. 9.5.



Figure 9.5: Strehl ratio predictions by corrective gain function and largeaperture approximation.

Normalizing these curves by their respective values at $\tau_2 = 0$ produces the set of curves shown in Fig. 9.6, which are presented next to Cicchiello's work for comparison. These simulated and approximated results do match Cicchiello's experimental results quite well. The approximation deviates from the experimental results with greater latency, tending to over-estimate the drop in Strehl ratio. This may be a result of the large-aperture approximation being originally intended for relatively small amplitudes of disturbance and becoming less accurate for larger disturbances,³ having an error of 10% for Strehl ratios around 0.3. Additionally, while the main structure of the disturbances is of a given size and frequency, there are other components of the distortions that are of

differing sizes, and for which given time periods of τ_1 and τ_2 will correspond to different phase periods of $\Delta \varphi$ and $\Delta \theta$ than for the main distortion component. Between these effects and the rough estimates made in producing these curves, it is remarkable how well the approximation agrees with the experimental results.



Figure 9.6: Normalized Strehl ratio predictions (left) compared to experimental results¹ (right).

In light of this, it may be concluded that the gain function found in section 8.6.3 for periodic correction with latency predicts the effects of such correction with sufficient accuracy that it may serve as a guideline in designing or selecting such systems for use in a given application. On the other hand, even as these numerical estimates agree with Cicchiello's experimental results, aspects of the mathematical analysis producing those estimates seem to be at odds with the conclusion Cicchiello drew from his experimental data. A verification or refutation of Cicchiello's conclusion is not the point of this study, and his conclusion was a reasonable one based on the results observed in his experiment and analysis, but it may be that using a wider range of values for τ_1 or τ_2 would have led him to a different conclusion.

9.4. Tracking Systems

The approximation of a wavefront as a sine wave was intended to capture the dominant qualities of a shear layer and is not one that could be applied to all forms of aberrating flow. However, the filter functions found by this approximation seem to work well in describing the effects and effectiveness of T/T correction on experimental data from shear layers that display optical activity at multiple length scales and associated frequencies. It is worth exploring whether this applies to other forms of optical distortion, such as that from atmospheric propagation and Kolmogorov turbulence described in section chapter 4.

As it so happens, this investigation has already been performed, after a fashion, as a byproduct of experiments by other researchers. Boeing performed a series of tests at the Advanced Concepts Laboratory (ACL) of MIT/Lincoln Laboratory in Lexington, MA, intended to evaluate the effectiveness of various tracking algorithms.⁴ A simplified schematic of the set up for these tests is shown in Fig. 9.7.



Figure 9.7: Scoring and Tracking cameras for Boeing test at ACL.⁴

On the left in this figure are sensors and processors receiving an image of a target or point source, seen through a series of phase screens which could be put in motion relative to the sensors and target. It should be noted that these screens were constructed with a distribution of phase distortions intended to replicate the statistical properties of atmospheric turbulence, as described in sections 4.2.2 and 4.3. Also on the left is a scoring laser that was projected through the screens along the same path as the image being received by the tracking sensor, but in the opposite direction. This scoring laser was then received by a scoring sensor and processor shown on the right in Fig. 9.7.

As the outgoing scoring laser followed the same path as the light of the incoming target image, the angle of deflection needed for a system to track the target should have been the same as the angle of deflection of the scoring beam. Performance of the tracking algorithms under consideration was evaluated by comparing the output of the tracking control system with the measured deflection of the scoring beam. The Boeing staff performing called this the coherence function, which was defined as

$$\gamma_{x-tr\,x-sc}^{2}(f) = \frac{\left|G_{x-tr\,x-sc}(f)\right|^{2}}{G_{xx-tr}(f)G_{xx-sc}(f)}.$$
(9.10)

Where $G_{x-tr\ x-sc}(f)$ represents the cross spectral density between the measured scoring beam deflection and the tracking control signal in the *x* direction. G_{xx-tr} and G_{xx-sc} represent the autospectral densities of the two outputs. A value of $\gamma^2 = 1$ indicates that the tracking signal perfectly follows the disturbances of a given frequency, while lower values indicate lesser degrees of effectiveness.

Figure 9.8 shows Boeing's results from two of these tests, performed with different relative velocities for the phase screens. They found that these tracking algorithms worked for low frequencies, but not for high frequencies, and the point of transition between these two frequency ranges was dubbed the "optical frequency." Furthermore, they found that this frequency corresponded roughly to the velocity at which the phase screens moved, divided by the diameter of the receiving aperture, expressed as v/D in their terminology shown in the figure, or U_C/A_p in the notation used elsewhere in this dissertation.



Figure 9.8: Results of Boeing tests at ACL.⁴

Tracking a target in this manner is effectively tilt correction by another name. The coherence function, γ^2 , corresponds to the function 1 – *G* from section 9.2, with high values indicating that the tilt-based distortions or deviations are successfully countered by the system and low values indicating that they are not. Figure 9.9 shows the results from Fig. 9.8, with the frequencies scaled by U_C/A_p to plot those results in terms of St_{Ap} . It also shows the function 1 - G for a two-dimensional aperture.



Figure 9.9: Scaled Boeing results compared with circular-aperture filter function for Z-tilt correction.

The 1 - G curve and the scaled experimental results agree fairly well. It should be noted that the optical frequency, v/D, defined by Boeing is a rough estimate. In section 8.4.2, the 50% cut-off for the two-dimensional aperture was found to be 0.93 of this value.

9.5. Application to Atmospheric Turbulence

The utility of these filter gain functions for different types of correction is that they can be applied to practically any spectrum produced by any set of physical conditions, whereas the commonly established rules for adaptive optics are based on specific physical and spectral characteristics of Kolmogorov turbulence in the free atmosphere. However, as generic rules, these gain functions should also apply to the atmospheric case as well.

9.5.1. Piston and T/T Correction

In their paper on the spectral requirements of corrective systems,⁵ Greenwood and Fried also considered the case of correcting piston and T/T over the entire aperture before providing localized correction in sub-sections of the aperture. In doing so they found that the average residual error, in the form of rms phase variance over the aperture after correction, (σ_{ω}) would be

$$\sigma_{\varphi}^2 = 1.075 (D/r_0)^{5/3} \tag{9.11}$$

when correcting for piston alone over an aperture of diameter D and would be

$$\sigma_{\sigma}^2 = 0.141 (D/r_0)^{5/3} \tag{9.12}$$

after correction of both piston and T/T. These equations are based specifically on the optical effects of Kolmogorov turbulence, which can characterized by the Fried parameter (r_0) described in section 4.3.1 and appearing in the two equations above.

The spectrum of optical aberrations used by Greenwood⁶ in characterizing the frequency that came to bear his name is

$$PSD(f) = 0.0326 \left(\frac{2\pi}{\lambda}\right)^2 f^{-8/3} \int_0^L C_n^2(s) |V(s)|^{5/3} ds .$$
(9.13)

Again, this is based on a parameter (C_n^2) which is used to characterize Kolmogorov turbulence, as well as the convective velocity (*V*) perpendicular to the optical path. As addressed in earlier chapters, the phase variance after correction should be equal to the product of the uncorrected spectrum and the effective filter gain of the correction, integrated over all frequencies.

$$\sigma_{\varphi}^{2} = \int_{0}^{\infty} G(f) PSD(f) df \qquad (9.14)$$

It should be noted that the spectrum of Eq. 9.13 is defined in such a way that the influence of each negative frequency is combined with its positive counterpart, so that the necessary integral runs from zero to infinity rather than from negative infinity to positive infinity.

An effective gain for piston-only correction over a circular aperture was found in section 8.4.4, in the form

$$G(St_{Ap}) = 1 - \frac{4J_1^2(\pi St_{Ap})}{(\pi St_{Ap})^2}.$$
(9.15)

The Strouhal number (St_{Ap}) is itself a ratio of the frequency, convective velocity, and diameter of the aperture. Integrating the product of Eqs. 9.13 and 9.15 over all frequencies and assuming C_n^2 and convective velocity are constants over the optical path indicates that

$$\sigma_{\varphi}^2 = 9.236 \left(\frac{D^{5/3} C_n^2 L}{\lambda} \right). \tag{9.16}$$

The Fried parameter is also defined in terms of an integral of C_n^2 over the optical path. Again assuming constant C_n^2 , this would be,

$$r_0 = \left(0.423 \left(\frac{2\pi}{\lambda}\right)^2 \int_0^L C_n^2(s) ds\right)^{-3/5} = 0.72939 \left(\frac{\lambda^2}{\pi^2 C_n^2 L}\right)^{3/5}.$$
 (9.17)

Based on this, Eq. 9.16 can be written in the form

$$\sigma_{\varphi}^{2} = 0.553 \left(\frac{D}{r_{0}}\right)^{5/3}.$$
(9.18)

This result has the form of the expression produced by Fried and Greenwood, but the coefficient of proportionality is quite different than that seen in their result of Eq. 9.11.

Upon further consideration, the gain functions developed in chapter 8 were all based on a disturbance in one dimension, wavefronts in the form of a sinusoid in x but with no variations in y, reflecting the expected characteristics of aero-optic distortions produced by a series of coherent structures in the flow. The analysis by Fried and Greenwood was for atmospheric distortions that produced variations in the received wavefront along every axis of orientation within the aperture. A first-order approximation of this might be to consider a wavefront of two overlapping effects, one in x and one in y. Again considering these distortions as sinusoids or summations of sinusoids, the variations seen in each axis would be uncorrelated components of the overall phase variance. The effects in x and y would have the same spectral characteristics, but would be "filtered" independently and their effect could be combined in the form

$$\sigma_{\varphi}^{2} = \sigma_{x\varphi}^{2} + \sigma_{y\varphi}^{2} = 0.553 \left(\frac{D}{r_{0}}\right)^{5/3} + 0.553 \left(\frac{D}{r_{0}}\right)^{5/3} = 1.106 \left(\frac{D}{r_{0}}\right)^{5/3}.$$
 (9.19)

As a first-order, back-of-the-envelope approach, this approximation fails to fully capture the isotropic characteristic of atmospheric distortion. The length scales associated with such distortions should be the same in all orientations, whereas the two-sinusoid approximation above would have somewhat longer length scales on the diagonals than along the x and y vectors. However, Eq. 9.19, the result of this approximation, in is within 3% of Fried and Greenwood's result shown in Eq. 9.11.

The same principles and approximations can be used with the T/T gain function developed in section 8.4.2 to find a prediction of phase variance after this form of correction,

$$\sigma_{\varphi}^{2} = 0.144 \left(\frac{D}{r_{0}}\right)^{5/3},$$
(9.20)

which deviates from the results arrived at by Fried and Greenwood by about 2%.

9.5.2. Latency

In a later paper,⁷ Fried returned to the subject of optical correction, this time focusing on the effect of a time delay in applying the correction. He arrived at a prediction of

$$\sigma_{\varphi}^{2} = 28.44 (f_{G} \Delta t)^{5/3}$$
(9.21)

for a time delay of Δt . The Greenwood frequency (f_G) is yet another parameter for characterizing atmospheric turbulence, which was addressed in section 4.3.2 and is defined as

$$f_G = \left[0.102 \left(\frac{2\pi}{\lambda} \right)^2 \int_{s_1}^{s_2} C_n^2(s) |V(s)|^{5/3} ds \right]^{3/5}.$$
 (9.22)

The gain function for latency in correction, developed in section 8.6.2, can be applied to the spectrum of Eq. 9.13 as was done for piston and T/T correction in the previous section. In this case, it is not necessary to consider and combine effects along two axes as was done in the previous section. Temporal effects occur primarily along the vector of the overall convective velocity. Performing this calculation and using the definition of the Greenwood frequency above predicts a residual phase variance of

$$\sigma_{\varphi}^{2} = 28.56 (f_{G} \Delta t)^{5/3}$$
(9.23)

for atmospheric correction with a time delay which is within 0.5% of Fried's result.

9.6. Conclusions

The gain functions of chapter 8 were found by approximating optical disturbances as a sinusoidal function. In the real world, it is rare to encounter disturbances that can accurately be represented in this manner. However, if this approximation does serve in a crude fashion, then these gain functions may serve as a sufficient rule-of-thumb guideline in design of corrective systems.

Beyond that, if the distortions or their effects can be broken up into components of varying frequency or associated length scale, as with a power spectrum or covariance function, then these gain functions may be quite useful in gauging the effect on differing components of the disturbances and total performance of a corrective system, even if sum of those components is decidedly non-sinusoidal. Application of the gain functions to experimental data and theoretical predictions in various cases, including atmospheric turbulence, bear this out.

- ¹ Cicchiello, J.M., and Jumper, E.J., "Far-field optical degradation due to near-field transmission through a turbulent heated jet," Applied Optics, Vol. 36, No. 25, Sep 1997, pp 6441-6452.
- ² Hugo, R.J. and Jumper, E.J., "Experimental measurement of a time-varying optical path difference using the small-aperture beam technique," Appl. Opt. 35, 4436–4447 (1996).
- ³ Mahajan, V.N., "Shrehl ratio for aberration in terms of their aberration variance," J. Opt. Soc. Am., Vol. 73, 860, 861 (1983).
- ⁴ Merritt, Paul; Peterson, Scott; Telgarsky, Rastislav; Brunson, Dick; Pringle, Ralph and O'Keefe, Shawn "Performance of Tracking Algorithms under Airborne Turbulence," SPIE Laser Weapons Technology Symposium, AeroSense, Orlando Florida, April 2001
- ⁵ Greenwood, D.P. and Fried, D.L., "Power spectra requirements for wave-front-compensative systems," J. Opt. Soc. Am., Vol. 66, No. 3, Mar. 1976, pp 193-206.
- ⁶ Greenwood, D.P., "Bandwidth specification for adaptive optic systems," J. Opt. Soc. Am., Vol 67, No. 3, Mar. 1977, pp. 390-393.
- ⁷ Fried, D.L., "Time-delay-induced mean-square error in adaptive optics," J. Opt. Soc. Am. A, Vol. 7, No. 7, July 1990, pp. 1224-1225.

CHAPTER 10:

CONCLUSIONS AND RECOMMENDATIONS

10.1. Context and Foundations

The work presented in this dissertation, and the contributions of this work, are born from a return to fundamental principles and a revelation of the limits of conventional wisdom. The conventional wisdom and standard terminology used to describe and characterize fluid-optic interactions until now have been based upon various simplifying assumptions regarding the flow and the mechanisms by which the flow produces optical distortions. These assumptions have proven to be fairly accurate in describing incompressible flows, such as the free atmosphere. Over decades, these characterizing descriptions and parameters have become so widely used in fields involving fluid-optic interactions that the underlying assumptions have largely passed out of mind.

The need to reconsider the fundamentals in fluid-optic interactions was revealed in Hugo's work at AEDC,¹ described in chapter 5. This was built upon by Fitzgerald² and Cicchiello³ in developing and refining the Weakly-Compressible Model (WCM). This model, described in chapter 6, has proven to be far more accurate than the previously accepted models in predicting the degree and magnitude of optical distortion produced by a compressible shear layer.

10.2. Development and Understanding

My work as a graduate student at Notre Dame began with assisting Chouinard⁴ in the experimental work described in section 7.3. These experiments confirmed the existence of organized structures and associated pressure fluctuations predicted by the WCM. After Chouinard's departure, I continued onward with measurements of the optical effects of these structures. In performing these measurements and looking at the associated power-density spectra, I did find confirmation that the optical effects of the flow were dominated by the coherent structures predicted by the WCM, and that the average size and frequency of those structures was determined by the growth of the shear layer.

The existence and near-periodic nature of these structures in the flow led to the use of a sine wave as a simplified simulation of the beam deflection or wavefront aberrations. Originally, this was not the focus of my research, but merely an attempt to increase my understanding of the process of OPD reconstruction and post-processing of the data, which included tip-tilt (T/T) removal. In performing this simplified simulation, I found explanations for some of the traits seen in the mean OPD_{rms} of the reconstructed wavefronts, particularly the effect of aperture size relative to the simulated structure size. Continuing forward with this work led to the set of filter functions describing various types of optical correction, which are presented in section 8.4 and derived in detail in appendix B.

Reviewing the literature and exploring the existing body of knowledge on wavefront correction, or more generally what is termed beam control, revealed to me that these insights and formulas had applications beyond descriptions and corrective systems

317

to deal with shear layers. As was presented in chapter 4, the most commonly used parameters in characterizing an optically-aberrating environment are based on assumptions of spectral characteristics and mechanisms of optical distortion that simply do not apply to compressible shear layers or to the wider field of aero-optics, of which compressible shear-layer flows are merely a part.

10.3. Contributions

The first contribution of this work is to sound a general warning for the adaptiveoptics community. As alluded to earlier, many researchers and engineers in the field of optical correction, when encountering aero-optic flow conditions and attempting to deal with the distortions produced by such flows, continue to use the standard parameters of C_n^2 , the Fried parameter, and the Greenwood frequency. As was addressed in chapter 4, a significant body of work and engineering guidelines for optical correction have been produced regarding atmospheric distortions, but these rules are not applicable if the underlying assumptions that went into their development are not applicable. Simply calling attention to this fact is a worthwhile contribution by itself.

The filter-function descriptions of corrective systems presented in this dissertation are not quite as neat as the guidelines for characterizing and correcting atmospheric propagation. Aero-optic flows tend to have a dominant range of frequencies, but one or two numbers will not tell an engineer everything he needs to know about the conditions, as the Fried parameter and Greenwood frequency do for atmospheric distortions. Then again, the original description of bandwidth requirements in corrective systems produced by Fried and Greenwood⁵ was also messily complicated, and did not produce the commonly-referenced Greenwood frequency until Greenwood applied a number of simplifications in a following paper. Similarly, the filter functions can be applied in greater or lesser detail, depending on how much detail is needed.

At the simplest level, the 3-dB cut-off frequency associated with each type of correction provides a guide to engineers regarding the necessary bandwidth in systems of this sort. Given an aperture size and a relative perpendicular velocity of the air to the optical path, the cut-off Strouhal number will indicate an upper limit to frequency affected by this type of correction. If the corrective system is unable to operate at this frequency, then it will fail to achieve maximum effectiveness. On the other hand, attempting to design and implement a system with a bandwidth significantly beyond this cut-off frequency can only add to the cost of the system without reaping any benefits in improved performance.

If more details concerning the effectiveness of the corrective system are desired, then the Taylor-series approximations can be applied. These take the form of relatively simple power-law descriptions, which adequately describe the effects of those forms of correction for virtually all frequencies below the cutoff frequency, while forms of correction other than G-tilt removal have virtually no effect on disturbances of a frequency higher than the cut-off.

Something to keep in mind about these filter functions is that they are meant to be applied to power-density spectra of a set of disturbances, but, judging from the comparison of this filter to the Boeing-SVS results in chapter 9, they can be applied universally to any such spectra. This is what sets them apart from the guidelines developed for correcting atmospheric propagation, since those rules are based on the disturbances having certain spectral and stochastic properties.

10.4. Recommendations for Future Work

The final proof of any hypothesis is whether or not it describes the physical world. The significance of a contribution to scientific knowledge and engineering applications is proven by whether or not others can make use of it. In chapter 9, the general guidelines and filter functions for tip-tilt correction were applied to data collected from a compressible shear layer, and compared to work done by other researchers. These comparisons indicate that the corrective filter functions used are applicable over a wide range of disturbance types; however, there is no end to the types of disturbance that may be found in the physical world. Likewise, there are many corrective systems already in existence to deal with these effects, and many more yet to be devised.

A major difference between corrective-system filter functions presented in this dissertation to the commonly-used guidelines developed for atmospheric propagation is a trade-off between simplicity and breadth of application. The rules established for atmospheric correction can characterize atmosphere-induced distortions with one or two numbers and likewise can predict the phase variance likely to remain after correction. Using a filter function to predict residual phase variance requires applying the filter to the power spectrum of the disturbance and then integrating over all frequencies. This step was used in developing the guidelines for atmospheric correction, but is normally skipped over in applying these rules due to implicit assumptions regarding the characteristics of the optical distortions. For the sake of utility, a means of characterizing any optically-

distorting flow in a few parameters should be sought after. In pursuing this, it may be necessary to limit the description to aero-optic flows of the sort that may be encountered around a vehicle in flight, but a set of simplified aero-optic rules would complement the existing rules for atmospheric-propagation. In fields of applied science, it is not uncommon to have one set of rules and guidelines for one regime of conditions, and another set for different conditions.

Section 8.5 addressed higher-order correction. Simulated correction was sufficient to establish cut-off frequencies for these forms of correction, but proved to have limitations in resolving disturbances of particularly high or low frequencies relative to this cut-off value. Likewise, derivations for analytical expressions of higher-order correction are prepared and presented in section 8.5 and section B.7 of appendix B, but carrying these derivations through to expressions simple enough for useful application is beyond the scope of this dissertation; however, taking this work to the more useful form would be a valuable goal for future work.

¹ Hugo, R. J., Jumper, E. J., Havener, G. & Stepanek, C. 1997 "Time-resolved wave front measurements through a compressible free shear layer. AIAA J. Vol. 35, pp 671–677.

² Fitzgerald, E.J., Jumper, E.J., "The optical distortion mechanism in a nearly incompressible free shear layer," J. Fluid Mech., Vol. 512, pp 153-189.

³ Cicchiello, J.M., Jumper, E.J., "Low-order representation of fluid-optic interactions associated with a shear layer," AIAA Paper 2001-0952, January 2001.

⁴ Chouinard, M., Asghar, A., Kirk, J.F., Siegenthaler, J.P., and Jumper, E.J., "An experimental verification of the weakly-compressible model," AIAA paper 02-0352

⁵ Greenwood, D.P., and Freid, D.L., Power Spectra Requirements for Wave-Front-Compensative Systems, J. Opt. Soc. Am., Vol. 66, March 1976, pp 193-206.

APPENDIX A:

THE REDESIGN AND CONSTRUCTION OF A NEW SHEAR LAYER FACILITY

A.1. Introduction

The Weakly-Compressible Shear Layer (WCSL) facility described in section 7.2 was useful in this study, but was also found to have limitations that could hinder further efforts. In particular, a wider test section was desired to accommodate larger beams and to investigate possible cross-stream effects in shear layers. In the process of designing and constructing a new facility with a larger cross-section, various modifications made to the original facility were refined and incorporated into the new facility, which is shown in the wireframe model and photo of Fig. A.1 below.



Figure A.1: Wire-frame design and photo of the second-generation WCSL facility.

A.2. Test Section

The primary goal of redesigning the tunnel was increase the width of the test section. The previous facility had a high-speed flow of approximately 0.8 M and a cross-sectional area of 1.5 inches by 3 inches (0.0029 m^2). Drawing from free air in the room, this indicates a flow velocity of 260 m/s and a density of 0.85 kg/m³. Indicating a mass flow rate of

$$0.85kg/m^3 \cdot 0.0029m^2 \cdot 260m/s = 0.64kg/s.$$
 (A.1)

The low-speed side of the original facility had a flow of about 0.09 M at the same static pressure (64 kPa) as the high-speed flow. This corresponded to a speed of 31 m/s, and a density of 0.76 kg/m³. With an area of 0.0085 m², this results in a mass flow rate of

$$0.76kg/m^3 \cdot 0.0085m^2 \cdot 31m/s = 0.20kg/s.$$
 (A.2)

Thus, the facility had a net mass flow rate of 0.84 kg/s. If the atmosphere being drawn from had a density of 1.2 kg/m^3 , then this is a volume flow rate of

$$\frac{0.84kg/s}{1.2kg/m^3} = 0.7m^3/s = 35.6ft^3/s = 2100ft^3/\min.$$
 (A.3)

The three pumps available to power the previous system were rated for 3310 ft³/min each, so it was judged that it should be possible to maintain the same flow conditions produced in the original WCSL with a cross-section three times as wide.

One of the refinements made over the course of using the previous facility was to replace the solid Plexiglas upper and lower walls of the test section with removable blocks, so that optically-flat windows could be inserted at points where optical measurements were to be taken. The modified first-generation facility still had a few structural members that could not be removed and blocked optical access at some points due to metal bolts running through the Plexiglas. The feature of removable sections was worked into this version of the facility without the sometimes-inconvenient cross-bracing. A wire-frame diagram and photo of this test section are shown in Fig. A.2. As previously indicated, it has the same dimensions as the test section of the previous version, except for the width being increased from 3.8 cm (1.5 inches) to 11.4 cm (4.5 inches).



Figure A.2: Wire-frame design and photo of test section for the second WCSL.

A.3. Diffuser

The existing diffuser attached to the wall of the laboratory, and to the port leading to the Allis Chalmer pumps behind the wall, expanded horizontally but not vertically. Thus, modifying this section to meet the new dimensions was simply a matter of cutting the diffuser at a point where its internal dimensions had expanded to match the desired 11.4 cm width, and wielding a new flange to the end with the necessary holes for attaching other portions of the new facility



Figure A.3: Wire-frame design of the diffuser for the second WCSL.

A.4. Sonic Throat

As the width of the test section was increased by a factor of three, so too was that of the sonic throat. No other modifications were necessary. A side-view schematic of the new sonic throat is shown in Fig. A.4.



Figure A.4: Sonic throat for the second WCSL.

A.5. Splitter-Plate Section

As noted above, the high and low-speed inlets of the first version of the WCSL facility were joined, having been fashioned out of one inlet with a plate inserted to partition it into two inlets. In the redesigned facility, the inlets for the two streams are separate pieces and a short section, shown in Fig. A.5, is placed between the inlets and the test section to bring the two flows into contact with a sharp trailing edge. A portion of the piece forming the splitter plate extends somewhat past the edges of this section so that it may fit into a socket formed when the high and low-speed nozzles are joined.



Figure A.5: Wire-frame design and photo of splitter-plate section for uniting the high and low-speed flows.

This is based on a modification to the earlier version of the facility that not used during the experiments in which the data presented in section 7.4.2 was gathered. Running parallel to the study and characterization of the shear layer described in this dissertation were attempts to implement flow control within the shear layer for the purpose of regularizing the structures shown to be present in section 7.3. From the standpoint of a corrective system, regularized disturbances are easier to deal with in that the next few cycles of such a disturbance can be predicted in advance.

A.6. High-Speed Inlet

The inlets for the original WCSL were produced by modifying a pre-existing inlet nozzle that had been used in other experiments. This was done by soldering a sheet of

metal inside nozzle, separating it into upper and lower sections, with the upper section forming a contraction drawing air from the room, while the lower section formed an expansion and drew air from a settling tank. As there was no pre-existing inlet nozzle to fit the dimensions of the new, wider WCSL. New inlets were designed from scratch, with the high-speed contraction and the low-speed expansion being separate pieces.

The new high-speed has an intake area of 53.3 cm by 80.0 cm (0.427 m^2) and discharges into the test section over an area of 11.4 cm by 7.6 cm (0.00871 m^2) , producing a contraction ratio of 49 to 1 over a length of 91.4 cm. A honeycomb of plastic tubes, often called a strawbox, has been placed in front of the inlet for safety and flow-straightening. A wire-frame diagram and photo of this nozzle are shown in Fig. A.6.



Figure A.6: Wire-frame design and photo of high-speed inlet nozzle for the second WCSL.

A.7. Low-Speed Inlet

For the low-speed portion of the shear layer, the original WCSL drew air from a settling tank through an expansion. The air in the settling tank was drawn from the room, through a set of valves to reduce the total pressure of this air. The original ball valve was eventually replaced with a set of "quiet valves", using sets of tubes to achieve the same total pressure loss with less noise. In the redesign of the facility, it was decided to do away with the settling tank entirely, replacing it with a series of quiet valves.

Thus, the new low-speed inlet is comprised of two sections, an expansion section, and a pressure drop section. The expansion section begins with an internal area of 6.86 cm by 13.4 cm (0.00917 m^2) and ends with an area of 11.4 cm by 22.3 cm (0.0254 m^2) for an expansion ratio of 2.78 to 1 over a length of 0.76 m. The low-speed inlet is bent to fit under the high-speed inlet, as can be seen in the pictures of the facility in Fig. A.1. Therefore, the flow is also made to turn by 28 degrees as it passes through this segment. A set of guide vanes inside this section, not shown in the diagrams of Fig. A.7 are used to assist this redirection of the flow and prevent separation.



Figure A.7: Wire-frame design of low-speed inlet expansion for the second WCSL.

The pressure-drop section is a straight rectangular tube, with a length of 0.813 m (2ft. 8 in) and an internal cross-sectional area of 8.9 cm by 15.2 cm (3.5 in. by 6 in). The desired drop in total pressure is produced by passing the air through a series of strawboxes one of which is shown in Fig. A.8. This material, forming a honeycomb of plastic tubes, is commonly used in the inlets of wind tunnels to straighten the incoming flow. A larger section of the same material can be seen in front of the high-speed inlet in

Fig. A.6. The portion shown in Fig. A.8 is 10.1 cm (4 in.) thick, and has been cut to the dimensions of the pressure-drop section interior so that it may be placed inside that section shown in Fig. A.9. These dimensions, are a little bigger than the intake area of the expansion section so that the straw-boxes will not be drawn into the facility.



Figure A.8: "Straw-box" and spacer used for inducing total pressure loss.



Figure A.9: Wire-frame design and photo of low-speed section (below the high-speed inlet) for inducing a drop in total pressure.

Also shown in Fig. A.8 is a 0.5-cm thick rectangular plastic frame with outer dimensions also sized to fit into the pressure-drop section, and inner dimensions that match the intake area of the expansion section. As with the quiet-valve added to the original facility and described in section 7.2.2, the desired reduction in total pressure is produced by friction as the air passes through the tubes of the straw-boxes and inlet and outlet effects as the air enters and leaves each box. The spacers are inserted between the straw-boxes to provide space between the exits of the tubes making up one straw-box and the inlets of the next, so that the effects associated with inlets and outlets can fully manifest.

The low-speed portion of the flow in the wider test section was calculated to have a mass-flow rate of about 0.66 kg/s at the desired flow conditions. The air in the room has a density of about 1.16 kg/m^3 . Air of that density moving through the area open to

flow in this section requires a velocity of about 54 m/s to match the mass-flow rate for the corresponding portion of the test section. This velocity is well below the transonic region for ambient, room-temperature air, and so compressibility effects may be neglected for this rough estimate. As was addressed in section 7.2.2, friction losses are found via the Darcy-Weisbach equation,¹

$$\Delta p = f \frac{L}{D} \rho \frac{V^2}{2}, \qquad (10.4)$$

in which D and L are the inner diameter and length of tube respectively. In this equation, f represents a friction factor, which may be found by from Moody diagrams or iteratively from the Colebrook-White equation,²

$$\frac{1}{\sqrt{f}} = -2\log_{10}\left(\frac{\varepsilon}{3.7D} + \frac{2.51}{\operatorname{Re}\sqrt{f}}\right),\tag{A.5}$$

in which ε is the roughness of the interior pipe surface and the Reynolds number is based on the diameter of the tube. The tubes making up the strawboxes have a diameter of about 3 mm and roughness for the plastic material of these tubes is estimated to be 5×10^{-8} m. Based on this and the two equations above, the drop in total pressure due to friction as the air passes through the tubes is expected to be about 1800 Pa.

The loss of total pressure associated with the inlet and outlet of a duct is found by

$$\Delta p = K \rho \frac{V^2}{2} \,. \tag{A.6}$$

The constant *K* is determined by the characteristics of the inlet or outlet. For any type of outlet, K = 1 and the expected total pressure loss is 1700 Pa. The spacers between strawboxes shown in Fig. A.8 are inserted to allow this exit effect to manifest. For inlets, *K* is primarily determined by how sharp of a turn air must make when entering the duct.

An inlet with sharp corners has an associated value of K = 0.5. For rounded inlets, K asymptotically approaches 0 as the radius curvature for the rounded edges increases relative to the radius of the duct. The inlets to the tubes making up a strawbox are not rounded. However, each tube is surrounded by other tubes, also drawing air. Therefore, air entering a given tube does not have to turn significantly to do so, as the air to the sides of the inlet enters the tubes to the sides of that tube. Because of this, the effects of the inlet on total pressure are likely to be negligible compared to the effects of the outlet and wall friction in this case.

This estimate suggests that one strawbox of the indicated dimensions should induce a drop in total pressure of about 3.5 kPa. To meet the desired conditions for flow in the test section, a drop of about 33 kPa is needed. This suggests that nine or ten such boxes in series should produce a total pressure drop close to this desired value. However, this does not take into account the fact that the pressure drop produced by the first box in the series will change the properties of the air entering the second box, which will change the properties for the third box, and so on. A more detailed analysis taking this into account indicated that a series of eight boxes would produce a drop in total pressure of 33.9 kPa.

The section shown in Fig. A.9 was constructed to hold eight strawboxes with associated spacers between each box. It was also constructed so that boxes could be removed if the actual pressure drop proved to be too high, and another section could be added onto the end with more pressure-loss inducing devices if the actual pressure drop proved to be too low. In fact, the drop in total pressure produced by eight strawboxes has proven to be perfectly suited to the needs of this facility.

335

² Colebrook, C.F. "Turbulent flow in pipes, with particular reference to the transition region between smooth and through pipe laws," Jour. Ist. Civil Engrs., London (Feb. 1929)

¹ Darcy, H.,. "Recherches Experimentales Relatives au Mouvement de L'Eau dans les Tuyaux," ("Experimental Research Relating to the Movement of Water in Pipes"), Mallet-Bachelier, Paris, 1857

APPENDIX B:

EXPANDED DERIVATIONS OF APERTURE FILTER FUNCTIONS

B.1. Introduction

In chapter 8 the effects of several types of correction were expressed in terms of gain functions in the form of the time-averaged mean-square *OPD* after correction divided by the time-averaged mean-square *OPD* of the uncorrected case.

$$G(St_{Ap}) = \frac{\left(\frac{OPD_{rms}}{OPD_{rms}}\right)^2}{\left(OPD_{rms}\right)^2}$$
(B.7)

The Strouhal number based on the aperture size and convective velocity perpendicular to the optical path also corresponds to the ratio of the aperture diameter to the period of the sinusoidal wavefront used in these derivations.

$$f = \frac{U_{\perp}}{\Lambda} \tag{B.8}$$

$$St_{Ap} = \frac{A_p f}{U_c} = \frac{A_p U_c / \Lambda}{U_c} = \frac{A_p}{\Lambda}$$
(B.9)

In many of these derivations, there are intermediate stages in which the equations contain a large number of terms that are eventually condensed into a smaller number of terms in the final result. These intermediate stages are not necessary for understanding of the results and so were skipped in most of the derivations in chapter 8. They are presented in this appendix for completeness.

B.2. Uncorrected Cases

All the derivations of chapter 8 used a sinusoidal waveform to represent the wavefront to be corrected, of the form

$$OPD(x,t) = K \sin\left[2\pi \left(f \cdot t - \frac{x}{\Lambda}\right)\right] = K \sin\left[2\pi \left(\frac{U_C}{\Lambda} \cdot t - \frac{x}{\Lambda}\right)\right].$$
 (B.10)

This was to be corrected, either over a one-dimensional aperture of length A_p , or a circular aperture of diameter A_p .

B.2.1. One-Dimensional Aperture

Over a one-dimensional aperture the OPD_{rms} at any moment in time is

$$OPD_{rms}(A_p, t) = \sqrt{\frac{1}{A_p} \int_{-A_p/2}^{A_p/2} [OPD(x, t)]^2 dx} .$$
(B.11)

In expressing these functions as a form of filter gain, it is more physically meaningful to work with mean-squared *OPD* rather than root-mean-squared *OPD*.

$$\left(OPD_{rms}(A_{p},t)\right)^{2} = \frac{1}{A_{p}} \int_{-Ap/2}^{Ap/2} [OPD(x,t)]^{2} dx.$$
(B.12)

Using the expression for *OPD* in Eq. B.10, $(OPD_{rms})^2$ is found to be

$$\left(OPD_{rms}(A_{p},t)\right)^{2} = \frac{1}{4} \frac{K^{2}}{\pi A_{p}} \begin{bmatrix} 2\pi A_{p} - \Lambda \cos\left(\frac{\pi (A_{p} - 2U_{C}t)}{\Lambda}\right) \sin\left(\frac{\pi (A_{p} - 2U_{C}t)}{\Lambda}\right) \\ -\Lambda \cos\left(\frac{\pi (A_{p} + 2U_{C}t)}{\Lambda}\right) \sin\left(\frac{\pi (A_{p} + 2U_{C}t)}{\Lambda}\right) \end{bmatrix}.$$
 (B.13)

A full cycle of the disturbance in Eq. B.10 occurs over the interval from t = 0 to $t = \Lambda/U_C$. Averaging over this period in time will produce the same result that would be approached asymptotically by averaging over increasingly longer periods of time:
$$\overline{\left(OPD_{rms}(A_p)\right)^2} = \frac{U_C}{\Lambda} \int_{0}^{\Lambda/U_C} \left(OPD_{rms}(A_p,t)\right)^2 dt = \frac{1}{2}K^2.$$
(B.14)

B.2.2. Two-Dimensional Aperture

A square or rectangular aperture is virtually identical to a one-dimensional aperture, provided the disturbances are aligned along one of the two axes for the aperture. However, the same can not be said for a circular aperture. The edge of a circular aperture with a diameter of A_p and centered at (x,y) = (0,0) is defined by

$$y = \pm \sqrt{\frac{A_p^2}{4} - x^2}$$
 (B.15)

Therefore, integrating over the aperture may be represented by integrating in y to the limits of Eq. B.15 and then over the range from $-A_p/2$ to $A_p/2$ in x. Over such an aperture the OPD_{rms} at any moment in time is

$$OPD_{rms}(A_p,t) = \sqrt{\frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} [OPD(x,t)]^2 \, dy dx} \,.$$
(B.16)

As with the one-dimensional apertures, it is more physically meaningful to work with mean-squared *OPD* rather than root-mean-squared *OPD*.

$$\left(OPD_{rms}(A_p,t)\right)^2 = \frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{-\sqrt{A_p^2/4-x^2}}^{\sqrt{A_p^2/4-x^2}} [OPD(x,t)]^2 dy dx.$$
(B.17)

The expression for *OPD* in Eq. B.10 can be used to represent a two-dimensional wavefront that only has variations in the *x* direction. Using this equation, $(OPD_{rms})^2$ is found to be based in part on Bessel functions of the first kind, represented in Eq. B.18 and following equations in this appendix by J_n for an nth order Bessel function.

$$\left(OPD_{rms}\left(A_{p},t\right)\right)^{2} = \frac{4K^{2}}{\pi A_{p}^{2}} \sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ + \frac{A_{p}\Lambda}{2} \sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right) \\ + \frac{A_{p}}{2} \cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right) \left[\pi A_{p} - \Lambda J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right)\right] \\ - \frac{A_{p}}{2} \cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \left[\pi A_{p} - \Lambda J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right)\right] \\ + \frac{A_{p}}{8}\left[\pi A_{p} - \Lambda J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right)\right] \\ + \frac{A_{p}}{8}\left[\pi A_{p} - \Lambda J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right)\right]$$
(B.18)

As in the one-dimensional case, a full cycle of the disturbance in Eq. B.10 occurs over the interval from t = 0 to $t = \Lambda/U_C$ and averaging over this period in time will produce the same result that would be approached asymptotically by averaging over increasingly longer periods of time. Interestingly, it also leads to the same end result found for the one-dimensional case.

$$\overline{\left(OPD_{rms}(A_p)\right)^2} = \frac{U_C}{\Lambda} \int_{0}^{\Lambda/U_c} \left(OPD_{rms}(A_p, t)\right)^2 dt = \frac{1}{2}K^2.$$
(B.19)

B.3. Piston Correction

The simplest form of correction may be that of adjusting the average phase or *OPD* of the wavefront in the aperture to zero. This average phase or displacement is called piston. There are also some applications, such as those where effects in the far field are of primary interest, for which the piston over an aperture is irrelevant, and only T/T or higher-order effects have impact. In such cases, simply having a finite aperture or beam diameter has a filtering effect equivalent to that of piston correction.

B.3.1. One-Dimensional Aperture

The piston-corrected wavefront is of the form

$$_{cor}OPD(x,t) = OPD(x,t) - A(t).$$
(B.20)

The coefficient A(t) is selected to remove the mean piston in the uncorrected wavefront at each point in time, so that

$$A(t) = \frac{1}{A_p} \int_{-A_p/2}^{A_p/2} OPD(x,t) dx.$$
 (B.21)

Using the sinusoidal wavefront of Eq. B.10

$$A(t) = \frac{K\Lambda}{2\pi A_p} \left[\cos\left(\frac{\pi (A_p - 2U_C t)}{\Lambda}\right) - \cos\left(\frac{\pi (A_p + 2U_C t)}{\Lambda}\right) \right].$$
(B.22)
$$\left(_{cor} OPD_{rms} (A_p, t) \right)^2 = \frac{1}{A_p} \int_{=Ap/2}^{Ap/2} (OPD(x, t) - A(t))^2 dx$$
$$\left[\pi^2 A_p^2 - \pi \Lambda A_p \sin\left(\frac{\pi A_p}{\Lambda}\right) \cos\left(\frac{\pi A_p}{\Lambda}\right) \right]$$

$$= \frac{1}{2\pi^{2}} \frac{K^{2}}{A_{p}^{2}} - \pi \Lambda A_{p} \sin\left(\frac{\pi A_{p}}{\Lambda}\right) \cos\left(\frac{\pi A_{p}}{\Lambda}\right) \\ - 4\pi \Lambda A_{p} \sin\left(\frac{\pi A_{p}}{\Lambda}\right) \cos\left(\frac{\pi A_{p}}{\Lambda}\right) \cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ + 4\pi \Lambda A_{p} \sin\left(\frac{\pi A_{p}}{\Lambda}\right) \cos\left(\frac{\pi A_{p}}{\Lambda}\right) \cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ + 4\pi \Lambda A_{p} \sin\left(\frac{\pi A_{p}}{\Lambda}\right) \cos\left(\frac{\pi A_{p}}{\Lambda}\right) \sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ + 4\pi \Lambda A_{p} \sin\left(\frac{\pi A_{p}}{\Lambda}\right) \cos\left(\frac{\pi A_{p}}{\Lambda}\right) \sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ - 8\Lambda^{2} \sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) \cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) \cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ \end{bmatrix}$$
(B.23)

$$\overline{\left(_{cor} OPD_{rms}(A_p)\right)^2} = \frac{U_C}{\Lambda} \int_0^{\Lambda/U_C} \left(_{cor} OPD_{rms}(A_p, t)\right)^2 dt$$
$$= \frac{1}{2\pi^2} \frac{K^2}{A_p^2} \left[\pi^2 A_p^2 - \Lambda^2 + \Lambda^2 \cos^2\left(\frac{\pi A_p}{\Lambda}\right)\right].$$
(B.24)

Using Eqs. B.9 and B.14,

$$G(St_{Ap}) = \frac{\overline{\left(_{cor} OPD_{rms}(A_{p})\right)^{2}}}{\overline{\left(OPD_{rms}(A_{p})\right)^{2}}} = 1 + \frac{\cos^{2}(\pi St_{Ap}) - 1}{(\pi St_{Ap})^{2}}.$$
 (B.25)

Plotting Eq. B.25 in decibels against a logarithmic scale (see Fig. 8.22 on page 258) indicates that this function may be approximated by a 2nd-order power function for small values of St_{Ap} . Using a Taylor-series approximation for a cosine:

$$G(St_{Ap}) \cong 1 + \frac{\left(1 - \frac{1}{2}(\pi St_{Ap})^{2} + \frac{1}{24}(\pi St_{Ap})^{4} + O(St_{Ap}^{6})\right)^{2} - 1}{(\pi St_{Ap})^{2}}$$
$$= 1 + \frac{\left(1 - (\pi St_{Ap})^{2} + \frac{1}{3}(\pi St_{Ap})^{4} + O(St_{Ap}^{6})\right) - 1}{(\pi St_{Ap})^{2}} = 1 + \frac{-(\pi St_{Ap})^{2} + \frac{1}{3}(\pi St_{Ap})^{4} + O(St_{Ap}^{6})}{(\pi St_{Ap})^{2}}$$
$$= \frac{\pi^{2}}{3}St_{Ap}^{2} + O(St_{Ap}^{4}) \cong 3.29St_{Ap}^{2}.$$
(B.26)

In the approximation above, $O(St_{Ap}^X)$, indicates a term on the order of St_{Ap} to the X power.

B.3.2. Two-Dimensional Aperture

Over a two-dimensional circular aperture, let the piston-corrected wavefront remain of the form in Eq. B.20, but mean piston and the coefficient A(t) will be defined over the aperture in both directions. Mean piston and the coefficient A(t) are defined by

$$A(t) = \frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{-\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} OPD(x, t) dy dx.$$
(B.27)

Using the sinusoidal wavefront of Eq. B.10

$$A(t) = \frac{2K\Lambda}{\pi A_p} \sin\left(\frac{2\pi U_C}{\Lambda}t\right) J_1\left(\frac{\pi A_p}{\Lambda}\right).$$
(B.28)

$$\left(_{cor} OPD_{rms}(A_p, t)\right)^2 = \frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{-A_p/2}^{\sqrt{A_p^2/4 - x^2}} (OPD(x, t) - A(t))^2 dy dx$$

$$2K^{2}\left[\frac{1}{4} + \sin^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right) + \cos^{4}\left(\frac{\pi U_{C}}{\Lambda}t\right) - \cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)\right]$$
$$-\frac{2K^{2}\Lambda}{\pi A_{p}}\left[\frac{1}{4}J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right) - \sin^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right)\right]$$
$$+\cos^{4}\left(\frac{\pi U_{C}}{\Lambda}t\right)J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right) - \cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right)\right]$$
$$-\frac{2K^{2}\Lambda^{2}}{\pi^{2}A_{p}^{2}}\sin^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)J_{1}^{2}\left(\frac{2\pi A_{p}}{\Lambda}\right).$$
(B.29)
$$\overline{\left(_{cor}OPD_{rms}(A_{p})\right)^{2}} = \frac{U_{C}}{\Lambda}\int_{0}^{\Lambda/U_{C}}\left(_{cor}OPD_{rms}(A_{p},t)\right)^{2}dt$$

$$= \frac{1}{2\pi^2} \frac{K^2}{A_p^2} \left[\pi^2 A_p^2 - 4\Lambda^2 J_1^2 \left(\frac{\pi A_p}{\Lambda} \right) \right].$$
(B.30)

Using Eqs. B.9 and B.19,

=

$$G(St_{Ap}) = \frac{\overline{\left(\frac{OPD_{rms}(A_{p})\right)^{2}}}{\overline{\left(OPD_{rms}(A_{p})\right)^{2}}} = 1 - \frac{4J_{1}^{2}(\pi St_{Ap})}{(\pi St_{Ap})^{2}}.$$
 (B.31)

Plotting Eq. B.31 in decibels against a logarithmic scale (see Fig. 8.22 on page 258) indicates that this function may be approximated by a 2nd-order power function for small values of St_{Ap} . Using a Taylor-series approximation for a first-order Bessel function of the first kind, Eq. B.31 may be approximated as:

$$G(St_{Ap}) \cong 1 - \frac{4\left(\frac{1}{2}(\pi St_{Ap}) - \frac{1}{16}(\pi St_{Ap})^{3} + O(St_{Ap}^{5})\right)^{2}}{(\pi St_{Ap})^{2}}$$
$$= 1 - \frac{4\left(\frac{1}{4}(\pi St_{Ap})^{2} - \frac{1}{16}(\pi St_{Ap})^{4} + O(St_{Ap}^{6})\right)}{(\pi St_{Ap})^{2}} = 1 - \left(1 - \frac{1}{4}(\pi St_{Ap})^{2} + O(St_{Ap}^{4})\right)$$
$$= \frac{\pi^{2}}{4}St_{Ap}^{2} + O(St_{Ap}^{4}) \cong 2.47St_{Ap}^{2}.$$
(B.32)

B.4. Z-Tilt Correction

The term "Z-tilt" is short for "Zernike-tilt" and is based on expressing the wavefront as a summation of Zernike polynomials. The Z_0 polynomial or mode is a constant value that corresponds to piston. The Z_1^1 and Z_1^{-1} modes are planes tilted in the *x* and *y* directions respectively.

B.4.1. One-Dimensional Aperture

For a one-dimensional aperture aligned in the *x*-direction, the tilt-corrected wavefront is of the form

$$_{cor}OPD(x,t) = OPD(x,t) - A(t) - xB(t)$$
. (B.33)

The coefficients A(t) and B(t) are fit to the wavefront so as to minimize the expression

$$\left(_{cor} OPD_{rms}(A_p, t)\right)^2 = \frac{1}{A_p} \int_{A_{p/2}}^{A_{p/2}} (OPD(x, t) - A(t) - xB(t))^2 dx.$$
(B.34)

These values can be found by taking the derivative of the equation above with respect to each coefficient, and setting those derivatives to zero.

$$\frac{\partial}{\partial A} \Big(_{cor} OPD_{rms}(A_p, t)\Big)^2 = \frac{1}{A_p} \int_{A_p/2}^{A_p/2} 2 \Big(OPD(x, t) - A(t) - xB(t) \Big) dx = 0.$$
(B.35)

$$\frac{\partial}{\partial B} \Big(_{cor} OPD_{rms}(A_p, t)\Big)^2 = \frac{1}{A_p} \int_{Ap/2}^{Ap/2} 2x \Big(OPD(x, t) - A(t) - xB(t) \Big) dx = 0. \quad (B.36)$$

In solving this set of two equations and two unknowns, the result for coefficient A(t), which represents piston, has the same expression and value as was found for piston-only correction in Eq. B.21. The tilt-coefficient, B(t), is

$$B(t) = \frac{12}{A_p} \int_{A_p/2}^{A_p/2} OPD(x, t) x dx .$$
 (B.37)

Using the sinusoidal wavefront of Eq. B.10

$$B(t) = \frac{3K\Lambda}{\pi^2 A_p^3} \begin{bmatrix} \Lambda \sin\left(\frac{\pi (2U_C t + A_p)}{\Lambda}\right) - \Lambda \sin\left(\frac{\pi (2U_C t - A_p)}{\Lambda}\right) \\ -\pi \cos\left(\frac{\pi (2U_C t + A_p)}{\Lambda}\right) - \pi \cos\left(\frac{\pi (2U_C t - A_p)}{\Lambda}\right) \end{bmatrix}.$$
 (B.38)

$$\begin{split} \left(_{cor} OPD_{rms} (A_{p}, t)\right)^{2} &= \frac{1}{A_{p}} \int_{-Ap/2}^{Ap/2} (OPD(x, t) - A(t) - xB(t))^{2} dx \\ &= \frac{1}{2} K^{2} - \frac{1}{2} \frac{K^{2} \Lambda}{\pi A_{p}} \sin\left(\frac{\pi A_{p}}{\Lambda}\right) \cos\left(\frac{\pi A_{p}}{\Lambda}\right) \left[\frac{1 + 4\cos^{4}\left(\frac{\pi U_{C}}{\Lambda}t\right) - 4\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)}{-4\sin^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)} \right] \\ &- \frac{K^{2} \Lambda^{2}}{\pi^{2} A_{p}^{2}} \left[\frac{3\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) - 12\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right) + 12\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{4}\left(\frac{\pi U_{C}}{\Lambda}t\right)}{+4\sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\sin^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)} - 36\sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{4}\left(\frac{\pi U_{C}}{\Lambda}t\right)}{\left[-12\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\sin^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)}{+36\sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)} \right] \\ &+ \frac{1}{2}\frac{K^{2} \Lambda^{4}}{\pi^{4} A_{p}^{4}} + 12\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right) + 36\sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)}{-12\sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\sin^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)\cos^{2}\left(\frac{\pi U_{C}}{\Lambda}t\right)}{-12\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{4}\left(\frac{\pi U_{C}}{\Lambda}t\right)} \right] \end{split}$$
(B.39)
$$\overline{\left(_{cor} OPD_{rms}(A_{p})\right)^{2}} = \frac{U_{C}}{\Lambda} \int_{0}^{AU_{C}} (OPD_{rms}(A_{p},t))^{2} dt \end{split}$$

$$= \frac{1}{2\pi^4} \frac{K^2}{A_p^4} \begin{bmatrix} \pi^4 A_p^4 - 3\Lambda^4 - \Lambda^2 \pi^2 A_p^2 + 3\Lambda^4 \cos^2\left(\frac{\pi A_p}{\Lambda}\right) \\ -2\Lambda^2 \pi^2 A_p^2 \cos^2\left(\frac{\pi A_p}{\Lambda}\right) + 6\Lambda^3 \pi A_p \sin\left(\frac{\pi A_p}{\Lambda}\right) \cos\left(\frac{\pi A_p}{\Lambda}\right) \end{bmatrix}.$$
 (B.40)

Using Eqs. B.9 and B.14,

$$G(St_{Ap}) = \frac{\overline{\left(\frac{c_{or} OPD_{rms}(A_{p})\right)^{2}}{(OPD_{rms}(A_{p}))^{2}}}}{\left(\frac{OPD_{rms}(A_{p})}{(\pi St_{Ap})^{2}}\right)^{2}} = 1 - \frac{1 + 2\cos^{2}(\pi St_{Ap})}{(\pi St_{Ap})^{2}} + 6\frac{\sin(\pi St_{Ap})\cos(\pi St_{Ap})}{(\pi St_{Ap})^{3}} - 3\frac{1 - \cos^{2}(\pi St_{Ap})}{(\pi St_{Ap})^{4}}.$$
 (B.41)

Plotting Eq. B.41 in decibels against a logarithmic scale (see Fig. 8.12 on page 246) indicates that this function may be approximated by a 4th-order power function for small values of St_{Ap} . Using Taylor-series approximations for trigonometric functions, Eq. B.41 can be approximated as

$$G(St_{Ap}) \approx \begin{bmatrix} 1 - \frac{1}{(\pi St_{Ap})^2} - \frac{3}{(\pi St_{Ap})^4} \\ -2 \frac{1 - (\pi St_{Ap})^2 + \frac{1}{3}(\pi St_{Ap})^4 - \frac{2}{45}(\pi St_{Ap})^6 + O(St_{Ap}^8)}{(\pi St_{Ap})^2} \\ + 6 \frac{(\pi St_{Ap}) - \frac{2}{3}(\pi St_{Ap})^3 + \frac{2}{15}(\pi St_{Ap})^5 - \frac{4}{315}(\pi St_{Ap})^7 + O(St_{Ap}^9)}{(\pi St_{Ap})^3} \\ + 3 \frac{1 - (\pi St_{Ap})^2 + \frac{1}{3}(\pi St_{Ap})^4 - \frac{2}{45}(\pi St_{Ap})^6 + \frac{1}{315}(\pi St_{Ap})^8 + O(St_{Ap}^{10})}{(\pi St_{Ap})^4} \end{bmatrix}$$

$$= \begin{bmatrix} 1 - \frac{1}{(\pi St_{Ap})^2} - \frac{3}{(\pi St_{Ap})^4} - \frac{2}{(\pi St_{Ap})^2} + \frac{6}{(\pi St_{Ap})^2} + \frac{3}{(\pi St_{Ap})^4} - \frac{3}{(\pi St_{Ap})^2} \end{bmatrix}$$
$$= \begin{bmatrix} -2\left(-1 + \frac{1}{3}(\pi St_{Ap})^2 - \frac{2}{45}(\pi St_{Ap})^4 + O(St_{Ap}^6)\right) \\ + 6\left(-\frac{2}{3} + \frac{2}{15}(\pi St_{Ap})^2 - \frac{4}{15}(\pi St_{Ap})^4 + O(St_{Ap}^6)\right) \\ + 3\left(\frac{1}{3} - \frac{2}{45}(\pi St_{Ap})^2 + \frac{1}{315}(\pi St_{Ap})^4 + O(St_{Ap}^6)\right) \end{bmatrix}$$

$$= \begin{bmatrix} 1 + \left(2 - \frac{2}{3}(\pi St_{Ap})^{2} + \frac{4}{45}(\pi St_{Ap})^{4} + O(St_{Ap}^{6})\right) \\ + \left(-4 + \frac{4}{5}(\pi St_{Ap})^{2} - \frac{8}{105}(\pi St_{Ap})^{4} + O(St_{Ap}^{6})\right) \\ + \left(1 - \frac{2}{15}(\pi St_{Ap})^{2} + \frac{1}{105}(\pi St_{Ap})^{4} + O(St_{Ap}^{6})\right) \end{bmatrix}$$
$$= \begin{bmatrix} \left(1 + 2 - 4 + 1\right) + \left(\pi St_{Ap}\right)^{2} \left(-\frac{2}{3} + \frac{4}{5} - \frac{2}{15}\right) \\ + \left(\pi St_{Ap}\right)^{4} \left(\frac{4}{45} - \frac{8}{105} + \frac{1}{105}\right) + O(St_{Ap}^{6}) \end{bmatrix}$$

$$=\frac{\pi^4}{45}St_{Ap}^4 + O(St_{Ap}^6) \cong 2.16St_{Ap}^2.$$
(B.42)

B.4.2. Two-Dimensional Aperture

Over a two-dimensional aperture, with the possibility of tilt in both the x and y directions, the tilt-corrected wavefront is of the form

$$_{cor}OPD(x,t) = OPD(x,t) - A(t) - xB(t) - yC(t)$$
. (B.43)

For Z-tilt correction the coefficients A(t), B(t), and C(t) are chosen as to minimize the equation

$$\left(_{cor} OPD_{rms}(A_p,t)\right)^2 = \frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{-\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} O(OPD(x,t) - A(t) - xB(t) - yC(t))^2 dy dx \cdot (B.44)$$

These values can be found by taking the derivative of the equation above with respect to each coefficient, and setting those derivatives to zero.

$$\frac{\partial}{\partial A} \Big(_{cor} OPD_{rms}(A_p, t)\Big)^2 = -2 \int_{-Ap/2}^{Ap/2} \int_{\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} OPD(x, t) dy dx + \frac{\pi}{2} A(t) A_p^2 = 0, \quad (B.45)$$

$$\frac{\partial}{\partial B} \Big(_{cor} OPD_{rms}(A_p, t)\Big)^2 = -2 \int_{-A_p/2}^{A_p/2} \int_{\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} S(x, t) dy dx + \frac{\pi}{32} B(t) A_p^4 = 0, \quad (B.46)$$

$$\frac{\partial}{\partial C} \Big(_{cor} OPD_{rms}(A_p, t)\Big)^2 = -2 \int_{-A_p/2}^{A_p/2} \int_{\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} y \cdot OPD(x, t) dy dx + \frac{\pi}{32} C(t) A_p^4 = 0.$$
(B.47)

In solving this set of three equations and three unknowns, the result for coefficient A(t), which represents piston, has the same expression and value as was found for piston-only correction in Eq. B.27. The tilt-coefficients, B(t) and C(t), are

$$B(t) = \frac{64}{\pi A_p^4} \int_{-A_p/2}^{A_p/2} \int_{\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} SPD(x, t) dy dx , \qquad (B.48)$$

$$C(t) = \frac{64}{\pi A_p^4} \int_{-A_p/2}^{A_p/2} \int_{-\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} y \cdot OPD(x, t) dy dx .$$
(B.49)

Using the sinusoidal wavefront of Eq. B.10, there is no tilt or other form of variation in the *y* direction, so C(t) = 0. Net T/T in the *x* direction is such that

$$B(t) = \frac{16K\Lambda\cos\left(\frac{2\pi U_{c}t}{\Lambda}\right)J_{0}\left(\frac{\pi A_{p}}{\Lambda}\right)}{\pi A_{p}^{2}} - \frac{32K\Lambda^{2}\cos\left(\frac{2\pi U_{c}t}{\Lambda}\right)J_{1}\left(\frac{\pi A_{p}}{\Lambda}\right)}{\pi^{2}A_{p}^{3}}.$$
 (B.50)

$$\left(_{cor} OPD_{rms}(A_p,t)\right)^2 = \frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{-\sqrt{A_p^2/4-x^2}}^{\sqrt{A_p^2/4-x^2}} (OPD(x,t) - A(t) - xB(t) - yC(t))^2 dydx$$

$$=2K^{2}\left[\frac{1}{4}-\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)+\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right)+\sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)\right]$$

$$+\frac{2\Lambda K^{2}}{\pi A_{p}}J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right)\left[-\frac{1}{4}+\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)-\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right)+\sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)\right]$$
$$-\frac{\Lambda^{2}K^{2}}{\pi^{2}A_{p}^{2}}\left[J_{0}^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)-4\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)J_{0}^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)+4\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right)J_{0}^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\right]$$
$$+\sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)J_{1}^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)$$
$$+\frac{\Lambda^{3}K^{2}}{\pi^{3}A_{p}^{3}}J_{0}\left(\frac{\pi A_{p}}{\Lambda}\right)J_{1}\left(\frac{\pi A_{p}}{\Lambda}\right)\left[1-4\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)+4\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right)\right]$$
$$-\frac{\Lambda^{4}K^{2}}{\pi^{4}A_{p}^{4}}J_{1}^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\left[1-4\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)+4\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right)\right].$$
(B.51)

$$\overline{\left(_{cor} OPD_{rms}(A_{p})\right)^{2}} = \frac{U_{c}}{\Lambda} \int_{0}^{\Lambda/U_{c}} OPD_{rms}(A_{p},t)^{2} dt$$
$$= \frac{1}{2\pi^{4}} \frac{K^{2}}{A_{p}^{4}} \left[\frac{\pi^{4}A_{p}^{4} - 16\Lambda^{2}\pi^{2}A_{p}^{2}J_{0}^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) - 64\Lambda^{4}J_{1}^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)}{+ 64\Lambda^{3}\pi A_{p}J_{0}\left(\frac{\pi A_{p}}{\Lambda}\right)J_{1}\left(\frac{\pi A_{p}}{\Lambda}\right) - 4\Lambda^{2}\pi^{2}A_{p}^{2}J_{1}^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)} \right].$$
(B.52)

Using Eqs. B.9 and B.19,

$$G(St_{Ap}) = \frac{\overline{\left(\frac{c_{or} OPD_{rms}(A_{p})\right)^{2}}}{\overline{\left(OPD_{rms}(A_{p})\right)^{2}}}$$
$$= 1 - \frac{16J_{0}^{2}(\pi St_{Ap}) + 4J_{1}^{2}(\pi St_{Ap})}{(\pi St_{Ap})^{2}} + 64\frac{J_{0}(\pi St_{Ap})J_{1}(\pi St_{Ap})}{(\pi St_{Ap})^{3}} - 64\frac{J_{1}^{2}(\pi St_{Ap})}{(\pi St_{Ap})^{4}}.$$
 (B.53)

Plotting Eq. B.53 in decibels against a logarithmic scale (see Fig. 8.14 on page 249) indicates that this function may be approximated by a 4th-order power function for small values of St_{Ap} . Using Taylor-series approximations for Bessel functions of the first kind, Eq. B.53 can be approximated as:

$$G(St_{Ap}) \cong \begin{bmatrix} 1 - 16 \frac{\left(1 - \frac{1}{2} (\pi St_{Ap})^{2} + \frac{3}{32} (\pi St_{Ap})^{4} - \frac{5}{576} (\pi St_{Ap})^{6} + O(St_{Ap}^{8})\right)}{(\pi St_{Ap})^{2}} \\ - 4 \frac{\left(\frac{1}{4} (\pi St_{Ap})^{2} - \frac{1}{16} (\pi St_{Ap})^{4} + \frac{5}{768} (\pi St_{Ap})^{6} + O(St_{Ap}^{8})\right)}{(\pi St_{Ap})^{2}} \\ + 64 \frac{\left(\frac{1}{2} (\pi St_{Ap}) - \frac{3}{16} (\pi St_{Ap})^{3} + \frac{5}{192} (\pi St_{Ap})^{5} - \frac{35}{18432} (\pi St_{Ap})^{7} + O(St_{Ap}^{9})\right)}{(\pi St_{Ap})^{3}} \\ - 64 \frac{\left(\frac{1}{4} (\pi St_{Ap})^{2} - \frac{1}{16} (\pi St_{Ap})^{4} + \frac{5}{768} (\pi St_{Ap})^{6} - \frac{7}{18432} (\pi St_{Ap})^{8} + O(St_{Ap}^{10})\right)}{(\pi St_{Ap})^{4}} \end{bmatrix}$$

$$\begin{bmatrix} 1 - \frac{16}{(\pi St_{Ap})^2} + 8 - 1 + \frac{32}{(\pi St_{Ap})^2} - 12 - \frac{16}{(\pi St_{Ap})^2} + 4 \end{bmatrix}$$

$$-16 \left(\frac{3}{32} (\pi St_{Ap})^2 - \frac{5}{576} (\pi St_{Ap})^4 + O(St_{Ap}^6) \right)$$

$$= \left(-4 \left(-\frac{1}{16} (\pi St_{Ap})^2 + \frac{5}{768} (\pi St_{Ap})^4 + O(St_{Ap}^6) \right) \right)$$

$$+ 64 \left(\frac{5}{192} (\pi St_{Ap})^2 - \frac{35}{18432} (\pi St_{Ap})^4 + O(St_{Ap}^6) \right)$$

$$- 64 \left(\frac{5}{768} (\pi St_{Ap})^2 - \frac{7}{18432} (\pi St_{Ap})^4 + O(St_{Ap}^6) \right)$$

$$= \begin{bmatrix} (\pi St_{Ap})^{2} \left(-\frac{3}{2} + \frac{1}{4} + \frac{5}{3} - \frac{5}{12} \right) \\ + (\pi St_{Ap})^{4} \left(\frac{5}{36} - \frac{5}{192} - \frac{35}{288} + \frac{7}{288} \right) + O(St_{Ap}^{6}) \end{bmatrix}$$
$$= \frac{\pi^{4}}{64} St_{Ap}^{4} + O(St_{Ap}^{6}) \cong 1.52St_{Ap}^{4}. \tag{B.54}$$

B.5. G-Tilt Correction

The term "G-tilt" is short for "gradient-tilt" and is based an average of the local gradient at each point on the wavefront.

B.5.1. One-Dimensional Aperture

As with Z-tilt, G-tilt correction for a one-dimensional aperture aligned in the *x*direction, the tilt-corrected wavefront is of the form

$$_{cor}OPD(x,t) = OPD(x,t) - A(t) - xB(t)$$
. (B.55)

However, G-tilt correction is based on the average of the local slopes over the wavefront, so the coefficient B(t) is defined as

$$B(t) = \frac{1}{A_p} \int_{-A_p/2}^{A_p/2} \frac{\partial}{\partial x} OPD(x,t) dx = \frac{1}{A_p} \left[OPD\left(\frac{A_p}{2}, t\right) - OPD\left(-\frac{A_p}{2}, t\right) \right].$$
(B.56)

Piston, which corresponds to the coefficient A(t), is often not part of G-tilt correction, in part because G-tilt is often detected and defined by its effects in the far field, in which mean piston has no effect. Technically, piston is also a separate quantity from Z-tilt, but was included in the derivations of sections B.4.1 and B.4.2 because it normal to define and deal with the Z_0 Zernike polynomial before dealing with the Z_1^1 and Z_1^{-1} polynomials. In the interest of producing a meaningful comparison between G-tilt and Z-tilt correction, A(t) will be defined as to remove mean piston, as was done in sections B.3.1 and B.4.1.

Using the sinusoidal wavefront of Eq. B.10,

$$B(t) = \frac{K}{A_p} \left[\sin\left(\frac{U_C}{\Lambda}t - \frac{A_p}{2\Lambda}\right) - \sin\left(\frac{U_C}{\Lambda}t + \frac{A_p}{2\Lambda}\right) \right].$$
(B.57)

$$\begin{split} \left(_{ov}OPD_{ms}(A_{p},t)\right)^{2} &= \frac{1}{A_{p}}\int_{-Ap/2}^{Ap/2}(OPD(x,t) - A(t) - xB(t))^{2} dx \\ &= \frac{K^{2}}{6} \begin{bmatrix} 5 - 2\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) + 8\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) + 8\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ -8\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right)\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) - 8\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ &+ \frac{K^{2}}{2}\frac{\Lambda}{\pi A_{p}}\sin\left(\frac{\pi A_{p}}{\Lambda}\right)\cos\left(\frac{\pi A_{p}}{\Lambda}\right) \begin{bmatrix} 3 - 12\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) + 12\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ + 4\sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \end{bmatrix} \\ &+ \sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos\left(\frac{\pi A_{p}}{\Lambda}\right) \begin{bmatrix} 3 - 12\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) + 12\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ + 4\sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \end{bmatrix} \\ &+ \sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) - 2\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) - 4\sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ &+ \sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) + 4\sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\sin^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ &+ \sin^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right) + 2\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ &- 2\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right) + 2\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{2}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ &- 2\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right)\cos^{4}\left(\frac{\pi U_{c}t}{\Lambda}\right) \\ &= \frac{K^{2}}{6\pi^{2}A_{p}^{2}} \begin{bmatrix} 4\pi^{2}A_{p}^{2} - 9\Lambda^{2} - \pi^{2}A_{p}^{2}\cos^{2}\left(\frac{\pi A_{p}}{\Lambda}\right) \\ &+ 6\Lambda\pi A_{p}\sin\left(\frac{\pi A_{p}}{\Lambda}\right)\cos\left(\frac{\pi A_{p}}{\Lambda}\right) \end{bmatrix}$$
 (B.59)

Using Eqs. B.9 and B.14,

$$G(St_{Ap}) = \frac{\overline{\left(\frac{cor}{cor}OPD_{rms}(A_{p})\right)^{2}}}{\overline{\left(OPD_{rms}(A_{p})\right)^{2}}}$$
$$= \frac{4}{3} - \frac{1}{3}\cos^{2}\left(\pi St_{Ap}\right) + 2\frac{\sin\left(\pi St_{Ap}\right)\cos\left(\pi St_{Ap}\right)}{\left(\pi St_{Ap}\right)} - 3\frac{1 - \cos^{2}\left(\pi St_{Ap}\right)}{\left(\pi St_{Ap}\right)^{2}}.$$
(B.60)

Plotting Eq. B.60 in decibels against a logarithmic scale (see Fig. 8.18) indicates that this function may be approximated by a 4th-order power function for small values of St_{Ap} . Using Taylor-series approximations for trigonometric functions, Eq. B.60 can be approximated as:

$$G(St_{Ap}) \cong \begin{bmatrix} \frac{4}{3} - \frac{1}{3} \left(1 - (\pi St_{Ap})^2 + \frac{1}{3} (\pi St_{Ap})^4 + O(St_{Ap}^6) \right) \\ + 2 \frac{\left((\pi St_{Ap}) - \frac{2}{3} (\pi St_{Ap})^3 + \frac{2}{15} (\pi St_{Ap})^5 + O(St_{Ap}^7) \right)}{(\pi St_{Ap})} \\ - 3 \frac{1 - \left(1 - (\pi St_{Ap})^2 + \frac{1}{3} (\pi St_{Ap})^4 - \frac{2}{45} (\pi St_{Ap})^6 + O(St_{Ap}^8) \right)}{(\pi St_{Ap})^2} \end{bmatrix}$$

$$= \begin{bmatrix} 1+2-3\\ -\frac{1}{3} \left(-\left(\pi St_{Ap}\right)^{2} + \frac{1}{3} \left(\pi St_{Ap}\right)^{4} + O\left(St_{Ap}^{6}\right) \right) \\ + 2 \left(-\frac{2}{3} \left(\pi St_{Ap}\right)^{2} + \frac{2}{15} \left(\pi St_{Ap}\right)^{4} + O\left(St_{Ap}^{6}\right) \right) \\ + 3 \left(\frac{1}{3} \left(\pi St_{Ap}\right)^{2} - \frac{2}{45} \left(\pi St_{Ap}\right)^{4} + O\left(St_{Ap}^{6}\right) \right) \end{bmatrix}$$

$$= \begin{bmatrix} (\pi St_{Ap})^{2} \left(\frac{1}{3} - \frac{4}{3} + 1\right) \\ + (\pi St_{Ap})^{4} \left(-\frac{1}{9} + \frac{4}{15} - \frac{2}{15}\right) + O(St_{Ap}^{6}) \end{bmatrix} = \frac{\pi^{4}}{45} St_{Ap}^{4} + O(St_{Ap}^{6}) \cong 2.16St_{Ap}^{4}. \quad (B.61)$$

B.5.2. Two-Dimensional Aperture

Over a two-dimensional aperture, with the possibility of tilt in both the x and y directions, the tilt-corrected wavefront is of the form

$$_{cor}OPD(x,t) = OPD(x,t) - A(t) - xB(t) - yC(t)$$
. (B.62)

As with one-dimensional G-tilt correction, the coefficient A(t) for piston is not inherently defined as part of this type of correction, but will be set to the values for removing mean piston that were found in sections B.3.2 and B.4.2. Coefficients B(t) and C(t) are defined as the mean gradient in the *x* and *y* directions respectively, averaged over the aperture.

$$B(t) = \frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} \frac{\partial}{\partial x} OPD(x, t) dy dx , \qquad (B.63)$$

$$C(t) = \frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{-\sqrt{A_p^2/4 - x^2}}^{\sqrt{A_p^2/4 - x^2}} \frac{\partial}{\partial y} OPD(x, t) dy dx .$$
(B.64)

Using the sinusoidal wavefront of Eq. B.10, there is no tilt or other form of variation in the *y* direction, so C(t) = 0. Net G-tilt in the *x* direction is such that

$$B(t) = 4 \frac{K}{A_p} J_1 \left(\frac{\pi A_p}{\Lambda} \right) \left[1 - 2\cos^2 \left(\frac{\pi U_C t}{\Lambda} \right) \right].$$
(B.65)

$$\begin{split} \left(_{cor} OPD_{rms}(A_{p},t)\right)^{2} &= \frac{4}{\pi A_{p}^{2}} \int_{-A_{p}/2}^{A_{p}^{2}/4-x^{2}} \int_{-A_{p}/2}^{\sqrt{A_{p}^{2}/4-x^{2}}} \int_{-A_{p}/2}^{\sqrt{A_{p}/2}} \int_{-A_{p}/2}^{\sqrt{A_{p}/2}}$$

$$-16K^{2} \frac{\Lambda^{2}}{\pi^{2} A_{p}^{2}} J_{1}^{2} \left(\frac{\pi A_{p}}{\Lambda} \right) \left[1 - 4\cos^{2} \left(\frac{\pi U_{C}t}{\Lambda} \right) + 4\cos^{4} \left(\frac{\pi U_{C}t}{\Lambda} \right) \right] + \sin^{2} \left(\frac{\pi U_{C}t}{\Lambda} \right) \cos^{2} \left(\frac{\pi U_{C}t}{\Lambda} \right) \right].$$
(B.66)

$$\overline{\left(_{cor} OPD_{rms}(A_p)\right)^2} = \frac{U_C}{\Lambda} \int_0^{\Lambda/U_C} \left(_{cor} OPD_{rms}(A_p, t)\right)^2 dt$$
$$= \frac{K^2}{2} \left[1 + J_1^2 \left(\frac{\pi A_p}{\Lambda}\right) + 8 \left(\frac{\Lambda}{\pi A_p}\right) J_0 \left(\frac{\pi A_p}{\Lambda}\right) J_1 \left(\frac{\pi A_p}{\Lambda}\right) - 20 \left(\frac{\Lambda}{\pi A_p}\right)^2 J_1^2 \left(\frac{\pi A_p}{\Lambda}\right) \right].$$
(B.67)

Using Eqs. B.9 and B.19,

$$G(St_{Ap}) = \frac{\overline{\left(\frac{c_{or} OPD_{rms}(A_{p})\right)^{2}}}{\left(OPD_{rms}(A_{p})\right)^{2}}$$
$$= 1 + J_{1}^{2} \left(\pi St_{Ap}\right) + 8 \frac{J_{0} \left(\pi St_{Ap}\right) J_{1} \left(\pi St_{Ap}\right)}{\left(\pi St_{Ap}\right)} - 20 \frac{J_{1}^{2} \left(\pi St_{Ap}\right)}{\left(\pi St_{Ap}\right)^{2}}.$$
(B.68)

Plotting Eq. B.68 in decibels against a logarithmic scale (see Fig. 8.20) indicates that this function may be approximated by a 4th-order power function for small values of St_{Ap} . Using Taylor-series approximations for Bessel functions of the first kind, Eq. B.68 can be approximated as:

$$G(St_{Ap}) \cong \begin{vmatrix} 1 + \left(\frac{1}{4}(\pi St_{Ap})^{2} - \frac{1}{16}(\pi St_{Ap})^{4} + O(St_{Ap}^{6})\right) \\ + 8 \frac{\left(\frac{1}{2}(\pi St_{Ap}) - \frac{3}{16}(\pi St_{Ap})^{3} + \frac{5}{192}(\pi St_{Ap})^{5} + O(St_{Ap}^{7})\right)}{(\pi St_{Ap})} \\ - 20 \frac{\left(\frac{1}{4}(\pi St_{Ap})^{2} - \frac{1}{16}(\pi St_{Ap})^{4} + \frac{5}{768}(\pi St_{Ap})^{6} + O(St_{Ap}^{8})\right)}{(\pi St_{Ap})^{2}} \end{vmatrix}$$

$$= \begin{bmatrix} 1+4-5+\left(\frac{1}{4}(\pi St_{Ap})^{2}-\frac{1}{16}(\pi St_{Ap})^{4}+O(St_{Ap}^{6})\right)\\ +8\left(-\frac{3}{16}(\pi St_{Ap})^{2}+\frac{5}{192}(\pi St_{Ap})^{4}+O(St_{Ap}^{6})\right)\\ -20\left(-\frac{1}{16}(\pi St_{Ap})^{2}+\frac{5}{768}(\pi St_{Ap})^{4}+O(St_{Ap}^{6})\right)\end{bmatrix}$$
$$= \begin{bmatrix} (\pi St_{Ap})^{2}\left(\frac{1}{4}-\frac{3}{2}+\frac{5}{4}\right)\\ +(\pi St_{Ap})^{4}\left(-\frac{1}{16}+\frac{5}{24}-\frac{25}{192}\right)+O(St_{Ap}^{6})\end{bmatrix} = \frac{\pi^{4}}{64}St_{Ap}^{4}+O(St_{Ap}^{6})\cong 1.52St_{Ap}^{4}. (B.69)$$

B.6. Temporal Effects

Not only does the type of correction used determine the effectiveness of the corrective system to various wavefront disturbances, but the timing of that correction can also be very important.

B.6.1. Periodic Correction

Digital systems are generally discrete in time, meaning inputs and outputs are updated at discrete moments in time, separated by some non-zero interval. Over the period between these updates, the inputs and outputs are generally held constant at the value of the last update. In applying correction, this may mean that the correction applied at one moment in time is then held constant for the period between updates, even as the optical disturbances and the flow causing those disturbances move and change.

To explore this phenomenon, let the wavefront be of the form

$$OPD(x,t) = K \sin\left(2\pi \left(\frac{x}{\Lambda} + \frac{U_c}{\Lambda}t\right)\right) = K \sin\left(2\pi \frac{x}{\Lambda} + \varphi(t)\right).$$
(B.70)

Let a perfect correction be applied at some time t_0 such that $\varphi(t_0) = \varphi_0$, and then let the correction be held constant over an interval following that time, so that the corrected wavefront over that interval is

$$_{cor}OPD(x,t,\varphi_0) = K\sin\left(2\pi\frac{x}{\Lambda} + \varphi(t)\right) - K\sin\left(2\pi\frac{x}{\Lambda} + \varphi_0\right).$$
(B.71)

The mean-squared OPD in that period, over a one-dimensional aperture, would then be

$$\begin{pmatrix} \sum_{cor} OPD_{rms}(A_p,t,\varphi_0) \end{pmatrix}^2 = \frac{1}{A_p} \int_{-Ap/2}^{Ap/2} OPD(x,t,\varphi_0) dx \\ = K^2 \begin{bmatrix} 1 + \frac{1}{8} \left(\frac{\Lambda}{A_p \pi}\right) \sin\left(\frac{2(-A_p \pi + \Lambda \varphi(t))}{\Lambda}\right) - \cos(\varphi(t) - \varphi_0) \\ - \frac{1}{4} \left(\frac{\Lambda}{A_p \pi}\right) \sin\left(\frac{-A_p 2\pi + \Lambda \varphi(t) + \Lambda \varphi_0}{\Lambda}\right) \\ + \frac{1}{8} \left(\frac{\Lambda}{A_p \pi}\right) \sin\left(\frac{2(-A_p \pi + \Lambda \varphi_0)}{\Lambda}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_p \pi}\right) \sin\left(\frac{2(A_p \pi + \Lambda \varphi(t))}{\Lambda}\right) \\ + \frac{1}{4} \left(\frac{\Lambda}{A_p \pi}\right) \sin\left(\frac{A_p 2\pi + \Lambda \varphi(t) + \Lambda \varphi_0}{\Lambda}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_p \pi}\right) \sin\left(\frac{2(A_p \pi + \Lambda \varphi(t))}{\Lambda}\right) \end{bmatrix}.$$
(B.72)

If this interval lasts for a period of time from t_0 to t_1 such that $\varphi(t_0) = \varphi_0$ and $\varphi(t_1) = \varphi_0 + \Delta \varphi$, with $d\varphi(t)/dt$ being a constant value over this interval as would be the case for frozenflow aberrating structures being carried by with a constant convective velocity, then the average mean-squared *OPD* over this interval will be

$$\overline{\left(_{cor} OPD_{rms}(A_{p},\varphi_{0},\Delta\varphi)\right)^{2}} = \frac{1}{\Delta\varphi} \int_{\varphi_{0}}^{\varphi_{0}+\Delta\varphi} (OPD_{rms}(A_{p},\varphi))^{2} d\varphi$$

$$\begin{bmatrix} 1 + \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{A_{p}\pi}{\Lambda}\right) \cos\left(\frac{A_{p}\pi}{\Lambda}\right) - 2\cos^{2}\left(\varphi_{0}\right) \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{A_{p}\pi}{\Lambda}\right) \cos\left(\frac{A_{p}\pi}{\Lambda}\right) \\ - \frac{2}{\Delta\varphi} \sin(\varphi_{0}) \cos(\varphi_{0}) \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{A_{p}\pi}{\Lambda}\right) \cos\left(\frac{A_{p}\pi}{\Lambda}\right) \\ + \frac{1}{\Delta\varphi} \sin(\varphi_{0} + \Delta\varphi) \cos(\varphi_{0}) \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{A_{p}\pi}{\Lambda}\right) \cos\left(\frac{A_{p}\pi}{\Lambda}\right) \\ + \frac{1}{\Delta\varphi} \sin(\varphi_{0}) \cos(\varphi_{0} + \Delta\varphi) \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{A_{p}\pi}{\Lambda}\right) \cos\left(\frac{A_{p}\pi}{\Lambda}\right) \\ + \frac{1}{\Delta\varphi} \sin(\varphi_{0}) \cos(\varphi_{0} + \Delta\varphi) - \frac{1}{\Delta\varphi} \sin(\varphi_{0} + \Delta\varphi) \cos(\varphi_{0}) \end{bmatrix} . (B.73)$$

Another necessary step is to take the average over all possible starting values of φ_0 for this interval:

$$\overline{\left(_{cor} OPD_{rms}(A_{p}, \Delta \varphi)\right)^{2}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\left(_{cor} OPD_{rms}(A_{p}, \varphi_{0}, \Delta \varphi)\right)^{2}} d\varphi_{0} = \frac{K^{2} (\Delta \varphi - \sin(\Delta \varphi))}{\Delta \varphi}.$$
 (B.74)

It is interesting to note that dependence on A_p simply drops out while performing the last averaging integral. The velocity U_C and length scale Λ are still factors in that $\Delta \varphi$ corresponds to the time interval between corrections ($\tau_1 = t_1 - t_0$) such that

$$\Delta \varphi = 2\pi \frac{U_C \tau_1}{\Lambda}.$$
(B.75)

Using the established OPD_{rms} for an uncorrected, one-dimensional aperture in Eq. B.14,

$$G(\Delta\varphi) = \frac{\left(\frac{1}{cor}OPD_{rms}(\Delta\varphi)\right)^{2}}{\left(OPD_{rms}(\Delta\varphi)\right)^{2}} = 2\frac{\left(\Delta\varphi - \sin(\Delta\varphi)\right)}{\Delta\varphi}.$$
 (B.76)

B.6.2. Latency

In any physical or electronic system there will be delays in calculating and implementing the correction. This is known as latency. To explore this phenomenon, let the sinusoidal wavefront of Eq. B.70 be used again, this time with the correction being applied at each moment, with no update period, but with some time delay, τ_2 , corresponding to a phase delay of $\Delta\theta$. The corrected wavefront would then be of the form

$$_{cor}OPD(x,t,\Delta\theta) = K\sin\left(2\pi\frac{x}{\Lambda} + \varphi(t)\right) - K\sin\left(2\pi\frac{x}{\Lambda} + \varphi(t) - \Delta\theta\right). \quad (B.77)$$

The mean-squared *OPD* at each moment in time over a one-dimensional aperture, would then be

$$\left(_{cor} OPD_{rms}(A_{p},t,\Delta\theta)\right)^{2} = \frac{1}{A_{p}} \int_{-A_{p}/2}^{A_{p}/2} (Cor OPD(x,t,\Delta\theta))^{2} dx$$

$$\left[1 - \frac{1}{8} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{2(A_{p}\pi - \Lambda\varphi(t))}{\Lambda}\right) - \cos(\Delta\theta) + \frac{1}{4} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{2A_{p}\pi - 2\Lambda\varphi(t) + \Lambda\Delta\theta}{\Lambda}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{2(A_{p}\pi - \Lambda\varphi(t) + \Lambda\Delta\theta)}{\Lambda}\right) + \frac{1}{4} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{2A_{p}\pi + 2\Lambda\varphi(t) - \Lambda\Delta\theta}{\Lambda}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{2(A_{p}\pi + \Lambda\varphi(t) - \Lambda\Delta\theta)}{\Lambda}\right) + \frac{1}{4} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{2A_{p}\pi + 2\Lambda\varphi(t) - \Lambda\Delta\theta}{\Lambda}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{2(A_{p}\pi + \Lambda\varphi(t) - \Lambda\Delta\theta)}{\Lambda}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{1}{8}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{1}{8}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_{p}\pi}\right) \sin\left(\frac{1}{8}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_{p}\pi}\right) - \frac{1}{8} \left(\frac{1}{8}\right) - \frac{1}{8} \left(\frac{\Lambda}{A_{p}\pi}\right) - \frac{1}{8} \left(\frac{1}{8}\right) - \frac{1}{8}$$

Averaging this over time by averaging over possible values of $\varphi(t)$ produces the result:

$$\overline{\left(_{cor} OPD_{rms}(A_{p}, \Delta\theta)\right)^{2}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\left(_{cor} OPD_{rms}(A_{p}, \varphi, \Delta\theta)\right)^{2}} d\varphi = K^{2} (1 - \cos(\Delta\theta)).$$
(B.79)

Again, dependence on A_p simply drops out while performing the last averaging integral. The velocity U_C and length scale Λ are still factors in that $\Delta\theta$ corresponds to the timedelay interval τ_2 such that

$$\Delta \varphi = 2\pi \frac{U_c \tau_2}{\Lambda}.$$
 (B.80)

Using the established OPD_{rms} for an uncorrected, one-dimensional aperture in Eq. B.14,

$$G(\Delta\theta) = \frac{\overline{\left(\frac{\cos OPD_{rms}(\Delta\theta)\right)^2}}}{\overline{\left(OPD_{rms}(\Delta\theta)\right)^2}} = 2(1 - \cos(\Delta\theta)).$$
(B.81)

B.6.3. Periodic Correction with Latency

Not all systems have periodic correction, but any system built in the real world has some degree of latency, though the degree of latency may not be significant for the application for which the system is intended. It may be beneficial, then, to take a look at the combined effects of periodic correction and latency. Using the sinusoidal wavefront of Eq. B.70, if the correction to be applied is for a time at which $\varphi(t) = \varphi_0$, but with some delay corresponding to a phase delay of $\Delta\theta$, then the corrected wavefront will be of the form:

$$_{cor}OPD(x,t,\Delta\theta) = K\sin\left(2\pi\frac{x}{\Lambda} + \varphi(t)\right) - K\sin\left(2\pi\frac{x}{\Lambda} + \varphi_0 - \Delta\theta\right).$$
(B.82)

The mean-squared *OPD* at each moment in time over a one-dimensional aperture, would then be

$$\begin{pmatrix} \sum_{cor} OPD_{rms}(A_{p},t,\varphi_{0},\Delta\theta) ^{2} = \frac{1}{A_{p}} \int_{-A_{p}/2}^{A_{p}/2} (\sum_{cor} OPD(x,t,\varphi_{0},\Delta\theta))^{2} dx \\ = \frac{1}{8} \left(\frac{\Lambda}{\pi A_{p}}\right) \sin\left(\frac{2(\pi A_{p}-\varphi(t)\Lambda)}{\Lambda}\right) - \cos(\varphi(t)-\varphi_{0}+\Delta\theta) \\ + \frac{1}{4} \left(\frac{\Lambda}{\pi A_{p}}\right) \sin\left(\frac{2\pi A_{p}-\varphi(t)\Lambda-\varphi_{0}\Lambda+\Lambda\Delta\theta}{\Lambda}\right) \\ - \frac{1}{8} \left(\frac{\Lambda}{\pi A_{p}}\right) \sin\left(\frac{2(\pi A_{p}-\varphi_{0}\Lambda+\Lambda\Delta\theta)}{\Lambda}\right) - \frac{1}{8} \left(\frac{\Lambda}{\pi A_{p}}\right) \sin\left(\frac{2(\pi A_{p}+\varphi(t)\Lambda)}{\Lambda}\right) \\ + \frac{1}{4} \left(\frac{\Lambda}{\pi A_{p}}\right) \sin\left(\frac{2\pi A_{p}+\varphi(t)\Lambda+\varphi_{0}\Lambda-\Lambda\Delta\theta}{\Lambda}\right) \\ - \frac{1}{8} \left(\frac{\Lambda}{\pi A_{p}}\right) \sin\left(\frac{2(\pi A_{p}+\varphi_{0}\Lambda-\Lambda\Delta\theta)}{\Lambda}\right) \\ - \frac{1}{8} \left(\frac{\Lambda}{\pi A_{p}}\right) \sin\left(\frac{\Lambda}{\pi A_{p}}\right) \\ - \frac{1}{8} \left(\frac{\Lambda}{\pi A_{p}}\right) \sin\left(\frac{\Lambda}{\pi A_{p}}\right) \\ - \frac{1}{8} \left(\frac{\Lambda}{\pi A_{p}}\right) \sin\left(\frac{\Lambda}{\pi A_{p}}\right) \\ - \frac{1}{8} \left(\frac{\Lambda}{\pi A_{p}}\right) \\ - \frac{1}{8} \left(\frac$$

Averaging over the period between updates of this correction,

Averaging over all possible starting points for the cycle, which are represented by values of φ_0 , simplifies this considerably:

$$\overline{\left(_{cor} OPD_{rms}(A_{p}, \Delta\varphi, \Delta\theta)\right)^{2}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\left(_{cor} OPD_{rms}(A_{p}, \varphi_{0}, \Delta\varphi, \Delta\theta)\right)^{2}} d\varphi_{0}$$
$$= \frac{K^{2} (\Delta\varphi + \sin(\Delta\theta) - \sin(\Delta\theta) \cos(\Delta\varphi) - \cos(\Delta\theta) \sin(\Delta\varphi))}{\Delta\varphi}.$$
(B.85)

Again, dependence on A_p simply drops out while performing the last averaging integral. Using the established OPD_{rms} for an uncorrected, one-dimensional aperture in Eq. B.14,

$$G(\Delta \varphi, \Delta \theta) = \frac{\overline{\left(\frac{\partial \varphi}{\partial rms}(\Delta \varphi, \Delta \theta)\right)^{2}}}{\left(\frac{\partial PD}{rms}(\Delta \varphi, \Delta \theta)\right)^{2}}$$
$$= \frac{2(\Delta \varphi + \sin(\Delta \theta) - \sin(\Delta \theta) \cos(\Delta \varphi) - \cos(\Delta \theta) \sin(\Delta \varphi))}{\Delta \varphi}.$$
(B.86)

B.6.4. Temporal Effects for Two-Dimensional Apertures

As noted in the preceding sections, all dependency on the aperture size drops out of these expressions for temporal effects during the later averaging stages of the derivations. This may apply to aperture shape as well, because the end results for performing these derivations for two-dimensional circular apertures are the same as for the one-dimensional apertures. These derivations are presented here for completeness.

Using the periodically corrected wavefront of Eq. B.71,

$$\begin{pmatrix} (_{cor} OPD_{rms}(A_{p},t,\varphi_{0}))^{2} = \frac{4}{\pi A_{p}^{2}} \int_{-A_{p}/2}^{A_{p}/2} \sqrt{A_{p}^{2}/4-x^{2}} OPD(x,y,t,\varphi_{0}))^{2} dy dx \\ = K^{2} \left[1 - \sin(\varphi(t)) \sin(\varphi_{0}) - \cos(\varphi(t)) \cos(\varphi_{0}) \right] \\ + K^{2} \frac{\Lambda}{\pi A_{p}} J_{1} \left(\frac{2\pi A_{p}}{\Lambda} \right) \left[1 - \cos^{2}(\varphi_{0}) - \cos^{2}(\varphi(t)) \\ - \sin(\varphi_{0}) \sin(\varphi(t)) + \cos(\varphi_{0}) \cos(\varphi(t)) \right]. \quad (B.87) \\ \overline{(_{cor} OPD_{rms}(A_{p},\varphi_{0},\Delta\varphi))^{2}} = \frac{1}{\Delta \varphi} \int_{\varphi_{0}}^{\varphi_{0}+\Delta \varphi} \left(\int_{cor} OPD_{rms}(A_{p},\varphi)^{2} d\varphi \right) \\ = \frac{K^{2}}{\pi \Delta \varphi} \left[\frac{\pi \Delta \varphi - \pi \cdot \sin(\varphi_{0}) (\cos(\varphi_{0}) - \cos(\varphi_{0} + \Delta \varphi_{0}))}{(+\pi \cdot \cos(\varphi_{0}) (\sin(\varphi_{0}) - \sin(\varphi_{0} + \Delta \varphi_{0}))} \right] \\ - K^{2} \frac{\Lambda}{\pi A_{p} \Delta \varphi} J_{1} \left(\frac{2\pi A_{p}}{\Lambda} \right) \left[2\sin(\varphi_{0}) \cos(\varphi_{0}) - \Delta \varphi + 2\Delta \varphi \cos^{2}(\varphi_{0}) \\ -\sin(\varphi_{0}) \cos(\varphi_{0} + \Delta \varphi) - \cos(\varphi_{0}) \sin(\varphi_{0} + \Delta \varphi) \right]. \quad (B.88)$$

$$\overline{\left(_{cor} OPD_{rms}(A_{p}, \Delta\varphi)\right)^{2}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\left(_{cor} OPD_{rms}(A_{p}, \varphi_{0}, \Delta\varphi)\right)^{2}} d\varphi_{0} = \frac{K^{2}(\Delta\varphi - \sin(\Delta\varphi))}{\Delta\varphi}.$$
 (B.89)

$$G(\Delta \varphi) = \frac{\overline{\left(\frac{\partial \varphi}{\partial rms}(\Delta \varphi)\right)^2}}{\left(\overline{OPD}_{rms}(\Delta \varphi)\right)^2} = 2\frac{(\Delta \varphi - \sin(\Delta \varphi))}{\Delta \varphi}.$$
 (B.90)

As noted, this is the same result found for a one-dimensional aperture in section B.6.1.

Using the wavefront that is corrected with latency in Eq. B.77,

$$\left(_{cor} OPD_{rms}(A_{p}, t, \Delta \theta_{0}) \right)^{2} = \frac{4}{\pi A_{p}^{2}} \int_{-A_{p}/2}^{A_{p}/2} \int_{-\sqrt{A_{p}^{2}/4 - x^{2}}}^{\sqrt{A_{p}^{2}/4 - x^{2}}} OPD(x, y, t, \Delta \theta) \right)^{2} dy dx$$
$$= K^{2} \left[1 - \cos(\Delta \theta) \right]$$

$$-K^{2} \frac{\Lambda}{\pi A_{p}} J_{1} \left(\frac{2\pi A_{p}}{\Lambda} \right) \begin{bmatrix} \cos^{2}(\varphi(t)) \cos^{2}(\Delta \theta) - \cos^{2}(\Delta \theta) - 2\cos^{2}(\varphi(t)) \cos(\Delta \theta) \\ + 2\cos(\varphi(t)) \cos(\Delta \theta) \sin(\varphi(t)) \sin(\Delta \theta) + \cos(\Delta \theta) \\ - 2\cos(\varphi(t)) \sin(\varphi(t)) \sin(\Delta \theta) \end{bmatrix}.$$
(B.91)

$$\overline{\left(_{cor} OPD_{rms}(A_{p}, \Delta\theta)\right)^{2}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\left(_{cor} OPD_{rms}(A_{p}, \varphi, \Delta\theta)\right)^{2}} d\varphi = K^{2} (1 - \cos(\Delta\theta)).$$
(B.92)

$$G(\Delta\theta) = \frac{\overline{\left(\frac{\partial \theta}{\partial rms}(\Delta\theta)\right)^2}}{\overline{\left(\frac{\partial \theta}{\partial rms}(\Delta\theta)\right)^2}} = 2(1 - \cos(\Delta\theta)).$$
(B.93)

As noted, this is the same result found for a one-dimensional aperture in section B.6.2.

Using the wavefront that is periodically corrected with latency in Eq. B.82,

$$\begin{pmatrix} \sum_{cor} OPD_{rms}(A_p, t, \varphi_o, \Delta\theta_0) \end{pmatrix}^2 = \frac{4}{\pi A_p^2} \int_{-A_p/2}^{A_p/2} \int_{-A_p/2}^{\sqrt{A_p^2/4 - x^2}} OPD(x, y, t, \varphi_0, \Delta\theta) \end{pmatrix}^2 dy dx$$
$$= K^2 \begin{bmatrix} 1 + \sin(\varphi(t))\cos(\varphi_0)\sin(\Delta\theta) - \sin(\varphi(t))\sin(\varphi_0)\cos(\Delta\theta) \\ -\cos(\varphi(t))\sin(\varphi_0)\sin(\Delta\theta) - \cos(\varphi(t))\cos(\varphi_0)\cos(\Delta\theta) \end{bmatrix}$$
$$-K^2 \left(\frac{\Delta}{\pi A_p}\right) J_1 \left(\frac{2\pi A_p}{\Lambda}\right) \begin{bmatrix} \cos^2(\varphi(t)) - \cos^2(\varphi_0) + \sin(\varphi(t)\sin(\varphi_0)\cos(\Delta\theta) \\ -\cos(\varphi(t)\sin(\varphi_0)\sin(\Delta\theta) - \sin(\varphi(t)\cos(\varphi_0)\sin(\Delta\theta) \\ -\cos(\varphi(t)\sin(\varphi_0)\sin(\Delta\theta) - \sin(\varphi(t)\cos(\varphi_0)\sin(\Delta\theta) \\ + 2\sin^2(\varphi_0)\cos^2(\Delta\theta) - \cos^2(\Delta\theta) \\ + 2\sin(\varphi_0)\cos(\varphi_0)\sin(\Delta\theta)\cos(\Delta\theta) \\ -\cos(\varphi(t))\cos(\varphi_0)\cos(\Delta\theta) \end{bmatrix}.$$
(B.94)

$$\overline{\left(_{cor} OPD_{rms}(A_{p}, \varphi_{0}, \Delta\varphi, \Delta\theta)\right)^{2}} = \frac{1}{\Delta\varphi} \int_{\varphi_{0}}^{\varphi_{0}+\Delta\varphi} OPD_{rms}(A_{p}, t, \varphi_{0}, \Delta\theta)^{2} d\varphi$$

$$=\frac{K^{2}}{\Delta\varphi}\begin{bmatrix}1+\sin(\Delta\theta)+\sin(\varphi_{0})\cos(\Delta\theta)\cos(\varphi_{0}+\Delta\varphi)-\cos(\varphi_{0})\cos(\Delta\theta)\sin(\varphi_{0}+\Delta\varphi)\\-\cos(\varphi_{0})\sin(\Delta\theta)\cos(\varphi_{0}+\Delta\varphi)-\sin(\varphi_{0})\sin(\Delta\theta)\sin(\varphi_{0}+\Delta\varphi)\end{bmatrix}$$

$$+\frac{1}{2}\frac{K^{2}}{\pi\Delta\varphi}\left(\frac{\Delta}{\pi A_{p}}\right)J_{1}\left(\frac{2\pi A_{p}}{\Lambda}\right)\left(\frac{2\pi A_{p}}{\Lambda}\right)\right)\left(\frac{2\pi A_{p}}{\Lambda}\right)\left(\frac{2\pi A_{p}}{\Lambda}\right)\left(\frac{2\pi A_{p}}{\Lambda}\right)\left(\frac{2\pi A_{p}}{\Lambda}\right)\left(\frac{2\pi A_{p}}{\Lambda}\right)\right)\left(\frac{2\pi A_{p}}{\Lambda}\right)\left(\frac{2\pi A_{p}}{\Lambda}\right$$

$$\overline{\left(_{cor} OPD_{rms}(A_{p}, \Delta \varphi, \Delta \theta)\right)^{2}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \overline{\left(_{cor} OPD_{rms}(A_{p}, \varphi_{0}, \Delta \varphi, \Delta \theta)\right)^{2}} d\varphi_{0}$$
$$= \frac{K^{2}(\Delta \varphi + \sin(\Delta \theta) - \sin(\Delta \theta) \cos(\Delta \varphi) - \cos(\Delta \theta) \sin(\Delta \varphi))}{\Delta \varphi}.$$
(B.96)

$$G(\Delta\varphi, \Delta\theta) = \frac{\overline{\left(\frac{\cos OPD_{rms}(\Delta\varphi, \Delta\theta)\right)^{2}}{(OPD_{rms}(\Delta\varphi, \Delta\theta))^{2}}}}{\frac{2(\Delta\varphi + \sin(\Delta\theta) - \sin(\Delta\theta)\cos(\Delta\varphi) - \cos(\Delta\theta)\sin(\Delta\varphi))}{\Delta\varphi}}.$$
 (B.97)

As noted, this is the same result found for a one-dimensional aperture in section B.6.3.

B.7. Higher-Order Correction

For this analysis, it will be assumed that the corrective fit to the wavefront will be produced by a sum of basis functions, with each function scaled by some time-varying value to produce the best fit to the wavefront at that time. If the *j*th basis function is represented by $R_j(x,y)$ and the associated scaling value by $S_j(t)$, then

$$Fit(x, y, t) = \sum_{j=1}^{N} \left(S_{j}(t) R_{j}(x, y) \right).$$
(B.98)

The residual disturbances in the wavefront after correction with this fit will be described by

$$_{cor}OPD(x, y, t) = OPD(x, y, t) - Fit(x, y, t) = OPD(x, y, t) - \sum_{j=1}^{N} \left(S_{j}(t)R_{j}(x, y) \right)$$
(B.99)

and

$$_{cor} OPD(t)_{rms}^{2} = \frac{1}{A} \int_{Ap} \left[OPD(x, y, t) - \sum_{j=1}^{N} \left(S_{j}(t) R_{j}(x, y) \right) \right]^{2} dx dy$$
(B.100)

where *A* is the area of the aperture.

B.7.1. The General Case

An optimal fit is one with values of $S_j(t)$ to minimize Eq. B.100 for each value of t. As was done with the T/T correction, the optimal values of these coefficients will correspond to values for which the derivative of Eq. B.100 with respect to that coefficient is zero.

$$\frac{\partial_{cor}OPD(t)_{rms}^2}{\partial S_k} = 0 = \frac{1}{A} \int_{A_p} 2 \left[OPD(x, y, t) - \sum_{j=1}^N \left(S_j(t) R_j(x, y) \right) \right] \left[-R_k(x, y) \right] dx dy .$$
(B.101)

This leads to the equation

$$\frac{1}{A} \int_{Ap} OPD(x, y, t) R_k(x, y) dx dy = \sum_{j=1}^{N} \left[S_j(t) \frac{1}{A} \int_{Ap} R_j(x, y) R_k(x, y) dx dy \right].$$
(B.102)

Using the following notation:

$$OPD(t)_{rms}^{2} = \frac{1}{A} \int_{Ap} [OPD(x, y, t)]^{2} dx dy, \qquad (B.103)$$

$$B_{j}(t) = \frac{1}{A} \int_{Ap} OPD(x, y, t) R_{j}(x, y) dx dy, \qquad (B.104)$$

$$C_{jk} = \frac{1}{A} \int_{Ap} R_j(x, y) R_k(x, y) dx dy, \qquad (B.105)$$

Eq. B.100 can be rewritten as

$$_{cor}OPD(t)_{rms}^{2} = OPD(t)_{rms}^{2} - 2\sum_{j=1}^{N} B_{j}(t)S_{j}(t) + \sum_{j=1,k=1}^{N} C_{jk}S_{j}(t)S_{k}(t)$$
(B.106)

and Eq. B.102 can be written as

$$B_{k} = \sum_{j=1}^{N} \left[S_{j}(t) C_{jk} \right].$$
(B.107)

The set of equations represented by Eq. B.107 can also be written in vector-andmatrix form with:

$$\vec{B}(t) = \begin{bmatrix} B_{1}(t) \\ B_{2}(t) \\ \dots \\ B_{N-1}(t) \\ B_{N}(t) \end{bmatrix}, \ \vec{S}(t) = \begin{bmatrix} S_{1}(t) \\ S_{2}(t) \\ \dots \\ S_{N-1}(t) \\ S_{N}(t) \end{bmatrix}, \ \vec{C} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1(N-1)} & C_{1N} \\ C_{21} & \ddots & & C_{2N} \\ \vdots & & \ddots & & \vdots \\ C_{(N-1)1} & & & \ddots & C_{(N-1)N} \\ C_{N1} & C_{N2} & \cdots & C_{N(N-1)} & C_{NN} \end{bmatrix},$$

as

$$\vec{B}(t) = \vec{C}\vec{S}(t). \tag{B.108}$$

Technically the matrix \vec{C} in Eq. B.108 should be \vec{C}^T , but by Eq. B.105, $C_{jk} = C_{kj}$ which indicates \vec{C} is a symmetric matrix, for which $\vec{C}^T = \vec{C}$. Likewise, Eq. B.100 can be written as

$$_{cor} OPD(t)_{rms}^{2} = OPD(t)_{rms}^{2} - 2\vec{B}(t)^{T}\vec{S}(t) + \vec{S}(t)^{T}\vec{C}\vec{S}(t) .$$
(B.109)

The solution to Eq. B.108 is quite clearly

$$\vec{S}(t) = \vec{C}^{-1}\vec{B}(t)$$
. (B.110)

Substituting this into Eq. B.108, the minimum achievable wavefront distortion after correction is

$$\begin{bmatrix} _{cor} OPD(t)_{rms}^{2} \end{bmatrix}_{min} = OPD(t)_{rms}^{2} - 2\vec{B}(t)^{T}\vec{C}^{-1}\vec{B}(t) + \vec{B}(t)^{T}\vec{C}^{-1}\vec{C}\vec{C}^{-1}\vec{B}(t)$$
$$= OPD(t)_{rms}^{2} - \vec{B}(t)^{T}\vec{C}^{-1}\vec{B}(t) = OPD(t)_{rms}^{2} - \sum_{j=1,k=1}^{N} C_{jk}^{-1}B_{j}(t)B_{k}(t).$$
(B.111)

This analysis was originally performed by Richard Hudgin¹ in the 1970's. He used ensemble averages to arrive at a general guideline for the necessary actuator spacing to achieve a given level of performance when dealing with atmospheric distortions. In dealing with aero-optic disturbances, time averages will serve the same function as ensemble averages

$$\overline{OPD(t)_{rms}^2} = \overline{OPD(t)_{rms}^2} - \overline{\vec{B}(t)}^T \vec{C}^{-1} \vec{B}(t).$$
(B.112)

In using the sinusoidal wavefront of

$$OPD(x, y, t) = K \sin\left(\frac{2\pi}{\Lambda}x + \varphi(t)\right)$$
 (B.113)

yet again, this corresponds to integrating a function over the rage $\varphi = -\pi$ to $\varphi = \pi$ and then dividing by 2π . From this,

$$\overline{_{cor} OPD(t)_{rms}^2} = \frac{1}{2} K^2 - \overline{\vec{B}(t)^T \vec{C}^{-1} \vec{B}(t)}, \qquad (B.114)$$

and the corrective gain is

$$G = \frac{\overline{OPD(t)_{rms}^{2}}}{\overline{OPD(t)_{rms}^{2}}} = 1 - 2\frac{\overline{\vec{B}(t)^{T}\vec{C}^{-1}\vec{B}(t)}}{K^{2}}.$$
 (B.115)

B.7.2. Local Piston Correction

If the overall aperture is divided into a rectangular grid of sub-apertures, each of some size ${}_{sub}A_p$, then the basis functions may be defined as

$$R_{j}(x) = \begin{cases} for\left(X_{j} - \frac{sub}{2}A_{p}\right) \leq x < \left(X_{j} + \frac{sub}{2}A_{p}\right) \\ and\left(Y_{j} - \frac{sub}{2}A_{p}\right) \leq y < \left(Y_{j} + \frac{sub}{2}A_{p}\right) \\ 0 & otherwise \end{cases}$$
(B.116)

The parameters X_j and Y_j represent the center of this basis function and presumably correspond to the location of an actuator driving the local piston correction in a physical system of this sort. Considering only a one-dimensional aperture and correction:

$$R_{j}(x) = \begin{cases} 1 & for\left(X_{j} - \frac{sub}{2}A_{p}\right) \le x < \left(X_{j} + \frac{sub}{2}A_{p}\right) \\ 0 & otherwise \end{cases}$$
(B.117)

and the center location (X_j) of each sub-aperture will then be

$$X_{j} = -\frac{A_{p}}{2} + \frac{sub}{2} \frac{A_{p}}{2} + sub} A_{p}(j-1).$$
(B.118)

From Eq. B.104,

$$B_{j}(t) = \frac{1}{A_{p}} \int_{-sub}^{sub} A_{p}/2 + X_{j}} K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right)(1) dx$$
$$= \frac{K}{2} \frac{\Lambda}{\pi A_{p}} \left[\cos\left(\frac{-\pi A_{p} + 2\pi_{sub}A_{p}(j-1) + \Lambda\varphi}{\Lambda}\right) - \cos\left(\frac{-\pi A_{p} + 2\pi_{sub}A_{p}(j) + \Lambda\varphi}{\Lambda}\right)\right]. \quad (B.119)$$

From Eq. B.105,

$$C_{jk} = \frac{1}{A_p} \int_{-sub}^{sub} \frac{A_p / 2 + X_j}{\int} dx = \frac{\left[sub A_p / 2 + X_j\right] - \left[-sub A_p / 2 + X_j\right]}{A_p} = \frac{sub A_p}{A_p} \quad (B.120)$$

for j = k, while $C_{jk} = 0$ for $j \neq k$, which can also be written as

$$\vec{C} = \frac{{}^{sub}A_p}{A_p}\vec{I}$$
(B.121)

where \vec{I} represents the identity matrix and

$$\vec{C}^{-1} = \frac{A_p}{_{sub}A_p}\vec{I}$$
(B.122)

where \vec{I} represents the identity matrix. Because of this, only cases where j = k will contribute to the summation in Eq. B.111, and so

$$_{cor}OPD(t)_{rms}^{2} = \frac{1}{2}K^{2} - \sum_{j=1}^{N}C_{jj}^{-1}(B_{j})^{2}.$$
 (B.123)

The average of the summation in Eq. B.123 can also be expressed as the sum of the averages

$$\sum_{j=1}^{N} \overline{C_{jj}^{-1}(B_{j})^{2}} = \sum_{j=1}^{N} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{A_{p}}{sub} \left(\frac{K}{2} \frac{\Lambda}{\pi A_{p}} \right)^{2} \left[\cos\left(\frac{-\pi A_{p} + 2\pi_{sub}A_{p}(j-1) + \Lambda\varphi}{\Lambda}\right) \right]^{2} d\varphi$$
$$- \cos\left(\frac{-\pi A_{p} + 2\pi_{sub}A_{p}(j) + \Lambda\varphi}{\Lambda}\right) \right]^{2} d\varphi$$
$$= \sum_{j=1}^{N} \frac{K^{2} \Lambda^{2} \left(1 - \cos\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right)^{2}\right)}{2\pi^{2}(A_{p})_{sub}A_{p}}.$$
(B.124)

It is noteworthy that the index, j, drops out of the equation during the integral, making the summation trivially easy. If an aperture of size A_p is divided evenly into subapertures of size $subA_p$, then the number of sub-apertures will be

$$N = \frac{A_p}{\sup_{sub} A_p}$$
(B.125)

and

$$\sum_{j=1}^{N} \overline{C_{jj}^{-1}(B_{j})^{2}} = \frac{A_{p}}{\sup A_{p}} \frac{K^{2} \Lambda^{2} \left(1 - \cos\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right)^{2}\right)}{2\pi^{2} (A_{p})_{sub} A_{p}} = \frac{K^{2} \Lambda^{2} \left(1 - \cos\left(\frac{\pi_{sub}A_{p}}{\Lambda}\right)^{2}\right)}{2\pi^{2} \sup A_{p}^{2}}.$$
 (B.126)

Just as the ratio A_p/Λ can be rewritten as a Strouhal number based on the length scale A_p ,

$$\frac{{}^{sub}A_p}{\Lambda} = \frac{{}^{sub}A_p}{U_C / f} = \frac{{}^{sub}A_p \cdot f}{U_C} = St_{subAp} .$$
(B.127)

From Eq. B.115, the corrective gain in terms of St_{subAp} is then

$$G(St_{subAp}) = \frac{\overline{(cor OPD(t)_{rms}^2)}}{\overline{OPD(t)_{rms}^2}} = 1 - \frac{\left(1 - \cos\left(\pi St_{subAp}\right)^2\right)}{\left(\pi St_{subAp}\right)^2}.$$
 (B.128)

B.7.3. First-Order Correction

A common form of B-spline basis functions are those produced by the Cox-de Boor recursion formula. In this formula the 0th-degree basis function is defined over the intervals between points defined by $x = u_i$, where $u_0 \le u_1 \le u_2 \le u_3 \le ... \le u_N$.

$$N_{i,0}(x) = \begin{cases} 1 & \text{if } u_i \le x < u_{i+1} \\ 0 & \text{otherwise} \end{cases}.$$
 (B.129)

This is clearly equivalent to the piston-correction basis function used in section B.7.2, except that function was defined around central points ($X_j = (u_i - u_{i+1})/2$) separated by uniform intervals of length _{sub} A_p .

In the Cox-de Boor formula, a basis function of order p > 0 is found from the order *p*-1 basis function by

$$N_{i,p}(x) = \frac{x - u_i}{u_{i+p} - u_i} N_{i,p-1}(x) + \frac{u_{i+p+1} - x}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(x).$$
(B.130)

The 1st-order basis function is then defined by

$$N_{i,1}(x) = \begin{cases} \frac{x - u_i}{u_{i+1} - u_i} & \text{if } u_i \le x < u_{i+1} \\ \frac{u_{i+2} - x}{u_{i+2} - u_{i+1}} & \text{if } u_{i+1} \le x < u_{i+2} \\ 0 & \text{otherwise} \end{cases}$$
(B.131)

The nature of this recursion is such that a basis function of order p extends across p+1 intervals. In terms of center points (X_i) and uniform sub-apertures, this would be.
$$R_{j}(x) = \begin{cases} \frac{x - X_{j} + _{sub} A_{p}}{_{sub} A_{p}} & \text{if } (X_{j} - _{sub} A_{p}) \le x < X_{j} \\ \frac{X_{j} + _{sub} A_{p} - x}{_{sub} A_{p}} & \text{if } X_{j} \le x < (X_{j} + _{sub} A_{p}), \\ 0 & \text{otherwise} \end{cases}$$
(B.132)

with center points at

$$X_{j} = -\frac{A_{p}}{2} +_{sub} A_{p}(j-1).$$
(B.133)

Using this 1st-order basis function,

$$B_{j}(t) = \frac{1}{A_{p}} \begin{bmatrix} \int_{X_{j}-sub}^{X_{j}} K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right) \left(\frac{x - X_{j} + sub}{sub}A_{p}\right) dx \\ + \int_{X_{j}}^{X_{j}+sub}A_{pj} \\ + \int_{X_{jj}}^{X_{j}+sub}A_{p} \int K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right) \left(\frac{X_{j} + sub}{sub}A_{p} - x\right) dx \end{bmatrix}$$
$$= \frac{\Lambda^{2}K}{\pi^{2}A_{p}\left(sub}A_{p}\right)} \begin{bmatrix} 2\sin\left(\frac{X_{j}\pi}{\Lambda}\right)\cos\left(\frac{X_{j}\pi}{\Lambda}\right)\cos\left(\varphi\right)\cos\left(\frac{sub}{\Lambda}A_{p}\pi\right) \\ - 2\sin\left(\frac{X_{j}\pi}{\Lambda}\right)\cos\left(\frac{X_{j}\pi}{\Lambda}\right)\cos(\varphi) + \sin(\varphi) \\ + 2\cos^{2}\left(\frac{X_{j}\pi}{\Lambda}\right)\sin(\varphi)\cos^{2}\left(\frac{sub}{\Lambda}A_{p}\pi\right) \\ - 2\cos^{2}\left(\frac{X_{j}\pi}{\Lambda}\right)\sin(\varphi) - \sin(\varphi)\cos^{2}\left(\frac{sub}{\Lambda}A_{p}\pi\right) \end{bmatrix}. \quad (B.134)$$

The matrix elements C_{jk} are equal to zero except when k = j, k = j+1, or k = j-1. In those cases:

$$C_{jj} = \frac{1}{A_p} \begin{bmatrix} \int_{X_{j-sub}A_{p_j}}^{X_{j-sub}A_p} \left(\frac{x - X_j + sub}{sub} A_p \right)^2 dx \\ \int_{X_{j-sub}A_{p_j}}^{X_{j-sub}A_{p_j}} \left(\frac{X_j + sub}{sub} A_p - x \right)^2 dx \\ + \int_{X_{j_j}}^{X_{j+sub}A_{p_j}} \left(\frac{X_j + sub}{sub} A_p - x \right)^2 dx \end{bmatrix} = \frac{2}{3} \frac{sub}{A_p} A_p, \quad (B.135)$$

$$C_{j(j+1)} = \frac{1}{A_p} \left[\int_{X_{j_j}}^{X_j + sub} \frac{A_p}{sub} A_p - x}{\int_{sub} A_p} \left(\frac{x - X_j}{sub} A_p \right) dx \right] = \frac{1}{6} \frac{sub}{A_p} A_p, \quad (B.136)$$

$$C_{j(j-1)} = \frac{1}{A_p} \left[\int_{X_j - sub A_{p_j}}^{X_{j_j}} \left(\frac{x - X_j + sub A_p}{sub A_p} \right) \left(\frac{X_j - x}{sub A_p} \right) dx \right] = \frac{1}{6} \frac{sub A_p}{A_p}.$$
 (B.137)

This type of matrix is called a tridiagonal matrix. Methods for inverting a tridiagonal matrix often involve somewhat complicated algorithms and most are intended to produce numerical results for a specific matrix, rather than a more general solution. However, inverting this particular matrix is made simpler by the facts that this matrix is symmetrical, and is a Toeplitz matrix, defined by having each diagonal defined by one value for all elements in that diagonal. It should be noted that for some physical implementations of a corrective system, effects along the edges of the aperture may result in a matrix that is not truly Toeplitz, having different values at or near the C_{11} and C_{NN} elements of the matrix. In this analysis, which is pursuing a generalized rule of thumb for the design and predicted effectiveness of such systems, a Toeplitz matrix will be assumed. For a symmetric, tridiagonal, Toeplitz matrix, the following analytical expression for the inverse matrix can be used.²

First, the *N*-by-*N* matrix \vec{C} is rewritten in the form

$$\vec{C} = \frac{1}{6} \frac{{}_{sub} A_p}{A_p} \begin{bmatrix} 4 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 4 & 1 & & \vdots \\ 0 & 1 & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 1 & 0 \\ \vdots & & & 1 & 4 & 1 \\ 0 & \cdots & \cdots & 0 & 1 & 4 \end{bmatrix}.$$
 (B.138)

Secondly, a parameter, β , is defined, such that

$$C_{jj} = \left(\frac{1}{6} \frac{{}_{sub} A_p}{A_p}\right) 4 = \left(\frac{1}{6} \frac{{}_{sub} A_p}{A_p}\right) (2\cosh\beta).$$
(B.139)

This indicates that

$$4 = 2\cosh\beta = e^{\beta} + e^{-\beta}, \qquad (B.140)$$

$$e^{2\beta} - 4e^{\beta} + 1 = 0, \qquad (B.141)$$

$$e^{\beta} = \frac{4 \pm \sqrt{16 - 4}}{2} = 2 \pm \sqrt{3},$$
 (B.142)

$$\beta = \ln(2 \pm \sqrt{3}). \tag{B.143}$$

The *N*-by-*N* inverse matrix is then described by

$$\left(\vec{C}^{-1}\right)_{jk} = 6 \frac{A_p}{\sup A_p} (-1)^{j+k} \frac{\cosh[(N+1-|k-j|)\beta] - \cosh[(N+1-j-k)\beta]}{2\sinh[\beta]\sinh[(N+1)\beta]}.$$
 (B.144)

Either value of β will work with this method, but the positive value, $\ln(2+\sqrt{3})$, will be used. The element for summation in Eq. B.111 can be written in terms of sines, cosines, and exponential functions as

$$C_{jk}^{-1}B_{j}B_{k} = \begin{bmatrix} -\sin\left(\frac{2\pi X_{k} + \varphi \Lambda + 2\pi_{sub}A_{p}}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda + 2\pi_{sub}A_{p}}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda + 2\pi_{sub}A_{p}}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ -\sin\left(\frac{2\pi X_{k} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda + 2\pi_{sub}A_{p}}{\Lambda}\right) \\ -\sin\left(\frac{2\pi X_{k} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda + 2\pi_{sub}A_{p}}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ -\sin\left(\frac{2\pi X_{k} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ -\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda + 2\pi_{sub}A_{p}}{\Lambda}\right) \\ -4\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda + 2\pi_{sub}A_{p}}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right) \\ + 2\sin\left(\frac{2\pi X_{k} + \varphi \Lambda}{\Lambda}\right)\sin\left(\frac{2\pi X_{j} + \varphi \Lambda - 2\pi_{sub}A_{p}}{\Lambda}\right)$$
(B.145)

Taking the time-average of this element by integrating over φ is possible, but unlike what was seen in the piston-correction case of section B.7.2, the result of this integral is not a simpler expression:

$$\overline{C_{jk}^{-1}B_jB_k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} C_{jk}^{-1}B_jB_k d\varphi$$

$$= \frac{3}{2\pi^4 A_p (_{sub} A_p)^3 [(45 + 26\sqrt{3})(7 + 4\sqrt{3})^N - 3 - 2\sqrt{3}]}$$
•
$$\begin{bmatrix} (2 + \sqrt{3})(2 + \sqrt{3})^{j+k} + (26 + 15\sqrt{3})(7 + 4\sqrt{3})^N (2 + \sqrt{3})^{-(j+k)} \\ -(2 + \sqrt{3})(2 + \sqrt{3})^{|j-k|} - (26 + 15\sqrt{3})(7 + 4\sqrt{3})^N (2 + \sqrt{3})^{-|j-k|} \end{bmatrix}$$

$$\begin{bmatrix} -4\sin\left(\frac{\pi X_j}{\Lambda}\right)\cos\left(\frac{\pi X_j}{\Lambda}\right)\sin\left(\frac{\pi X_k}{\Lambda}\right)\cos\left(\frac{\pi X_k}{\Lambda}\right)\cos^4\left(\frac{\pi_{sub} A_p}{\Lambda}\right) \\ -4\cos^2\left(\frac{\pi X_j}{\Lambda}\right)\cos^2\left(\frac{\pi X_k}{\Lambda}\right)\cos^4\left(\frac{\pi_{sub} A_p}{\Lambda}\right) + \cos^4\left(\frac{\pi_{sub} A_p}{\Lambda}\right) \\ +2\cos^2\left(\frac{\pi X_j}{\Lambda}\right)\cos^2\left(\frac{\pi X_j}{\Lambda}\right)\sin\left(\frac{\pi X_k}{\Lambda}\right)\cos\left(\frac{\pi X_k}{\Lambda}\right)\cos^2\left(\frac{\pi_{sub} A_p}{\Lambda}\right) \\ +8\sin\left(\frac{\pi X_j}{\Lambda}\right)\cos\left(\frac{\pi X_j}{\Lambda}\right)\sin\left(\frac{\pi X_k}{\Lambda}\right)\cos\left(\frac{\pi X_k}{\Lambda}\right)\cos^2\left(\frac{\pi_{sub} A_p}{\Lambda}\right) \\ +8\cos^2\left(\frac{\pi X_j}{\Lambda}\right)\cos^2\left(\frac{\pi x_k}{\Lambda}\right)-4\cos^2\left(\frac{\pi X_k}{\Lambda}\right)\cos^2\left(\frac{\pi x_k}{\Lambda}\right) \\ -4\cos^2\left(\frac{\pi X_j}{\Lambda}\right)\cos\left(\frac{\pi X_j}{\Lambda}\right)\sin\left(\frac{\pi X_k}{\Lambda}\right)\cos\left(\frac{\pi X_k}{\Lambda}\right) \\ -4\sin\left(\frac{\pi X_j}{\Lambda}\right)\cos\left(\frac{\pi X_j}{\Lambda}\right)\sin\left(\frac{\pi X_k}{\Lambda}\right)\cos\left(\frac{\pi X_k}{\Lambda}\right) \\ -4\cos^2\left(\frac{\pi X_j}{\Lambda}\right)\cos^2\left(\frac{\pi X_j}{\Lambda}\right) \\ -4\cos^2\left(\frac{\pi X_j}{\Lambda}\right)\cos^2\left(\frac{\pi X_k}{\Lambda}\right) \\ +2\cos^2\left(\frac{\pi X_j}{\Lambda}\right) + 2\cos^2\left(\frac{\pi X_k}{\Lambda}\right) \\ +2\cos^2\left(\frac{\pi X_j}{\Lambda}\right) + 2\cos^2\left(\frac{\pi X_k}{\Lambda}\right) + 1$$
(B.146)

Combining Eqs. B.111, B.115, B.133, and B.146 should produce an estimate of this type of correction for all values of *N*. However, performing the necessary summations and rendering the result into a simplified form for ease of use is left for the future.

B.7.4. Third-Order Correction

The 3rd-order basis function produced with the Cox-de Boor recursion formula is

$$R_{j}(x) = \begin{cases} \frac{1}{6} \frac{\left(x - X_{j} + 2_{sub} A_{p}\right)^{3}}{\left(s_{sub} A_{p}\right)^{3}} & \text{if } (X_{j} - 2_{sub} A_{p}) \leq x \\ < (X_{j} - s_{sub} A_{p}) \\ = \begin{cases} -\frac{1}{6} \frac{1}{\left(s_{sub} A_{p}\right)^{3}} \left[3x^{3} + \left(6_{sub} A_{p} - 9X_{j}\right)x^{2} \\ + \left(9X_{j}^{2} - 12_{sub} A_{p}X_{j}\right)x \\ + \left(-3X_{j}^{3} + 6_{sub} A_{p}X_{j}^{2}\right) \\ -4_{sub} A_{p}^{3} \\ + \left(-4_{sub} A_{p}^{3}\right)^{2} \\ + \left(9X_{j}^{2} + 12_{sub} A_{p}X_{j}\right)x \\ + \left(9X_{j}^{2} + 12_{sub} A_{p}X_{j}\right)x \\ + \left(-3X_{j}^{3} - 6_{sub} A_{p}X_{j}^{2}\right) \\ + \left(-3X_{j}^{3} - 6_{sub} A_{p}X_{j}^{2}\right) \\ - \frac{1}{6} \frac{\left(x - X_{j} - 2_{sub} A_{p}\right)^{3}}{\left(s_{ub} A_{p}\right)^{3}} & \text{if } (X_{j} + s_{ub} A_{p}) \leq x \\ < (X_{j} + s_{ub} A_{p}) \\ - \frac{1}{6} \frac{\left(x - X_{j} - 2_{sub} A_{p}\right)^{3}}{\left(s_{ub} A_{p}\right)^{3}} & \text{if } (X_{j} + s_{ub} A_{p}) \leq x \\ < (X_{j} + s_{ub} A_{p}) \\ - \frac{1}{6} \frac{\left(x - X_{j} - 2_{sub} A_{p}\right)^{3}}{\left(s_{ub} A_{p}\right)^{3}} & \text{otherwise} \end{cases}$$

$$(B.147)$$

Using this 3rd-order basis function,

$$B_{j}(t) = \frac{1}{A_{p}} \left[+ \sum_{\substack{X_{j} - 2_{sub}A_{p_{j}} \\ X_{j} - 2_{sub}A_{p_{j}}}}^{X_{j} - 2_{sub}A_{p_{j}}} K \sin\left(\frac{2\pi}{\Lambda}x + \varphi\right) \left(-\frac{1}{6} \frac{1}{\left(s_{sub}A_{p}\right)^{3}} \right) \left[\frac{3x^{3} + \left(6_{sub}A_{p} - 9X_{j}\right)x^{2}}{\left(-8x_{j}^{2} - 12_{sub}A_{p}X_{j}\right)x} + \frac{x_{j}}{\left(-3X_{j}^{3} + 6_{sub}A_{p}X_{j}\right)x} \right] dx + \left(-3X_{j}^{3} + 6_{sub}A_{p}X_{j}^{2} \right) dx + \left(-3x_{j}^{3} + 6_{sub}A_{p}X_{j}^{2} \right) dx + \left(-3x_{j}^{3} - 6_{sub}A_{p}X_{j}\right)x^{2} + \left(9X_{j}^{2} + 12_{sub}A_{p}X_{j}\right)x + \left(-3X_{j}^{3} - 6_{sub}A_{p}X_{j}^{2} \right) dx + \left(-3X_{j}^{3} - 6_{sub}A_{p}X_{j}^{3} \right) dx + \left(-3X_{j}^{3} - 6_{sub}A_{p}X_{j}^{3} \right) dx + \left(-3X_{j}^{3} - 6_{sub}A_{p}X_{j}^{3} \right) d$$

The matrix elements C_{jk} are equal to zero except when k = j, k = j+1, k = j+2, k = j+3, k = j-1, k = j-2, or k = j-3. In those cases:

$$C_{jj} = \frac{1}{A_{p}} \left[\sum_{X_{j}=2_{ub}A_{p_{j}}}^{X_{j}=2_{ub}A_{p_{j}}} \left[\frac{1}{6} \frac{\left(x - X_{j} + 2_{sub}A_{p}\right)^{3}}{\left(s_{ub}A_{p}\right)^{3}} \right]^{2} dx + \sum_{X_{j}=2_{ub}A_{p_{j}}}^{X_{j}=2_{ub}A_{p_{j}}} \left(-\frac{1}{6} \frac{1}{\left(s_{ub}A_{p}\right)^{3}} \right)^{2} \left[\frac{3x^{3} + \left(6_{sub}A_{p} - 9X_{j}\right)x^{2}}{\left(-4s_{ub}A_{p}X_{j}\right)x} \right]^{2} dx + \left(-3X_{j}^{3} + 6s_{ub}A_{p}X_{j}\right)x + \left(-3X_{j}^{3} + 6s_{ub}A_{p}X_{j}\right)x^{2} + \left(-3X_{j}^{3} + 6s_{ub}A_{p}X_{j}\right)x^{2} + \left(-4s_{ub}A_{p}^{3}\right)^{2} \right]^{2} dx + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}\right)x^{2} + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}\right)x + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}\right)x + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x^{2} + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x^{2} + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x^{2} + \left(-3X_{j}^{3} - 6s_{ub}A_{p}X_{j}^{2}\right)x^{$$

$$C_{j(j+1)} = \frac{1}{A_{p}} \left[+ \sum_{x_{j}=aaA_{p_{j}}}^{x_{j}} - \frac{1}{36} \frac{1}{(aabA_{p})^{3}} \right] \left[+ \left[\frac{9X_{j}^{2}}{(-12_{aab}A_{p}X_{j})} x_{j}^{2} + \left[\frac{9X_{j}^{2}}{(aabA_{p})^{3}} \right] x_{j}^{2} + \left[\frac{(x - X_{j+1})^{3}}{(aabA_{p})^{3}} \right] dx \\ + \left[\frac{-3X_{j}^{3}}{(-4_{aab}A_{p}^{3})} \right] dx \\ + \left[\frac{-3X_{j}^{3}}{(-4_{aab}A_{p}^{3})} \right] dx \\ + \left[\frac{9X_{j}^{2}}{(-4_{aab}A_{p}^{3})} \right] dx \\ + \left[\frac{9X_{j}^{2}}{(-9X_{j})} \right] x_{j}^{2} + \left[\frac{9X_{j}^{2}}{(-9X_{j})} \right] x_{j}^{2} \\ + \left[\frac{9X_{j}^{2}}{(-9X_{j})} \right] x_{j}^{2} \\ + \left[\frac{9X_{j}^{2}}{(-3X_{j}^{3})} \right] x_{j}^{2} + \left[\frac{9X_{j}^{2}}{(-12_{aab}A_{p}X_{j+1})} \right] x_{j}^{2} \\ + \left[\frac{9X_{j}^{2}}{(-3X_{j}^{3})} \right] x_{j}^{2} + \left[\frac{9X_{j}^{2}}{(-3X_{j}^{3})} \right] x_{j}^{2} \\ + \left[\frac{-3X_{j}^{3}}{(-6_{aab}A_{p}X_{j}^{2})} \right] x_{j}^{2} + \left[\frac{9X_{j}^{2}}{(-12_{aab}A_{p}X_{j+1})} \right] x_{j}^{2} \\ + \left[\frac{-3X_{j}^{3}}{(-6_{aab}A_{p}X_{j}^{2})} \right] x_{j}^{2} + \left[\frac{-3X_{j}^{3}}{(-4_{aab}A_{p}^{3})} \right] x_{j}^{2} + \left[\frac{-3X_{j}^{3}}{(-4_{aab}A_{p}^{3})} \right] x_{j}^{2} \\ + \left[\frac{-3X_{j}^{3}}{(-6_{aab}A_{p}X_{j+1})} \right] x_{j}^{2} \\ + \left[\frac{-3X_{j}^{3}}{(-6_{$$

$$=\frac{397}{1680}\frac{{}^{sub}A_p}{A_p},\tag{B.150}$$

$$C_{j(j+2)} = \frac{1}{A_{p}} \left[+ \int_{X_{j}+ub}^{X_{j}+ub}A_{p_{j}}} \frac{1}{36} \frac{1}{\left(\frac{1}{sub}A_{p}\right)^{3}} + \int_{\left(\frac{9X_{j}^{2}}{+12}sub}A_{p}X_{j}\right)^{3}} \frac{1}{4} \left(\frac{9X_{j}^{2}}{+12}sub}A_{p}X_{j}\right)^{3} \right] \left[\frac{\left(\frac{x-X_{j+2}}{+2}\right)^{3}}{\left(\frac{x-X_{j+2}}{+2}\right)^{3}} \frac{1}{4}dx}{\left(\frac{x-X_{j}}{+4}sub}A_{p}^{3}\right)^{3}} \right] \left(\frac{-3X_{j}^{3}}{-6}sub}A_{p}X_{j}^{2} + \frac{1}{4}sub}A_{p}^{3}\right)^{3} \right] \left(\frac{x-X_{j}}{-2}sub}A_{p}^{3} - \frac{1}{4}sub}A_{p}^{3} - \frac{1}{4}sub}A_{p}^{3$$

$$=\frac{1}{42}\frac{{}^{sub}A_p}{A_p},\tag{B.151}$$

$$C_{j(j+3)} = \frac{1}{A_p} \left[\int_{X_j + 2_{sub} A_{p_j}}^{X_j + z_{sub} A_{p_j}} - \frac{1}{36} \left[\frac{\left(x - X_j - 2_{sub} A_p \right)^3}{\left(_{sub} A_p \right)^3} \right] \left[\frac{\left(x - X_{j+3} + 2_{sub} A_p \right)^3}{\left(_{sub} A_p \right)^3} \right] dx \right]$$
$$= \frac{1}{5040} \frac{_{sub} A_p}{A_p}.$$
(B.152)

As with the 1st-order case of the previous section, this produces a symmetric, Toeplitz matrix, and analytical solutions for the inverse of such matrices exist.³

First, the matrix \vec{C} is rewritten in the form

$$\vec{C} = \frac{1}{5040} \frac{{}_{sub}A_p}{A_p} \begin{bmatrix} 2416 & 1191 & 120 & 1 & 0 & \cdots & 0\\ 1191 & 2416 & 1191 & 120 & \ddots & \vdots\\ 120 & 1191 & 2416 & \ddots & 0\\ 1 & 120 & \ddots & 120 & 1\\ 0 & \ddots & 2416 & 1191 & 120\\ \vdots & \ddots & 120 & 1191 & 2416 & 1191\\ 0 & \cdots & 0 & 1 & 120 & 1191 & 2416 \end{bmatrix}.$$
(B.153)

Secondly, a parameter, β , is defined, by the roots of an equation constructed from the elements of this matrix, such that

$$0 = \frac{1}{2}2416 + 1191\cos(\beta) + 120\cos(2\beta) + \cos(3\beta).$$
 (B.154)

The roots of Eq. B.154 are:

$$\beta_1 = \arccos\left((-5166 + 71\sqrt{11905})^{1/3} + \frac{301}{(-5166 + 71\sqrt{11905})^{1/3}} - 20\right), \quad (B.155)$$

$$\beta_{2} = \arccos \left(-\frac{1}{2} (-5166 + 71\sqrt{11905})^{1/3} + \frac{301}{2(-5166 + 71\sqrt{11905})^{1/3}} - 20 \right) + i\frac{1}{2} \sqrt{3} \left((-5166 + 71\sqrt{11905})^{1/3} - \frac{301}{(-5166 + 71\sqrt{11905})^{1/3}} \right) \right), \quad (B.156)$$

$$\beta_{3} = \pi - \arccos \left(-\frac{1}{2} (-5166 + 71\sqrt{11905})^{1/3} + \frac{301}{2(-5166 + 71\sqrt{11905})^{1/3}} + 20 + i\frac{1}{2}\sqrt{3} \left((-5166 + 71\sqrt{11905})^{1/3} - \frac{301}{(-5166 + 71\sqrt{11905})^{1/3}} \right) \right).$$
(B.157)

A set of functions of an integer $\boldsymbol{\chi}$ are then defined:

$$u_{1} = -\left(\frac{\sin((3+\chi)\beta_{1})}{F'(\beta_{1})} + \frac{\sin((3+\chi)\beta_{2})}{F'(\beta_{2})} + \frac{\sin((3+\chi)\beta_{3})}{F'(\beta_{3})}\right),$$
(B.158)

$$u_6 = -u_1(\chi - 1), \tag{B.159}$$

$$u_5 = -120u_1(\chi - 1) + u_6(\chi - 1), \qquad (B.160)$$

$$u_4 = -1191 u_1(\chi - 1) + u_5(\chi - 1), \qquad (B.161)$$

$$u_3 = -2416u_1(\chi - 1) + u_4(\chi - 1), \qquad (B.162)$$

$$u_2 = -1191u_1(\chi - 1) + u_3(\chi - 1), \qquad (B.163)$$

where F' is the derivative of Eq. B.154 with respect to β

$$F'(\beta) = -3\sin(3\beta) - 240\sin(2\beta) - 1191\sin(\beta).$$
 (B.164)

A matrix $\vec{M}(\chi)$ is then defined by

$$\vec{M}(\chi) = \begin{bmatrix} u_4(-\chi - 3) & u_5(-\chi - 3) & u_6(-\chi - 3) \\ u_4(-\chi - 4) & u_5(-\chi - 4) & u_6(-\chi - 4) \\ u_4(-\chi - 5) & u_5(-\chi - 5) & u_6(-\chi - 5) \end{bmatrix}.$$
 (B.165)

The inverse of the matrix \vec{C} can then be defined as

$$\left(\vec{C}^{-1}\right)_{jk} = \begin{cases} -\left[\vec{M}(j-1)\vec{M}^{-1}(N)\vec{M}(N-k)\right]_{1,3} & \text{if } k > j \\ -\left[\vec{M}(j-1)\vec{M}^{-1}(N)\vec{M}(N-k)\right]_{1,3} & \text{if } k \le j \end{cases}$$

$$(B.166)$$

where $[]_{1,3}$ indicates the 3rd element of the 1st row of the matrix inside the brackets.

As in the previous sections, it should be possible to combine this inverse matrix with Eqs. B.111, B.115, and B.148 to find the effective frequency-dependent gain of this form of correction. However, finding an analytic form or approximation simple enough for convenient use will be left for future work in this field.

¹ Hudgin, R., "Wavefront compensation error due to finite corrector-element size," J. Opt. Soc. Am., Vol. 67, No. 3, March 1977.

² Hu, G.Y., O'Connell, R.F., "Analytical inversion of symmetric tridiagonal matrices," J. Phys. A 29 (1996), pp 1511-1513.

³ Lavis, D.A., Southern, B.W., "The Inverse of a Symmetric Banded Toeplitz Matrix," Reports on Math. Phys. Vol. 39, pp 137-146, 1997