

Principles of Long Baseline Stellar Interferometry

Edited by Peter R. Lawson

Course Notes from the 1999 Michelson Summer School
August 15 –19, 1999

JPL Publication 00-009 07/00
<http://sim.jpl.nasa.gov/michelson/iss.html>

Contents

I	Introduction to Long-Baseline Stellar Interferometry	1
1	Why Build Stellar Interferometers?	3
	<i>by H.A. McAlister, Georgia State University</i>	
1.1	The Challenge	3
1.2	The Opportunity	5
1.3	Towards the Future	7
2	Elementary Theory of Interferometry	9
	<i>by A.F. Boden, California Institute of Technology</i>	
2.1	Introduction	9
2.2	A Simple Interferometer and a Monochromatic Source	10
2.3	Polychromatic Sources	12
2.4	Phase Reference of the Interferometer	15
2.5	Image Synthesis by Discrete Visibility Measurements	17
2.6	Visibilities of Various Flavors	18
2.7	Visibility of Common Source Morphologies	21
	2.7.1 Point Source	22
	2.7.2 Uniform Disk	23
	2.7.3 Multiple Stellar Systems	25
2.8	Summary	27
3	Beam Combination and Fringe Measurement	31
	<i>by W.A. Traub, Smithsonian Astrophysical Observatory</i>	
3.1	Elements of Beam Combination	32
	3.1.1 Single Aperture and Point Source	32
	3.1.2 Two Apertures and Point Source	38
	3.1.3 Two Apertures and Binary Star	40
	3.1.4 Two Apertures and Uniform Disk	40
	3.1.5 The Pocket Interferometer	41
	3.1.6 Two-Aperture Beam Pattern on Sky	42
	3.1.7 van Cittert–Zernike Theorem	43
3.2	Beam Combination in Practice	43
	3.2.1 Michelson’s Stellar Interferometer	43
	3.2.2 Image-Plane and Pupil-Plane Combination	45
	3.2.3 Beam splitter Phase-Shift	48
	3.2.4 Multiplexing Three or More Apertures	50
3.3	Visibility Loss Effects	51

3.3.1	Spectral Bandpass	52
3.3.2	Wavefront Tilt	52
3.3.3	Intensity Mismatch	53
3.3.4	Optical Surface Figure Errors	54
3.3.5	Polarization Effects	54
3.4	Visibility Enhancement Methods	55
3.4.1	Adaptive Optics	55
3.4.2	Fiber Optics	56
3.4.3	Integrated Optics	56
3.4.4	Nulling	56
4	Noise and Sensitivity in Interferometry	59
	<i>by C.H. Townes, University of California, Berkeley</i>	
4.1	Introduction	59
4.2	Wavefront Aberrations	59
4.3	Thermal and Quantum Noise	60
4.3.1	Noise Power Fluctuations	60
4.3.2	Signal-to-Noise Ratio	60
4.4	Heterodyne versus Direct-Detection Interferometry	61
4.4.1	Signal-to-Noise Ratio for Fringe Measurements	61
4.4.2	Bandwidth Considerations	62
4.4.3	Arrays with Multiple Telescopes	63
4.4.4	Coherence Time and Fringe Tracking	64
4.5	Conclusion	64
5	Observing Through the Turbulent Atmosphere	71
	<i>by A. Quirrenbach, University of California, San Diego</i>	
5.1	Introduction	71
5.2	The Kolmogorov Turbulence Model	72
5.2.1	Eddies in the Turbulent Atmosphere	72
5.2.2	The Structure Function for Kolmogorov Turbulence	72
5.2.3	Structure Function and Power Spectral Density of the Refractive Index	73
5.3	Wave Propagation Through Turbulence	74
5.3.1	The Effects of Turbulent Layers	74
5.3.2	Calculation of the Phase Structure Function	75
5.3.3	Phase-Coherence Function and Fried Parameter	76
5.4	The Effect of Turbulence on Images	76
5.4.1	Optical Image Formation	76
5.4.2	Diffraction-Limited Images and Seeing-Limited Images	77
5.4.3	The Significance of the Fried Parameter r_0	78
5.4.4	Strehl Ratio	79
5.4.5	Taylor Hypothesis and τ_0	80
5.4.6	Anisoplanatism	80

5.4.7	Scintillation	82
5.4.8	Turbulence and Wind Profiles	83
5.5	Appendix: Some Useful Facts from Fourier Theory	84
6	Design of Stellar Interferometers	87
	<i>by T.A. ten Brummelaar, Georgia State University</i>	
6.1	Introduction	87
6.2	What We Would Like to Measure	87
6.3	Logistics	88
6.4	Concrete and Steel	89
6.5	The Problem with the Atmosphere	90
6.6	Polarization	92
6.7	Internal Optical Quality	94
6.8	Diffraction	94
6.9	Dispersion	94
6.10	Controlling the Beast	95
6.11	Conclusion	96
II	Measurement and Calibration of Fringe Parameters	101
7	Visibility Estimation and Calibration	103
	<i>by M.M. Colavita, Jet Propulsion Laboratory</i>	
7.1	Introduction	103
7.2	Fringe Scanning and Matched Filtering	103
7.3	Visibility Estimation and Signal-to-Noise Ratio	104
7.4	Estimator Biases	106
7.5	Atmospheric Biases	108
	7.5.1 Spatial Wavefront Errors	108
	7.5.2 Temporal Errors	109
	7.5.3 Finite Coherence	109
	7.5.4 Mismatched Stroke	111
7.6	Conclusion	111
8	Phase and Group Delay Estimation	113
	<i>by P.R. Lawson, Jet Propulsion Laboratory</i>	
8.1	Motivation	114
8.2	Phase and Group Delay	115
8.3	Model of the Fringe	117
	8.3.1 Coherence Envelope	118
	8.3.2 Channeled Spectrum	119
8.4	Methods of Tracking Fringe Phase	123
	8.4.1 Phase Tracking with Four Quarter-Wavelength Steps	124
	8.4.2 Phase Tracking with a Linear Pathlength Sweep	125

8.4.3	Simultaneous Phase Measurements at Several Wavelengths	126
8.4.4	Triangle vs Sawtooth Modulations	127
8.5	Methods of Tracking Group-Delay	127
8.5.1	Channeled Spectrum: Fast Fourier Transform	129
8.5.2	Channeled Spectrum: Lomb-Scargle Periodogram	133
8.5.3	Channeled Spectrum: Least-Squares Fit	133
8.5.4	Multi-Wavelength Phasor Measurements: Fast or Direct Fourier Transform	134
8.5.5	Multi-Wavelength Phasor Measurements: “Optimal” Estimator	135
8.6	Variance of Phase and Group Delay Estimates	136
8.6.1	Variance of Phase	136
8.6.2	Variance of Group Delay	136
8.6.3	Phase and Group Delay Variance Compared	138
8.7	Conclusion	138
9	Phase Referencing	143
	<i>by A. Quirrenbach, University of California, San Diego</i>	
9.1	Introduction	143
9.2	Principles of Phase Referencing	144
9.2.1	The Mark III Interferometer	144
9.2.2	Visibility Estimation and Signal-to-Noise Ratio	146
9.2.3	Phase-Referenced Visibility Averaging	147
9.2.4	Limb Darkening of Arcturus	148
9.2.5	Further Applications of Same-Source Phase Referencing	149
9.2.6	Off-Source Phase Referencing	152
9.3	Phase Decorrelation Mechanisms	154
9.3.1	Phase Errors and Coherence Losses	154
9.3.2	Photon Noise in the Tracking Channel	154
9.3.3	Color and Visibility Dependence of the Effective Tracking Wavelength	155
9.3.4	Stroke Mismatch	155
9.3.5	Fringe Jumps	156
9.3.6	Dispersion	156
9.3.7	Anisoplanatism	157
9.3.8	Differential Refraction	157
9.3.9	Diffraction	160
III	Astrometry	163
10	Wide-Angle Astrometry	165
	<i>by D.J. Hutter, U.S. Naval Observatory</i>	
10.1	Introduction	165
10.2	Wide-Angle Astrometry	167

10.3	Baseline Metrology	167
10.4	Constant Term	168
10.5	Compensation of Atmospheric Effects	171
10.6	Current Status	172
11	Ground-Based Narrow-Angle Astrometry	177
	<i>by M.M. Colavita, Jet Propulsion Laboratory</i>	
11.1	Introduction	177
11.2	Atmospheric Effects	178
11.3	Other Errors	178
11.4	Implementing a Narrow-Angle Measurement	179
11.5	Conclusion	181
IV	Optical Synthesis Imaging	183
12	Interferometry with Two Telescopes	185
	<i>by H.M. Dyck, U.S. Naval Observatory</i>	
12.1	Introduction	185
12.2	Simple Concepts of Aperture Synthesis	185
12.3	Optical Delay and (u, v) Plane Tracks	189
12.4	Simple Models of Astrophysical Sources	190
	12.4.1 Binary Stars	192
	12.4.2 Stellar Angular Diameters	193
	12.4.3 Circumstellar Shells	194
12.5	Cautions and Caveats	196
12.6	Comparisons	196
13	An Introduction to Closure Phases	203
	<i>by J.D. Monnier, Smithsonian Astrophysical Observatory</i>	
13.1	Introduction	203
	13.1.1 Telescope Errors: Complex Gain	203
	13.1.2 Phase Referencing	206
13.2	The Closure Quantities	207
	13.2.1 Closure Phase and the Bispectrum	207
	13.2.2 Closure Amplitudes	208
	13.2.3 Measuring Closure Phases	209
	13.2.4 Closure Phase Relations	213
	13.2.5 Simple Cases	214
	13.2.6 Summary	216
13.3	Imaging	216
	13.3.1 Converting Bispectrum to Complex Visibilities	217
	13.3.2 Imaging Goals	218
	13.3.3 “Standard” Aperture Synthesis Imaging	219

13.3.4	Including Closure Phase Information	221
13.3.5	Imaging Results from Keck Aperture Masking	223
13.4	Outstanding Issues	224
13.4.1	Overview of Problems	224
13.4.2	New Possibilities	225
13.5	Summary of Important Points	226
14	Interferometer Design for Synthesis Imaging	231
	<i>by D. Mozurkewich, Naval Research Laboratory</i>	
14.1	Good Fringes vs Good Science	231
14.1.1	Alternative 1: Integrate Forever	234
14.1.2	Alternative 2: Sources with High Visibility	235
14.1.3	Alternative 3: Wavelength Bootstrapping	237
14.1.4	Alternative 4: Baseline Bootstrapping	239
14.1.5	Alternative 5: Guide Star Methods	240
14.2	Design of Interferometric Arrays	242
14.2.1	Optimal Design of Interferometric Arrays	242
14.2.2	Partially Redundant Arrays	242
14.2.3	Array Design with Vacuum Feed Systems	243
14.2.4	NPOI Configuration	243
14.2.5	Polarization and Beam Rotation	244
14.3	Beam Combination and Modulation	245
14.3.1	Pupil-Plane Combination: Passive Detection	245
14.3.2	Image-Plane Combination: Spatial Modulation	246
14.3.3	Pupil-Plane Combination: Temporal Modulation	247
14.3.4	Demodulating Multiple Baselines	248
14.4	Beam Combination Techniques	251
14.4.1	All-On-One Combination	251
14.4.2	Pairwise Combination	251
14.4.3	Partial Pairwise Combination	253
14.5	Comparison of Beam Combination Techniques	254
15	Data Reduction for Synthesis Imaging	257
	<i>by J.T. Armstrong, Naval Research Laboratory</i>	
15.1	Preliminaries	257
15.2	Planning the Observations	258
15.2.1	Stellar Diameters	259
15.2.2	Stellar Surface Structure	261
15.2.3	Binary-Star Separation	262
15.2.4	Circumstellar Material	263
15.3	Closure Phase and Calibration	263
15.4	Model Fitting and Imaging	265
15.5	Current Capabilities and Limitations	269

15.6	Future Developments	269
V	Nulling Interferometry	273
16	Nulling Interferometry and Planet Detection	275
	<i>by E. Serabyn, Jet Propulsion Laboratory</i>	
16.1	Introduction	275
16.2	Exozodiacal Light	276
16.3	Nulling Interferometry Basics	277
16.4	Symmetry and Stability Requirements	282
16.5	Rooftop-Based Rotational Shearing Interferometers	285
16.6	Experimental Results	287
16.7	Multi-Baseline Nulling	289
VI	Stellar Astronomy and Astrophysics	293
17	Binary Stars	295
	<i>by W.I. Hartkopf, U.S. Naval Observatory</i>	
17.1	Reason One: Binaries as Scales	296
	17.1.1 Part 1	296
	17.1.2 Part 2	296
17.2	Reason Two: Binaries as Yardsticks	301
17.3	Reason Three: Binaries and Stellar Evolution	301
17.4	Reason Four: Binaries in Other Guises	302
17.5	Reason Five: Binaries as “Vermin”	304
VII	Future Ground-Based Interferometry	309
18	Future Ground-Based Interferometry	311
	<i>by S.T. Ridgway, NOAO & Georgia State University</i>	
18.1	Existing Facilities and Scientific Output	311
18.2	Prospects for Astrophysics with Interferometers	313
18.3	Next Generation Ground-Based Array	315
	18.3.1 Near Infrared Capability	315
	18.3.2 Baseline of ~1000 m	315
	18.3.3 General Array Capabilities	315
	18.3.4 Number of Telescopes and Array Geometry	316
	18.3.5 Adaptive Optics and Aperture Size	317
	18.3.6 Budget Estimate	318
18.4	Astrophysical Potential of an IRVLA	318
18.5	Comments and Conclusion	320

Appendices	322
A Notes on the History of Stellar Interferometry	325
<i>by P.R. Lawson, Jet Propulsion Laboratory</i>	
A.1 Fizeau and Michelson	325
A.1.1 American References to European Work in Stellar Interferometry . .	326
A.1.2 Could Michelson Have Been Influenced by Fizeau Prior to 1890? . .	327
A.2 A Timeline of Stellar Interferometry: 1868–2000	329
B Suggested Reading	333
B.1 General References	333
B.2 Long Baseline Stellar Interferometers	335

Preface

Up until the latter half of the twentieth century, high angular resolution astronomy has been limited by the unsteadiness of images observed through the turbulent atmosphere. Although we are now able with adaptive optics to partially compensate for atmospheric seeing, even the largest telescopes are only able to resolve six or seven of the largest and nearest stars.

Further progress in high angular resolution astronomy can only be obtained through methods of interferometry. When the light from an *array* of telescopes is combined in an interferometer, the attainable resolution is limited not by the diameter of individual telescopes, but by the longest baseline spanning the array. Whereas interferometry has been well developed for use at radio wavelengths, with researchers involved both in national facilities and international collaborations, interferometry at optical and infrared wavelengths has yet remained relatively unexploited. Major new optical/infrared interferometers are now under construction by groups in both the United States and Europe and will soon open the field to new and exciting science. Longer-term plans are also underway by the National Aeronautics and Space Administration and the European Space Agency for the development of space-borne interferometers for astrometry, imaging, and planet detection. The technology for optical/infrared interferometry continues to develop rapidly, and exciting opportunities await astronomers and astrophysicists now entering this field.

Michelson Fellowship Program

The Michelson Fellowship Program, funded through NASA's Origins program and the Space Interferometry Mission, seeks to support the scientific community in building expertise in optical and infrared interferometry. The Michelson Fellowship Program brings together students and researchers in all stages of their career, through fellowships at the post-doctoral and graduate level, as well as through undergraduate research opportunities, and summer schools.

1999 Michelson Summer School

The 1999 Michelson Interferometry Summer School was held 9–13 August, 1999, at the California Institute of Technology in Pasadena, California. The school was attended by 43 graduate students, 13 post-doctoral level researchers, and 15 professional scientists. This was the first Summer School within the Michelson Fellowship Program and followed on from the Summer School on Optical/IR Interferometry hosted by the U.S. Naval Observatory in Flagstaff, Arizona, in October 1998. The lectures were given by staff of the Jet Propulsion Laboratory, the US Naval Observatory, the Naval Research Laboratory, and faculty and

staff from the University of California (at Berkeley and San Diego), Harvard University, and Georgia State University.

The lectures emphasized the fundamentals of astronomical interferometry, focusing principally on the engineering aspects of stellar interferometers. Subjects that were reviewed included the design of interferometric arrays, strategies for combining starlight, and the principles of observing, data reduction, modeling, and synthesis imaging.

The lectures were supplemented with day trips to the Mount Wilson Observatory and the Palomar Observatory. This provided the opportunity to visit the Infrared Spatial Interferometer, the CHARA Array, and the Palomar Testbed Interferometer, and the possibility for further discussions with the lecturers.

Overview of the Course Notes

These course notes document the lecture series from the 1999 Summer School. The lecture material was chosen to emphasize the development of ground-based interferometry and to include an introduction to the future possibilities of space missions within NASA's Origins Program. The material contained in the course notes is, however, primarily concerned with ground-based interferometry. Plans for space-based interferometry, although not described here, are well represented in the JPL publications describing the Space Interferometry Mission and the Terrestrial Planet Finder (see Appendix B).

Although most subjects are covered in detail in their respective chapters, certain subjects, such as the use of fiber optics, spatial filters, and adaptive optics, are mentioned only in passing. Some omissions were inevitable due to the limited scope of the Summer School, and so resources for further reading are included in the Appendices.

The course notes are divided into 7 parts and 18 chapters covering the major themes presented during the school. When reading from one chapter to another, please bear in mind that the notation is only consistent within individual chapters.

PETER R. LAWSON

Addendum

These course notes have been revised for re-issue on CDROM. The revision corrected minor errors in the text, updated the website addresses included in Chapter 1 and Appendix B, and improved the quality of the corresponding PDF files. I am most grateful to Chris Hawley at JPL for helping me resolve questions concerning L^AT_EX 2_ε, dvips, and PDF file conversion. This revision of the Course Notes in PDF format is also available at the Michelson Fellowship Program Website at <http://sim.jpl.nasa.gov/library/coursenotes.html>.

— PRL, December 6, 2000.

Acknowledgments

I would like to begin by thanking my wife Laurence for her patience and encouragement, my daughter Chloé for enduring a somewhat more hectic routine in August 1999, and my son Felix for staying where he was until after the dust settled—born August 30, 1999.

I offer my warmest thanks to the lecturers and authors who contributed to this volume. As a group they laughed at my early deadlines but were in time helpful, kind, and accommodating. I am grateful for their time and cooperation.

I would particularly like to thank Ken Johnston for his encouragement and for the financial support of staff from the U.S. Naval Observatory.

I was fortunate to have the advice of Mark Colavita, Mike Shao, and Steve Unwin in sorting out priorities, speakers, subjects, times, and overall content during the early stages of planning.

I take particular pleasure in thanking Jo Pitesky, Danika Jensen, and Debi Fambro, for helping bring the Summer School together. I am grateful to Jo and Danika for their efforts in shaping the Summer School, including (but not limited to) registration, accommodations, transportation, tours, lecture facilities, banquet arrangements, coffee breaks, video-taping, and audio-visual equipment. I am grateful to Debi for providing the travel arrangements for lecturers and students, and for her remarkable courtesy, efficiency, and patience.

For exceptional service through the JPL travel authorization office, I am most grateful to Laura Lee and Mariza Arnot.

My thanks also to Brian Knapp and Kalee Perrin for their efforts in the labour-intensive task of scanning viewgraphs into PDF format, and Arlene Dondoyano and Kate Weisberg at NASA Ames for copying the lectures onto real-video format.

The tours and facilities at the Mount Wilson Observatory were made possible through the Mount Wilson Institute with the generous help of Bob Jastrow, Mary Cragg, Sean Hoss, and the docents Mary Brown, Gale Gant, Don Nicholson, and Greg Smith. Access to and subsequent re-alignment of the CHARA Array was kindly provided by Hal McAlister, Theo ten Brummelaar, Steve Ridgway, Nils Turner, Laszlo Sturmman, and Joey Seymore. Access to the Infrared Spatial Interferometer was provided courtesy of C.H. Townes, Bill Danchi, John Monnier, and David Hale.

Facilities at the Palomar Observatory were made possible with the help of Wal Sargent, Bob Brucato, and Bob Thicksten. Tours of the Observatory and the Palomar Testbed

Interferometer were provided courtesy of staff at the Jet Propulsion Laboratory, including Andrew Booth, Andy Boden, Gerard van Belle, Kent Wallace, and Mark Colavita.

I would like to thank the Michelson Fellowship Program for its generous financial support for the editing of this volume. The Course Notes were edited with the assistance of Sunjay Moorthy, Mary Young, and Larry Palkovic at JPL Publications. The real or virtual book you hold was assembled using L^AT_EX 2_ε, and B^IB_TE_X with a standard `book` class modified to include PostScript fonts.

It is a great pleasure to thank Rudolf Danner and Mark Colavita for allowing what was previously only “a rewarding detour from my usual responsibilities” to become an integral part of my work at JPL.

The inside front cover photograph of Albert Michelson (standing in front of the Maryland Hotel) and inside back cover photograph of Francis Pease (using the 20-ft interferometer on the 100-in telescope) were provided courtesy of the Archives, California Institute of Technology.

The Michelson Fellowship Program is funded through NASA’s Origins program and the Space Interferometry Mission at the Jet Propulsion Laboratory. This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

PETER R. LAWSON
Jet Propulsion Laboratory
Pasadena, California
July 7, 2000

Part I

Introduction to Long-Baseline Stellar Interferometry

Chapter 1

Why Build Stellar Interferometers?

HAROLD A. MCALISTER

CENTER FOR HIGH ANGULAR RESOLUTION ASTRONOMY
GEORGIA STATE UNIVERSITY, ATLANTA, GEORGIA

1.1 The Challenge

Optical interferometers offer tremendous challenges and opportunities to suit the scientific tastes of those wishing to build instruments at the frontiers of technology that will enable science of an unprecedented nature. The minimum problem posed by interferometry is the combination at a beam splitter of light from a pair of telescopes whose baseline projected onto the sky determines the achievable resolution. In order to produce fringes at the beam splitter, one must match the paths followed by the two beams to a micron or so in length and hold them stable to a fraction of a wavelength of light. Herein lies the technical challenge. Numerous factors conspire to make the creation of fringes a devilishly difficult chore. Path delays and wavefront tilts are induced by the atmosphere even before light is collected by the telescopes, and then all the downstream subsystems further degrade the problem.

Thus, in order to produce fringes, interferometers are necessarily complex and nested systems possessing numerous sophisticated subsystems. The light collectors themselves, be they siderostats or telescopes, have complex controls and must be engineered to maintain stiffness and smoothness of operation. Their sheer numbers pose problems of maintenance that must not be underestimated. Optical delay lines are complex instruments occupying large physical spaces and are a primary facilities burden and cost driver for an interferometer. On the other hand, the technical basis for delay lines is quite mature and highly functional, and the devices work really very well. Beam combination schemes are an especially complicated issue, particularly when the number of collecting elements grows beyond

Table 1.1: Current Ground-Based Optical/Infrared Interferometers

Facility Acronym	Operating Institution(s)	Site Location	No. of Collecting Elements	Element Aperture (cm)	Maximum Baseline (m)	Operating Wavelength (microns)	Operating Status
GI2T	Obs. Côte d'Azur	Calern, FR	2	150	70	0.4–0.8 & >1.2	since 1985
ISI	UC Berkeley	Mt. Wilson, US	3	165	30+	10	since 1990
COAST	Cambridge U	Cambridge, UK	5	40	22	0.4–0.95 & 2.2	since 1991
SUSI	Sydney U	Narrabri, AU	13	14	640	0.4–0.66	since 1991
IOTA	CfA/U Mass	Mt. Hopkins, US	3	45	38	0.5–2.2	since 1993
NPOI	USNO/NRL	Anderson Mesa, US	6	60	435	0.45–0.85	since 1995
PTI	JPL/Caltech	Mt. Palomar, US	2	40	110	1.5–2.4	since 1995
MIRA-I	NAO Japan	Tokyo, Japan	2	25	4	0.8	since 1998
CHARA	Georgia St. U	Mt. Wilson, US	6	100	350	0.45–2.4	since 1999
KI	CARA	Mauna Kea, US	2(4)	1,000(180)	140	2.2–10	initial 2001
VLTI	ESO	Cerro Paranal, Chile	4(3)	820(180)	200	0.45–12	initial 2001
LBT	U Arizona, Italy, et al.	Mt. Graham, US	2	840	23	0.4–400	initial 2005?

half a dozen or so, and offer plenty of room for creative ingenuity. Alignment and stability offer interesting problems, and interferometers possess a depressingly large number of optical surfaces, each one of which represents loss of light and corruption of that which it passes downstream.

Many of these hardware subsystems have to be actively controlled and must work together with other subsystems. There is no shortage of really fascinating hardware to be designed, fabricated, installed, aligned, controlled and maintained in an interferometer. Thousands of lines of code must be written, tested and probably continuously debugged. Obviously, there is plenty of opportunity for failures along the way, but optimists would regard these risks part of the allure of interferometry.

In addition to the hardware/software issues, interferometers require new tools and algorithms for optimally scheduling observations and calibrating and archiving the data. Perhaps the most challenging problem in this area is the task of producing images of high fidelity and reliability.

Finally, there is the challenge of obtaining adequate funding to develop instruments in an area still regarded by most as developmental. In particular, one must avoid over-heightening expectations as to the kind of science forthcoming in the near term. To a considerable extent, speckle interferometry suffered in the 1970s and early 1980s from an exaggeration of its potential.

Many people find this sobering litany of problems to be the real meat of interferometry, and, rather than being discouraged, some very clever scientists and engineers have devoted much of their careers to solving the basic technical issues. Because of their efforts, we now have a modest retinue of interferometers around the world poised to provide a substantial body of science.

It has taken more than a century to extrapolate the basic physics of interferometry into the working (or nearly so) instruments listed in Tables 1.1 and 1.2. I include only ground-based

Table 1.2: Additional Information on Ground-Based Projects

Facility Acronym	Facility Name	Website (http://)
GI2T	Grand Interféromètre à 2 Télescopes	www.obs-nice.fr/fresnel/gi2t/en/
ISI	Infrared Spatial Interferometer	isi.ssl.berkeley.edu/
COAST	Cambridge Optical Aperture Synthesis Telescope	www.mrao.cam.ac.uk/telescopes/coast/
SUSI	Sydney University Stellar Interferometer	www.physics.usyd.edu.au/astron/astron.html
IOTA	Infrared/Optical Telescope Array	cfa-www.harvard.edu/cfa/oir/IOTA/
NPOI	Navy Prototype Optical Interferometer	ad.usno.navy.mil/npoi/
PTI	Palomar Testbed Interferometer	huey.jpl.nasa.gov/palomar/
MIRA-I	Mitaka optical-Infrared Array	tamago.mtk.nao.ac.jp/mira/
CHARA	Center for High Angular Resolution Astronomy (CHARA) Array	www.chara.gsu.edu/CHARA/
KI	Keck Interferometer	huey.jpl.nasa.gov/keck/
VLTi	Very Large Telescope Interferometer	www.eso.org/projects/vlti/
LBT	Large Binocular Telescope	medusa.as.arizona.edu/lbtwww/lbt.html

facilities in the tables (and in this discussion), but there is, of course, considerable activity and momentum in space-borne interferometers as well.

Current instruments explore four degrees of freedom (number and aperture of collecting elements, maximum baseline, and wavelength regime) with a level of incompleteness consistent with available funding. Each of these facilities has its own approach to solving the phasing problem, but all interferometers possess certain similarities. Some of these instruments are principally devoted to rather specific scientific problems such as absolute astrometry for NPOI and stellar diameters for SUSI. The reader can explore websites to see how each of these instruments has responded to the technical challenges of producing fringes.

Several of these instruments are lineal descendents of the Mark III interferometer that operated on Mount Wilson during the 1980s and presented solutions to most of the requisite technical challenges. The Mark III also produced important scientific results of high accuracy to lend confidence in the value of interferometry.

1.2 The Opportunity

What science can these instruments pursue? Will they live up to their promise? Will their scientific products engender the confidence of the scientific community to invest precious resources in next-generation instruments? Are we even building the right instruments now?

We can only hazard a guess at the first of these four questions. But it is a well-informed guess. Unquestionably, current interferometers possess wonderful resolution. The longest baseline facility now in existence is the Sydney University Stellar Interferometer (SUSI). Its limiting resolution of 100 micro-arcsec is a gain of four orders of magnitude over traditional ground-based direct imaging through photography and more than two orders of magnitude gain over adaptive optics corrected telescopes and over the Hubble Space Telescope. But SUSI and other long-baseline interferometers are extremely limited in sensitivity, and all interferometers inherently suffer from extremely narrow fields of view. A comparison with HST is thus a bit cavalier in terms of comparative sensitivity and field size.

The current generation of interferometers will primarily contribute to stellar astronomy. Their ability to play an important role in extragalactic astronomy (for example, through the direct imaging of broad-line regions of quasars) is made infeasible by their small apertures and/or relatively short baselines. Signal-to-noise ratio is a precious commodity to an interferometer where exposure times are limited by the atmospheric redistribution time t_0 , typically a few tens of milliseconds.

So, while interferometry will not soon satisfy the needs of the extragalactic community, the resolution and accuracy brought to bear on problems of stellar astrophysics will yield substantial new science. Fundamental new data for stars will be forthcoming in unprecedented quantity and quality. These data will include effective temperatures, surface fluxes, masses and luminosities for stars well distributed over spectral type and luminosity class. Sensitivity limitations will maintain the elusiveness of white dwarfs and the lower end of the main sequence, but, for the first time, tens of thousands of objects populating the majority of the H-R diagram will be accessible to high-resolution studies.

To be most useful, this flood of new data must be well calibrated. Measurements of the physical parameters for stars require accuracies at the couple of percent level in order to best challenge astrophysical theory. Resolution *and* accuracy are together the key to having the greatest scientific impact. Here, again, the history of speckle interferometry comes to mind in which casual calibration of potentially simple things like pixel scale led to results of little or no use even though significant amounts of large telescope time were consumed in their production.

We do not need thousands of new stellar masses accurate to 10%, but we do need hundreds accurate to 1%. Similarly, stellar limb darkening does not require confirmation but does need to be measured with sufficient accuracy to confront theory. Interferometrists need to establish more collaborations with theorists in selecting the optimal utilization of these wonderful new instruments.

In addition to the proliferation of basic data for stars, interferometers can and will contribute to a wide variety of problems. For single stars, such problems include the measurement of limb darkening, determination of linear diameters for stars with accurate parallaxes, studying phenomena associated with star formation (including dynamic phenomena) and pre-main sequence objects, measuring absolute rotation, stellar flares, p-mode oscillations and the pulsations of Cepheid and Mira variables (to include the direct geometric calibration of the period-luminosity relation for Cepheids), and phenomena associated with hot stars (shells, winds, etc.) and cool giants and supergiants. For binary stars, in addition to resolving the majority of the spectroscopic binaries and providing masses in large numbers, interferometric surveys for duplicity will be carried to new levels of completeness and close binary phenomena will be detected and maybe even imaged. Low mass companions, including those of planetary mass, may be astrometrically detected in binaries.

The first really interesting images from interferometry will involve the detection of surface features on normal stars and phenomena in the close-in environments of young stellar

objects. Interferometers will witness the eruptions of novae and perhaps even the explosion of a supernova.

We can predict, with great longing, many of the research enterprises to be opened by interferometry in the coming years. But we must keep in mind that we are dealing with multiple orders of magnitude increase in resolution. In a letter to the author in 1990, UCLA astronomer Daniel Popper remarked

History has taught us that whenever a new technique enters a new realm of observational phase space, the most striking and productive results tend to be those not anticipated by even the most prescient thinkers.

Professor Popper, who maintained the very highest research standards throughout his long and exceptionally productive career and who was known and respected for his very careful and critical approach to science, clearly felt that the unexpected discoveries to be made by interferometry will be the true hallmarks of the field.

1.3 Towards the Future

The current generation of projects may be the stepping-stones to an “Optical/IR Very Large Array.” For this to happen, significant science must be forthcoming in the near term from our present investment in the field. In this context, “significant” implies quality as well as quantity in support of pressing problems in stellar astrophysics. Imaging of relatively complex objects must be demonstrated, and this is a challenge due to the small number of collecting telescopes in current arrays. We can anticipate very little extragalactic results except for calibrations, based upon galactic objects. which extend to extragalactic realms and into cosmology. The field needs more partnerships to pool intellectual and financial resources, more involvement of theorists, and the training of more “black-belt” interferometrists.

One might look forward to great successes from present-day efforts so that by, say 2010, considerable momentum will exist towards the design and construction of an interferometer comprised of several dozen 4–6 meter aperture telescopes (each equipped with adaptive optics) distributed over kilometer-plus baselines. At that time, interferometry will truly have come of age and the words of Dan Popper will entice us to new realms of exploration and discovery.

Chapter 2

Elementary Theory of Interferometry

A.F. BODEN

IPAC, CALIFORNIA INSTITUTE OF TECHNOLOGY, AND
JET PROPULSION LABORATORY
PASADENA, CALIFORNIA

We introduce and discuss the elementary theory of astronomical interferometry. We derive the basic quantitative formalism for interferometric observables from incoherent astronomical sources with a particular emphasis upon optical interferometry. Concrete examples of the theory are given in the context of common model source morphologies.

2.1 Introduction

It is deceptively simple to describe interferometers as instruments that measure interference (or other properties associated with the interference) of an electromagnetic field. The motivation to consider interferometry of astronomical sources is fundamentally pragmatic; we are compelled to consider astronomical interferometers because interferometers provide access to high angular resolution information at a small fraction of the price of conventional single-aperture telescopes with similar angular resolution. This is not to suggest that interferometers replace more conventional astronomical instrumentation, only that they provide a cost-effective means to address certain scientific questions.

Herein we will introduce and develop the basic theory of astronomical interferometry. Starting from general properties of the electromagnetic field, we will consider the response of an idealized interferometer to idealized astronomical sources. We will further apply this theory to common idealized source morphologies. Given the context in which these lecture notes appear, we will primarily make these developments with optical interferometers in

mind; interferometers that operate over finite pass-bands in the optical (or near-optical) part of the electromagnetic spectrum. Optical astronomical interferometers have historically been used to study stars—a natural match given most stars emit a large fraction of their radiation at optical (near-optical) wavelengths.

2.2 A Simple Interferometer and a Monochromatic Source

Consider a model two-aperture interferometer as depicted in Figure 2.1. Two identical apertures A_1 and A_2 are located at three-space positions \mathbf{x}_1 and \mathbf{x}_2 respectively, and thus are separated by a displacement $\mathbf{B} \equiv \mathbf{x}_2 - \mathbf{x}_1$. \mathbf{B} is typically known as the *baseline* of the interferometer. Each aperture is pointed at a single celestial point source located at relative position \mathbf{S} from the centerline of the array pair; the pointing direction is given by the unit vector $\hat{\mathbf{s}} \equiv \mathbf{S}/|\mathbf{S}|$. Because optical photons (to an extremely good approximation) do not interact with each other, in analyzing the interferometer we can consider the harmonic decomposition of the light from the astronomical source. We therefore start by considering the source as monochromatic with wavelength λ . We'll also assume the celestial source is a sufficient distance that the phase-fronts of the incident optical radiation field are planar. At positions \mathbf{x}_1 and \mathbf{x}_2 the monochromatic optical fields from the source have a simple form as (the real part of) an exponential.*:

$$\phi_1 \sim e^{i\mathbf{k}\cdot\mathbf{x}_1} e^{-i\omega t} = e^{-ik\hat{\mathbf{s}}\cdot\mathbf{x}_1} e^{-i\omega t}$$

and

$$\begin{aligned} \phi_2 \sim e^{i\mathbf{k}\cdot\mathbf{x}_2} e^{-i\omega t} &= e^{-ik\hat{\mathbf{s}}\cdot\mathbf{x}_2} e^{-i\omega t} \\ &= e^{-ik\hat{\mathbf{s}}\cdot\mathbf{x}_1} e^{-ik\hat{\mathbf{s}}\cdot\mathbf{B}} e^{-i\omega t}. \end{aligned}$$

Without loss of generality we can absorb the common phase factor $e^{-ik\hat{\mathbf{s}}\cdot\mathbf{x}_1}$ into whatever normalization we choose for the optical fields, hence:

$$\begin{aligned} \phi_1 &\sim e^{-i\omega t}, \\ \phi_2 &\sim e^{-ik\hat{\mathbf{s}}\cdot\mathbf{B}} e^{-i\omega t}. \end{aligned} \tag{2.1}$$

Equation 2.1 merely codifies the fact that the relative phase of the radiation incident on the two apertures is a function of the geometry of the viewing situation—in particular the relative angle of the incoming phase fronts and the baseline vector \mathbf{B} .

*Herein, in so far as possible, we take our nomenclature for electromagnetic fields from Jackson (1998) In particular, a plane parallel monochromatic electromagnetic field of frequency ν propagating in free space in a direction $\hat{\mathbf{n}}$ is written as:

$$\phi \sim A e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)},$$

with

$$\begin{aligned} \omega &= 2\pi\nu = 2\pi c/\lambda, \\ k &= \omega/c = 2\pi\nu/c, \\ \mathbf{k} &= k\hat{\mathbf{n}}, \end{aligned}$$

and \mathbf{k} (k) as the wave vector (number) of the field. Consult Jackson (1998) Chapter 7 for additional details.

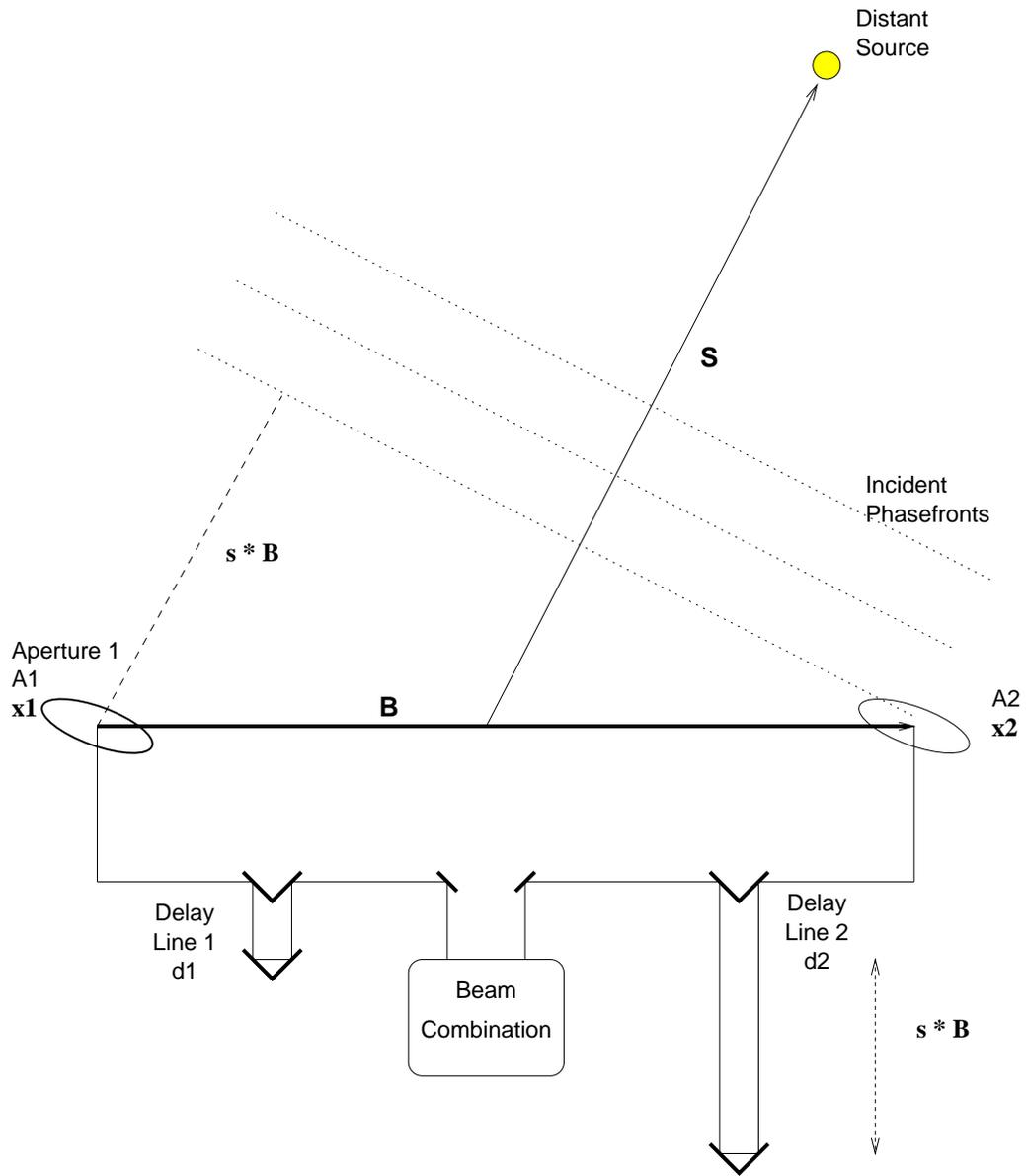


Figure 2.1: Idealized Interferometer.

As shown in Figure 2.1, we imagine that after collection the optical fields are propagated to a power-linear detector where they combined, and the resulting output is measured. Post-collection the optical fields are propagated over distinct distances d_1 and d_2 , incurring an additional relative phase. Imagining that the optical fields are directly combined, at the detector the fields have phases that are:

$$\begin{aligned}\phi_1 &\sim e^{ikd_1} e^{-i\omega t}, \\ \phi_2 &\sim e^{ikd_2} e^{-ik_0 \hat{\mathbf{s}} \cdot \mathbf{B}} e^{-i\omega t}.\end{aligned}\tag{2.2}$$

With no significant efficiency differences in the two interferometer arms, a direct[†] combination of the two optical fields results in a net field whose phase is given by:

$$\phi_{net} = \phi_1 + \phi_2 \sim e^{-i\omega t} \left(e^{ikd_1} + e^{ikd_2} e^{-ik\hat{\mathbf{s}} \cdot \mathbf{B}} \right),$$

from which the resulting time-averaged detected power is (proportional to):

$$P \propto \phi_{net}^* \phi_{net} = 2(1 + \cos k(\hat{\mathbf{s}} \cdot \mathbf{B} + d_1 - d_2)).$$

To be more concrete, let us specify the incident source flux power F in units of energy incident per unit time per unit cross-sectional area, and the collecting area of the apertures as A . Then (up to efficiency factors) the detected power is given by:

$$P = 2AF(1 + \cos k(\hat{\mathbf{s}} \cdot \mathbf{B} + d_1 - d_2))\tag{2.3}$$

$$= 2AF(1 + \cos kD).\tag{2.4}$$

In the space of relative delay $D \equiv \hat{\mathbf{s}} \cdot \mathbf{B} + d_1 - d_2$, P varies harmonically between zero and $2AF$ (the total collected power of the two apertures) with period λ ; this is plotted in Figure 2.2.

Equation 2.3 has the form of an infinite series of power oscillations or *interference fringes*, as a function of the optical delay D , or equivalently $d_1 - d_2$. Because $\hat{\mathbf{s}}$ can be interpreted as an angle on the sky with dimensions of radians, adjacent fringe crests projected on the sky are separated by an angle given by:

$$\Delta s = \frac{\lambda}{B}.\tag{2.5}$$

2.3 Polychromatic Sources and Interferometers of Finite Bandwidth

The interference fringes in Equation 2.3 were infinite; we saw interference regardless of the values of d_1 and d_2 . From a practical standpoint life is not quite this kind. In general we

[†]Direct combination in this context means combination without additional phase asymmetry between the two arms. This is an idealization which simplifies the mathematics at the expense of ignoring a relatively unimportant phase factor present in most optical interferometers which use beam splitters for beam combination.

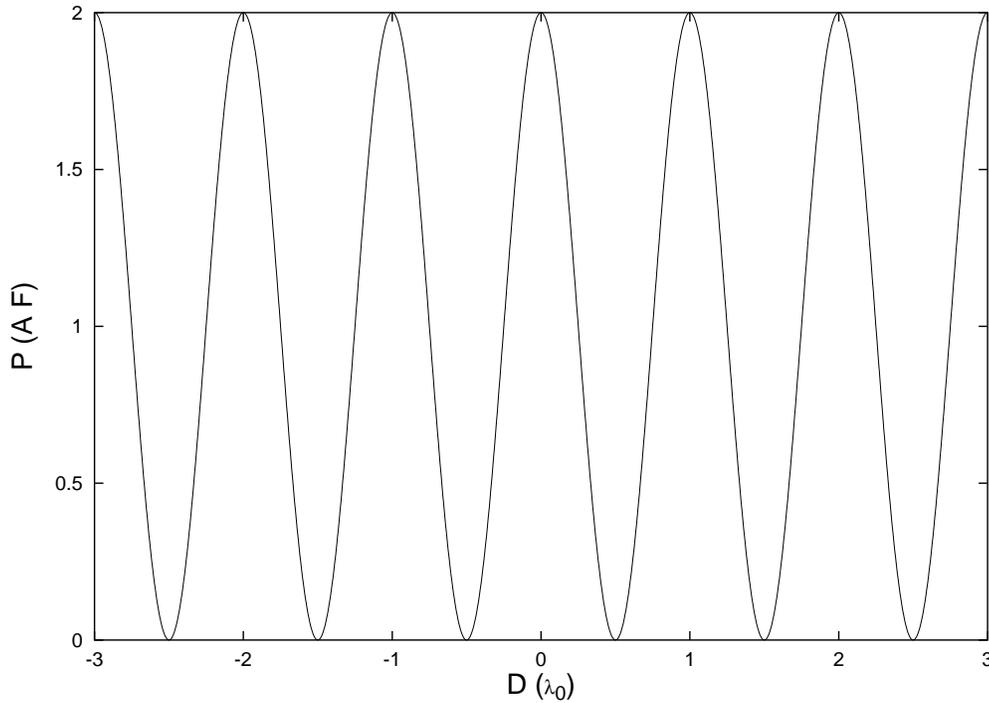


Figure 2.2: Monochromatic Power.

can only build interferometers with finite passbands, and this causes some headaches. Let's see how this goes.

Consider now a more general source with spectral intensity F_ν (dimensions of incident power per cross-sectional area per unit frequency), and an interferometer that has a finite frequency response given by $\eta(\nu)$. Because astronomical sources give us individual frequencies that are mutually incoherent, the total detected power becomes a sum of the detected power at each frequency after Equation 2.3 (writing the integration in the frequency domain):

$$P = \int d\nu 2AF_\nu \eta(\nu) [1 + \cos kD]. \quad (2.6)$$

It is illustrative to consider a specific instance. First, as a simplifying assumption we take the source spectral power to be constant, $F_{\nu-0}$, over the system bandwidth. Next, take a specific bandwidth pattern—a “top hat” pattern with constant throughput η_0 over a frequency (wavelength) band $\nu_0 \pm \Delta\nu/2$ ($\lambda_0 \pm \Delta\lambda/2$). Then Equation 2.6 becomes:

$$\begin{aligned} P &= 2AF_{\nu-0}\eta_0 \int_{\nu_0-\Delta\nu/2}^{\nu_0+\Delta\nu/2} d\nu (1 + \cos 2\pi\nu\tau) \\ &= 2AF_{\nu-0}\eta_0 \left[\nu + \frac{\sin 2\pi\nu\tau}{2\pi\tau} \right]_{\nu_0-\Delta\nu/2}^{\nu_0+\Delta\nu/2} \\ &= 2AF_{\nu-0}\eta_0\Delta\nu \left[1 + \frac{\sin \pi\Delta\nu\tau}{\pi\Delta\nu\tau} \cos 2\pi\nu_0\tau \right] \end{aligned}$$

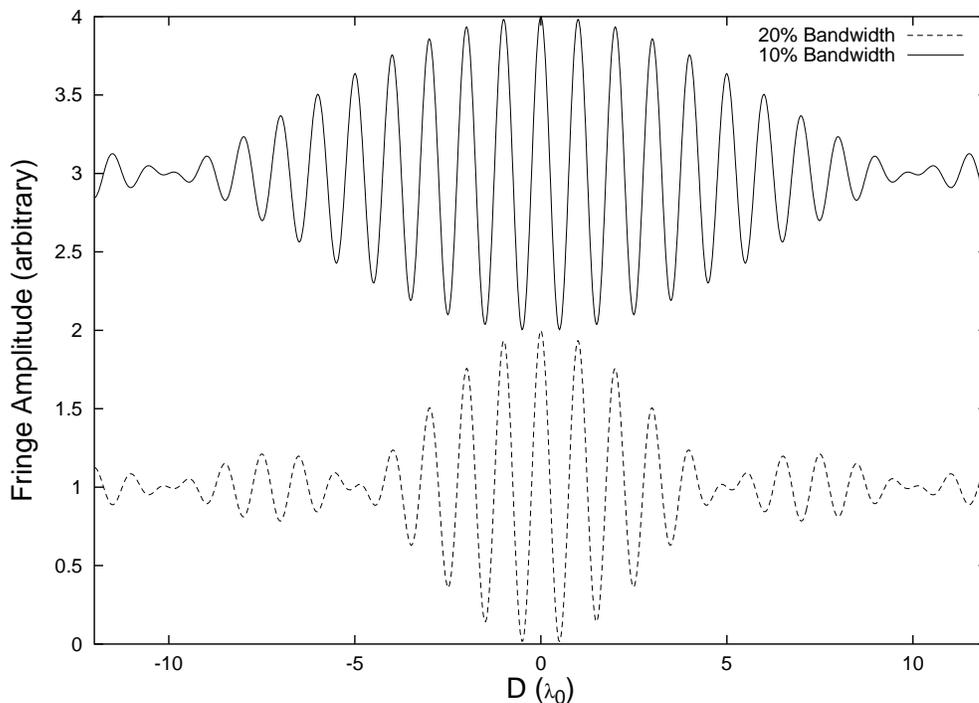


Figure 2.3: Polychromatic Fringe Coherence.

$$\begin{aligned}
 &= 2AF_{\lambda-0}\eta_0\Delta\lambda \left[1 + \frac{\sin \pi\Delta\lambda/\lambda_0^2 D}{\pi\Delta\lambda/\lambda_0^2 D} \cos k_0 D \right] \\
 &= 2AF_{\lambda-0}\eta_0\Delta\lambda \left[1 + \frac{\sin \pi D/\Lambda_{coh}}{\pi D/\Lambda_{coh}} \cos k_0 D \right] \quad (2.7)
 \end{aligned}$$

(with $\tau \equiv D/c$ —dimensions of time). This result is qualitatively similar to Equation 2.3. First it is noteworthy that the leading coefficient of Equation 2.7 is the total collected power by both apertures in a bandwidth $\Delta\nu$ (with efficiency η_0). Further, as in Equation 2.3 the term in brackets has a positive-definite oscillatory behavior in D at a frequency ν_0 /wavelength λ_0 —the center-band frequency. But rather than the fringes being observed at all D as suggested by Equation 2.3, we find the fringes modulated by a sinc function symmetrically centered at $D = 0$, and becoming small as $|\pi\Delta\lambda/\lambda_0^2 D| > 1$. The sinc-modulation of the interference fringes has a characteristic scale or *coherence length* of:

$$\Lambda_{coh} \equiv \frac{\lambda_0^2}{\Delta\lambda}. \quad (2.8)$$

It is noteworthy that the sinc function is the Fourier transform of the top-hat function we took for the system bandpass.

In Figure 2.3 we give two illustrative examples of the oscillatory argument of Equation 2.7, showing fringe patterns at 20% (red— $\Lambda_{coh} = 5\lambda_0$) and 10% (blue— $\Lambda_{coh} = 10\lambda_0$) fractional bandwidths.

Equation 2.7 and Figure 2.3 are typical of the types of fringe envelopes that one can expect from stellar interferometers; it is usually a reasonable first approximation to assume both the source spectral flux and the system throughput are quasi-constant over a finite frequency (wavelength) interval.

The fringe envelopes from interferometers can typically be written as:

$$1 + M(\Lambda_{coh}, D) \cos k_0 D, \quad (2.9)$$

where $M(\Lambda_{coh})$ is the fringe envelope modulation function, typically given by the Fourier transform (modulus) of the system bandpass, and k_0 is the (possibly weighted) center-band wave number. In this context Λ_{coh} sets the physical scale for the precision with which the delays must be matched in the interferometer. Conversely, if it can be measured the relative delay $d_1 - d_2$ becomes a proxy for the source astrometry (Shao *et al.* 1990).

2.4 Phase Reference of the Interferometer; Off-Axis and Extended Sources

In general the response of the interferometer to a point-source at location $\hat{\mathbf{s}}_0$ in the sky is given by something like Equation 2.7; fringes with frequency ω_0 modulated by a fringe envelope that is a function of the system bandwidth (in delay space). In particular, we see the fringe envelope position is given by the product $k_0 D = k_0(\hat{\mathbf{s}}_0 \cdot \mathbf{B} + d_1 - d_2)$, and we are motivated to minimize this product. It is conventional to *define* (or control) the relative delay $d_2 - d_1$ to be equal to (a model of) $\hat{\mathbf{s}}_0 \cdot \mathbf{B}$, then we are guaranteed we are at maximum of the fringe envelope function for a source at $\hat{\mathbf{s}}_0$. In this context $\hat{\mathbf{s}}_0$ becomes our *phase reference* or *phase tracking center*.

Now we can ask what is the response from a point source at $\hat{\mathbf{s}}$ offset slightly from the reference position $\hat{\mathbf{s}}_0$:

$$\hat{\mathbf{s}} = \hat{\mathbf{s}}_0 + \Delta \mathbf{s}.$$

If we write the fringe envelope function as $M(\Lambda_{coh}, \Delta D)$ (which goes to unity in the monochromatic limit), the output power from the interferometer is (after Equation 2.3):

$$\begin{aligned} P &= 2AF (1 + M(\Lambda_{coh}, \Delta D) \cos k_0(\hat{\mathbf{s}} \cdot \mathbf{B} - \hat{\mathbf{s}}_0 \cdot \mathbf{B})) \\ &= 2AF (1 + M(\Lambda_{coh}, \Delta D) \cos k_0(\Delta \mathbf{s} \cdot \mathbf{B})) \\ &= 2AF (1 + M(\Lambda_{coh}, \Delta D) \cos k_0(\Delta D)), \end{aligned} \quad (2.10)$$

with $\Delta D \equiv \Delta \mathbf{s} \cdot \mathbf{B}$. In this construction the sky position $\hat{\mathbf{s}}_0$ as defined by the relative delay $d_2 - d_1$ *defines* the phase reference of the interference fringes on the sky[‡].

[‡]In fact, the relative delay $d_2 - d_1$ defines a *circle* on the celestial sphere around the baseline vector \mathbf{B} . The peak of the optical aperture reception pattern $A(\hat{\mathbf{s}})$ breaks the circular symmetry of the dot product and defines the phase reference $\hat{\mathbf{s}}_0$.

Extended Sources

Any source with finite surface temperature has the potential for being resolved, so we must consider the possibility of resolved sources. Let's describe the source intensity as a function of position $\hat{\mathbf{s}}$ in the sky as $F(\hat{\mathbf{s}})$. Typically F has units of power incident per unit area per solid angle on the sky, and for the moment let's take this with respect to one particular wavelength λ . It is also necessary to characterize the throughput or collection efficiency of the interferometer telescopes as a function of sky position. In terms of what we've had before we'll write this as $A(\hat{\mathbf{s}}, \hat{\mathbf{s}}_0)$, assuming that the telescopes are boresighted on the phase tracking center $\hat{\mathbf{s}}_0$. It is convenient to take the units of A to be effective cross-sectional area, such that a product of $A(\hat{\mathbf{s}}, \hat{\mathbf{s}}_0)F(\hat{\mathbf{s}}) d\Omega$ forms a received power differential.

In the assumption that the radiation from different locations on the source is incoherent, the detected power from an extended source can be computed as an incoherent sum of power from the source decomposed into infinitesimal point sources. Such a model can be written a straightforward extension of the point source model from Equation 2.10:

$$\begin{aligned} P(\hat{\mathbf{s}}_0, \mathbf{B}) &= \int d\Omega A(\hat{\mathbf{s}}, \hat{\mathbf{s}}_0)F(\hat{\mathbf{s}}, \hat{\mathbf{s}}_0) (1 + M(\Lambda_{coh}, \Delta D) \cos k(\Delta\mathbf{s} \cdot \mathbf{B})) \\ &\rightarrow \int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s}) (1 + \cos k(\Delta\mathbf{s} \cdot \mathbf{B})), \end{aligned} \quad (2.11)$$

where I have suppressed the factor of 2 into the magnitude of A , and dropped the envelope function as a notational convenience, regressing to a monochromatic source.

It is interesting, and in fact evocative of how optical interferometers measure fringes in practice, to consider the detected power when a small additional phase is added to one of the delay line arms. To be definite, let's call the delay offset δ with dimensions of length like D , and define it to be positive when a positive delay is added to delay line 1. In this case the detected power becomes:

$$\begin{aligned} P(\hat{\mathbf{s}}_0, \mathbf{B}, \delta) &= \int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s}) (1 + \cos k(\Delta\mathbf{s} \cdot \mathbf{B} + \delta)) \\ &= \int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s}) \\ &\quad + \cos k\delta \int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s}) \cos k(\Delta\mathbf{s} \cdot \mathbf{B}) \\ &\quad - \sin k\delta \int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s}) \sin k(\Delta\mathbf{s} \cdot \mathbf{B}). \end{aligned} \quad (2.12)$$

It is conventional to introduce the *complex visibility* V of the brightness distribution B with respect to the phase reference $\hat{\mathbf{s}}_0$ and aperture function A as:

$$V(k, \mathbf{B}) \equiv \int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s})e^{-ik\Delta\mathbf{s} \cdot \mathbf{B}}. \quad (2.13)$$

Using V we can write the detected power concisely as:

$$\begin{aligned} P(\hat{\mathbf{s}}_0, \mathbf{B}, \delta) &= \int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s}) + Re\{V\} \cos k\delta + Im\{V\} \sin k\delta \\ &= P_0 + Re\{Ve^{ik\delta}\}, \end{aligned} \quad (2.14)$$

where I have written the (two) aperture-integrated power as a constant P_0 :

$$P_0 \equiv \int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s}).$$

To see why Equation 2.14 is considered progress, let's look a little closer at V . To make things definite, let's take a coordinate system where $\hat{\mathbf{s}}_0 = (0, 0, 1)$. So long as either the field of view of the interferometer telescopes is small or the the source brightness is of limited angular extent, $\Delta\mathbf{s}$ is approximately perpendicular to $\hat{\mathbf{s}}_0$ and can be written in terms of angles α and β (units of radians):

$$\Delta\mathbf{s} \approx (\alpha, \beta, 0),$$

and the visibility becomes:

$$V(k, \mathbf{B}) = \int d\alpha d\beta A(\alpha, \beta) F(\alpha, \beta) e^{-ik(\alpha B_x + \beta B_y)}.$$

It is further conventional to define *spatial frequencies* u and v [§]:

$$\begin{aligned} u &\equiv \frac{B_x}{\lambda} = \frac{kB_x}{2\pi}, \\ v &\equiv \frac{B_y}{\lambda} = \frac{kB_y}{2\pi}, \end{aligned} \tag{2.15}$$

for which V becomes:

$$V(u, v) = \int d\alpha d\beta A(\alpha, \beta) F(\alpha, \beta) e^{-2\pi i(\alpha u + \beta v)}. \tag{2.16}$$

As written, $V(u, v)$ is a complex quantity with dimensions of power (as given by the product of A and F).

2.5 Image Synthesis by Discrete Visibility Measurements

The form of Equation 2.16 is clearly that of a two-dimensional Fourier transform of the (aperture efficiency modulated) brightness distribution with u and v assuming roles of spatial frequencies (units of fringe cycles per radian on the sky). That the interferometer response is related to the Fourier transform of the brightness distribution under certain assumptions (source incoherence, small-field approximation) is typically known as the van Cittert–Zernike theorem; the interested reader can find more thorough discussions of the van Cittert–Zernike theorem in Born and Wolf (1999) and Thompson, Moran, and Swenson (1986).

[§]It is further conventional to orient coordinates so u and v represent spatial frequencies in convenient astronomical coordinates like right ascension and declination, but for the present purpose our choice of coordinate rotation is arbitrary

Since Fourier transforms are straightforwardly invertible, the visibility (provided we can measure it) can be used to compute the source brightness distribution:

$$F(\alpha, \beta) = \left(\int du dv V(u, v) e^{2\pi i(\alpha u + \beta v)} \right) / A(\alpha, \beta). \quad (2.17)$$

Equation 2.17 suggests the canonical synthesis imaging program: collect a set of visibility measurements that in some sense approximate the visibility surface over the (u, v) coordinate plane. In practice a set of discrete interferometer baselines \mathbf{B}_i targeted on a common phase-tracking center $\hat{\mathbf{s}}_0$ yields a set of discrete visibility measurements $V_i(u_i, v_i)$. This discrete visibility field can then be inverted by means of a discrete Fourier transform operation to obtain a bandwidth-limited estimate of the parent brightness distribution. The accuracy of the synthesized image is naturally a function of the coverage of the (u, v) plane.

To be concrete, we can describe a sampling function $S(u, v)$ that has the form of a sum of delta functions at the sampled locations (u_i, v_i) :

$$S(u, v) \equiv \sum_i \delta(u - u_i) \delta(v - v_i). \quad (2.18)$$

Utilizing this sampling function Equation 2.17 can be written:

$$F_d(\alpha, \beta) = \left(\int du dv V(u, v) S(u, v) e^{2\pi i(\alpha u + \beta v)} \right) / A(\alpha, \beta). \quad (2.19)$$

Radio Astronomers typically refer to F_d as the *dirty* brightness distribution or image, in that it is apparently related to the true brightness distribution F by the convolution of an effective point-spread function (PSF) or synthesized beam*:

$$F_d(\alpha, \beta) = F(\alpha, \beta) * p(\alpha, \beta),$$

with

$$p(\alpha, \beta) = \int du dv S(u, v) e^{2\pi i(u\alpha + v\beta)}.$$

For our discussion suffice it to say that there are deconvolutional methods to estimate $F(\alpha, \beta)$ from $F_d(\alpha, \beta)$ and $p(\alpha, \beta)$ in the presence of noise; the interested reader is referred to the NRAO Summer School Proceedings (Perley *et al.*, 1989).

2.6 Visibilities of Various Flavors and Their Physical Interpretations

The development of Equation 2.13 might leave the reader with an impression that fringe visibility is a mathematical artifice useful only for image inversion. This is incorrect, and it is instructive to consider the properties of the visibility in a physical context as well as a mathematical one.

*Convolution in the spatial domain is multiplication in the spatial frequency domain

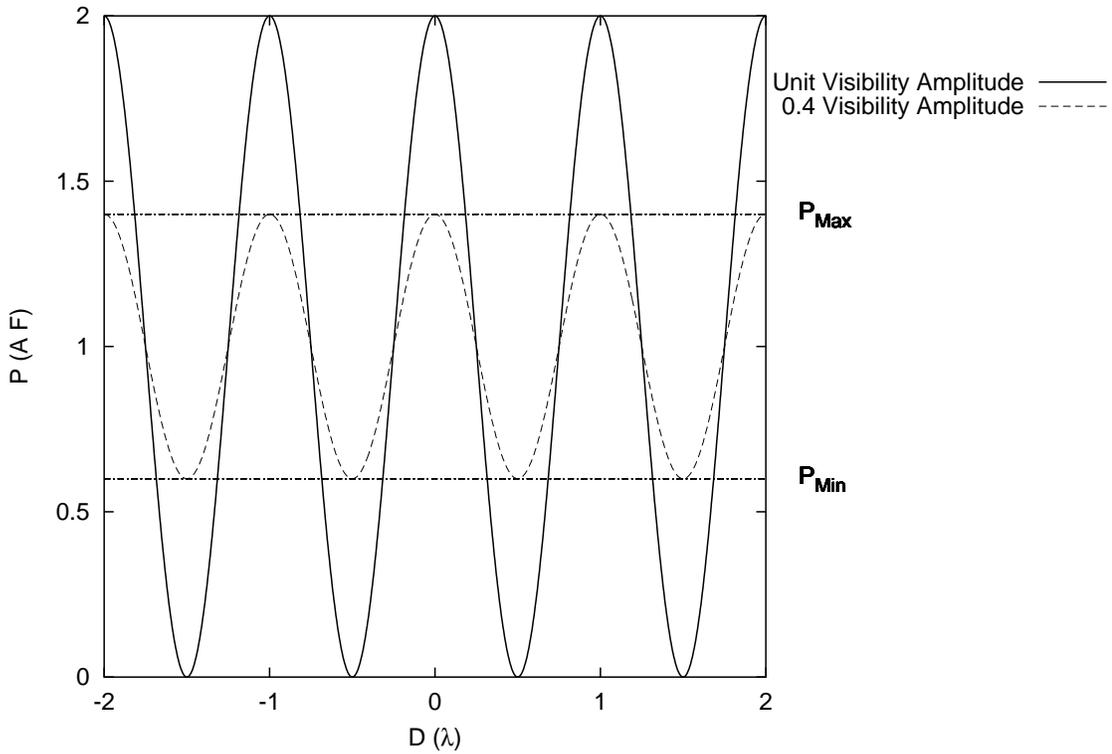


Figure 2.4: Normalized Visibility Amplitude (Michelson Visibility).

First, we remind the reader that as defined to this point the visibility is *dimensional*—it has dimensions of power (Equations 2.13 and 2.16). In fact, the form of Equation 2.14 makes it clear that the modulus or *amplitude* of the complex visibility describes the amount of power the interferometer measures in delay-space fringes. In the sense we used the word in the introduction, the visibility quantitatively captures the coherent response of the interferometer to the astronomical source. In the sense of language used in the development of the van Cittert–Zernike theorem, the visibility captures (one component of) the spatial coherence function of the astronomical source.

Optical interferometers typically measure the normalized fringe power, the fringe power *relative* to the total power collected from the source. For instance, in his classical studies on stellar diameters at Mt. Wilson, Michelson (Michelson, 1920; Michelson and Pease, 1921) defined the visibility of his fringes as the apparent contrast between light and dark areas (power P_{Max} and P_{Min} respectively) of fringes visible in his telescope eyepiece. This is quantified as the Michelson fringe visibility

$$\mathcal{V}_M \equiv \frac{P_{\text{Max}} - P_{\text{Min}}}{P_{\text{Max}} + P_{\text{Min}}},$$

which, of course, is dimensionless and contained in the interval $[0,1]$. Michelson’s construction is depicted in Figure 2.4. As we will demonstrate below, for sources that are unresolved to the interferometer the fringes oscillate with a peak-to-peak amplitude of the full received power ($2AF$). As the source increases in apparent size they become resolved by the in-

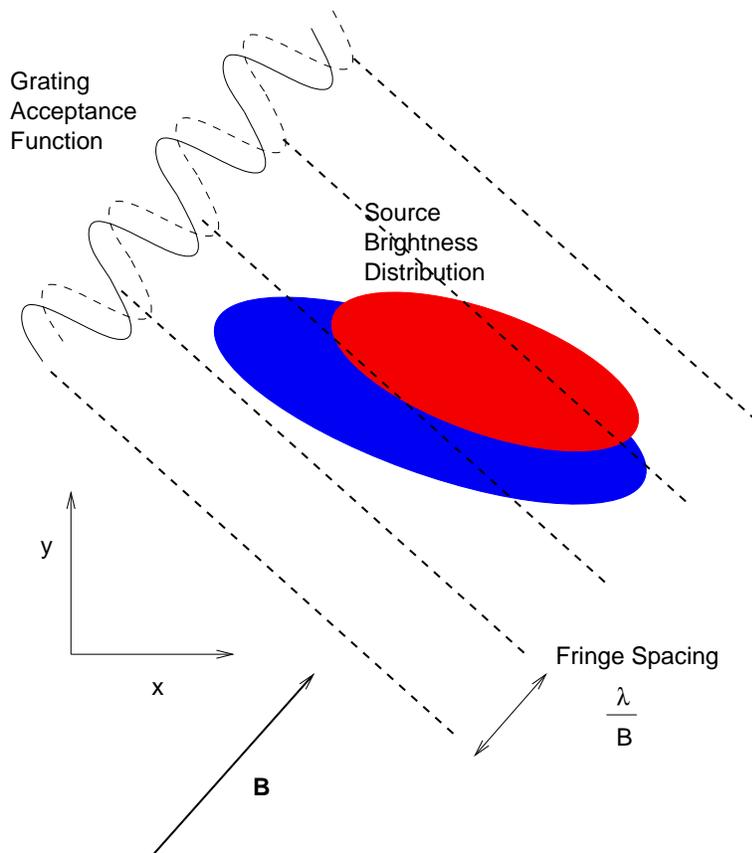


Figure 2.5: Interferometer Response to Extended Source.

terferometer and the fringes decrease in amplitude. We shall make these statements more quantitative in Section 2.7.

Following Michelson's example, we commonly work in a normalized, dimensionless visibility, given as an extension of Equation 2.13:

$$\mathcal{V}(k, \mathbf{B}) \equiv \frac{\int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s})e^{-ik\Delta\mathbf{s}\cdot\mathbf{B}}}{\int d\Omega A(\Delta\mathbf{s})F(\Delta\mathbf{s})} = \frac{V(k, \mathbf{B})}{P_0}, \quad (2.20)$$

in which case the detected power given by Equation 2.14 takes the form.

$$P = P_0 \left(1 + \text{Re}\{\mathcal{V}e^{ik\delta}\} \right). \quad (2.21)$$

Comparison of the Michelson visibility and Equation 2.21 makes it clear that $\mathcal{V}_M = |\mathcal{V}|$; the modulus of \mathcal{V} is similarly contained in the interval $[0,1]$.[†]

Figure 2.5 gives a depiction of how I think of the interferometric visibility. Given some arbitrary source intensity distribution on the sky, the instantaneous interferometer response/visibility is given by a sum of the received power from the source multiplied by

[†]This had to be true—the modulus of the exponential kernel in Equation 2.20 is contained in the interval $[0,1]$, and we normalize by the received power from the source.

a cosine grating acceptance function referenced to the instantaneous phase center. This grating function oscillates along the (projected) baseline direction with angular frequency λ/B , and is constant in the direction normal to the baseline (up to the angular extent defined by the collecting aperture acceptance function A). The grating function is further multiplied by a fringe coherence envelope along the baseline direction in the polychromatic case; this is not shown in the figure. A small change in the path difference between the two arms of the interferometer translates the phase tracking center and shifts the reference point of the grating acceptance function, with corresponding changes in the received fringe power. Both the baseline direction and angular frequency of the fringes in the grating acceptance function are conveniently captured in the spatial frequencies u and v .

In terms of predicting the response of the interferometer, rather than evaluating this grating-modulated source distribution over some continuum of phase centers, we find it quantitatively convenient to describe the morphology of the source in terms of a complex visibility that simultaneously captures both the even (cosine) and odd (sine) components of the source morphology relative to a fixed phase center. This is given not by the real, cosine grating function, but by a complex exponential grating function containing both even and odd components. When the interferometer is phased at the reference center it responds to the even (cosine) component, and when it is phased $\pm\pi/2$ radians away from the reference center it responds to the odd (sine) component. Between these two situations the interferometer sees an admixture of the even and odd components.

As a final remark for readers familiar with the mathematics of quantum mechanics, I have often found it constructive to think of the interferometer grating acceptance function as a particular basis vector in an Hilbert space, with the interferometer response in any given configuration given by a projection of the source brightness distribution onto the particular Hilbert basis vector. Like quantum mechanics, a full description of a general source morphology requires both even and odd components in the space. In this analogy the act of image synthesis reduces to estimating the properties of the source morphology having measured some (finite) set of these Hilbert components, and employing some *a priori* knowledge of the source morphology (e.g. positivity, bandwidth limitations, etc.). Perhaps someday I'll develop this analogy with quantum mechanics further in the context of a monograph.

2.7 Visibility of Common Source Morphologies

To close our discussion it is interesting and constructive to consider the visibilities of some of the more common source morphologies. It serves to codify application of the visibility formalism developed above, and we can discuss some of the general properties of interferometric visibility in the context of applications. Further, for optical interferometry these developments are instructive and necessary to interpret observations. We will, characteristically, be dealing with some idealized examples amenable to analytical treatment (cf.

Michelson 1890). However, these offer significant insight into the strengths and weaknesses of the interferometer as an imaging instrument.

2.7.1 Point Source

Many sources are sufficiently distant and isolated that they may be considered point-like, or at least approximately so. To close the loop on our original calculations as well as providing a point of departure for multiple sources systems it is interesting to treat the point source within the visibility framework.

The brightness distribution of a point source at source coordinates (α_0, β_0) relative to the phase reference is simple to write down in terms of Dirac delta functions:

$$F_0 \delta(\alpha - \alpha_0) \delta(\beta - \beta_0).$$

The total collected power from such a source is:

$$P_0 = \int d\alpha d\beta A(\alpha, \beta) F_0 \delta(\alpha - \alpha_0) \delta(\beta - \beta_0) = A(\alpha_0, \beta_0) F_0.$$

The complex visibility of such a point source is computed after Equation 2.16 as:

$$\begin{aligned} V(u, v) &= \int d\alpha d\beta A(\alpha, \beta) F_0 \delta(\alpha - \alpha_0) \delta(\beta - \beta_0) e^{-2\pi i(\alpha u + \beta v)} \\ &= A(\alpha_0, \beta_0) F_0 e^{-2\pi i(\alpha_0 u + \beta_0 v)} = P_0 e^{-2\pi i(\alpha_0 u + \beta_0 v)} \end{aligned} \quad (2.22)$$

the total received power times a phase. Of course, the normalized visibility for the point source is trivially:

$$\mathcal{V} = e^{-2\pi i(\alpha_0 u + \beta_0 v)}, \quad (2.23)$$

a pure phase—of course. Note that for the on-axis (on-reference) source α_0 and β_0 are zero, and the normalized visibility is unity. Of course, the normalized visibility amplitude (modulus) of the point source is always unity.

To compute the interferometer detected power for this source we can insert the complex visibility, Equation 2.22 (normalized visibility, Equation 2.23) into the detected power equation, Equation 2.14 (Equation 2.21), yielding:

$$\begin{aligned} P &= P_0 \left(1 + \text{Re}\{e^{-2\pi i(\alpha_0 u + \beta_0 v)} e^{ik\delta}\} \right) \\ &= P_0 \left(1 + \text{Re}\{e^{-ik\mathbf{\Delta s} \cdot \mathbf{B}} e^{ik\delta}\} \right) \\ &= P_0 (1 + \cos k(\mathbf{\Delta s} \cdot \mathbf{B})), \end{aligned} \quad (2.24)$$

where we have identified (α_0, β_0) with $\mathbf{\Delta s}$, and set $\delta = 0$ in the last equality to demonstrate consistency with Equation 2.10.

2.7.2 Uniform Disk

Longer interferometric baselines offer unprecedented angular resolution—so much resolution in fact that sources conventionally taken as point sources become resolved. One important class of such objects is nearby stars; resolving and measuring the angular diameters of stars is one of the bread-and-butter science topics for optical interferometers (Michelson and Pease, 1921; Hanbury Brown *et al.*, 1974; Mozurkewich *et al.*, 1991; van Belle *et al.*, 1999).

A reasonable approximation to the brightness distribution of a resolved star is the model of a uniform disk. (Interesting physics arises from considering deviations from the uniform disk model; Quirrenbach *et al.* (1996); Hajian *et al.* (1998) discuss stellar limb darkening as measured by optical interferometers.) We can write the model for an axisymmetric disk in terms of polar coordinates as:

$$F(\rho) = F_0(\rho < \theta/2),$$

ρ being an angular offset on the celestial sphere away from the nominal center of the source, and θ being the diameter of the source. In such a model the total power collected from the source is trivially $P_0 = A_0 F_0 \pi \theta^2 / 4$ —this in fact defines the value of the surface brightness F_0 (dimensions of incident power per square angle on the sky per collecting area). After Equation 2.16 (and assuming the angular extent of the disk is much smaller than the angular size of the aperture response function), the complex visibility of the disk at source position (α_0, β_0) is:

$$V(u, v) = e^{-2\pi i(\alpha_0 u + \beta_0 v)} \int d\alpha d\beta A_0 F e^{-2\pi i(u\alpha + v\beta)},$$

where we have taken advantage of the phase property of the Fourier transform under coordinate translations to arrange convenient integration variables; we saw an example of this phase property for the point source in Equation 2.22.

Fourier Transform of the Axisymmetric Function

To compute the visibility for the disk we need to consider the two-dimensional Fourier transform of an axisymmetric function. Taking $f = f(\rho)$, we wish to evaluate

$$F(u, v) = \int d\alpha d\beta f(\rho) e^{-2\pi i(u\alpha + v\beta)} = \int d\rho d\theta \rho f(\rho) e^{-2\pi i\rho(u \cos \theta + v \sin \theta)},$$

with $\alpha = \rho \cos \theta$ and $\beta = \rho \sin \theta$. It is convenient to drop u and v in favor of some angular spatial frequency variables:

$$u \equiv v_r \cos \phi, \quad v \equiv v_r \sin \phi,$$

v_r is a radial spatial frequency; like u and v it has dimensions of fringe cycles per radian on the sky. In this transformation F becomes:

$$F(v_r, \phi) = \int d\rho d\theta \rho f(\rho) e^{-2\pi i\rho v_r (\cos \theta \cos \phi + \sin \theta \sin \phi)} = \int d\rho d\theta \rho f(\rho) e^{-2\pi i\rho v_r \cos(\theta - \phi)}.$$

F is by construction axially symmetric, so we are free to take $\phi = 0$ without loss of generality. The θ -integral can now be performed, as:

$$\int_0^{2\pi} d\theta e^{ix\cos\theta} = 2\pi J_0(x),$$

J_0 being the zeroth-order Bessel function of the first kind. This allows us to finally write:

$$F(v_r) = 2\pi \int_0^\infty d\rho \rho f(\rho) J_0(-2\pi\rho v_r) = 2\pi \int_0^\infty d\rho \rho f(\rho) J_0(2\pi\rho v_r) \quad (2.25)$$

as even-ordered Bessel functions are even functions and odd-ordered Bessel functions are odd functions. Equation 2.25 is the general form of a two-dimensional Fourier transform of an axially symmetric function; we have used the axial symmetry to trade the two-dimensional transform for a one-dimensional transform with a different (slightly more complicated) transform kernel. This transform is commonly known as a Hankel (or Fourier-Bessel) transform.

Back to the uniform disk, after Equation 2.25 the visibility of the disk is evidently:

$$\begin{aligned} V(v_r) &= e^{-2\pi i(\alpha_0 u + \beta_0 v)} 2\pi \int_0^{\theta/2} d\rho \rho F_0 J_0(2\pi\rho v_r) \\ &= e^{-2\pi i(\alpha_0 u + \beta_0 v)} \frac{8P_0}{\theta^2} \int_0^{\theta/2} d\rho \rho J_0(2\pi\rho v_r). \end{aligned}$$

This is straightforwardly evaluated from:

$$\int_0^x dx' x' J_0(x') = x J_1(x).$$

Then:

$$\begin{aligned} V(v_r) &= e^{-2\pi i(\alpha_0 u + \beta_0 v)} \frac{8P_0}{\theta^2} \int_0^{x=2\pi v_r \theta/2} \frac{dx x}{(2\pi v_r)^2} J_0(x) \\ &= e^{-2\pi i(\alpha_0 u + \beta_0 v)} 2P_0 \frac{J_1(\pi v_r \theta)}{\pi v_r \theta}. \end{aligned} \quad (2.26)$$

Recall v_r is a radial spatial frequency ($v_r^2 = u^2 + v^2 = B_\perp^2/\lambda^2$), making the visibility:

$$V(B_\perp, \lambda, \theta) = e^{-2\pi i(\alpha_0 u + \beta_0 v)} 2P_0 \frac{J_1(\pi\theta B_\perp/\lambda)}{\pi\theta B_\perp/\lambda}. \quad (2.27)$$

Trivially the normalized visibility is given by:

$$\mathcal{V}(B_\perp, \lambda, \theta) = e^{-2\pi i(\alpha_0 u + \beta_0 v)} \frac{2J_1(\pi\theta B_\perp/\lambda)}{\pi\theta B_\perp/\lambda}. \quad (2.28)$$

Because of noise properties, optical interferometers typically measure squared normalized visibility. Trivially the squared normalized visibility amplitude (modulus) for the uniform disk is:

$$\mathcal{V}^2(B_\perp, \lambda, \theta) = \left(\frac{2J_1(\pi\theta B_\perp/\lambda)}{\pi\theta B_\perp/\lambda} \right)^2. \quad (2.29)$$

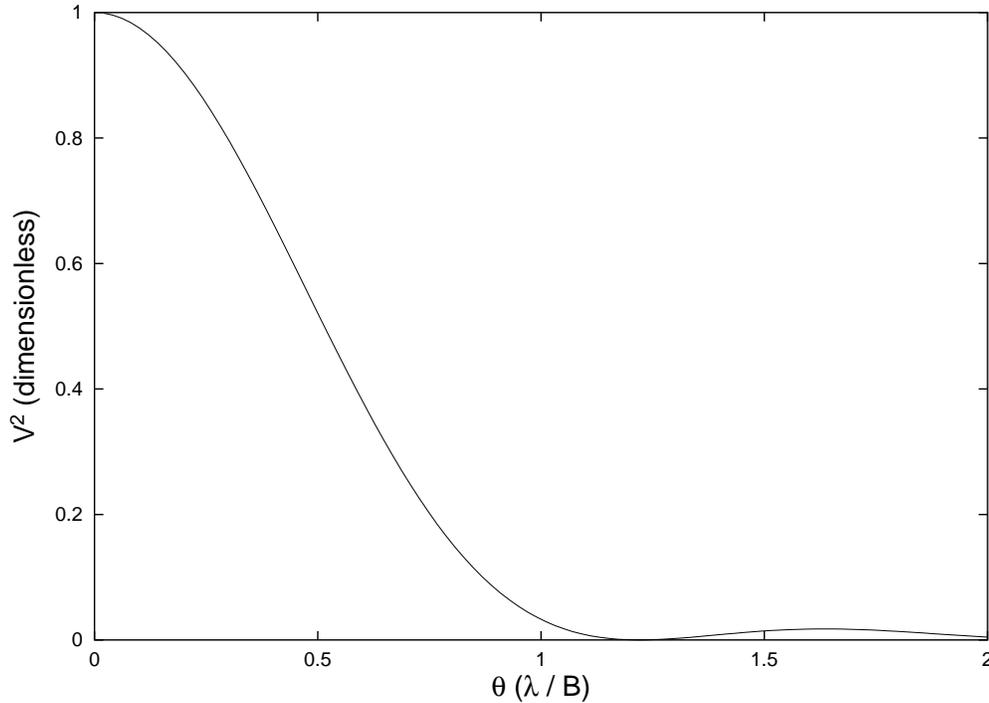


Figure 2.6: Squared Normalized Visibility Amplitude for the Uniform Disk. The disk diameter θ is plotted in units of the interferometer fringe spacing λ / B_{\perp} .

Figure 2.6 gives a plot of \mathcal{V}^2 as a function of disk diameter θ (in units of the fringe spacing λ / B_{\perp}). In the limit that the disk is much smaller than the fringe spacing $\mathcal{V}^2 \approx 1$, the disk is unresolved by the interferometer, and the visibility reduces to the results for the unresolved point source developed above. As the disk becomes an appreciable fraction of the fringe spacing the visibility \mathcal{V} becomes less than 1; in general terms we speak of a source being *resolved* by the interferometer when the normalized visibility amplitude (or \mathcal{V}^2) is measurably less than 1. The visibility actually goes to zero—fringes disappear—for the disk at a diameter of $\theta \approx 1.22 \lambda / B_{\perp}$. After this first null the fringes reappear, but at very low amplitude.

2.7.3 Multiple Stellar Systems

Nature often forms stars in multiple systems, and we are therefore motivated to consider multiple stellar systems.

We can consider the interferometer response to a multiple system as a collection of quasi-uniform stellar disks that lie in the aperture acceptance pattern. Labeling the parameters of the j th source with a subscript, up to aperture efficiency factors, the total received power from the system is simply the sum of the powers from the visible sources:

$$P_0 = \sum_j P_j.$$

Because of the linear properties of the Fourier transform (physically, because the light from the individual sources is incoherent), we can compose the system's complex visibility as the sum of the complex visibilities of the constituent disks at source positions (α_j, β_j) (Equation 2.27):

$$V = \sum_j V_j = \sum_j P_j \frac{2J_1(\pi\theta_j B_\perp/\lambda)}{\pi\theta_0 B_\perp/\lambda} e^{-2\pi i(u\alpha_j + v\beta_j)},$$

straightforwardly making the normalized visibility:

$$\mathcal{V} = \frac{\sum_j P_j \frac{2J_1(\pi\theta_j B_\perp/\lambda)}{\pi\theta_0 B_\perp/\lambda} e^{-2\pi i(u\alpha_j + v\beta_j)}}{\sum_j P_j} = \frac{\sum_j P_j \mathcal{V}_j}{\sum_j P_j}, \quad (2.30)$$

with \mathcal{V}_j given by Equation 2.28 in a uniform disk model.

Equation 2.30 doesn't offer much physical insight, therefore a concrete example is in order. Binary stars are a traditional and important target of optical interferometers (Michelson, 1920; Herbison-Evans *et al.*, 1971; Hummel *et al.*, 1995; Hummel *et al.*, 1998; Boden *et al.*, 1999), with the visibility acting as a proxy for the relative astrometry between the two components. Straightforward application of Equation 2.30 to a two component binary system yields:

$$\begin{aligned} \mathcal{V}_{\text{binary}} &= \frac{P_1 \mathcal{V}_1 + P_2 \mathcal{V}_2}{P_1 + P_2} \\ &= e^{-2\pi i(u\alpha_1 + v\beta_1)} \frac{|\mathcal{V}_1| + r |\mathcal{V}_2| e^{-2\pi i(u\Delta\alpha + v\Delta\beta)}}{1 + r}, \end{aligned} \quad (2.31)$$

having defined $r \equiv P_2/P_1$, and relative source coordinates $\Delta\alpha \equiv \alpha_2 - \alpha_1$ and $\Delta\beta \equiv \beta_2 - \beta_1$. When the observable is $\mathcal{V}_{\text{binary}}^2$, this is given straightforwardly by the squared modulus of Equation 2.31:

$$\begin{aligned} \mathcal{V}_{\text{binary}}^2 &= \mathcal{V}_{\text{binary}}^* \mathcal{V}_{\text{binary}} \\ &= \frac{\mathcal{V}_1^2 + r^2 \mathcal{V}_2^2 + 2r |\mathcal{V}_1| |\mathcal{V}_2| \cos(2\pi(u\Delta\alpha + v\Delta\beta))}{(1 + r)^2} \\ &= \frac{\mathcal{V}_1^2 + r^2 \mathcal{V}_2^2 + 2r |\mathcal{V}_1| |\mathcal{V}_2| \cos(2\pi \mathbf{B} \cdot \mathbf{s}_{\text{binary}}/\lambda)}{(1 + r)^2}, \end{aligned} \quad (2.32)$$

with $\mathbf{s}_{\text{binary}} \equiv (\Delta\alpha, \Delta\beta)$. Note that there are corrections due to finite bandwidth effects when $\mathbf{B} \cdot \mathbf{s}_{\text{binary}}$ is more than a few fringe spacings; the exact form of these corrections can depend on details of the fringe measurement process.

In the limit of point-like, equal-amplitude components $\mathcal{V}_{\text{binary}}^2$ reduces to:

$$\mathcal{V}_{\text{binary}}^2 \rightarrow \frac{1 + \cos(2\pi \mathbf{B} \cdot \mathbf{s}_{\text{binary}}/\lambda)}{2}.$$

Clearly as $\mathbf{s}_{\text{binary}} \rightarrow 0$ the two components of the binary system are unresolved by the interferometer. With increasing $\mathbf{s}_{\text{binary}}$, as $\mathbf{B} \cdot \mathbf{s}_{\text{binary}} \rightarrow \lambda/4$, $\mathcal{V}_{\text{binary}}^2 \rightarrow 1/2$, and the binary

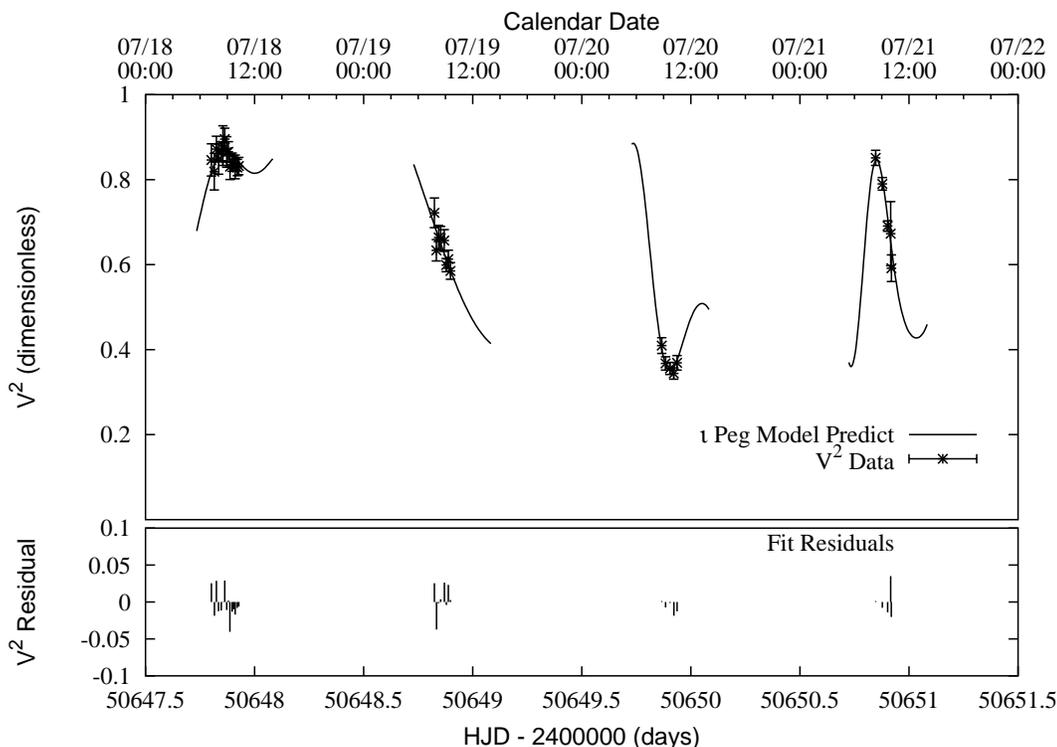


Figure 2.7: Squared Normalized Visibility Amplitude on the binary star ι Pegasi (HD 210027). Palomar Testbed Interferometer near-infrared \mathcal{V}^2 measurements of ι Peg are shown from four consecutive nights in July 1997. A model based on Equation 2.32 (and incorporating finite bandwidth effects) is fit to the \mathcal{V}^2 measurements to derive an orbit model for ι Peg (from Boden *et al.* 1999).

system becomes “resolved” by the interferometer. It is noteworthy that this happens at an order-of-magnitude similar separation as the resolution of two point sources by a filled aperture telescope; $s_{\text{binary}} = \lambda/B$ for a Rayleigh resolution criterion (cf. Jenkins and White 1957).

Even when the binary star is quasi-static, Equation 2.32 describes sinusoidal variations of the fringe visibility (squared modulus) with varying $\mathbf{B} \cdot \mathbf{s}_{\text{binary}}$; for ground based interferometers this variation occurs as a consequence of Earth rotation. Figure 2.7 depicts real \mathcal{V}^2 measurements on a binary star ι Pegasi, used to derive an orbit model for the system (Boden *et al.*, 1999).

2.8 Summary

In this Chapter we have discussed the response of interferometers to idealized astronomical sources. We have developed this theory in the context of typical detection strategies for optical interferometers: direct (*homodyne*) combination of the optical fields and detection by power-linear detectors. This is to be contrasted to the typical heterodyne, amplitude-linear

detection technologies used in radio interferometers (see Thompson, Moran, and Swenson 1986). Despite the technology differences between between radio and optical interferometers, a common characterization of source properties, namely the source visibility, suffices to give a qualitative and quantitative description of the interferometer response.

Visibility is a complex quantity whose amplitude (modulus) describes the intensity of the interferometric fringes, and whose phase describes the position of the fringes relative to a phase center. Sources that produce fringes (in the space of relative delay between the two interferometer arms) with an amplitude equal to the full received power of the source are said to have unit normalized visibility amplitude and are unresolved by the interferometer. Conversely, fringes with amplitudes less than the received power have normalized visibility amplitude less than one, and sources that produce such fringes are said to be resolved by the interferometer.

In ordinary circumstances the source visibility can be computed as a simple Fourier transform of the source brightness morphology, and an inverse Fourier transform of the source visibility function yields the source morphology. Techniques based on this relationship are given the term *synthesis imaging*, and have been employed for many years in radio interferometry (Perley *et al.*, 1989). Recently optical interferometers have begun making their first few forays into synthesis imaging (e.g. Baldwin *et al.* 1996). However, many optical interferometers today are limited to measurements of (squared) visibility amplitude, with no (useful) phase information available. Given our knowledge of expected source visibilities for different morphologies (e.g. uniform stellar disk, binary star), even such visibility amplitude measurements can be used to infer interesting properties of astronomical sources.

Acknowledgments

The work described in this manuscript was performed at the Infrared Processing and Analysis Center, California Institute of Technology, and the Jet Propulsion Laboratory under contract with the National Aeronautics and Space Administration. Thanks go to P.R. Lawson and M.M. Colavita for discussing various parts of the manuscript.

References

- J.E. Baldwin, M.G. Beckett, R.C. Boyesen, D. Burns, D.F. Buscher, G.C. Cox, C.A. Haniff, C.D. Mackay, N.S. Nightingale, J. Rogers, P.A.G. Sheuer, T.R. Scott, P.G. Tuthill, P.J. Warner, D.M.A. Wilson, and R.W. Wilson, “The first images from an optical aperture synthesis array: mapping of Capella with COAST at two epochs,” *Astron. Astrophys.* **306**, L13–L16 (1996).
- G.T. van Belle, B.F. Lane, R.R. Thompson, A.F. Boden, M.M. Colavita, P.J. Dumont, D.W. Mobjley, D. Palmer, M. Shao, G. Vasisht, J.K. Wallace, M.J. Creech-Eakman, C.D. Koresko, S.R. Kulkarni, X.P. Pan, and J. Gubler, “Radii and effective temperatures for G, K and M giants and supergiants,” *Astron. J.* **117**, 521–533 (1999).

- A.F. Boden, C.D. Koresko, G.T. van Belle, M.M. Colavita, P.J. Dumont, J. Gubler, S.R. Kulkarni, B.F. Lane, D. Mobley, M. Shao, and J.K. Wallace, “The visual orbit of Iota Pegasi,” *Astrophys. J.* **515**, 356–364 (1999).
- M. Born and E. Wolf, *Principles of Optics*, 7 (expanded) edn. (Cambridge, UK: Cambridge University Press, 1999).
- R. Hanbury Brown, J. Davis, and L.R. Allen, “The angular diameters of 32 stars,” *Mon. Not. R. Astron. Soc.* **167**, 121–136 (1974).
- A.R. Hajian, J.T. Armstrong, C.A. Hummel, J.A. Benson, D. Mozurkewich, T.A. Pauls, D.J. Hutter, N.M. Elias, K.J. Johnston, L.J. Rickard, and N.M. White, “Direct confirmation of stellar limb darkening with the Navy Prototype Optical Interferometer,” *Astrophys. J.* **496**, 484–489 (1998).
- D. Herbison-Evans, R. Hanbury Brown, J. Davis, and L.R. Allen, “A study of α Virginis with an intensity interferometer,” *Mon. Not. R. Astron. Soc.* **151**, 161–176 (1971).
- C.A. Hummel, J.T. Armstrong, D.F. Buscher, D. Mozurkewich, A. Quirrenbach, and M. Vivekanand, “Orbits of small angular scale binaries resolved with the Mark III Interferometer,” *Astron. J.* **110**, 376–390 (1995).
- C.A. Hummel, D. Mozurkewich, J.T. Armstrong, A.R. Hajian, N.M. Elias, and D.J. Hutter, “Navy Prototype Optical Interferometer observations of the double stars Mizar A and Matar,” *Astron. J.* **116**, 2536–2548 (1998).
- J.D. Jackson, *Classical Electrodynamics*, 3 edn. (New York: John Wiley and Sons, 1998).
- F. Jenkins and H. White, *Fundamentals of Optics* (New York: McGraw-Hill, 1957).
- A.A. Michelson, “On the application of interference methods to astronomical measurements,” *Phil. Mag.* **30**, 1–21 (1890).
- A.A. Michelson, “On the application of interference methods to astronomical measurements,” *Astrophys. J.* **51**, 257–262 (1920).
- A.A. Michelson and F.G. Pease, “Measurement of the diameter of α Orionis with the interferometer,” *Astrophys. J.* **53**, 249–259 (1921).
- D. Mozurkewich, K.J. Johnston, R.S. Simon, P.F. Bowers, R. Gaume, D.J. Hutter, M.M. Colavita, M. Shao, and X.P. Pan, “Angular diameter measurements of stars,” *Astron. J.* **101**, 2207–2219 (1991).
- R.A. Perley, F.R. Schwab, and A.H. Bridle, *Synthesis Imaging in Radio Astronomy* ASP Conf. Ser., **6** (Provo, UT: Brigham Young Univ. Press, 1989).
- A. Quirrenbach, D. Mozurkewich, D.F. Buscher, C.A. Hummel, and J.T. Armstrong, “Angular diameter and limb darkening of Arcturus,” *Astron. Astrophys.* **312**, 160–166 (1996).
- M. Shao, M.M. Colavita, B.E. Hines, J.L. Hershey, J.A. Hughes, D.J. Hutter, G.H. Kaplan, K.J. Johnston, D. Mozurkewich, R.S. Simon, and X.P. Pan, “Wide-angle astrometry with the Mark III interferometer,” *Astron. J.* **100**, 1701–1711 (1990).
- A.R. Thompson, J.M. Moran, and G.W. Swenson Jr., *Interferometry and Synthesis in Radio Astronomy* (New York: John Wiley and Sons, 1986).

Chapter 3

Beam Combination and Fringe Measurement

WESLEY A. TRAUB

SMITHSONIAN ASTROPHYSICAL OBSERVATORY
CAMBRIDGE, MASSACHUSETTS

This chapter discusses how to calculate the interference of wavefronts for telescopes and interferometers, how to separate astrophysical effects from instrumental effects, and how to optimize visibility measurements.

We focus on homodyne detection, a technique in which stellar wavefront segments are combined with each other in real time to make an image or a fringe pattern, whose intensity is then detected.

The alternative is heterodyne detection, a technique in which each stellar wavefront segment is combined with a local oscillator signal to make a beat frequency which is recorded, and at a later time a collection of simultaneously recorded beat signals is combined to form an image or a fringe pattern. The heterodyne technique and its applicability to wavelengths of about $10\ \mu\text{m}$ and longer is discussed by Townes in Chapter 4.

However, despite their different detection techniques, homodyne and heterodyne interferometry still share identical basic principles, and much of the material in this chapter applies equally to both.

3.1 Elements of Beam Combination

In this section we present simple examples of beam combination, for one- and two-aperture configurations and for three basic types of sources. We distinguish multiple apertures (this section) from multiple telescopes (the following section). From the discussion it should be clear how to generalize each result to more realistic or complex situations. In simple cases it is relatively straightforward to see how to invert observed data to infer source properties, but in general the van Cittert–Zernike theorem is needed, as discussed at the end of this section.

3.1.1 Single Aperture and Point Source

It is instructive to start thinking about interferometers by first considering the case of a single telescope. In fact, an ideal single telescope is also an ideal interferometer. If it were possible to build a single telescope with a diameter as large as the baseline of an interferometer, we would never even consider building an interferometer, since interferometers are intrinsically more complex than telescopes. Also, many of the technical aspects of interferometers derive directly from the need to duplicate the action of a single large mirror, so it is useful to understand how a single collecting element works.

The simplest possible case is that of a plane wave incident on a single aperture, as shown in Figure 3.1. The collimating element is drawn as a lens, but any equivalent optical system, such as a Cassegrain telescope, will do. The aperture is stopped to a diameter D by a baffle in the (x, y) plane. The symmetry axis (z) in Figure 3.1 is a line from the center of the star through the optical center of the telescope lens.

Throughout this chapter we will assume that the optical systems are completely free of geometrical optical aberrations. For a useful book on telescope optics see Schroeder (2000).

The incident light from a real star is a stream of photons arriving at random times from a range of random angles within the angular diameter of the star. A single photon from this stream effectively exists over an arbitrarily large area on the surface of a sphere centered on its emitting atom, prior to detection. The photon simultaneously senses the presence of all the details of the collecting aperture, which can be of arbitrary shape, size, and degree of topological connectedness.

It helps at this point to forget about photons and think instead about waves. It also helps to think about a wave as a series of little spherical waves which repeatedly generate themselves at all points across the wavefront, and propagate outwards, but which in free space only end up propagating in the forward direction because this is the only direction in which the little waves constructively interfere. These are called Huygens’ wavelets in optics texts, where the term “wavelet” means little wave. It does not mean a wave packet in the modern mathematical sense of the word “wavelet”!

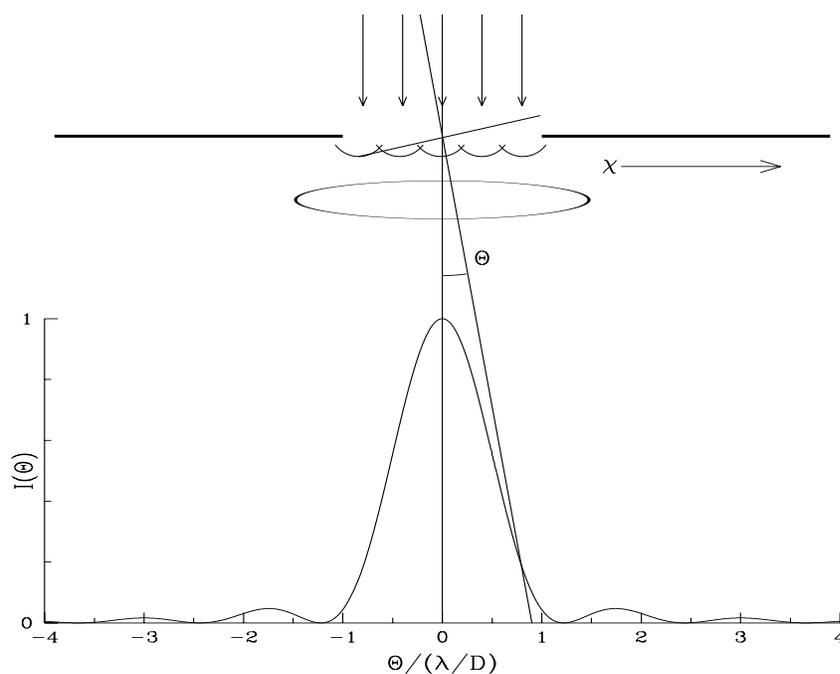


Figure 3.1: Delta-function star and single telescope. Input wavefronts arrive at the aperture from angle $\theta = 0$ only, but output wavefronts exist at all angles, albeit concentrated near the input angle. The intensity pattern plotted here is the circular aperture diffraction pattern $I_{\text{tel}}(\theta) = [2J_1(\pi\theta D/\lambda)/(\pi\theta D/\lambda)]^2$.

The incident idealized photon is also monochromatic, and therefore has essentially infinite extent in the direction of propagation. The corresponding classical wave has the same extent. If we think of an emitting atom at the surface of the star where the photon originated as a classical oscillator, then we may define a wavefront from this oscillator as the collection of all points on the outwardly propagating wave which were generated at the same time. There are therefore an infinite series of concentric wavefronts being emitted over time, and we may choose any one of them to consider as the wave moves through our optical system. I find it useful to think of the wavefront as the surface on which the wave has its maximum positive electric field strength. Since the wave is periodic, successive wavefronts are separated by one wavelength.

As the wavefront crosses the plane of the entrance pupil, the wavelets in the center of the pupil continue as before, propagating a plane wavefront. However at the edge of the pupil there is no longer a reinforcement of wavelets from the part of the incident wavefront that is now blocked, and the wavelets inside the pupil will start to spread transversely into the geometric optics shadow region.

For an alternative mental model of wavefront propagation, one could also think of row after row of people marching shoulder to shoulder in a huge parade, with the people in

the middle being kept to the forward direction by their neighbor's shoulders. When the marching rows encounter a wide gate open in a long wall, the people in the center are still constrained to march forward by their neighbors, but the few folks nearest the gate edges will feel unconstrained on one side, and without guidance their paths will begin to divert from that of the main group.

Actually the row-of-people analogy is not quite as far fetched as it sounds, because in principle each of the marching people has a de Broglie wavelength given by $\lambda = h/mc$, so to the extent that these people are mindless particles with no other influences on them, they will behave like diffracting particles. More practically, a beam of electrons, neutrons, protons, atoms, or molecules will follow the same rules, as has been demonstrated many times.

In astronomical telescopes we are only interested in what happens to the wavelets at large distances (compared to a wavelength) from the diffracting edges, i.e., in the far-field diffraction pattern. Fortunately for us, this is relatively simple to calculate, given that each wavelet has a known amplitude and phase, and given that we have a rule for combining electric field amplitudes and another rule for converting the total amplitude to an intensity, which is the measurable quantity of interest.

Another measurable quantity is the state of polarization of the electric field of the detected photon. In general we will ignore polarization because the net effect of polarization on diffraction is most often a second-order effect. There is one important exception, however, in the case of reflection from mirrors, which we will discuss later in this chapter.

In Figure 3.1, we show an aperture followed by a lens and a focal plane. According to geometric optics, the star will be focussed in the focal plane, and the image will be a perfect replica of the star, here a delta-function. However according to physical optics we must find the image by adding up the amplitudes of the wavelets which make it through the aperture and propagate in the direction of any point in the focal plane.

Recall that an ideal lens merely acts to convert an input direction of propagation into an output location in the focal plane. A pinhole does the same thing, but a lens is better in the sense that it converts an incident ray anywhere on its surface into a ray heading for a single position in the focal plane. It is also helpful to recall that the ray through the center of the lens is not deviated, so input direction equals output direction. Thus in Figure 3.1 we show intensity in the focal plane in terms of an output angle θ . This is equivalent to position in the focal plane, divided by the focal length of the imaging optics.

At a given point θ in the focal plane, the total amplitude is the sum of all the wavelet amplitudes at the aperture heading in the direction θ , allowing for their relative phases. From Maxwell's equations we know that the electric field amplitude is sinusoidal in space and time, so that in one dimension and for one polarization, the electric field amplitude E can be written conveniently as the real part of the complex amplitude $A(z, t) = e^{i(\omega t - kz + \phi)}$. We use conventional notation where z is linear distance in the direction of propagation, t

is time, $k = 2\pi/\lambda = nk_0 = 2\pi n/\lambda_0$ is the magnitude of the wave vector (or propagation vector), n is the index of refraction, nz is the “optical path,” $\lambda = \lambda_0/n = c/n\nu$ is the wavelength in the medium, ν is the temporal frequency of oscillation, $\omega = 2\pi\nu$ is the angular frequency, and ϕ is the phase of the wave (Born and Wolf, 1999).

Let us define the position of measurement of the amplitude to be $z = 0$, the time of measurement to be $t = 0$, and the medium of propagation to be vacuum, so $n = 1$. This reduces the complex amplitude to $e^{i\phi}$.

At first sight it may appear that we have defined away all the interesting physics, but the seemingly ridiculous simplicity of $e^{i\phi}$ is in reality the heart of the problem of calculating interference effects. We have merely stripped away the non-essential parts.

From a quantum-mechanical point of view, the $e^{i\phi}$ term is a propagator of a probability amplitude from one place and time to another, where ϕ represents the change in phase (modulo 2π) of the probability amplitude along the minimum path.

The phase is calculated across a tilted surface in the pupil, oriented at an angle θ with respect to the incoming wavefront. There are an infinite number of such tilted surfaces. The relative strength of an outgoing wavefront parallel to one of these surfaces is determined by adding up all the wavelets on that surface. The phase at each point is 2π times the distance between the input and output wavefronts, in units of wavelength. Let us focus on the aperture’s x -dimension for the moment. The phase of a wavelet is

$$\phi(z) = 2\pi x \sin(\theta)/\lambda \quad (3.1)$$

$$\simeq 2\pi x\theta/\lambda \quad (3.2)$$

where $x \sin(\theta)$ is the distance between the incoming wavefront from direction $\theta = 0$ and the outgoing wavefront at angle θ , and we assume $\theta \ll 1$.

The net output amplitude from the telescope in the direction θ is $A_{\text{tel}}(\theta)$, which we calculate as the algebraic sum of all wavelets across the pupil.

$$A_{\text{tel}}(\theta) = \sum(\text{wavelets}) \quad (3.3)$$

$$= \int_{\text{pupil}} e^{i\phi(x)} dx \quad (3.4)$$

$$= \int_{-D/2}^{+D/2} e^{i(2\pi x\theta/\lambda)} dx \quad (3.5)$$

$$= \frac{\lambda}{2\pi i\theta} \left[e^{+i\pi\theta D/\lambda} - e^{-i\pi\theta D/\lambda} \right] \quad (3.6)$$

$$= \frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda} D \quad (3.7)$$

The measured intensity I is the squared magnitude of the amplitude.

$$I_{\text{tel}}(\theta) = |A_{\text{tel}}|^2 \quad (3.8)$$

$$I_{\text{tel}}(\theta) = \left[\frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda} \right]^2 D^2 \quad (3.9)$$

The intensity pattern is thus a $\text{sinc}(X) \equiv \sin(X)/X$ function, with a strong central peak and small secondary peaks, as shown in Figure 3.1. The first zero is the solution of $I_{\text{tel}}(\theta_{\text{tel}}) = 0$ and is given by

$$\theta_{\text{tel}} = \lambda/D. \quad (3.10)$$

The corresponding result for a two-dimensional circular aperture is found by replacing \int_D by \int_{circle} with the result

$$I_{\text{tel}}(\theta) = \left[\frac{2J_1(\pi\theta D/\lambda)}{\pi\theta D/\lambda} \right]^2 D^2 \quad (3.11)$$

where $J_1(X)$ is the Bessel function of first order, roughly similar to a damped sine function. Numerically $J_1(X)$ for any real X can be calculated using the BESSJ1 routine in Numerical Recipes (Press *et al.*, 1992). The first zero-intensity angle is the solution of $I_{\text{tel}}(\theta_{\text{tel}}) = 0$ and is given by $X = 1.22\pi$ or

$$\theta_{\text{tel}} = 1.22\lambda/D \quad (3.12)$$

which is the famous relation for an unobstructed circular aperture. The full-width at half-maximum (FWHM) of the intensity pattern is roughly approximated by the value of θ_{tel} , so this value is often loosely referred to as the diameter of the diffraction-limited image.

Here are several useful variations on the same theme.

Constant Phase

Suppose we add a constant phase ϕ_0 across the aperture. Then we get

$$A_{\text{tel}}(\theta) = \int_{-D/2}^{+D/2} e^{i(2\pi x\theta/\lambda + \phi_0)} dx \quad (3.13)$$

$$= \frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda} e^{i\phi_0} D \quad (3.14)$$

and I_{tel} is unchanged.

Star Off-Axis

Suppose that the star moves off the telescope axis, or equivalently that the telescope is pointed away from the star by an angle θ_0 . The input wavefronts are then tilted by θ_0 , and the summing of phases, determined by the distance from the input and output wavefronts, yields

$$A_{\text{tel}}(\theta) = \int_{-D/2}^{+D/2} e^{i(2\pi x(\theta - \theta_0)/\lambda)} dx \quad (3.15)$$

$$= \frac{\sin(\pi(\theta - \theta_0)D/\lambda)}{\pi(\theta - \theta_0)D/\lambda} D \quad (3.16)$$

and we see that the intensity pattern will be shifted to a new center at position θ_0 in the focal plane, just as one would expect from geometric optics.

Phase Step: Speckle

Suppose we add a phase step of π across half of the aperture. This simulates the action of the turbulent atmosphere in a very simple case. The net amplitude with this phase step is

$$A_{\text{tel}}(\theta) = \int_0^{+D/2} e^{i(2\pi x\theta/\lambda + \pi/2)} dx + \int_{-D/2}^0 e^{i(2\pi x\theta/\lambda - \pi/2)} dx \quad (3.17)$$

$$= -\frac{\sin^2(\pi\theta D/2\lambda)}{\pi\theta D/2\lambda} D \quad (3.18)$$

and the intensity is

$$I_{\text{tel}}(\theta) = \left[\frac{\sin^2(\pi\theta D/2\lambda)}{\pi\theta D/2\lambda} \right]^2 D^2. \quad (3.19)$$

This pattern has two main peaks offset from the axis by about $\pm\lambda/D$, and with widths about λ/D , plus small secondary peaks. The effect of the phase step is to split the unperturbed image into two pieces, each of which looks rather similar to the original image. The perturbed images are called *speckles*. If more phase steps are added, more speckles will appear, but each will still be a more or less faithful copy of the diffraction-limited unperturbed case. This is the basis for speckle interferometry. One of the simplest techniques to recover the original image from a speckle pattern is called “shift and add,” which seeks to superpose the speckles on a common axis, post detection. From the example here, one can see why this technique enjoys some success.

Central Obscuration

Suppose that the telescope aperture has an outer width D and a central obscuration of inner width d . In the one-dimensional case, the electric field amplitude in the focal plane is then

$$A_{\text{tel}}(\theta) = \int_{-D/2}^{-d/2} e^{i(2\pi x\theta/\lambda)} dx + \int_{+d/2}^{+D/2} e^{i(2\pi x\theta/\lambda)} dx \quad (3.20)$$

$$= \frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda} D - \frac{\sin(\pi\theta d/\lambda)}{\pi\theta d/\lambda} d \quad (3.21)$$

and the light intensity is given by

$$I_{\text{tel}}(\theta) = \left[\frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda} D - \frac{\sin(\pi\theta d/\lambda)}{\pi\theta d/\lambda} d \right]^2. \quad (3.22)$$

By analogy, the corresponding expression for the intensity from a two-dimensional circular aperture of diameter D , with central obscuration (e.g., a secondary mirror) of diameter d is given by

$$I_{\text{tel}}(\theta) = \left[\frac{2J_1(\pi\theta D/\lambda)}{\pi\theta D/\lambda} D - \frac{2J_1(\pi\theta d/\lambda)}{\pi\theta d/\lambda} d \right]^2. \quad (3.23)$$

Comparing this with the non-obscured case, we see that for a telescope primary mirror of a given diameter D and a finite secondary diameter d , the intensity distribution has a slightly narrower central core (i.e., slightly better angular resolution) but at the expense of significantly stronger diffraction rings around the central core. The reason that the core is narrower is that the inner portion of the aperture, i.e., the low angular resolution part, has been removed, leaving the outer portions which are responsible for the high angular resolution performance. The sidelobes increase for the reason that, at angles outside the central core, the missing central portion means that there are fewer central wavelets available to provide phase cancellation with the edge wavelets. Thus more of the edge wavelet power appears in an aliased form outside the central core.

Some additional interesting possibilities that could be calculated include tapering the transmission factor of the pupil near the edges so as to *apodise* (“remove the feet”), or reduce, the secondary diffraction rings, or adding a phase screen across the aperture in order to delay the phase near the outer edges and thereby alter the phase of the diffraction rings.

With these examples worked out, we can now graduate quickly to several other key cases of interest to interferometry.

3.1.2 Two Apertures and Point Source

The two-aperture calculation proceeds as with the one-aperture case, but with the single opening replaced by two openings of equal diameter D , separated by a baseline B , as shown in Figure 3.2. The configuration is that of a pointed interferometer, and can be achieved by masking the primary mirror of a large telescope as indicated in the Figure; this is nominally what Michelson did at the 100-inch telescope, and it also describes the original MMT mirrors, and the current LBT. In a very large interferometer, which is too large to mount on a single pointed platform, one can literally join the focal planes of two or more separate telescopes, in which case additional issues of path length, magnification, field rotation, and polarization must be considered.

For two apertures the amplitude A_{int} of the electric field in the focal plane of a two-element interferometer is given by

$$A_{\text{int}}(\theta) = \sum(\text{wavelets}) \quad (3.24)$$

$$= \int_{\text{pupil}} e^{i\phi(x)} dx \quad (3.25)$$

$$= \int_{+B/2-D/2}^{+B/2+D/2} e^{i(2\pi x\theta/\lambda)} dx + \int_{-B/2-D/2}^{-B/2+D/2} e^{i(2\pi x\theta/\lambda)} dx \quad (3.26)$$

$$= \frac{\sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda} \cos(\pi\theta B/\lambda) 2D \quad (3.27)$$

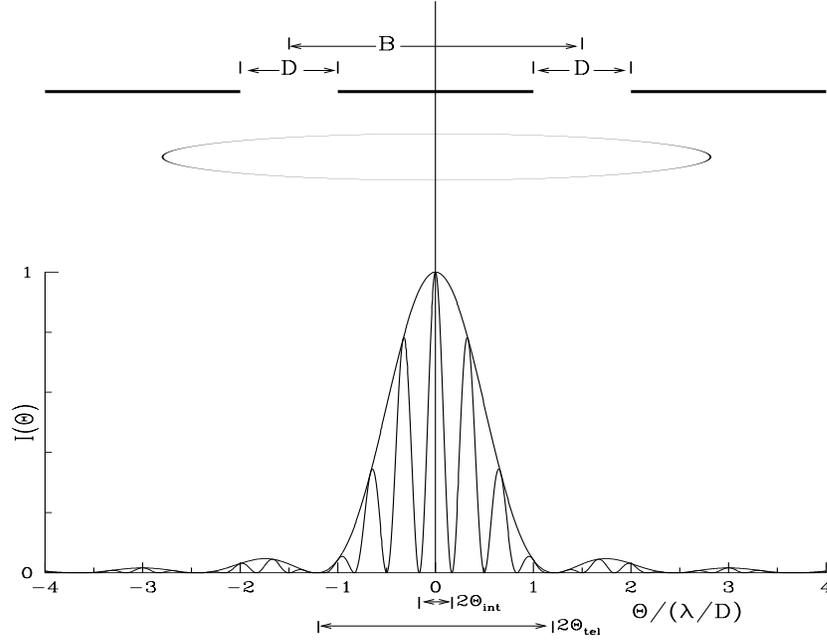


Figure 3.2: Two apertures and a single delta-function source. The two apertures form a pair of superposed diffraction-limited images, crossed by interference fringes. The figure is drawn for the case $B/D = 3$, so the expected number of fringes in the packet is $N_{\text{packet}} = 7.3$, which is about what is seen here.

and the intensity is given by

$$I_{\text{int}}(\theta) = 2I_{\text{tel}}(\theta) [1 + \cos(2\pi\theta B/\lambda)] \quad (3.28)$$

which is the product of two terms, the broad envelope of a single-telescope diffraction pattern, and the rapidly varying interference term which depends only on the distance between the telescopes, as shown in Figure 3.2. Clearly the I_{tel} term can be either the one-dimensional one shown here, or a two-dimensional J_1 function derived above.

The first zero of the intensity pattern is the solution of $I_{\text{int}}(\theta_{\text{int}}) = 0$ and is given by

$$\theta_{\text{int}} = \lambda/2B \quad (3.29)$$

which is also the width (FWHM) of one of the narrow fringes, and therefore the angular resolution limit of the interferometer.

Let us define a *fringe packet* as the central lobe of the telescope diffraction pattern. The envelope of the fringe packet has an angular width of $2\theta_{\text{tel}}$ between first zeros. The number of fringes N_{packet} in a fringe packet is given by

$$N_{\text{packet}} = 2.44B/D. \quad (3.30)$$

3.1.3 Two Apertures and Binary Star

Suppose we have a binary system comprising two equal-magnitude stars separated by θ_{bin} , centered on the axis of a two-aperture interferometer. The amplitudes and intensities from each star must be treated independently, because the sources are not coherent (photons from one star have no knowledge of photons from the other star). In the focal plane we then have two independent interferometer intensity patterns which add to give a binary-star intensity

$$I_{\text{bin}} = I_{\text{int}}(\theta - \theta_{\text{bin}}/2) + I_{\text{int}}(\theta + \theta_{\text{bin}}/2) \quad (3.31)$$

If the separation is small compared to the width of the fringe packet, i.e., $\theta_{\text{bin}} \ll 2\theta_{\text{tel}}$, then we can factor out the envelope shape and find

$$I_{\text{bin}} \simeq 2I_{\text{tel}}(\theta) [1 + \cos(2\pi(\theta + \theta_{\text{bin}}/2)B/\lambda) + 1 + \cos(2\pi(\theta - \theta_{\text{bin}}/2)B/\lambda)] \quad (3.32)$$

$$= 4I_{\text{tel}}(\theta) [1 + V_{\text{bin}} \cos(2\pi\theta B/\lambda)]. \quad (3.33)$$

The coefficient of the interference modulation term is known as the *fringe visibility*, or simply the *visibility*, and is given in this case by

$$V_{\text{bin}} = \cos(\pi\theta_{\text{bin}}B/\lambda). \quad (3.34)$$

The visibility has its first zero $V_{\text{bin}} = 0$ when the binary separation is

$$\theta_{\text{bin}} = \lambda/2B. \quad (3.35)$$

The generalization of these results to unequal magnitudes and a two-dimensional configuration is straightforward but messy.

3.1.4 Two Apertures and Uniform Disk

A real star has a finite diameter, and each of the photons emitted from its surface is independent of all other photons (unless there is maser activity taking place, as does happen in the atmospheres of some stars, under appropriate conditions). As with the binary star case, the intensity in the focal plane of a telescope or interferometer is then given by the superposition of appropriately shifted and scaled intensity patterns. For a uniformly bright disk in one-dimension (UD1) whose width is θ_{UD1} , and where we assume that the disk is small compared to the fringe packet width ($\theta_{\text{UD1}} \ll 2\theta_{\text{tel}}$), we add up the incoherent fringe patterns as follows.

$$I_{\text{UD1}}(\theta) = \sum_{\text{disk}} (\text{intensities}) \quad (3.36)$$

$$= \int_{\text{disk}} I_{\text{int}}(\theta - \theta_x) d\theta_x \quad (3.37)$$

$$= \int_{-\theta_{\text{UD1}}/2}^{+\theta_{\text{UD1}}/2} 2I_{\text{tel}}(\theta - \theta_x) [1 + \cos(2\pi(\theta - \theta_x)B/\lambda)] d\theta_x \quad (3.38)$$

$$\simeq 2I_{\text{tel}}(\theta) [1 + V_{\text{UD1}} \cos(2\pi\theta B/\lambda)] \quad (3.39)$$

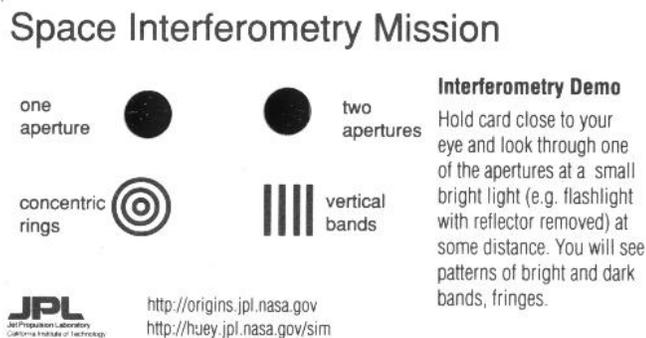


Figure 3.3: The Space Interferometry Mission (SIM) pocket interferometer card. The apertures on the actual card are a single circular hole (.) in a piece of black film on the “one aperture” side, and two holes (..) on the “two apertures” side. With one aperture, a point source will appear as concentric rings, as shown, but with the outer rings progressively fainter than the center one. With two apertures, the central bright peak is crossed by strong vertical modulation bands, as is the secondary ring, and (faintly) the third ring.

Here the fringe visibility of a uniform disk in one-dimension is given by

$$V_{\text{UD1}} = \frac{\sin(\pi B \theta_{\text{UD1}} / \lambda)}{\pi B \theta_{\text{UD1}} / \lambda}. \quad (3.40)$$

By analogy we immediately see that the intensity pattern for a two-dimensional round uniform disk (UD) is given by a similar equation where the visibility is

$$V_{\text{UD}} = \frac{2J_1(\pi B \theta_{\text{UD}} / \lambda)}{\pi B \theta_{\text{UD}} / \lambda}. \quad (3.41)$$

The visibility has its first zero $V_{\text{UD}} = 0$ when the star diameter is

$$\theta_{\text{UD}} = 1.22\lambda / 2B. \quad (3.42)$$

Note that $2J_1(X)$ is similar to $\sin(X)$ in that the central lobe is positive, the first secondary lobe is negative, the second secondary lobe is positive, and so on. In the fringe pattern, this means that the fringes in the alternate lobes have their signs inverted with respect to the extrapolated fringes from neighboring lobes. In other words, counting the central lobe as number 0, the even lobes have phase = 0, and the odd lobes have phase = π .

3.1.5 The Pocket Interferometer

The SIM pocket demonstration card shown in Figure 3.3 is an excellent one-dimensional, two-aperture, mask, which when held close to your eye provides a complete interferometer. Here the telescope is your eye lens, and the focal plane is your retina. The diameter of each aperture is $D \simeq 0.07$ mm, so the telescope diffraction pattern has a width $\theta_{\text{tel}} = 1.22\lambda / D \simeq$

2000 arcsec, which is about the angular diameter of the sun or moon. The separation of the apertures is $B \simeq 0.25$ mm, so the angular width of a fringe with this interferometer is $\theta_{\text{int}} = \lambda/2B \simeq 200$ arcsec, which is about the width of a Mag-LiteTM filament at a distance of roughly a foot (note however that the filament is much longer than it is wide). The number of fringes in a fringe packet (the angular width of the single-telescope diffraction pattern) is $N_{\text{packet}} = 2.44B/D \simeq 8$ in this case, independent of the light source, of course.

The Mag-Lite should be used with the normally-present flashlight reflector removed, so that you see just the filament itself. The Mag-Lite will appear to be a small, unresolved star if it is viewed at arm's length or farther, with the filament rotated (i.e., the flashlight rolled around your line of sight) so that its narrow dimension is parallel to the baseline \vec{B} of the pocket interferometer, thus making a one-dimensional system. In this case you will see a central bright lobe crossed by about eight fringes, oriented perpendicular to the baseline. You will also see the same straight fringes crossing the first and second side lobes of the single-telescope diffraction pattern.

The Mag-Lite can be made into a large-diameter star by rolling it 90° about the line of sight, so that the long axis of the filament is now parallel to \vec{B} . In this case you will see a central lobe which is smooth, with no fringes. What has happened is that multiple fringe patterns are now superposed with a range of shifts, and the fringe pattern is washed out, as predicted by the equations above.

3.1.6 Two-Aperture Beam Pattern on Sky

If you were to think like a radio astronomer, you would imagine the antenna pattern to be projected out from the receiver horn of each antenna and thence from the array as a whole and onto the sky. As you move the antenna, or change the phase at an array element, the pattern sweeps across the sky. The received signal is the convolution of the moving pattern and the sources in the sky. A sinusoidal pattern projects out the Fourier component of that spacing of fringes on the object, and therefore the Fourier component of the intensity distribution across the sky.

You can see from this view that if all possible fringe spacings and orientations could be swept across the object, and if their phases (i.e., the relative locations of the central fringe peaks) could be recorded, then these measured quantities would essentially fill the two-dimensional Fourier plane with complex values (an amplitude and phase pair). A Fourier transformation of these values would then yield a perfect image of the source. This is not only a good mental picture, but it is also the basis of the van Cittert–Zernike theorem in the following section.

3.1.7 van Cittert–Zernike Theorem

The famous van Cittert–Zernike theorem was developed from the work of van Cittert in 1934 and Zernike in 1938. This theorem (Born and Wolf, 1999) formalizes the heuristic discussion in the preceding section, and it is the basis of any attempt to reconstruct an object from interferometer measurements.

Suppose that two apertures are separated by baseline vector \vec{B} . Suppose that the source has an intensity distribution on the sky given by $I(\vec{\alpha})$, where $\vec{\alpha}$ is a two-dimensional sky coordinate. Then the *complex degree of coherence* μ is given by

$$\mu(\vec{B}) = \int I(\vec{\alpha})e^{-ik\vec{B}\cdot\vec{\alpha}}d\vec{\alpha} / \int I(\vec{\alpha})d\vec{\alpha} \quad (3.43)$$

where $k = 2\pi/\lambda$, and the integrals are over the field of view of the diffraction-limited single-aperture beam, a cone of half-angle θ_{tel} . The *degree of coherence* is the modulus

$$\text{visibility} = |\mu| \quad (3.44)$$

and the phase is the argument

$$\text{phase} = \arg(\mu). \quad (3.45)$$

The inverse relation

$$I(\vec{\alpha}) / \int I(\vec{\alpha})d\vec{\alpha} = \int \mu(\vec{B})e^{+ik\vec{B}\cdot\vec{\alpha}}d\vec{B} \quad (3.46)$$

recovers the image from the suite of visibility measurements. In this equation, the integral on the right is over all possible baseline positions.

3.2 Beam Combination in Practice

This section addresses some practical aspects of beam combination, including Michelson’s pioneering stellar interferometer, optical configurations for large ground-based interferometers, and multiplexing methods.

3.2.1 Michelson’s Stellar Interferometer

An interesting aspect of Michelson’s original stellar interferometer is that although it was mounted on a conventional telescope structure, the actual wavefront collecting mirrors were not part of the telescope optics, but rather were a pair of 45° flats riding on an external rail so that the baseline length B could be adjusted to be up to several times larger than the telescope primary itself. The beam combination itself used a smaller baseline B_0 which had used the telescope primary to bring together the separate beams and form superposed images of the star in the focal plane, as shown in Figure 3.4.

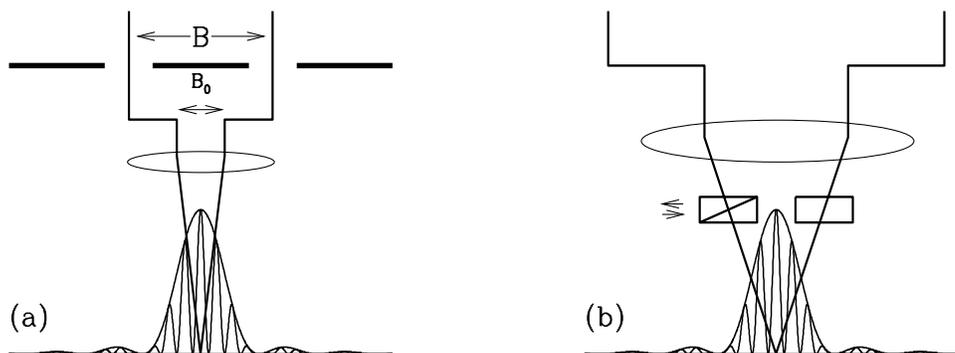


Figure 3.4: (a) Michelson interferometer schematic, showing the external collection baseline B , and the internal combination baseline B_0 . The fringe spacing is determined by B_0 . (b) Detail schematic of variable-thickness wedge (left) and tilt plate (right) used by Michelson to equalize the optical paths at all wavelengths and to precisely superpose two images of the target star, respectively.

Thus the coherence of the star is measured by the collecting baseline B , so this value governs the visibility of the source. The modulation pattern in the focal plane is set by the combining baseline B_0 . This is a sufficiently important distinction that we write it out:

$$B = \text{collection baseline} \quad (3.47)$$

$$B_0 = \text{combination baseline} \quad (3.48)$$

So, with this distinction in mind, and without further derivation, we write the intensity in the focal plane of the interferometer as

$$I_{\text{int}}(\theta) = 2I_{\text{tel}}(\theta) [1 + V_{\text{UD}} \cos(2\pi\theta B_0/\lambda)] \quad (3.49)$$

where the B_0 -dependent cosine term expresses the modulation of the envelope I_{tel} , and where the B -dependent V_{UD} term expresses the degree of modulation

$$V_{\text{UD}} = \frac{2J_1(\pi B\theta_{\text{UD}}/\lambda)}{\pi B\theta_{\text{UD}}/\lambda}. \quad (3.50)$$

So B_0 can be made to be any convenient value. Michelson used $B_0 = 1.14$ m, so the fringe width is $\theta_{\text{int}} = \lambda/2B_0 = 0.045$ arcsec. Let us assume that Michelson's viewing eye had an angular resolution $\theta_{\text{eye}} = 1.22\lambda/(5 \text{ mm}) \simeq 25$ arcsec. To resolve the fringes with his eye he would therefore need additional angular magnification M from the telescope eyepiece, where nominally $M = \theta_{\text{eye}}/\theta_{\text{int}} \simeq 25/0.045 \simeq 600$ times, and, in fact, that is what he reported.

Two further details are worth noting. First, Michelson inserted a plane-parallel plate of glass in one of the beams, within arm's reach, and tilted it in order to precisely superpose

the two star images, thereby effectively making the wavefronts parallel at the entrance pupil, and compensating for small alignment errors in the relay flats and bending of the structure. (Remember, position in the focal plane corresponds to angle at the incoming wavefront.)

Second, Michelson inserted two opposing glass wedges into the other beam, and slid these past each other so as to give a variable thickness of glass, and thereby compensate for the variable effective thickness of the tilt plate. This compensation ensures that wavefronts from different wavelengths arrive at the same time from both beams, and ensures that the broad band of wavelengths produces sharp fringes across the intensity envelope.

3.2.2 Image-Plane and Pupil-Plane Combination

There are two fundamentally different types of beam combination at the back end of an interferometer, and all ground- and space-based interferometers use one or the other of these methods. Deciding which one to use depends to some extent on the personal style of the designer. In principle, with an ideal instrument (noiseless detector, etc.), the ultimate signal-to-noise ratios from both methods should be identical.

Image-Plane Interferometry

Image-plane interferometry is the method of combining two beams in which each beam is focussed to make an image of the sky, and the images are superposed, so that interference fringes will form across the combined image. This is also called *Fizeau interferometry*, after Fizeau who originally suggested using a two-slit mask across the aperture of a conventional telescope to resolve stellar diameters. It is also the method used by Michelson in his stellar interferometer, as described above.

A generalized image-plane interferometer, for ground-based observations with long baselines, is sketched in Figure 3.5(a). Let ζ be the angle between the baseline vector \vec{B} and the stellar wavefront above the atmosphere. Then a stellar wavefront arrives at one telescope with an external vacuum path difference $z_{\text{ext.}} = B \sin(\zeta)$ compared to its arrival at the other telescope. To compensate, a delay line is introduced into one arm of the interferometer, giving an internal vacuum delay $z_{\text{int.}}$. The phase difference between the two beams is then $\phi = 2\pi\Delta z/\lambda$, where $\Delta z = (z_{\text{ext.}} - z_{\text{int.}})$ is the optical path difference (OPD) between the wavefronts, and the delay line is continuously adjusted to keep this quantity close to zero, as the Earth rotates.

The fringe intensity is displayed as a function of angular position θ in the focal plane, and is given by

$$I_{\text{int}}(\theta) = 2I_{\text{tel}}(\theta) [1 + V \cos(2\pi(\theta B_0 + \Delta z)/\lambda)] \quad (3.51)$$

where I_{tel} is the envelope shape, V is the visibility of the star, θB_0 is the fringe modulation

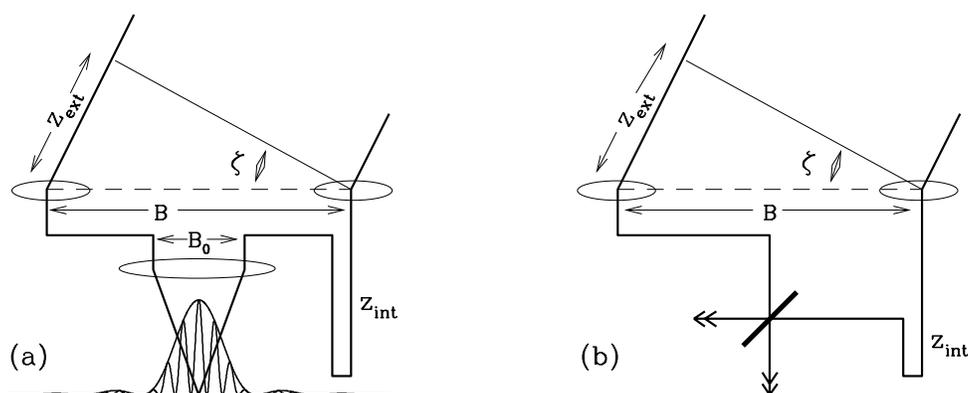


Figure 3.5: (a) Image-plane interferometer schematic, showing the external path difference $z_{ext.}$, internal path difference $z_{int.}$, and the non-zero beam-combination baseline B_0 . (b) Pupil-plane interferometer schematic, showing a half-silvered beam-combiner plate and zero distance between combining beams.

term and Δz is the fringe position term. Appropriate magnification and a multi-element detector are used to detect the fringes.

When the delay line is adjusted to give an OPD of $\Delta z = 0$, then the peak intensity of the fringe pattern is centered in the envelope. If the delay line is moved off of the zero OPD position, the envelope will stay fixed but the fringes will move across the envelope. If a finite bandwidth or multiple narrow wavelength bands are present, then at non-zero OPD values the fringe peaks from different wavelengths will be non-coincident, and for large values of OPD the fringes will blur out completely, as described in the section below, under “spectral bandpass.”

Spectral dispersion can be used with image-plane interferometry by introducing a prism or grating to disperse the light along the direction of the fringes. Since the fringe spacing is proportional to λ , the dispersed fringes will have a fan-like appearance with the red fringes at the wide end of the fan, and blue at the narrow end.

Pupil-Plane Interferometry

Pupil-plane interferometry is the method of combining two beams in which parallel beams are superposed, using a half-silvered mirror or equivalent, and the two resulting output beams are each focussed on single detector pixels. This is called *Michelson interferometry*, after the eponym’s original 1893 interferometer, which showed that the speed of light is independent of the observer’s velocity.

A ground-based pupil-plane interferometer is sketched in Figure 3.5(b). In pupil-plane interferometry the combining beams are completely overlapped, so the combining baseline is zero, $B_0 = 0$. The delay-line compensation is identical to that in the preceding case.

If the OPD in Figure 3.5(b) is adjusted to be zero, then by symmetry the overlapped beams emerging from either side of the beam splitter should have equal intensities, since each is the sum of one reflected and one transmitted beam. Another way to express this is to say that the beam splitter has the property that the phase difference between transmitted and reflected beams is exactly $\pi/2$, which we shall prove in the following section entitled “beam splitter phase shift,” but which we will simply accept for the moment.

The phase difference between the combined beams is then $\phi = 2\pi\Delta z/\lambda \pm \pi/2$, where $\Delta z \equiv z_{\text{ext.}} - z_{\text{int.}}$ is the OPD as before, and where the additional $\pm\pi/2$ is the beam splitter phase shift.

The fringe intensity follows a similar expression as above, but the $\pi/2$ term changes the $+\cos$ term to $\pm\sin$. Thus as the OPD is varied, the combined beam intensities will vary with opposite signs, giving

$$I_{\text{int}}(t) = 2I_{\text{tel}} [1 \pm V \sin(2\pi\Delta z(t)/\lambda)]. \quad (3.52)$$

Here we have integrated over angle in the focal plane so that $I_{\text{tel}} = \int I_{\text{tel}}(\theta)d\theta$. We have explicitly assumed that the phase difference is time-modulated, rather than being spatially modulated as in the image-plane case. If the time modulation is a triangle or ramp function, then $\Delta z(t) = vt$ over part of the modulation cycle. The measured amplitude of the time-modulated signal gives the visibility V directly.

Note that the $\pi/2$ term, or equivalently the sine dependence, is very frequently ignored by practitioners and textbooks alike, but it is nevertheless a salient feature of pupil-plane interferometry.

If a finite spectral bandwidth is present, then, just as in the image-plane case, if the OPD is adjusted to be non-zero, the fringe peaks of the different wavelengths will fail to overlap perfectly, and the fringes will blur out, as described in the “spectral bandpass” section below.

Spectral dispersion can be used with pupil-plane interferometry by adding a prism or grating just before detection, so that adjacent wavelengths fall on adjacent detector pixels. The resulting display is called a channel spectrum, because for non-zero path differences the spectrum will be wavelength-modulated by a sinusoidal intensity pattern (opposite in the two output beams) with the appearance of channels in an otherwise smooth spectrum. The depth of modulation at each wavelength gives the visibility directly.

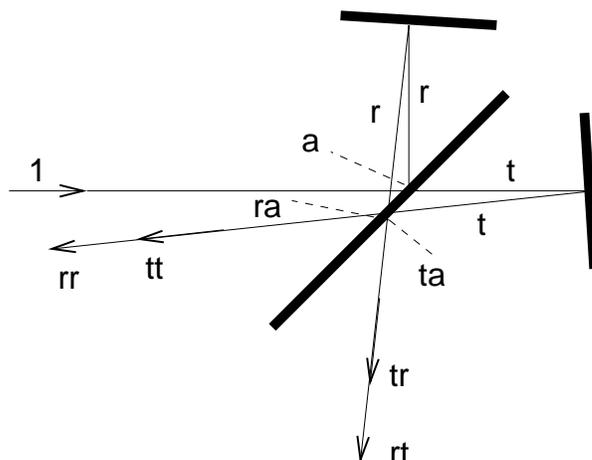


Figure 3.6: Beam splitter experiment showing a unit-intensity beam incident on a thin beam splitter, with two mirrors arranged so as to give equal length arms. The emerging beams are parallel and overlapping. Losses due to absorption or scattering are indicated.

3.2.3 Beam splitter Phase-Shift

In this section we prove the remarkable fact that the phase difference between reflected and transmitted beams from a beam splitter is $\pi/2$.

Suppose that we have a thin, symmetric beam splitter, such as a thin metal layer suspended in space or sandwiched between two identical sheets of glass. Suppose that the relative amplitude of a wavefront reflected from this beam splitter is r with phase shift δ_r , the relative amplitude of the transmitted wavefront is t with phase shift δ_t , and the relative amplitude absorbed or scattered is a .

Suppose that we set up a lab experiment as shown in Figure 3.6. The incident beam has amplitude $A_0 = 1$ from one side of the beam splitter, and zero from the other side.

The incident beam is split into a reflected complex amplitude $re^{i\delta_r}$, a transmitted complex amplitude $te^{i\delta_t}$, and an absorbed amplitude a . The corresponding relative intensities are reflectance $R = |r|^2$, transmittance $T = |t|^2$, and absorptance $A = |a|^2$. Each split beam is then reflected by a perfectly reflecting mirror and returned to the beam splitter with identical delay and phase shift on reflection in each arm; these terms will factor out, so we ignore their effect here to keep the equations uncluttered.

The returned beams then each split again as before, and are partially absorbed as well. The emerging amplitudes are then given by

$$A_1 = re^{i\delta_r}te^{i\delta_t} + te^{i\delta_t}re^{i\delta_r} \quad (3.53)$$

$$= 2rte^{i(\delta_r+\delta_t)} \quad (3.54)$$

$$A_2 = re^{i\delta_r}re^{i\delta_r} + te^{i\delta_t}te^{i\delta_t} \quad (3.55)$$

$$= r^2e^{i2\delta_r} + t^2e^{i2\delta_t} \quad (3.56)$$

The absorbed amplitudes are

$$A_3 = a \quad ; \quad A_4 = ra \quad ; \quad A_5 = ta. \quad (3.57)$$

The corresponding intensities of the incident, reflected, and absorbed beams are

$$I_0 = 1 \quad (3.58)$$

$$I_1 = 4RT \quad (3.59)$$

$$I_2 = R^2 + 2RT \cos^2(\delta_r - \delta_t) + T^2 \quad (3.60)$$

$$= R^2 - 2RT + T^2 + 4RT \cos^2(\delta_r - \delta_t) \quad (3.61)$$

$$I_3 + I_4 + I_5 = A + RA + TA \quad (3.62)$$

Conservation of energy requires that $I_0 = I_1 + I_2 + I_3 + I_4 + I_5$. Inserting the above values and simplifying, we find that the phase shifts are required to obey

$$\cos^2(\delta_r - \delta_t) = 0 \quad (3.63)$$

$$|\delta_r - \delta_t| = \pi/2 \quad (3.64)$$

Thus a thin beam splitter will have a $\pi/2$ phase shift between the reflected and transmitted beams, independent of the reflection, transmission, and absorption in the beam splitter. (I suspect, but have not shown, that this result still applies to finite-thickness non-absorbing beam splitters, but that the result fails for finite-thickness absorbing asymmetric beam splitters.)

By substituting $\delta_r = \delta_t \pm \pi/2$ back into the amplitude equations, it is easy to show that the output beams both have the same phase, $+2\delta_t$, which is interesting, but has no immediate application.

By repeating the entire derivation with unequal arm lengths it is also easy to see that the output beam intensities are complementary, i.e., that the intensities add to a constant value. This result does have great value, because it means that in a real pupil-plane interferometer, the sum of the output intensities can be used to normalize unavoidable intensity fluctuations due to atmospheric or other perturbations. Since these fluctuations often exceed photon-counting (Poisson) fluctuations on relatively bright stars, the technique of intensity normalization is a valuable tool for maximizing the observed signal-to-noise ratio.

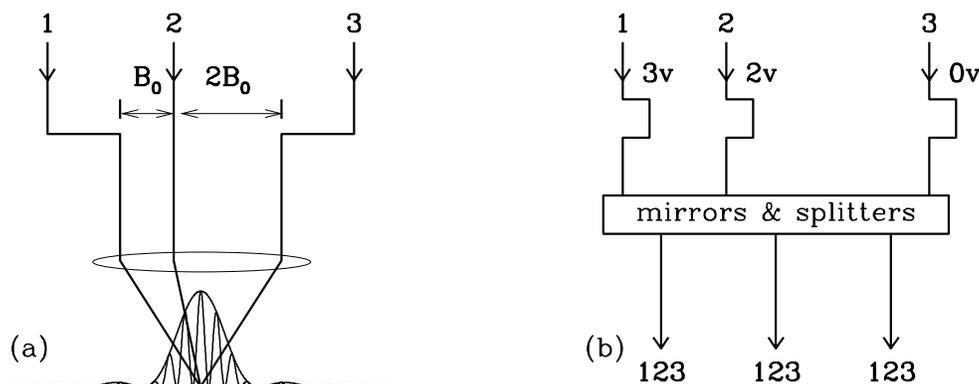


Figure 3.7: (a) A multiplexing scheme for a pupil-plane interferometer showing three combination baselines $1B_0, 2B_0, 3B_0$ for three beams. The input baselines B_{ij} can be arbitrary, and are independent of the output baselines. (b) A multiplexing scheme for an image-plane interferometer, showing three combination Doppler-shift velocities $3v, 2v, 0v$ for three beams. Here too, the input baselines can be arbitrary.

3.2.4 Multiplexing Three or More Apertures

For N apertures or telescopes, there are $N(N - 1)/2$ baselines among the apertures. The N beams can be combined in $N(N - 1)/2$ pairs, with each pair being detected as discussed above. However the N beams can also be combined all at once on a single detector pixel, if care is taken to encode each different pair of telescope beams with a different modulation frequency so that the component pairs can be extracted in post-processing.

There are two advantages to multiplexing. First, in the case where detector noise is greater than photon shot noise, it is advantageous to put the largest possible signal on the fewest possible detector elements. Second, in the case where phase-closure measurements are being made (which requires three or more telescopes), it is advantageous to have all the beams traversing the same paths as much as possible, to avoid unmeasured path changes in the optics due to temperature changes, etc., and this leads naturally to having all beams superposed.

A multiplexing image-plane interferometer can be made by arranging the output beam separations in a minimum redundancy array, so that the spatial frequencies in the image plane all have different values. This is illustrated in Figure 3.7(a) where the separations are B_0 and $2B_0$, so that the squared visibilities at each of the output spatial frequencies are in the proportions $f_{12} \sim 1$, $f_{23} \sim 2$, and $f_{31} \sim 3$, for example.

The power spectral density of the spatial intensity distribution is defined as

$$PSD = |FFT(\text{fringe pattern})|^2. \quad (3.65)$$

A plot of the PSD against spatial frequency will give distinct peaks at each of these frequencies, proportional to the image content at the corresponding external baselines B_{12} , B_{23} , and B_{31} . Alternatively a two-dimensional \vec{B}_0 pattern could be used, and a two-dimensional FFT extraction performed.

A multiplexing pupil-plane interferometer can be made by arranging the input beam delays to have different delay-line speeds, so that the Doppler shifts of the combined beams are each different. Figure 3.7(b) illustrates this with delay line velocities in the ratios $v_1 \sim 3$, $v_2 \sim 2$, $v_3 \sim 1$, so that the temporal PSD of any of the combined beams will contain distinct peaks at the output temporal frequencies in the ratios $f_{12} \sim 1$, $f_{23} \sim 2$, and $f_{31} \sim 3$, and the power in each peak is proportional to the visibility squared.

3.3 Visibility Loss Effects

High-quality measurements require that the observer minimize and calibrate the instrumental losses of visibility. Some of these effects can be minimized by proper design of the interferometer, and some by operation; all effects can be calibrated out of the data, in principle. A recent examination of instrumental sources of visibility loss for the IOTA interferometer, including more effects than listed here, is given by Porro *et al.* (1999).

The Strehl ratio S is defined as the ratio of (a) the measured peak intensity $I_{\text{meas}}(\text{max})$ of an image formed by a real optical system, including optical aberrations, and frequently including atmospheric seeing; and (b) the idealized peak intensity $I_{\text{ideal}}(\text{max})$ of an image formed by an ideal optical system, including only the effects of diffraction, and not including atmospheric seeing. Thus the Strehl ratio for images is

$$S = I_{\text{meas}}(\text{max})/I_{\text{ideal}}(\text{max}). \quad (3.66)$$

The Strehl ratio concept is also applicable to the fringe modulation in an interferogram, for either spatially or temporally displayed fringes. By analogy we write the Strehl ratio for fringes as

$$S = V_{\text{meas}}(\text{max})/V_{\text{ideal}}(\text{max}). \quad (3.67)$$

To estimate the combined effect of different sources of visibility, or of Strehl ratio, from the star and the instrument, the general practice is simply to multiply the various factors together, because we assume that they are all independent. Although this cannot be strictly valid, it is a very good approximation for small perturbations.

Atmospheric fluctuations can also cause visibility losses, and these can be more troublesome than instrumental losses because potentially they are larger in magnitude and variable in time; these effects are discussed by Quirrenbach in Chapter 5.

The results in this section will be stated without derivation; however, using the principles outlined above, the derivations could be supplied by the reader.

3.3.1 Spectral Bandpass

All stellar measurements use a finite range of wavelengths, or bandwidth. Any fringe packet, whether it is displayed spatially or temporally, will suffer a reduction in modulation amplitude at the edges of the packet, where the different wavelengths will produce opposing peaks and valleys. Suppose that the spectral bandpass is rectangular, and has a center and full-width at half-maximum (FWHM) of $(\lambda, \Delta\lambda)$ wavelengths, or $(\sigma, \Delta\sigma)$ wavenumbers, where $\sigma = 1/\lambda$ and $\Delta\sigma = \Delta\lambda/\lambda^2$ and where $\Delta\lambda/\lambda = \Delta\sigma/\sigma$.

In this case the visibility decreases with distance from the zero path-difference point according to

$$V_{\text{bandpass}}(\Delta z) = \frac{\sin(\pi\Delta z\Delta\sigma)}{\pi\Delta z\Delta\sigma} \quad (3.68)$$

$$= \frac{\sin(\pi\Delta z\Delta\lambda/\lambda^2)}{\pi\Delta z\Delta\lambda/\lambda^2}. \quad (3.69)$$

Here the path difference is given by $\Delta z = z_{\text{ext.}} - z_{\text{int.}}$ as discussed in the previous section.

The first zero of this function is at $\Delta z\Delta\sigma = 1$, from which we find that the number of fringes N_{bandpass} between envelope zero-crossings in a finite-bandpass wave packet is

$$N_{\text{bandpass}} = 2\lambda/\Delta\lambda. \quad (3.70)$$

In the general case there is a Fourier-transform relation between the bandpass shape and the fringe packet shape. This is illustrated in Figure 3.8, for the case of a real K-band filter, which is approximately rectangular. Note the sidelobe ringing, which results from aliased beating of the various wavelength fringe patterns outside the main lobe.

In Figure 3.8(a) we measure the filter width to be $\Delta\lambda \simeq 0.40 \mu\text{m}$, so in the fringe packet we expect $N_{\text{bandpass}} \simeq 11$ fringes, and this is in fact about what we see in corresponding fringe packet in Figure 3.8(b).

In a pupil-plane interferometer, as the delay line is scanned through the white-light point, the pattern in Figure 3.8(b) is exactly the observed modulation of intensity vs time. In the extreme case of no spectral filtering, the wave packet will tend towards a single spike delta-function. At the other extreme of a very narrow filter, the wave packet will be very many wavelengths wide.

3.3.2 Wavefront Tilt

If two wavefronts of width D are tilted by an angle α , then the interference pattern will be smeared and visibility reduced. The one-dimensional visibility factor from this effect is

$$V_{\text{tilt}} = \frac{\sin(\pi D\alpha/\lambda)}{\pi D\alpha/\lambda} \quad (3.71)$$

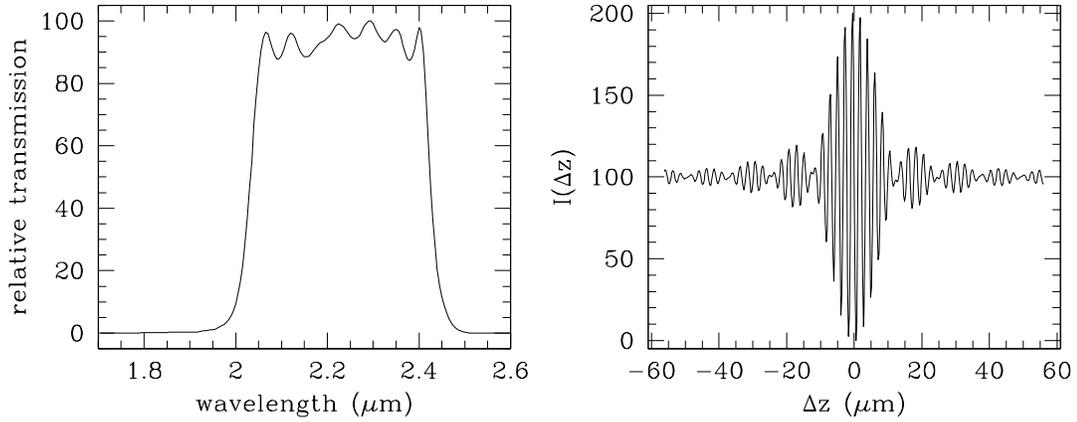


Figure 3.8: (a) Measured K filter transmission profile. (b) Calculated wave packet shape of a pupil-plane interferometer temporal scan through the zero-path-difference point, with the K filter shown. Note that the wave packet is sine-modulated, not cosine-modulated, as is appropriate for an ideal beam splitter.)

for a slit or rectangular aperture. The visibility factor from a two-dimensional circular aperture of diameter D is

$$V_{\text{tilt}} = \frac{2J_1(\pi D\alpha/\lambda)}{\pi D\alpha/\lambda}. \quad (3.72)$$

If you wish to have $V_{\text{tilt}} > 0.90$, say, then you need to be sure that the wavefronts combine at an angle $\alpha < 0.3\lambda/D$. For example, in a pupil- or image-plane interferometer, where in either case a star image will be formed, this amounts to a star-image-overlap criterion of about 25% of a diffraction-limited spot.

Likewise, a star tracker system, which directly controls wavefront tilt on a continuous basis, will have to perform at least as well, i.e., to 25% of the telescope's diffraction limit, to ensure that the measured visibilities do not fluctuate appreciably. The star-tracker system is thus a crucial part of an interferometer, and it can be a challenging task to achieve an optimum design.

3.3.3 Intensity Mismatch

If the relay optics fail to perfectly overlap the beams from each telescope, or if the beam combiner has unequal reflection and transmission factors, or if the combined beams come from different diameter telescopes and therefore have different intensities in the overlap region, then we will have a reduction in visibility from any of these factors. Let the intensity ratio between one beam and another be written as $I_1/I_2 = \rho$. The visibility factor from this effect is

$$V_{\text{mismatch}} = \frac{2}{\rho^{+1/2} + \rho^{-1/2}}. \quad (3.73)$$

This is a relatively tolerant effect. The reason is that amplitudes vary as the square root

of intensity, so two amplitudes will always be relatively closer in numerical value than the corresponding two intensities.

For example, if the beam combiner has an intensity reflection $R = 60\%$ and a transmission $T = 40\%$, then $\rho = 1.5$, so $V_{\text{mismatch}} \simeq 0.98$, and there is very little loss of visibility.

3.3.4 Optical Surface Figure Errors

If the combining wavefronts each have a root-mean-square (rms) perturbation of δ with respect to a perfect wavefront, and if the perturbations are randomly distributed across the wavefront, and uncorrelated between the two wavefronts, then the Strehl ratio and the fringe visibility will be degraded (Born and Wolf, 1999) according to

$$V_{\text{surfaces}} \simeq e^{-(2\pi\delta/\lambda)^2}. \quad (3.74)$$

If there are N surfaces with rms of δ_0 each, then

$$\delta \simeq N^{1/2}\delta_0. \quad (3.75)$$

For example, if we use a common optical polishing criterion which balances quality and cost, each flat mirror will have a peak-to-valley (pv) surface flatness of $\lambda_0/20$, where $\lambda_0 = 0.632 \mu\text{m}$ is the laser measurement wavelength in the optical shop. Experience with several measured mirrors suggests that pv and rms are related by a factor $\text{pv}/\text{rms} \simeq 5.5$. Suppose these mirrors are used at an average angle of incidence of 45° . The reflected wavefront will be two times worse than the mirror itself. Combining these factors, we get the wavefront rms at 45° incidence from a single $\lambda_0/20$ pv mirror as

$$\delta_0 = \frac{2(\lambda_0/20)}{5.5 \cos(45^\circ)} \quad (3.76)$$

$$= \lambda_0/39 \quad (3.77)$$

for each reflection. Suppose that there are $N = 14$ mirrors in a typical ground-based interferometer. The net wavefront rms after N reflections will then be $\delta \simeq \lambda_0/10$, and the visibility from this alone will be $V_{\text{surfaces}} \simeq e^{-(\pi/5)^2(\lambda_0/\lambda)^2}$. If the test and operating wavelengths are the same, then we find $V_{\text{surfaces}} \simeq 0.67$. This is a significant loss, and it shows that the cumulative effect of even rather good optical surfaces can strongly affect an interferometer. If the operating wavelength is longer, then the visibility is improved. See, e.g., Porro *et al.* (1999) for a complete discussion.

3.3.5 Polarization Effects

We tend to ignore polarization, perhaps because our eyes are not sensitive to it, but anyone with the type of polarized sunglasses that you could buy at one time will tell you that reflected light, from the sky or pavement or automobile hood, can be highly polarized. It should be no surprise then that polarization can reduce fringe visibility, as we now show.

Suppose we consider a typical flat mirror which reflects light incident at 45° from the normal. The electric vector components which are perpendicular and parallel to the plane of incidence are called the s and p components respectively. A typical overcoated silver mirror will cause the difference between the s and p phases to change by about 30° in the visible, with the change at other wavelengths varying roughly as λ^{-1} , i.e., less in the infrared.

If an interferometer can be built so that the reflections in each arm follow the same sequence of changes of direction, and if the corresponding mirrors at each reflection are of the same type, then both beams will experience the same phase shifts, and the respective s and p components will combine independently in the focal plane and produce identical fringe packets. This is the principle behind the layout of the IOTA interferometer (Traub, 1988), for example.

However if the sequence of reflections is different, or the mirrors are not the same, then $s - p$ differences can occur. Suppose that the $s - p$ shift between the two beams is ϕ_{sp} . Then the interferogram will have a visibility term V_{pol} , where

$$V_{pol} = |\cos(\phi_{sp}/2)|. \quad (3.78)$$

Note that if $\phi_{sp} = \pi$, then the interferograms from each polarization will be modulated such that the peaks of one occur in the valleys of the other, and the net modulation will be zero. Thus it is possible to make the net interference effect disappear completely! See Traub (1988) for more details.

3.4 Visibility Enhancement Methods

Among the many fascinating methods that have been invented to enhance the instrumental visibility, we briefly discuss four which are of particular interest for current ground- and space-based interferometers: adaptive optics, single-mode fiber optics, single-mode integrated optics, and nulling.

3.4.1 Adaptive Optics

If an adaptive optics (AO) system is used at a telescope the distortion of an input wavefront can be measured and corrected in real time. A compensating distortion is applied to a mirror, so that the resulting wavefront is (in principle) perfectly flat. The measurement is done using either a natural guide star (NGS) or laser guide star (LGS) as a wavefront reference. The technology has been dramatically demonstrated at large ground-based telescopes, where a typical image-width reduction of roughly a factor of 10, and a central intensity increase of a factor of $10^{1.5}$ can be achieved. To date, the technique has not yet been applied to an interferometer, but it will be required when interferometer mirror diameters much exceeding 1 m are used.

The advantage of AO is that large telescope diameters can be used, independent of the atmospheric coherence length. Two references are Roddier (1999) and Hardy (1998).

3.4.2 Fiber Optics

Single-mode fiber optics may be used within an interferometer (a) to select essentially the plane-wave part of a wavefront, (b) to split a guided wave into any desired intensity ratio, and (c) to interferometrically combine two guided waves. Wavelengths longer than the cutoff wavelength will excite a single electromagnetic mode of a fiber waveguide; the cutoff wavelength is a simple function of core radius, core index, and cladding index. Fiber couplings are formed by arranging the cores of two fibers to run parallel to each other with roughly one core diameter thickness of cladding material between the cores. Under these conditions the core excitation hops periodically between one core and the other. By adjusting the interaction length one can achieve any desired degree of transfer, including roughly 50:50, at a particular wavelength.

The advantage of single-mode fibers is that when only the plane-wave part of the wavefront is used, the fluctuations in visibility due to random atmospheric warping of the wavefront are dramatically reduced. Typical visibility uncertainties go from 5% with a classical beam combiner to 0.5% with a fiber-optic combiner. Atmospheric effects blur the image and therefore reduce the intensity coupled into each fiber at any given instant. Thus the flux in each fiber must be monitored. This can be easily done by tapping off a portion of the flux with a coupler. Two references are Coudé du Foresto *et al.* (1997) and Delage and Reynaud (2000).

3.4.3 Integrated Optics

The methods of integrated optics (IO) allow single-mode waveguides to be manufactured in-situ on the surface of a plane glass substrate, using integrated-circuit techniques, with all of the advantages of fiber optics, plus the advantage of small size and reduced cost of production. IO will be tested at ground-based interferometers in the near future. Recent progress in this field is described by Malbet *et al.* (1999) and Haguenauer *et al.* (2000).

3.4.4 Nulling

Nulling interferometry is a technique in which a phase shift of π is added to one wavefront segment, so that when it interferes with another segment of the same wavefront, perfect cancellation is achieved on-axis. Thus a bright central star can be dimmed by many orders of magnitude relative to the surrounding off-axis material, such as a planetary system.

There are several techniques which can be used to create the π phase shift. Note that this is not the same as moving a mirror $1/4$ wavelength, because for other wavelengths the phase

shift is different. One technique is to use roof reflectors (pairs of mirrors at 90°) to achieve a reversal of sign of the electric vector. Another technique is to introduce a precise thickness of glass whose index will act to retard all wavelengths by very nearly one-half wavelength. Mirror and lens combinations can also be used. To date broad-band nulling of 1 part in 10^4 in the visible has been achieved, using pairs of roof reflectors. Nulling interferometry is discussed in Serabyn in Chapter 16. Four further references are Bracewell and MacPhie (1979), Hinz *et al.* (1998), and Serabyn *et al.* (1999).

References

- M. Born and E. Wolf, *Principles of Optics*, 7 (expanded) edn. (Cambridge, UK: Cambridge University Press, 1999).
- R.N. Bracewell and R.H. MacPhie, "Searching for nonsolar planets," *Icarus* **38**, 136–147 (1979).
- L. Delage and F. Reynaud, "Analysis and control of polarization effects on phase closure and image acquisition in a fibre-linked three-telescope stellar interferometer," *J. Opt. A: Pure Appl. Opt.* **2**, 147–153 (2000).
- V. Coudé du Foresto, S. Ridgway, and J.-M. Mariotti, "Deriving object visibilities from interferograms obtained with a fiber stellar interferometer," *Astron. Astrophys. Sup. Ser.* **121**, 379–392 (1997).
- P. Haguenauer, J.P. Berger, K. Rousset-Perraut, P. Kern, F. Malbet, I. Schanen-Duport, and P. Benech, "Integrated optics for astronomical interferometry. III. Optical validation of a planar optics two-telescope beam combiner," *Appl. Opt.* **39**, 2130–2139 (2000).
- J.W. Hardy, *Adaptive Optics for Astronomical Telescopes* (Oxford, UK: Oxford University Press, 1998).
- P.M. Hinz, J.R.P. Angel, W.F. Hoffmann, D.W. McCarthy Jr, P.C. McGuire, Matt Cheselka, J.L. Hora, and N.J. Woolf, "Imaging circumstellar environments with a nulling interferometer," *Nature* **395**, 251–253 (1998).
- F. Malbet, P. Kern, I. Schanen-Duport, J.-P. Berger, K. Rousset-Perraut, and P. Benech, "Integrated optics for astronomical interferometry. I. Concept and astronomical applications," *Astron. Astrophys. Sup. Ser.* **138**, 135–145 (1999).
- I.L. Porro, W.A. Traub, and N.P. Carleton, "Effect of telescope alignment on a stellar interferometer," *Appl. Opt.* **38**, 6055–6067 (1999).
- W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2 edn. (Cambridge, UK: Cambridge University Press, 1992).
- F. Roddier, ed., *Adaptive Optics in Astronomy* (Cambridge, UK: Cambridge University Press, 1999).
- D.J. Schroeder, *Astronomical Optics*, 2 edn. (San Diego, CA: Academic Press, 2000).

E. Serabyn, J.K. Wallace, G.J. Hardy, E.G.H. Schmidtlin, and H. Nguyen, “Deep nulling of visible laser light,” *Appl. Opt.* **38**, 7128–7132 (1999).

W.A. Traub, “Polarization effects in stellar interferometers,” in *High Resolution Imaging by Interferometry*, F. Merkle, ed. Proc. ESO Conf. **29**, 1029–1038 (1988).

Chapter 4

Noise and Sensitivity in Interferometry

CHARLES H. TOWNES

UNIVERSITY OF CALIFORNIA AT BERKELEY
BERKELEY, CALIFORNIA

4.1 Introduction

The quality of an interferometer may be measured by a number of parameters—these include the precision of visibility measurements, sensitivity in measuring weak sources, precision in measuring fringe phases, range of baselines available, and general flexibility. Some of these factors will be discussed here, with emphasis on noise phenomena and their effect on sensitivity.

4.2 Wavefront Aberrations

Although complete interference can occur if the wavefronts are both perfectly planar as they strike the two telescopes, they are likely to be distorted over a given telescope aperture by imperfect seeing, which can give misleadingly low values of visibility. Heterodyne detection selects and detects only the components of the wavefront of the stellar radiation that are in phase with the wavefront of the laser local oscillator (cf. Kingston 1978), and thus tends to prevent this difficulty. A similar result can be achieved when direct, rather than heterodyne, detection is used by spatial filtering with a glass fiber to obtain a single geometric mode. However, this usually entails some loss of signal.

4.3 Thermal and Quantum Noise

4.3.1 Noise Power Fluctuations

The fundamental noise for an ideal direct detector, which detects both polarizations, is due to thermal radiation striking the detector, which for an ideal detector produces a noise power fluctuation of

$$N_d = h\nu \sqrt{\frac{2\Delta\nu}{t} \frac{1 - \varepsilon}{e^{h\nu/kT} - 1}}, \quad (4.1)$$

where $\Delta\nu$ is bandwidth, t the averaging time, T the temperature of optics and atmosphere through which the signal is received, and ε the fractional transmission of radiation reaching the telescope. This expression is valid for a photodiode; photoconductors have more noise by a factor of $\sqrt{2}$. Direct detection does not determine the phase of a wave, and hence noise due to the uncertainty principle is not present; in principle, noise is due only to fluctuations in the number of quanta in the radiation received. These fluctuations are also present in heterodyne detection, but at IR frequencies are generally much smaller than the uncertainty principle noise, and hence are omitted from Equation 4.2 below, where heterodyne detection is discussed.

The fundamental noise power for a heterodyne detector, (which detects only one polarization, i.e., that of the local oscillator) is equivalent to an average of one quantum per second per unit bandwidth in the same polarization as the local oscillator. For an ideal photodiode, the noise power fluctuation is hence

$$N_h = h\nu \sqrt{\frac{2\Delta\nu}{t}}, \quad (4.2)$$

where $h\nu$ is the quantum energy, $\Delta\nu$ is the single sideband bandwidth in Hz, and t the post-detection averaging time in seconds (cf. Teich 1970, Kingston 1978, Townes 1984). Since heterodyne detection has the ability to measure the phase of a wave, and phase is complementary to energy or number of quanta, this noise is an inescapable result of quantum mechanics and the uncertainty principle (cf. Serber and Townes 1960; Kimble and Walls 1987). The ratio of uncertainties in number of photons to those in phase of the wave can be changed while still satisfying the uncertainty principle (*ibid.*), but such possibilities are not very practical for heterodyne detectors and are not considered here.

4.3.2 Signal-to-Noise Ratio

For a useful signal, equivalent power of noise fluctuations must be substantially less than the signal power P_ν from a source observed. The signal-to-noise ratio for the two cases, assuming ideal detectors with 100% quantum efficiency, is

$$(S/N)_h = \frac{P_\nu}{h\nu} \sqrt{2\Delta\nu t}, \quad (4.3)$$

and

$$(S/N)_d = \frac{P_\nu}{h\nu} \sqrt{\frac{2\Delta\nu t (e^{h\nu/kT} - 1)}{1 - \varepsilon}}, \quad (4.4)$$

where P_ν is the power in each polarization per unit bandwidth (Hertz). Thus, even when direct detection has bandwidths as narrow as heterodyne detection, it appears to have a substantial advantage by a factor:

$$\sqrt{\frac{e^{h\nu/kT} - 1}{1 - \varepsilon}}, \quad (4.5)$$

and indeed it does under some circumstances. For a wavelength of 10 μm , room temperature T of 293 K, and transmission $\varepsilon=0.9$ (characteristic of atmospheric transmission at 10 μm), this factor can be as large as 37. For visible or near-IR radiation the sensitivity advantage of direct detection is usually overwhelming, even though direct-detection interferometers have complex enough optics that the net transmission is usually substantially less than 0.9—sometimes as low as 0.05. However, there are a number of other considerations, which for wavelengths as large as 10 μm can in some cases give a substantial advantage to heterodyne detection, and in other cases to direct detection. These will be discussed below.

4.4 Heterodyne versus Direct-Detection Interferometry

4.4.1 Signal-to-Noise Ratio for Fringe Measurements

The signal-to-noise for measurement of fringe power in an interferometer with heterodyne detection is (Townes, 1984; cf. also Johnson, 1974, for detailed discussion of detection with a photoconductor)

$$(S/N)_{\text{Fringe}} \equiv \left(\frac{\text{vis} \times P_\nu}{h\nu} \right)^2 (t_0 t)^{1/2} \Delta\nu, \quad (4.6)$$

where

P_ν \equiv the power per unit bandwidth of source, as in Equation 4.3

$h\nu$ \equiv the quantum energy

vis \equiv the visibility or fraction of the source power which provides interference

t_0 \equiv the total observing time

t \equiv the length of time atmospheric fluctuations do not change the fringe phase by more than approximately one radian.

$\Delta\nu$ \equiv the single sideband IF bandwidth, as in Equation 4.3

The signal-to-noise for fringe amplitude, and hence for visibility determination, is proportional to the square root of that for fringe power.

The signal-to-noise for a direct-detection interferometer involves somewhat more complex considerations. This is because the relative pathlengths for the signals from the two telescopes must be accurately tracked, and this tracking usually involves radiation of a different wavelength and bandwidth from that used for visibility measurements. If the relative pathlengths can be tracked accurately, then the theoretical signal-to-noise for fringes is comparable to that for power, but reduced by the visibility. It can in principle be made large by long-term averaging. However, the pathlengths themselves fluctuate quite rapidly on timescales of roughly 0.01 to 1.0 seconds due to atmospheric seeing, and there must be enough sensitivity to determine the phase of interference between light from the two telescope sources during a time as short as these fluctuations. If the signal-to-noise approaches unity or less for these short times, no measurement of fringe intensity can be made. Long-term observations for averaging and improving the signal-to-noise are then also not possible.

Equations 4.4 and 4.5 assume essentially perfect detectors and optics. Actually, detectors normally have quantum efficiencies in the range 0.2 to 0.8, which reduces sensitivity by these factors. In addition, transmission of signals through the optical systems involves losses, which for stellar interferometers typically reduces signals by factors between 0.05 and 0.8, depending on the number of mirrors or other optical components involved.

Heterodyne detection sensitivity actually obtained on interference fringes is rather close to the theoretical limit given above. For fringe amplitude, the Infrared Spatial Interferometer (ISI) is within a factor of about 4 of this limit, with much of this factor due to present detector quantum efficiencies being between about 0.25 and 0.40 rather than unity. This closeness to the theoretical limit is in part because heterodyne detection conveniently eliminates unwanted radiation outside a chosen bandwidth and partly because only a very small source at a long distance can produce interference, so more local and extraneous radiation does not provide a false interference signal. However, in contrast to fringe observation, the detection of power radiated into a single telescope suffers from variations in stray radiation as does direct detection, discussed below.

4.4.2 Bandwidth Considerations

For bandwidths narrower than about 1 cm^{-1} ($3 \times 10^{10} \text{ Hz}$), the dominant noise in direct detection is typically associated with detector dark current and readout fluctuations (or “read” noise), combined with some stray radiation. For narrow bandwidths, the highest sensitivity normally achieved in direct-detection ground-based astronomical systems is $\sim 2 \times 10^{-15} \text{ W}$ for a one second averaging time.* This equals the theoretical noise for heterodyne detection with a bandwidth of about $5 \times 10^9 \text{ Hz}$. Of course, heterodyne detection also does not give perfect theoretical performance, largely because the quantum efficiency of detectors is not 100%. However, its noise does decrease with decreasing bandwidth in accordance with theory rather than reaching a lower limit, as is characteristic of direct

*J.H. Lacy, private communication, 1998. Also personal experience of C.H. Townes.

detection. Hence, for bandwidths appreciably narrower than about 5×10^9 Hz, heterodyne detection is typically the more sensitive, since its noise continues to decrease with decreasing bandwidth. This is important for certain applications such as measuring spectral lines (which are often as narrow as 5×10^7 Hz).

For continuum radiation, direct detection has the advantage of being able to use broad bandwidths. The single-sideband bandwidths for efficient heterodyne detectors at $10 \mu\text{m}$ are presently $\sim 0.1 \text{ cm}^{-1}$, or 3×10^9 Hz. However, quantum-well detectors have been made with substantially larger bandwidths, and those as large as $\sim 0.5 \text{ cm}^{-1}$ can be envisioned. Direct detection can in principle include essentially all the mid-IR radiation transmitted by the atmosphere, or a bandwidth as large as 600 cm^{-1} . However, without further division of the radiation such a broad band of wavelengths results in a rather broad range of resolutions, with varying visibilities. Hence a range of not more than 10%, which is $1 \mu\text{m}$ in wavelength range or 100 cm^{-1} , is the maximum considered here. A bandwidth of $10\text{--}20 \text{ cm}^{-1}$ would be normal, giving a 1–2% range of resolution. Such bandwidths provide a substantial apparent advantage over the narrower-band heterodyne detection. However, for broad bandwidths, fluctuations in radiation from the sky and optics usually dominate the noise rather than fundamental quantum fluctuations or detector noise. For a $1\text{-}\mu\text{m}$ bandwidth, experience shows that non-fundamental noise prevents the detection of power below about 5×10^{-14} W for an averaging time of 1 s.[†] This is about 40 times larger than what would be obtained with the theoretical limit given above, and for this case much of the expected gain from broad bandwidths is negated. In principle, the sensitivity may possibly come closer to theoretical values, but this represents a long-standing challenge to experimentalists.

4.4.3 Arrays with Multiple Telescopes

Interferometers frequently use multiple telescopes to obtain many baselines simultaneously; for example the VLA radio interferometer uses 27 telescopes (or 351 baselines). This requires sending individual signals from each telescope to 26 different interference measurements or correlators. For heterodyne detection, such a system involves no further loss in signal-to-noise, because after detection the signal can be amplified and divided without introduction of any significant noise, as is done in radio interferometry. However, a direct-detection interferometer would normally divide the signal of each telescope into 26 equal parts. This would reduce the signal by the same factor and negate much of the theoretical signal-to-noise advantage noted above. In addition, bringing the signals together from separate telescopes for correlation is also more tractable in the case of heterodyne detection, requiring only electrical cables rather than the evacuated light-pipes and multiple optical

[†]For 10% bandwidth at $11.7 \mu\text{m}$ and 1-s averaging time, the MIRAC2 camera used on the UKIRT telescope has a sensitivity (1σ uncertainty) of 5×10^{-13} W (Hoffman et al. 1998). For $1\text{-}\mu\text{m}$ spectral width and 1-s averaging time, specification for the MICS mid-IR camera on the UKIRT telescope is 1.2×10^{-13} W, and for $0.1\text{-}\mu\text{m}$ spectral width, it is 1.3×10^{-14} for a 1-s averaging time (Miyata et al., 1999). When used as a $10\text{-}\mu\text{m}$ camera on the Keck telescope, the LWS has a sensitivity 5.9×10^{-14} W for a $1\text{-}\mu\text{m}$ bandwidth and 1-s averaging time (Keck website LWS Instrument Document).

components which are needed for direct detection. These considerations lead us to conclude that a many-telescope multiple baseline system for the mid-IR region is probably simpler, cheaper, and more flexible with heterodyne detection than with direct detection.

4.4.4 Coherence Time and Fringe Tracking

A prominent task for direct-detection interferometry is to accurately track the relative delay between any two beams which are to interfere. As noted above, this is generally done at a wavelength different from that used for fringe measurement and with a relatively broad bandwidth. However, a useful signal for delay compensation must be obtained in a time as short as atmospheric fluctuations, hence in about 0.01 s. If this signal is detected well, this short time has no direct effect on the sensitivity of measurement, and the delay compensation is helpful in decreasing noise because the phase of interference does not then fluctuate, as it may for heterodyne detection. However, if the signal is lost, visibilities cannot be measured. In contrast, the narrow bandwidth and IF delays of heterodyne detection allow the use of calculated delays without direct tracking, and hence the time for averaging a signal can be arbitrarily long regardless of signal strength. This is another basic convenience of narrow bandwidths. For an averaging time of one hour instead of 0.01 s, this can boost the relative advantage of heterodyne interferometry in detecting fringe power by as much as 600 (cf. Equation 4.6), or in detecting fringe amplitude by a factor of $\sqrt{600} \simeq 25$. The averaging time can also be increased to many hours by many nights of observing, as is often done in radio astronomy. Direct detection can avoid this disadvantage, however, if there is a bright star close enough to be within the same isoplanatic patch as the object being measured, since then the bright star can be used for delay line tracking. And there are some other techniques, not yet generally used, which can ameliorate this sometimes large disadvantage of direct detection. These can involve use of multiple wavelength bands, each possibly as large as atmospheric dispersion allows, to track the pathlength variations, or multiple tracking units, each involving very short distances.

4.5 Conclusion

If a relatively wide bandwidth and single baseline are used, and if there is a strong guide star within the same isoplanatic patch as the object observed, then for wavelengths as short as 10 μm direct detection has a large advantage. However, if many baselines are used and there is no strong guidestar, or if bandwidths less than about 1 cm^{-1} are used, heterodyne detection has a substantial advantage. Both techniques are useful, each has its own optimum functions, and achievement of ideal performance in either one is challenging.

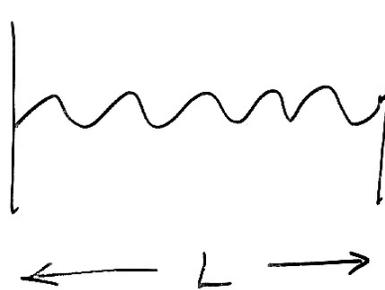
Various expressions and factors discussed above are outlined in the following brief notes and tables.

1.

number of quanta per mode in thermal equilibrium

$$\bar{n} = \frac{\sum_{m=0}^{\infty} m e^{-\frac{m h \nu}{kT}}}{\sum_{m=0}^{\infty} e^{-\frac{m h \nu}{kT}}} = \frac{1}{e^{\frac{h \nu}{kT}} - 1}$$

Power flow



$$\frac{c}{\nu} = \lambda = \frac{2L}{n} \quad \therefore n = \frac{2L}{c} \nu$$

Assume 1 photon per mode, Then $\Delta n = \frac{2L}{c} \Delta \nu$

Round trip time $t = \frac{2L}{c} \quad \therefore \frac{\Delta n}{t} = \Delta \nu$

or, for 2 polarizations, the total

$$\begin{aligned} \text{photon flux} &= 2h\nu \Delta \nu \quad \text{or, at thermal} \\ \text{equilibrium} &= \frac{2h\nu \Delta \nu}{e^{\frac{h\nu}{kT}} - 1} \end{aligned}$$

Figure 4.1: The photon flux per unit mode or frequency interval.

2.

For fractional loss ϵ in path to $T \rightarrow 0$,

$$\text{photon flux} = \frac{2 h \nu \Delta \nu \epsilon}{e^{h \nu / k T} - 1}$$

For signal power $P_{\nu} \Delta \nu$ in each polarization

$$\begin{aligned} S/N &= \frac{2 P_{\nu} \Delta \nu t}{h \nu \sqrt{\frac{2 \Delta \nu \epsilon t}{e^{h \nu / k T} - 1}}} \\ &= \frac{P_{\nu}}{h \nu} \sqrt{(e^{h \nu / k T} - 1) \frac{2 \Delta \nu t}{\epsilon}} \end{aligned}$$

where t is time length of measurement in seconds.

N.B. This is for a photodiode
For a photoconductor, S/N is less by $\sim \frac{1}{\sqrt{2}}$

Figure 4.2: Ideal signal-to-noise ratio for direct detection.

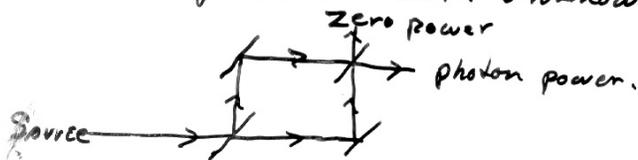
3.

Heterodyne Detection

Uncertainty principle requires $\Delta\phi\Delta n \geq 1/2$

or, for $\Delta n = \nu n$, $\Delta\phi \geq \frac{1}{2\nu n}$

This is true for absolute phase, not for relative phase of two waves, with absolute phase of both unknown. e.g. interferometer



"Squeezing"

Heterodyne detection

$$\text{Current } i \propto (E_{L0} + E_s + E_0)^2$$

$$\approx E_{L0}^2 + 2E_{L0}E_s + 2E_{L0}E_0$$

where E_{L0} = field of local oscillator

E_s = " " signal

E_0 = zero point fluctuation field

$$\text{Power} \propto E_{L0}^2 (E_s^2 + E_0^2)$$

where usually $E_0^2 \gg E_s^2$ and it produces 1 photon/sec/unit bandwidth

\therefore Noise power = $2h\nu \Delta\nu$ for two sidebands of width $\Delta\nu$.

Figure 4.3: Noise power in heterodyne detection associated with the uncertainty principle, assuming no special procedures such as "squeezing" are used.

4.

For signal power P_r per unit bandwidth

$$\text{Heterodyne } S/N = \frac{P_r}{h\nu} \sqrt{2\Delta\nu t}$$

where t is time length of measurement in seconds.

∴ Fundamental laws give S/N advantage to direct detection of

$$\text{if } \frac{e^{\frac{h\nu}{kT}} - 1}{\epsilon} > 1$$

$$\sqrt{\frac{e^{\frac{h\nu}{kT}} - 1}{\epsilon}}$$

Table

$$\lambda \quad \sqrt{e^{\frac{h\nu}{kT}} - 1} \quad \text{for } T = 283^\circ\text{K}$$

1 cm	0.071
100 μm	0.81
11 μm	10
3 μm	4.8×10^3
0.6 μm	2.6×10^{18}

Figure 4.4: Signal-to-noise ratios for ideal heterodyne detection and a numerical comparison with direct detection as a function of wavelength.

Practicalities

5.

Band width

pros and cons
factor of $\sqrt{\frac{100}{1/9}}$

For 100 cm^{-1} bandwidth, $\Delta T < 10^{-6} \text{ K}$
for direct detection sensitivity

Spectral problems

Non fundamental noise

leakage radiation - telescope noise

" current in detectors

L.O. scattering and cure

Present limits for direct detection at $10 \mu\text{m}$

For 1 sec ave.

limit $\sim 2 \times 10^{-15}$ watts for minimum $\Delta\nu$
= theoretical value for $\Delta\nu \approx 100 \text{ cm}^{-1}$

for $\Delta\nu$ actually = 100 cm^{-1} , limit $\sim 5 \times 10^{-14}$ W.

" " = 10 cm^{-1} , " $\sim 1.3 \times 10^{-14}$ W

Other direct vs. heterodyne differences

Isolation of plane waves or point sources
by heterodyne detection

Isolation of fringe signal

Decrease vs. no decrease of signal with
multiple telescopes

Difference in practical averaging time

Complexities of each.

Figure 4.5: Various practical considerations which affect the relative performance of direct and heterodyne detection.

References

- W. F. Hoffman, J. L. Hora, G. G. Fazio, L. K. Deutsch, and A. Dayal, “MIRAC2: a mid-infrared array camera for astronomy,” in *Infrared Astronomical Instrumentation*, A.M. Fowler, ed., Proc. SPIE **3354**, 647–658 (1998).
- M. A. Johnson, *An Infrared Stellar Interferometer Using Heterodyne Detection*, Ph.D. dissertation, University of California at Berkeley (1974).
- H. J. Kimble and D.F. Walls eds., “Squeezed states of the electromagnetic field,” *J. Opt. Soc. Am. B.* **4**, 1450–1741 (1987).
- R.H. Kingston, *Detection of Optical and Infrared Radiation*, Springer Series in Optical Sciences, **10** (Springer-Verlag, 1978).
- T. Miyata, H. Kataza, Y. Okamoto, T. Junabe, T. Onaka, T. Yamashita, K. Nakamura, and H. Shibai, “MICS: a new mid-infrared camera and spectrometer for ground-based astronomy,” *Pub. Astron. Soc. Pac.* **111**, 750–764 (1999).
- R. Serber and C.H. Townes, “Limits on electromagnetic amplification due to complementarity,” in *Quantum Electronics*, C.H. Townes, ed., 223–255 (Columbia Univ. Press, 1960).
- M.C. Teich, in *Semiconductors and Semimetals*, R. K. Willardson and A. C. Beer, eds., 361 (New York: Academic Press, 1970).
- C. H. Townes, “Spatial interferometry in the mid-infrared region,” *Journal of Astrophysics and Astronomy* **5**, 111–130 (1984).

Chapter 5

Observing Through the Turbulent Atmosphere

ANDREAS QUIRRENBACH

UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA

5.1 Introduction

Atmospheric turbulence is a major contributor to the difficulty of optical and infrared interferometry from the ground. In fact, many of the key design parameters of an interferometer—site selection, operating wavelength, aperture size, bandwidths of the angle-tracking and fringe-tracking servo loops, coherent integration time, inclusion of adaptive optics—are driven largely by the boundary conditions set by the atmosphere. The sensitivity of interferometers, and the precision of astrometric measurements, depend strongly on the “seeing.” It is therefore important to understand how turbulence is generated in the atmosphere, and how its effects on the propagation of light can be quantified. This tutorial is intended to give a brief overview of these topics. Their application to the design and operation of interferometers is discussed in other contributions to this volume.

The organization of the present article is as follows: Section 5.2 introduces the Kolmogorov turbulence model, which gives some physical insight into the generation of turbulence, and which is widely used as a quantitative model to describe the spatial variations of quantities such as density and refractive index in turbulent media. Section 5.3 deals with the propagation of waves through turbulence. In this section, the *Fried parameter* r_0 is defined as a numeric measure of the integrated turbulence strength. Section 5.4 discusses optical image formation and the effects of turbulence on images. It is shown that r_0 is

directly related to the width of seeing-limited images. The parameter τ_0 , which measures the coherence time, and the concept of anisoplanatism and the coherence angle θ_0 are introduced. The appendix, Section 5.5, summarizes a few useful facts from Fourier theory.

Due to space constraints, the treatment of the subject matter in this article is necessarily short. More detailed accounts of atmospheric turbulence and its effect on high-angular resolution astronomy can be found in the articles by Roddier (1981 and 1989) and by Fried (1994), as well as in the excellent book by Hardy (1998). The text by Léna, Lebrun, and Mignard (1998) contains a very short summary of atmospheric turbulence in the context of image formation. The book by Born and Wolf (1999) is still the standard resource for general information on the principles of optics.

5.2 The Kolmogorov Turbulence Model

5.2.1 Eddies in the Turbulent Atmosphere

The properties of fluid flows are characterized by the well-known Reynolds number $Re = VL/\nu$, where V is the fluid velocity, L a characteristic length scale, and ν the kinematic viscosity of the fluid. For air, $\nu = 1.5 \cdot 10^{-5} \text{ m}^2 \text{ s}^{-1}$, so that atmospheric flows with winds of a few m s^{-1} and length scales of several meters to kilometers have $Re \gtrsim 10^6$ and are therefore almost always turbulent. The turbulent energy is generated by eddies on a large scale L_0 , which spawn a hierarchy of smaller eddies (see also Figure 5.1). Dissipation is not important for the large eddies, but the kinetic energy of the turbulent motion is dissipated in small eddies with a typical size l_0 . The characteristic size scales L_0 and l_0 are known as the *outer scale* and the *inner scale* of the turbulence. There is considerable debate over typical values of L_0 ; it may be a few tens to hundreds of meters in most cases. l_0 is of order a few millimeters.

In the so-called *inertial range* between l_0 and L_0 , there is a universal description for the turbulence spectrum, i.e., the strength of the turbulence as a function of the eddy size, or of the spatial frequency κ . This somewhat surprising result is the underlying reason for the importance of this simple turbulence model, which was developed by Kolmogorov, and is generally known as *Kolmogorov turbulence*. In the following section, a simple argument based on dimensional analysis will be used to derive the structure function for the Kolmogorov model.

5.2.2 The Structure Function for Kolmogorov Turbulence

The only two relevant parameters (in addition to l_0 and L_0) that determine the strength and spectrum of Kolmogorov turbulence are the rate of energy generation per unit mass ε , and the kinematic viscosity ν . The units of ε are $\text{J s}^{-1} \text{ kg}^{-1} = \text{m}^2 \text{ s}^{-3}$, and those of ν are $\text{m}^2 \text{ s}^{-1}$. Under the assumption that the turbulence is homogeneous and isotropic, the

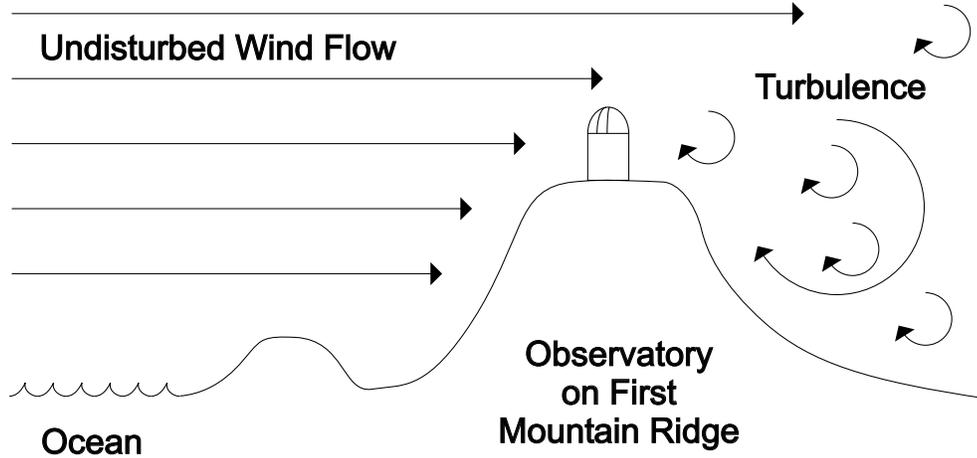


Figure 5.1: Schematic of turbulence generation in the wake of obstacles. Most world-class observatories are located on the first mountain ridge near the coast (or on mountains on islands), with prevailing winds from the ocean.

structure function of the turbulent velocity field, $D_v(R_1, R_2)$, depends only on $|R_1 - R_2|$, and can therefore be written as

$$\begin{aligned} D_v(R_1, R_2) &\equiv \langle |v(R_1) - v(R_2)|^2 \rangle \\ &= \alpha \cdot f(|R_1 - R_2| / \beta) \quad , \end{aligned} \quad (5.1)$$

where f is some dimensionless function of a dimensionless argument. It is immediately clear that the dimensions of α are velocity squared, and those of β length. Since α and β depend only on ε and ν , it follows from dimensional analysis that

$$\alpha = \nu^{1/2} \varepsilon^{1/2} \quad \text{and} \quad \beta = \nu^{3/4} \varepsilon^{-1/4} \quad . \quad (5.2)$$

In addition, the structure function must be independent of ν in the inertial range. This is possible only if

$$D_v(R_1, R_2) = \alpha \cdot (|R_1 - R_2| / \beta)^{2/3} = C_v^2 \cdot |R_1 - R_2|^{2/3} \quad , \quad (5.3)$$

where C_v^2 is a constant. We have thus derived the important result mentioned above, namely a universal description of the turbulence spectrum. It has only one parameter C_v^2 , which describes the turbulence strength.

5.2.3 Structure Function and Power Spectral Density of the Refractive Index

The turbulence, whose velocity field is characterized by Equation 5.3, mixes different layers of air, and therefore carries around “parcels” of air with different temperature. Since these “parcels” are in pressure equilibrium, they have different densities ρ , and therefore different indices of refraction n . It can be shown that the temperature fluctuations also follow Kolmogorov’s law with a new parameter C_T^2 of their own:

$$D_T(R_1, R_2) = C_T^2 \cdot |R_1 - R_2|^{2/3} \quad . \quad (5.4)$$

From the ideal gas law, and $N \equiv (n - 1) \propto \rho$, it follows that the structure function of the refractive index is

$$D_n(R_1, R_2) = D_N(R_1, R_2) = C_N^2 \cdot |R_1 - R_2|^{2/3} \quad , \quad (5.5)$$

with

$$C_N = (7.8 \cdot 10^{-5} P[\text{mbar}]/T^2[\text{K}]) \cdot C_T \quad . \quad (5.6)$$

We note that Equation 5.5 contains a complete description of the statistical properties of the refractive index fluctuations, on length scales between l_0 and L_0 . It is possible to calculate related quantities such as the power spectral density Φ from the structure function D . Using the relation between the structure function and the covariance (Equation 5.51), and the Wiener–Khinchin Theorem (Equation 5.49), we obtain

$$C_N^2 \cdot R^{2/3} = D_N(R) = 2 \int_{-\infty}^{\infty} d\kappa (1 - \exp(2\pi i\kappa R)) \Phi(\kappa) \quad . \quad (5.7)$$

Calculating $\Phi(\kappa)$ from this relation is a slightly non-trivial task*; the result is

$$\Phi(\kappa) = \frac{\Gamma(\frac{5}{3}) \sin \frac{\pi}{3}}{(2\pi)^{5/3}} C_N^2 \kappa^{-5/3} = 0.0365 C_N^2 \kappa^{-5/3} \quad . \quad (5.8)$$

We have thus obtained the important result that the power spectrum of Kolmogorov turbulence follows a $\kappa^{-5/3}$ law in the inertial range.†

5.3 Wave Propagation Through Turbulence

5.3.1 The Effects of Turbulent Layers

We now look at the propagation of a wavefront $\psi(x) = \exp i\phi(x)$ through a turbulent layer of thickness δh at height h . The phase shift produced by refractive index fluctuations is

$$\phi(x) = k \int_h^{h+\delta h} dz n(x, z) \quad , \quad (5.9)$$

where $k = 2\pi/\lambda$. For layers that are much thicker than the individual turbulence cells, many independent variables contribute to the phase shift, which therefore has Gaussian statistics according to the Central Limit Theorem.

The task at hand is now using the statistical properties of the refractive index fluctuations, which were calculated in Section 5.2.3, to derive the statistical properties of the wavefront.

*See Tatarski (1961). Note that his definition of the power spectral density has an additional factor $\frac{1}{2\pi}$, and that his ω corresponds to $2\pi\kappa$.

†Note: We have defined $R = |R_1 - R_2|$ and κ as one-dimensional variables, and consequently used a one-dimensional Fourier transform in Equation 5.7. Sometimes three-dimensional quantities \vec{R} and $\vec{\kappa}$ are used instead. Then a three-dimensional Fourier transform with volume element $4\pi |\vec{\kappa}|^2 d|\vec{\kappa}|$ has to be used in Equation 5.7, and the result is a power spectrum $\Phi(|\vec{\kappa}|) \propto |\vec{\kappa}|^{-11/3}$.

We first express the coherence function $B_h(r)$ of the wavefront after passing through the layer at height h in terms of the phase structure function:

$$\begin{aligned}
B_h(r) &\equiv \langle \psi(x)\psi^*(x+r) \rangle \\
&= \langle \exp i [\phi(x) - \phi(x+r)] \rangle \\
&= \exp \left(-\frac{1}{2} \langle |\phi(x) - \phi(x+r)|^2 \rangle \right) \\
&= \exp \left(-\frac{1}{2} D_\phi(r) \right) .
\end{aligned} \tag{5.10}$$

Here we have used the fact that $[\phi(x) - \phi(x+r)]$ has Gaussian statistics with zero mean, and the relation

$$\langle \exp(\alpha\chi) \rangle = \exp \left(\frac{1}{2} \alpha^2 \langle \chi^2 \rangle \right) \tag{5.11}$$

for Gaussian variables χ with zero mean, which can easily be verified by carrying out the integral over the distribution function.

5.3.2 Calculation of the Phase Structure Function

The next step is the computation of $D_\phi(r)$. We start with the covariance $B_\phi(r)$, which is by definition:

$$\begin{aligned}
B_\phi(r) &\equiv \langle \phi(x)\phi(x+r) \rangle \\
&= k^2 \int_h^{h+\delta h} \int_h^{h+\delta h} dz' dz'' \langle n(x, z')n(x+r, z'') \rangle \\
&= k^2 \int_h^{h+\delta h} dz' \int_{h-z'}^{h+\delta h-z'} dz B_N(r, z) .
\end{aligned} \tag{5.12}$$

Here we have introduced the new variable $z = z'' - z'$, and the covariance $B_N(r, z)$ of the refractive index variations. For δh much larger than the correlation scale of the fluctuations, the integration can be extended from $-\infty$ to ∞ , and we obtain

$$B_\phi(r) = k^2 \delta h \int_{-\infty}^{\infty} dz B_N(r, z) . \tag{5.13}$$

Now we can use Equation 5.51 again, first for $D_\phi(r)$, then for $D_N(r, z)$ and $D_N(0, z)$, and get:

$$\begin{aligned}
D_\phi(r) &= 2[B_\phi(0) - B_\phi(r)] \\
&= 2k^2 \delta h \int_{-\infty}^{\infty} dz [B_N(0, z) - B_N(r, z)] \\
&= 2k^2 \delta h \int_{-\infty}^{\infty} dz \left[(B_N(0, 0) - B_N(r, z)) - (B_N(0, 0) - B_N(0, z)) \right] \\
&= k^2 \delta h \int_{-\infty}^{\infty} dz [D_N(r, z) - D_N(0, z)] .
\end{aligned} \tag{5.14}$$

Inserting from Equation 5.5 gives

$$\begin{aligned}
 D_\phi(r) &= k^2 \delta h C_N^2 \int_{-\infty}^{\infty} dz \left[(r^2 + z^2)^{1/3} - |z|^{2/3} \right] \\
 &= \frac{2\Gamma(\frac{1}{2})\Gamma(\frac{1}{6})}{5\Gamma(\frac{2}{3})} k^2 \delta h C_N^2 r^{5/3} \\
 &= 2.914 k^2 \delta h C_N^2 r^{5/3} .
 \end{aligned} \tag{5.15}$$

This is the desired expression for the structure function of phase fluctuations due to Kolmogorov turbulence in a layer of thickness δh .

5.3.3 Phase-Coherence Function and Fried Parameter

We are now in a position to put everything together. Inserting Equation 5.15 into Equation 5.10, we get

$$B_h(r) = \exp \left[-\frac{1}{2} (2.914 k^2 C_N^2 \delta h r^{5/3}) \right] . \tag{5.16}$$

Integration over the whole atmosphere, and taking into account the zenith angle z , gives:

$$B(r) = \exp \left[-\frac{1}{2} \left(2.914 k^2 (\sec z) r^{5/3} \int dh C_N^2(h) \right) \right] . \tag{5.17}$$

We now define the *Fried parameter* r_0 by

$$r_0 \equiv \left[0.423 k^2 (\sec z) \int dh C_N^2(h) \right]^{-3/5} \tag{5.18}$$

and can write

$$B(r) = \exp \left[-3.44 \left(\frac{r}{r_0} \right)^{5/3} \right] , \quad D(r) = 6.88 \left(\frac{r}{r_0} \right)^{5/3} . \tag{5.19}$$

We have thus derived fairly simple expressions for the phase-coherence function and the phase structure function. They depend only on the Fried parameter r_0 , which in turn is a function of turbulence strength, zenith angle, and wavelength. The significance of the Fried parameter will be discussed further in Section 5.4.3.

5.4 The Effect of Turbulence on Images

5.4.1 Optical Image Formation

The complex amplitude A of a wave ψ diffracted at an aperture P with area Π is given by Huygens' principle, which states that each point in the aperture can be considered as the center of an emerging spherical wave. In the far field (i.e., in the case of Fraunhofer

diffraction), the spherical waves are equivalent to plane waves, and we can write down the expression for the amplitude:

$$A(\alpha) = \frac{1}{\sqrt{\Pi}} \int dx \psi(x)P(x) \exp(-2\pi i\alpha x/\lambda) . \quad (5.20)$$

Here we describe the aperture P by a complex function $P(x)$; in the simple case of a fully transmissive aperture without aberrations $P(x) \equiv 1$ inside the aperture, and $P(x) \equiv 0$ outside. Introducing the new variable $u \equiv x/\lambda$ we can write

$$A(\alpha) = \frac{1}{\sqrt{\Pi}} FT[\psi(u)P(u)] . \quad (5.21)$$

The normalization in equations 5.20 and 5.21 has been chosen such that the illumination in the focal plane is given by the square of the wave amplitude:

$$S(\alpha) = |A(\alpha)|^2 = \frac{1}{\Pi} \left| FT[\psi(u)P(u)] \right|^2 . \quad (5.22)$$

Applying the Wiener–Khinchin Theorem (Equation 5.49) to this equation we get

$$S(f) = \frac{1}{\Pi} \int du \psi(u)\psi^*(u+f)P(u)P^*(u+f) . \quad (5.23)$$

This equation can be used to describe the spatial frequency content $S(f)$ of images taken through the turbulent atmosphere if ψ is identified with the wavefront after passing through the turbulence. Taking long exposures (in practice this means exposures of at least a few seconds) means averaging over different realizations of the atmosphere:

$$\begin{aligned} \langle S(f) \rangle &= \frac{1}{\Pi} \int du \langle \psi(u)\psi^*(u+f) \rangle P(u)P^*(u+f) \\ &= B_\psi(f) \cdot T(f) . \end{aligned} \quad (5.24)$$

Here we have introduced the *telescope transfer function*

$$T(f) = \frac{1}{\Pi} \int du P(u)P^*(u+f) . \quad (5.25)$$

Equation 5.24 contains the important result that for long exposures the optical transfer function is the product of the telescope transfer function and the atmospheric transfer function, which is equal to the phase-coherence function $B_\psi(f)$.

5.4.2 Diffraction-Limited Images and Seeing-Limited Images

The resolving power R of an optical system can very generally be defined by the integral over the optical transfer function. For the atmosphere/telescope system we get

$$R \equiv \int df S(f) = \int df B(f)T(f) . \quad (5.26)$$

In the absence of turbulence, $B(f) \equiv 1$, and we obtain the *diffraction-limited* resolving power of a telescope with diameter D :

$$\begin{aligned} R_{\text{tel}} &= \int df T(f) = \frac{1}{\Pi} \int \int dudf P(u)P^*(u+f) \\ &= \frac{1}{\Pi} \left| \int du P(u) \right|^2 = \frac{\pi}{4} \left(\frac{D}{\lambda} \right)^2 . \end{aligned} \quad (5.27)$$

The last equality assumes a circular aperture and shows the relation of R to the familiar Rayleigh criterion $1.22 \cdot \lambda/D$. The advantage of using R over the Rayleigh criterion is that R is a well-defined quantity for arbitrary aperture shapes and in the presence of aberrations.

For strong turbulence and rather large telescope diameters, $T = 1$ in the region where B is non-zero, and we get the *seeing-limited* resolving power:

$$\begin{aligned} R_{\text{atm}} &= \int df B(f) = \int df \exp \left[- \left(3.44 \left(\frac{\lambda f}{r_0} \right)^{5/3} \right) \right] \\ &= \frac{6\pi}{5} \Gamma\left(\frac{6}{5}\right) \left(3.44 \left(\frac{\lambda}{r_0} \right)^{5/3} \right)^{-6/5} = \frac{\pi}{4} \left(\frac{r_0}{\lambda} \right)^2 . \end{aligned} \quad (5.28)$$

Here we have used Equation 5.19 with $r = \lambda f$ for the phase-coherence function $B(f)$.

5.4.3 The Significance of the Fried Parameter r_0

A comparison of Equations 5.27 and 5.28 elucidates the significance of the Fried parameter, and reveals the reason for the peculiar choice of the numerical parameter 0.423 in Equation 5.18: *The resolution of seeing-limited images obtained through an atmosphere with turbulence characterized by a Fried parameter r_0 is the same as the resolution of diffraction-limited images taken with a telescope of diameter r_0 .* Observations with telescopes much larger than r_0 are seeing-limited, whereas observations with telescopes smaller than r_0 are essentially diffraction-limited. It can also be shown that the mean-square phase variation over an aperture of diameter r_0 is about 1 rad^2 (more precisely, $\sigma_\phi^2 = 1.03 \text{ rad}^2$). These results are captured in the extremely simplified picture that describes the atmospheric turbulence by r_0 -sized “patches” of constant phase, and random phases between the individual patches. While this picture can be useful for some rough estimates, one should keep in mind that Kolmogorov turbulence has a continuous spectrum ranging from l_0 to L_0 .

The scaling of r_0 with wavelength and zenith angle implied by Equation 5.18 has far-reaching practical consequences. Since

$$r_0 \propto \lambda^{6/5} , \quad (5.29)$$

it is much easier to achieve diffraction-limited performance at longer wavelengths. For example, the number of degrees of freedom (the number of actuators on the deformable mirror and the number of subapertures in the wavefront sensor) in an adaptive optics system must be of order $(D/r_0)^2 \propto \lambda^{-12/5}$. An interferometer works well only if the wavefronts from the individual telescopes are coherent (i.e., have phase variances not larger than about

1 rad²); therefore the maximum useful aperture area of an interferometer is $\propto \lambda^{12/5}$ (unless the wavefronts are corrected with adaptive optics). Equation 5.29 implies that the width of seeing-limited images, $\theta \simeq 1.2 \cdot \lambda/r_0 \propto \lambda^{-1/5}$, varies only slowly with λ ; it is somewhat better at longer wavelengths. In addition, we see from Equation 5.18 that $r_0 \propto (\sec z)^{-3/5}$; the effects of seeing increase with air mass.

From this discussion it should be clear that the magnitude of r_0 —given by the integral over C_N^2 —is a crucial parameter for high-resolution observations. At good sites, such as Mauna Kea, r_0 is of order 20 cm at 500 nm, which corresponds to an image FWHM of 0."6. The scaling of r_0 with λ (Equation 5.29) implies that in the mid-infrared ($\lambda \gtrsim 10 \mu\text{m}$) even the 10 m Keck Telescopes are nearly diffraction-limited, whereas a 1.8-m telescope has $D/r_0 \sim 2$ at $\lambda = 2 \mu\text{m}$ and $D/r_0 \sim 5$ at $\lambda = 800 \text{ nm}$. It should be noted that at any given site r_0 varies dramatically from night to night; it may be a factor of 2 better than the median or a factor of 5 worse. In addition, the seeing fluctuates on all time scales down to minutes and seconds; this has to be taken into account in calibration procedures.

5.4.4 Strehl Ratio

The quality of an imaging system, or of the wavefront after propagation through turbulence, is often measured by the *Strehl ratio* S , defined as the peak intensity in the image of a point source divided by the peak intensity in a diffraction-limited image taken through the same aperture. For a circular aperture with an aberration function $\psi(\rho, \theta)$, which describes the wavefront distortion (in μm or nm) as a function of the spherical coordinates (ρ, θ) , the Strehl ratio is given by:

$$S = \frac{1}{\pi^2} \left| \int_0^1 \int_0^{2\pi} \rho d\rho d\theta e^{ik\psi(\rho,\theta)} \right|^2 . \quad (5.30)$$

From this equation it is immediately clear that $0 \leq S \leq 1$, that $S = 1$ for $\psi = \text{const.}$, that $S \ll 1$ for strongly varying ψ , and that for any given (varying) ψ the Strehl ratio tends to be larger for longer wavelengths (smaller k). In the case of atmospheric turbulence, only the statistical properties of ψ are known. If the rms wavefront error $\sigma_\phi \equiv k \sigma_\psi$ is smaller than about 2 rad, S can be approximated by the so-called *extended Marechal approximation*:

$$S = e^{-\sigma_\phi^2} . \quad (5.31)$$

We have seen above (Equation 5.19 and Section 5.4.3) that

$$\sigma_\phi^2 = 1.03 \left(\frac{D}{r_0} \right)^{5/3} . \quad (5.32)$$

Equations 5.31 and 5.32 show that the Strehl ratio for a telescope with diameter $D = r_0$ is $S = 0.36$; for $D \gtrsim r_0$ the Strehl ratio decreases precipitously with telescope diameter. (Equivalently S decreases sharply with decreasing wavelength, since $r_0 \propto \lambda^{6/5}$.)

If $S \gtrsim 0.1$ in an imaging application, deconvolution algorithms can usually be applied to obtain diffraction-limited images, but the dynamic range and signal-to-noise ratio are worse

than for $S \sim 1$. For example, because of spherical aberration, the Hubble Space Telescope has $S \simeq 0.1$ without corrective optics. Before the installation of COSTAR and WFPC2 in the first servicing mission, the imaging performance of HST was severely affected by the flawed optics, although diffraction-limited images could be obtained with image restoration software. In an interferometer, the maximum fringe contrast is roughly proportional to the Strehl ratio if no corrective measures (adaptive optics or mode filtering with pinholes or single-mode fibers) are taken.

5.4.5 Taylor Hypothesis and τ_0

So far we have discussed the spatial structure of atmospheric turbulence and its effects on image formation. Now we turn to the question of temporal changes of the turbulence pattern. The time scale for these changes is usually much longer than the time it takes the wind to blow the turbulence past the telescope aperture. According to the *Taylor hypothesis of frozen turbulence*, the variations of the turbulence caused by a single layer can therefore be modeled by a “frozen” pattern that is transported across the aperture by the wind in that layer. If multiple layers contribute to the total turbulence, the time evolution is more complicated, but the temporal behavior of the turbulence can still be characterized by a time constant

$$\tau_0 \equiv r_0/v \quad , \quad (5.33)$$

where v is the wind speed in the dominant layer. With typical wind speeds of order 20 m/s, $\tau_0 \simeq 10$ ms for $r_0 = 20$ cm. The wavelength scaling of τ_0 is obviously the same as that of r_0 , i.e., $\tau_0 \propto \lambda^{6/5}$.

Observations with exposure time $t \gg \tau_0$ average over the atmospheric random process; these are the *long exposures* for which Equations 5.24 and 5.28 are applicable. In contrast, *short exposures* with $t \ll \tau_0$ produce images through a single instantaneous realization of the atmosphere; these *speckle images* contain information at high spatial frequencies up to the diffraction limit, which can be extracted from series of such images with computer processing (e.g., bispectrum analysis). The parameter τ_0 is also of great importance for the design of adaptive optics systems and interferometers. All control loops that have to reject atmospheric fluctuations—AO control loops, angle trackers, fringe trackers—must have bandwidths larger than $1/\tau_0$. Together r_0 and τ_0 set fundamental limits to the sensitivity of these wavefront control loops: a certain number of photons must arrive per r_0 -sized patch during the time τ_0 for the wavefront sensor (or fringe sensor) to work. This implies that the sensitivity scales with $r_0^2 \cdot \tau_0 \propto \lambda^{18/5}$ (for equal photon flux per bandpass).

5.4.6 Anisoplanatism

The light from two stars separated by an angle θ passes through different patches of the atmosphere and therefore experiences different phase variations. This *angular anisoplanatism* limits the field corrected by adaptive optics systems and causes phase decorrelation for off-axis objects in interferometers. To calculate the effect of anisoplanatism, we trace

back the rays to two stars separated by an angle θ from the telescope pupil. They coincide at the pupil, and their separation $r(d)$ at a distance d is $\theta \cdot d$. At zenith angle z , the distance is related to the height h in the atmosphere by $d = h \sec z$. To calculate the phase variance between the two rays, we insert this relation in

$$D_\phi(r) = 2.914 k^2 \sec z \delta h C_N^2 r^{5/3} \quad (5.34)$$

(see Equation 5.15) and obtain

$$\begin{aligned} \langle \sigma_\theta^2 \rangle &= 2.914 k^2 (\sec z) \int dh C_N^2(h) (\theta h \sec z)^{5/3} \\ &= 2.914 k^2 (\sec z)^{8/3} \theta^{5/3} \int dh C_N^2(h) h^{5/3} \\ &= \left(\frac{\theta}{\theta_0} \right)^{5/3}, \end{aligned} \quad (5.35)$$

where we have introduced the *isoplanatic angle* θ_0 , for which the variance of the relative phase is 1 rad^2 :

$$\theta_0 \equiv \left[2.914 k^2 (\sec z)^{8/3} \int dh C_N^2(h) h^{5/3} \right]^{-3/5}. \quad (5.36)$$

By comparing the definitions for the Fried parameter r_0 and for θ_0 , (Equations 5.18 and 5.36), we see that

$$\theta_0 = 0.314 (\cos z) \frac{r_0}{H}, \quad (5.37)$$

where

$$H \equiv \left(\frac{\int dh C_N^2(h) h^{5/3}}{\int dh C_N^2(h)} \right)^{3/5} \quad (5.38)$$

is the *mean effective turbulence height*. Equations 5.36 and 5.37 show that the isoplanatic angle is affected mostly by high-altitude turbulence; the anisoplanatism associated with ground layers and dome seeing is very weak. Moreover, we see that θ_0 scales with $\lambda^{6/5}$, but it depends more strongly on zenith angle than r_0 . For $r_0 = 20 \text{ cm}$ and an effective turbulence height of 7 km , Equation 5.37 gives $\theta_0 = 1.8 \text{ arcsec}$. For two stars separated by more than θ_0 the short-exposure point-spread functions (or point-spread functions generated by adaptive optics) are different.[‡] In contrast the long-exposure point-spread functions, which represent averages over many realizations of the atmospheric turbulence, are nearly identical even over angles much larger than θ_0 .

[‡]It should be pointed out that these calculations of anisoplanatism give somewhat too pessimistic results. The reason is that a large fraction of the phase variance between the two rays considered is a piston term which doesn't lead to image motion or blurring. (Note, however, that the piston term has to be taken into account in interferometry.) Moreover, anisoplanatism is less severe for low spatial frequencies, which most adaptive optics systems correct much better than high spatial frequencies. The degradation of the Strehl ratio with off-axis angle is therefore not quite as bad as suggested by inserting Equation 5.35 in Equation 5.31.

5.4.7 Scintillation

The geometric optics approximation of light propagation that was used in Section 5.3 is only valid for propagation pathlengths shorter than the *Fresnel propagation length* $d_F \equiv r_0^2/\lambda$. (In other words, the Fresnel scale $r_F \equiv \sqrt{\lambda L}$, where L is the distance to the dominant layer of turbulence, must be smaller than the Fried scale r_0 .) For $r_0 = 20$ cm and $\lambda = 500$ nm, $d_F = 80$ km, and the geometric approximation is a good first-order approach at good sites for visible and infrared wavelengths (since $d_F \propto \lambda^{7/5}$ for Kolmogorov turbulence). However, if the propagation length is comparable to d_F or longer, the rays diffracted at the turbulence cells interfere with each other, which causes intensity fluctuations in addition to the phase variations. This phenomenon is called *scintillation*; it is an important error source in high-precision photometry unless the exposure times are very long. Since scintillation is an interference phenomenon, it is highly chromatic. This effect can be easily observed with the naked eye: bright stars close to the horizon twinkle strongly and change color on time scales of seconds.

Although scintillation is weak for most applications of adaptive optics and interferometry, it has to be taken into account under some circumstances. For example, high-performance adaptive optics systems designed for the direct detection of extrasolar planets have to correct the wavefront errors so well that intensity fluctuations become important. In interferometers that use fringe detection schemes based on temporal pathlength modulation and synchronous photon detection, scintillation noise has to be considered when very small fringe amplitudes are to be measured.

The effects of scintillation can be quantified by determining the relative intensity fluctuations $\delta I/I$; for small amplitudes $\delta I/I = \delta \ln I$. A calculation similar to the one in Section 5.3 gives the variance of the log intensity fluctuations:

$$\sigma_{\ln I}^2 = 2.24 k^{7/6} (\sec z)^{11/6} \int dh C_N^2(h) h^{5/6} . \quad (5.39)$$

This expression is valid only for small apertures with diameter $D \ll r_F$. For larger apertures, scintillation is reduced by averaging over multiple independent subapertures. This changes not only the amplitude of the intensity fluctuations, but also the functional dependence on zenith angle, wavelength, and turbulence height. The expression

$$\sigma_{\ln I}^2 \propto D^{-7/3} (\sec z)^3 \int dh C_N^2(h) h^2 , \quad (5.40)$$

which is valid for $D \gg r_F$ and $z \lesssim 60^\circ$, shows the expected strong decrease of the scintillation amplitude with aperture size; note that it is independent of the observing wavelength. For larger zenith angles the assumption $\delta \ln I \ll 1$ is no longer valid, the fluctuations increase less strongly with $\sec z$ than predicted by Equation 5.40, and eventually saturate.

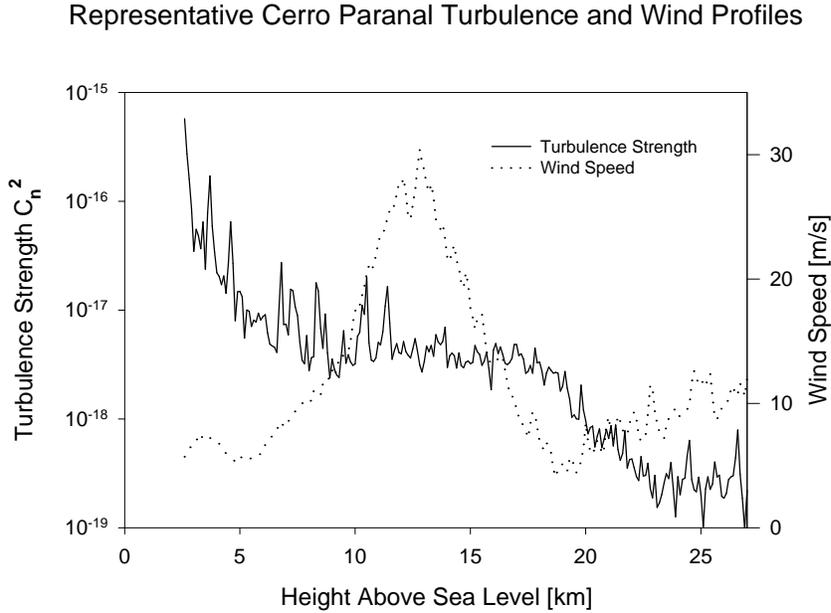


Figure 5.2: Turbulence and wind profiles measured on Cerro Paranal, Chile. The turbulence is strongest close to the ground (2635 m above sea level). The wind speed is highest at an altitude of ~ 10 to 15 km. Wind shear often leads to additional layers of strong turbulence at high altitude (only weakly present in this data set).

5.4.8 Turbulence and Wind Profiles

We have seen in the preceding sections that the most important statistical properties of seeing can be characterized by a few numbers: the Fried parameter r_0 , the coherence time τ_0 , the isoplanatic angle θ_0 , and the scintillation index $\sigma_{\ln I}$. For the design and performance evaluation of high-angular-resolution instruments it is of great importance to have reliable statistical information on these parameters. Therefore extensive seeing monitoring campaigns are normally conducted before decisions are made about the site selection for large telescopes and interferometers, or about the construction of expensive adaptive optics systems. Having access to the output of a continuously running seeing monitor which gives the instantaneous value of r_0 (and ideally also of the other seeing parameters) is also very convenient for debugging and for optimizing the performance of high-resolution instruments.

From Equations 5.18, 5.33, 5.35, and 5.39 it is obvious that all seeing parameters can easily be calculated from moments

$$\mu_m \equiv \int dh C_N^2(h) h^m \quad (5.41)$$

of the turbulence profile $C_N^2(h)$; and (in the case of τ_0) from moments

$$v_m \equiv \int dh C_N^2(h) v^m(h) \quad (5.42)$$

of the wind profile $v(h)$. More complicated analyses such as performance estimates of adaptive optics systems with laser guide stars and of multi-conjugate AO systems also rely on knowledge of $C_N^2(h)$ and $v(h)$. In-situ measurements of these profiles with balloon flights and remote measurements with SCIDAR[§] or related methods are therefore needed to fully characterize the atmospheric turbulence. Figure 5.2 shows profiles measured on Cerro Paranal, the site of the European Southern Observatory's Very Large Telescope observatory. The decrease of C_N^2 with height is typical for most sites; frequently wind shear at altitudes near 10 km creates additional layers of enhanced turbulence. The highest wind speeds normally occur at heights between 9 and 12 km. Extensive sets of observed turbulence and wind profiles, combined with the analytic methods sketched in this article and numerical simulations, form a firm basis for the evaluation of astronomical sites, and for the design of interferometers and adaptive optics systems.

5.5 Appendix: Some Useful Facts from Fourier Theory

For reference, this appendix lists a few useful results from Fourier theory without proofs. In the notation adopted, $g \iff G$ means “ G is the Fourier transform of g ,” and it is understood that $g \iff G$ and $h \iff H$. H^* is the complex conjugate of H . Introductions into Fourier theory and more details can be found in many textbooks, for example the one by Bracewell (1999).

The *convolution* $g * h$ and *correlation* $\text{Corr}(g, h)$ of two functions g and h are defined by:

$$g * h \equiv \int_{-\infty}^{\infty} d\tau g(t - \tau)h(\tau) \quad (5.43)$$

and

$$\text{Corr}(g, h) \equiv \int_{-\infty}^{\infty} d\tau g(t + \tau)h(\tau) \quad (5.44)$$

A special case of the latter is the correlation of a function with itself, the *covariance*:

$$B_g \equiv \text{Corr}(g, g) \quad (5.45)$$

For complex functions, the *coherence function* is defined by:

$$B_g \equiv \text{Corr}(g, g^*) \quad (5.46)$$

The customary use of the same symbol B for covariance and coherence function is somewhat unfortunate, but should not be too confusing.

The famous *Convolution Theorem* and *Correlation Theorem* are:

$$g * h \iff G(f)H(f) \quad (5.47)$$

[§]The SCIDAR technique is based on auto-correlating pupil images of double stars.

and

$$\text{Corr}(g, h) \iff G(f)H^*(f) \quad . \quad (5.48)$$

The special case of the Correlation Theorem for the covariance is the *Wiener–Khinchin Theorem*:

$$B_g = \text{Corr}(g, g) \iff |G(f)|^2 \quad . \quad (5.49)$$

The *structure function* D_g of a function g is defined by:

$$D_g(t_1, t_2) \equiv \langle |g(t_1) - g(t_2)|^2 \rangle \quad . \quad (5.50)$$

If g describes a homogeneous and isotropic random process, D_g depends only on $t = |t_1 - t_2|$. By expanding the square in Equation 5.50, we see that in this case

$$D_g(t) = 2(B_g(0) - B_g(t)) \quad . \quad (5.51)$$

Finally, *Parseval's Theorem* states that the total power in a time series is the same as the total power in the corresponding spectrum:

$$\text{Total Power} \equiv \int_{-\infty}^{\infty} dt |g(t)|^2 = \int_{-\infty}^{\infty} df |G(f)|^2 \quad . \quad (5.52)$$

References

- M. Born and E. Wolf, *Principles of Optics*, 7 edn. (Cambridge, UK: Cambridge University Press, 1999).
- R.N. Bracewell, *The Fourier Transform and its Applications*, 3 edn. (New York: McGraw-Hill, 1999).
- D.L. Fried, "Atmospheric Turbulence Optical Effects: Understanding the Adaptive-Optics Implications," in *Adaptive Optics for Astronomy*, D.M. Alloin and J.-M. Mariotti eds., NATO ASI Series **423** (Dordrecht, Netherlands: Kluwer Academic, 1994), 25–57.
- J.W. Hardy, *Adaptive Optics for Astronomical Telescopes* (Oxford, UK: Oxford University Press, 1998).
- P. Léna, F. Lebrun, and F. Mignard, *Observational Astrophysics*, 2 edn. (Berlin, Germany: Springer-Verlag, 1998).
- F. Roddier, "The effects of atmospheric turbulence in optical astronomy," *Prog. Opt.* **19**, 281–376 (1981).
- F. Roddier, "Optical Propagation and Image Formation through the Turbulent Atmosphere," in *Diffraction-Limited Imaging with Very Large Telescopes*, D.M. Alloin and J.-M. Mariotti eds., NATO ASI Series Vol. **274** (Dordrecht, Netherlands: Kluwer Academic, 1989), 33–52.
- V.I. Tatarski, *Wave Propagation in a Turbulent Medium* (New York: McGraw-Hill, 1961).

Chapter 6

Overview of the Design of Stellar Interferometers

THEO A. TEN BRUMMELAAR

CENTER FOR HIGH ANGULAR RESOLUTION ASTRONOMY
GEORGIA STATE UNIVERSITY, ATLANTA, GEORGIA

6.1 Introduction

An optical interferometer is a large and complex beast, comprised of many active and passive systems. There are many engineering challenges in the design and construction of these instruments including, but not confined to, physical stability, atmospheric seeing, path-length equalization, fringe tracking, dynamic control, and dispersion/diffraction problems.

The engineering driving forces for all interferometers are similar. This, combined with our habit of copying each other, results in the fact that most interferometers resemble one another. In this paper I will attempt to discuss some of the known design problems and describe solutions that the community has used to date.

A good overview of what is required in order to build an interferometer was given by Tango and Twiss (1980). This is a “must read” for any student of the subject.

6.2 What We Would Like to Measure

Let us begin by considering the fringe equation using V to represent the measured visibility, V_{obj} the object’s real visibility, and ν the wavenumber ($1/\lambda$) of the spectral channel with

bandwidth $\Delta\nu$. The visibility phase is Φ_{obj} , the phase introduced by the atmosphere is Φ_{atm} , the current baseline is given by B and the observation is at elevation angle θ . Given all of that, we can write

$$I(t, \nu) = 1 + V(t, \nu) \times \frac{\sin [\pi x(t, \nu) \Delta\nu]}{\pi x(t, \nu) \Delta\nu} \times \cos [2\pi x(t, \nu) \nu + \Phi_{\text{obj}} + \Phi_{\text{atm}}(t)] \quad (6.1)$$

where the observed visibility is

$$V(t, \nu) = V_{\text{obj}} \times \eta_{\text{atm}}(t, \nu) \times \eta_{\text{inst}}(t, \nu). \quad (6.2)$$

Here the reduction in visibility due the atmosphere (η_{atm}) and the instrument itself (η_{inst}) have been added. The total optical path length difference is given by

$$x(t, \nu) = B \cos [\theta(t)] + x_{\text{vac}}(t) + n_{\text{air}}(\nu)x_{\text{air}}(t) + n_{\text{glass}}(\nu)x_{\text{glass}}(t) \quad (6.3)$$

using x_{air} to represent the internal path through air and x_{glass} the internal path through glass along with their respective refraction coefficients. The time dependence of these parameters has been explicitly shown in these equations.

Almost all of the extra terms in these equations have the effect of reducing the measured visibility, and therefore also the signal-to-noise ratio of the scientific measurement. Each of these also imply an electro-optical subsystem within the array in order to reduce the visibility loss. For example: atmospheric wavefront distortions must be removed, or reduced, with tip/tilt servos, or more complex adaptive optics systems; path-length variations need to be tracked using a fringe tracker and delay-line combination; and the amount of air and glass paths needs to be carefully controlled in order to reduce differential dispersion effects.

I will try and deal with the most important areas of visibility loss and their solution, or reduction.

6.3 Logistics

In order to do all the science we would like to do, an interferometer requires baselines several hundred meters long, or even larger. The light from each telescope needs to be brought into a central beam-combining facility, which itself must include space for delay lines large enough to compensate for the geometric delays imposed by such long baselines. All of this means a logistic nightmare for the designer of the system. Large amounts of land are required, hopefully flat, and the other elements of the infrastructure, like power, communications and transport, must somehow fit in around the interferometer.

One would like to have a large piece of flat land at high altitude. Unfortunately mountains tend to have more vertical real estate than desired, adding to the logistics problem. It has been the trend to date to place interferometers on an existing observatory site. This tends to further complicate the placement of the various systems, since the existing facilities are rarely transparent, but fortunately it means that most of the other necessary things are

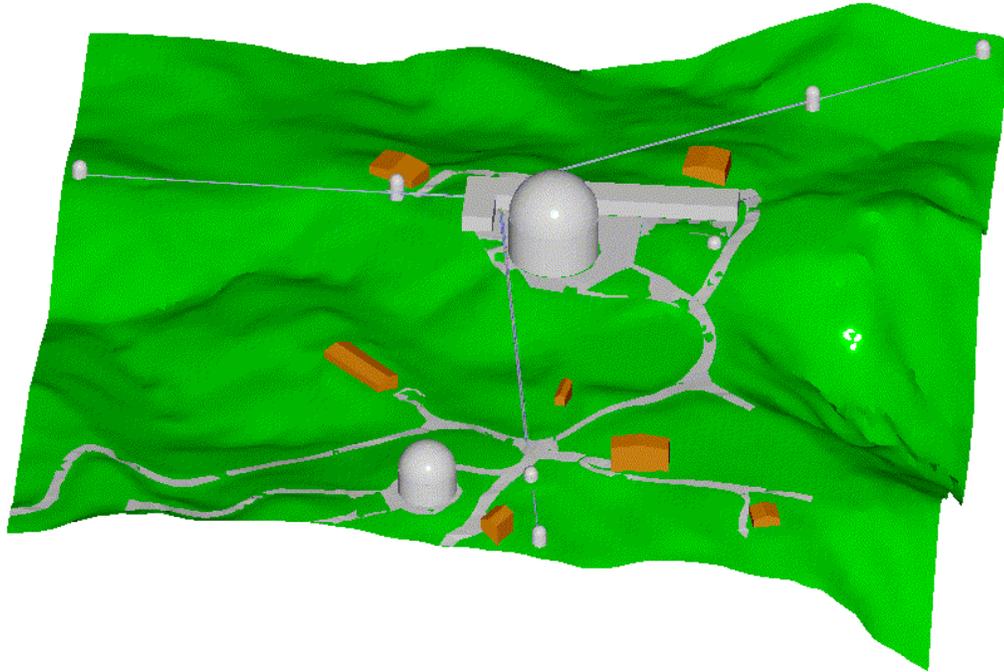


Figure 6.1: Computer aided drawing of the terrain at Mount Wilson, with beam transport pipes for the CHARA Array indicated.

already in place, like a place to sleep, roads, and so on. As an example, the arrangement of the arms of the CHARA Array at Mount Wilson is shown in Figure 6.1.

Until we have solved all the problems associated with using optical fibers for beam transport, you will need to have a line of site between each input aperture and the central building. Deciding where to put each aperture will become a matter of juggling (u, v) plane coverage, light pipes and existing structures and dirt. There is no simple solution to this other than having good survey data, patience and a good three-dimensional modeling package.

6.4 Concrete and Steel

An interferometer has engineering tolerances of the same order as the wavelength at which it will operate. This means everything must either have an active optical system for correction or be stable at the micron level. It is for this reason that you will find a lot of concrete and steel at any interferometric site. The depth of typical concrete foundations is shown in Figure 6.2, which is a photograph taken at SUSI in the early stages of construction.

Not only do you need to use large inertial masses but you need to isolate them from any local sources of vibration. Thus you must ensure that the concrete holding up the building is not the same as that below the optical tables or delay line. You don't want people walking, and therefore vibrating, the flooring below the optical systems. Telescope enclosures should not



Figure 6.2: Preparations for pouring the siderostat foundations at SUSI.

be coupled to telescope support systems. Not being careful about these things will mean you end up with a very sensitive seismic monitor and not an interferometer.

6.5 The Problem with the Atmosphere

The atmosphere has an unfortunate habit of changing all the time and it doesn't always behave as theory says it should. Our most common measurements of the condition of the atmosphere are the coherence length r_0 and the coherence time τ_0 .

In the case of τ_0 Davis and Tango (1996) have shown, based on the definition of coherence time given by Buscher (1988), that

$$\frac{\overline{C(\Delta t)}}{C(0)} = \frac{6}{5\tau} \left[\gamma(3/5, \tau^{5/3}) - \tau^{-1} \gamma(6/5, \tau^{5/3}) \right] \quad (6.4)$$

where Δt is the sample time, the ratio $\overline{C(\Delta t)}/C(0)$ is the correlation loss factor due to this finite sample time, $\tau = \Delta t/\tau_0$, and $\gamma(a, x)$ is a partial gamma function. Davis and Tango (1996) used this equation to measure τ_0 and thereby calibrate out the effects of a finite sample time. To do this you must measure the visibility simultaneously at many sample times and fit these measurements to Equation 6.4. It is then possible to extrapolate back to zero sample time.

The value of τ_0 is in the range of a few milliseconds in the visible during normal seeing conditions and can be as high as several tens of milliseconds in excellent seeing and in the

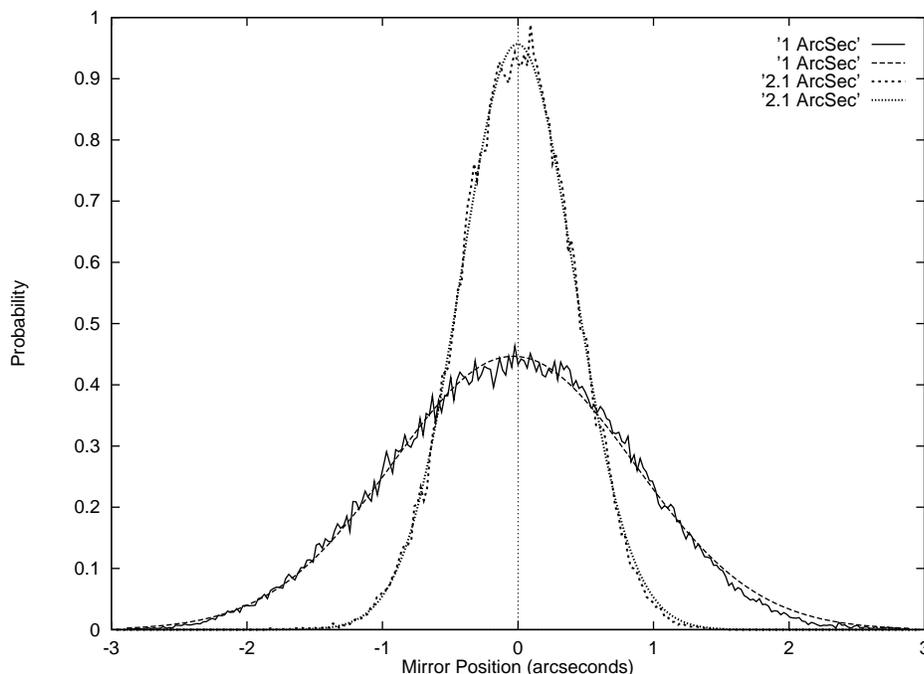


Figure 6.3: Histogram of tip/tilt mirror motion while tracking a star during relatively good (1 arcsec) and bad (2.1 arcsec) seeing. Measurements like this help establish the seeing, both internal and external, during a visibility measurement. These data are taken from ten Brummelaar and Tango (1994).

infrared. This basically sets the time constant for the servo systems within the interferometer. The adaptive optics must run at the rate of hundreds of Hertz in order to keep up with the motion of the atmosphere.

It is also possible to measure r_0 using an already existing subsystem within the interferometer. One possibility is to use the motion of the delay line when locked onto fringes, but this requires locking the system onto fringe phase and will not work if you are scanning through the entire fringe packet or using group-delay tracking. A second, and easier, option is to use the data from the tip/tilt servo. See for example the data in Figure 6.3.

Unfortunately, while many studies of atmospheric seeing have been done over the years (cf. ten Brummelaar *et al.* 1994; Buscher 1994; Haniff *et al.* 1994; Davis *et al.* 1995) it is still not clear how one interprets the spatial seeing data. For the time being, one simply hopes that the optical systems work well, that the internal seeing is negligible, and that things do not change too much between observations of science object and calibrator.

It is, of course, important that the seeing internal to the instrument does not further reduce the visibility. To this end it is common to use “light-pipes” to direct the light from the input aperture, either a siderostat or a telescope, into the central beam-synthesis facility. Often these light-pipes are evacuated in order to remove most of the air and improve the seeing, and will therefore require air-tight windows at either end.

There are several other things we can do to reduce the internal seeing once the light reaches the delay-line and beam combination building. For example, the entire delay-line can be placed in a vacuum system. This is advantageous also from the point of view of differential dispersion discussed below, but is expensive and possibly difficult to deal with. An alternative is to use the “building within a building.” In this scheme the optical systems are placed within an inner enclosure, while the space between the outer and inner walls is air conditioned. The inner area is a large passive thermal mass and can remain stable for long periods of time.

6.6 Polarization

Each reflection from a mirror surface introduces a phase shift between polarization states. This means that the reflections used in each arm of an interferometer must be the same: reflection symmetry should not be broken. Reflection symmetry is also important in order to have the same image rotation in each arm. This normally means you end up with more reflections than you would like, but there is little one can do about this. For example, starlight passing through the central siderostat at COAST, shown in Figure 6.4, undergoes an additional two reflections so that its component s and p polarizations experience the same reflections as light from the other siderostats.



Figure 6.4: View of the array layout at COAST. The s and p polarizations of the light from each siderostat experience the same reflections prior to arriving at the beam combiner.

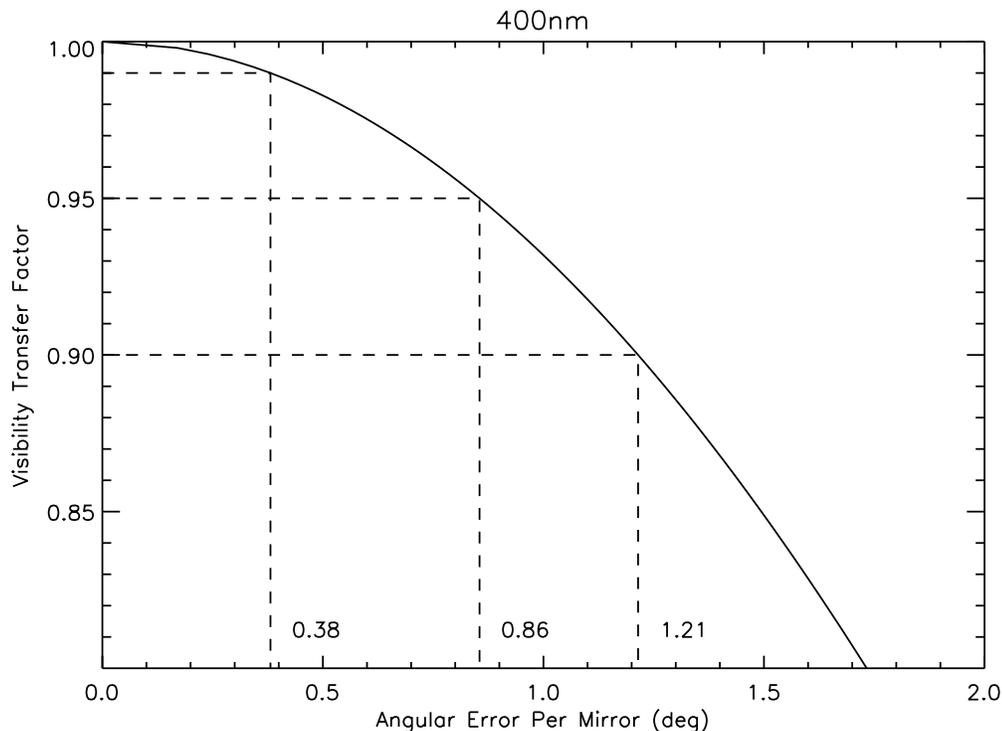


Figure 6.5: Plot of the visibility loss due to the phase difference between polarization states introduced by various differential reflection angles. Assumes 20 mirrors in each arm and is based on the work of Traub (1988).

The reflection angles in each arm must be kept the same to within quite small tolerances. If differential polarization-dependent phase changes are introduced into the different paths within an interferometer, visibility will be reduced. As an example of this effect, Figure 6.5 shows the visibility loss factor for a range of different reflection angle errors for a wavelength of 400 nm. In order to keep the visibility losses to less than 1%, the positioning of each mirror must be good to within a small fraction of a degree. Note that this figure is based on a polarization model, and the predictions of these models are often substantially different from real-world measurements.

One strategy for reducing errors due to polarization is to separate the two polarization states in the beam-combining area. For example, in the Sydney University Stellar Interferometer (Davis *et al.*, 1999) one polarization is used for the tip/tilt detection system while the other is used for the visibility measurement. Alternatively, one could use one polarization state for fringe tracking and the other for fringe measurement. A danger here is that there will be a small phase difference between the s and p polarization states, even if the reflection paths are symmetric, so that the centers of the fringe envelopes will not be in the same location.

6.7 Internal Optical Quality

Clearly the quality of the optics within the interferometer will have a significant effect on the measured visibility amplitudes. Each optic within the system will introduce a phase variance into the beam, reducing Strehl and also visibility. One normally assumes that these phase errors are random and add in an rms way. However, if all mirrors are made by the same process, this may not be true. Nevertheless, the internal image quality is relatively easy to measure and therefore calibrate.

A second important area of consideration when specifying the internal optics is optical throughput. There are often as many as twenty, or even more, reflections in an interferometer and each one contributes to signal loss. Coatings with reflectivities as high as 0.98 are available, but $0.98^{20} = 0.67$. Furthermore it is highly unlikely that these mirrors will stay this reflective; dust will always gather, and the light you do finally get to the back end of the instrument will be further divided into various subsystems. It is not uncommon to have as little as 5% or less of the light that enters the system go towards a scientific measurement. Another problem with some of the more fancy optical coatings is that they can have serious polarization effects. Plain silver or aluminum coatings seem to be the most common.

It may be possible to use an adaptive optics system to correct the internal image quality. In fact this may be one of the first applications for adaptive optics in interferometry. The deformable mirror can be set once using high intensity light from an internal source and held in position while the stellar light is sent through the system.

6.8 Diffraction

The long paths required in an interferometer imply that diffraction effects are almost unavoidable. Differential paths must be introduced to compensate for the external path introduced by the projected baseline and therefore differential diffraction will result in reduced visibilities (Hrynevych, 1992). Any beam reduction will make these problems worse. Furthermore, the combination of atmospheric turbulence and diffraction is not well understood.

In the case of differential diffraction, one must either rely on the calibration object, or re-image the input apertures within the beam combiner. The former assumes a relatively stable system and atmosphere. The latter is preferable but very hard to do.

6.9 Dispersion

There have been numerous studies of dispersion effects in stellar interferometry (eg. Tango 1990; ten Brummelaar 1995; Lawson and Davis 1996; Davis *et al.* 1998). The problem here

is that unequal path lengths within the instrument through vacuum, air and glass result in differential dispersion and thus a reduction in visibility. In fiber-based systems dispersion is a more serious problem, and it is therefore essential to use fibers of carefully matched lengths.

Dispersion is relatively easy to model approximately, although more difficult to model accurately. Compensation for dispersion can be done over a modest bandwidth using glass wedges. However, even with vacuum delay lines and a correction system, dispersion makes it very important to have a calibrator object as close as possible to the science target so that any dispersive effects will be the same in both measurements.

6.10 Controlling the Beast

Clearly an interferometer requires a very complex real-time control system, and the necessary software and hardware are by no means trivial. Most existing interferometers use a distributed control system and a mixture of real-time and non-real-time operating systems. In this way, individual subsystems and their controllers can be developed and tested in isolation. They are then linked together using some form of master control computer and user interface.



(a)



(b)

Figure 6.6: Well trained sequencers: (a) W.J. Tango at SUSI, (b) R.C. Boysen at COAST.

One of the most difficult aspects, apart from actually making each device function, is connecting them all together, sequencing, and error recovery. In most existing systems, apart from a few notable exceptions, the sequencer is a well trained observer, and error recovery depends on the knowledge this observer has of the system and its parts.

It has not been uncommon for the control code, and user interfaces, to lag behind the hardware in the development of an interferometer. The importance of starting control code development early in the design process cannot be overlooked.

6.11 Conclusion

There is of course no single or correct way to build an interferometer, but it is true to say that all of those built so far have many common elements, as is shown by the interferometers illustrated in Figures 6.7 and 6.8. I will conclude then with a list of things you will probably find on a “vanilla” interferometer:

- It will probably be located on an existing observatory site.
- It will have baselines several hundred meters long.
- The light from the input apertures will be brought into the beam-combining laboratory using evacuated light pipes.
- The lab will be located downwind of the array if possible.
- There will be many tons of concrete and steel.
- The input apertures, which to date have usually been siderostats, will in future interferometers probably be large telescopes in a partially redundant array.
- The optical system will be symmetric.
- The facility will include a long building to house the delay lines for each telescope; the delay lines being either in long vacuum pipes or in air in a building-within-a-building.
- There will be at least a tip/tilt system, perhaps even a full-blown adaptive optics system, located (preferably) at the telescopes.
- The beams will be brought onto a single table and combined together, either pair-wise or all at once.
- Some system for measuring the differential delay of each pair of beams will exist.
- The beam combiner will include beam splitters and/or fibers, and the light will be divided up for use in various subsystems, for tip/tilt detection (which must be done as close as possible to the beam combiner), fringe tracking, and so on.
- The back end will also include a great deal of optics for alignment, including a laser, a pin hole, a white-light source, a theodolite, TV cameras of many flavors, and lots of small pieces of paper with targets.
- It will not be easy.

Acknowledgments

Photographs of the Sydney University Stellar Interferometer courtesy of John Davis. Figure 6.7(a) courtesy of Ken Johnston. All other photographs courtesy of Peter Lawson.



(a)



(b)



(c)



(d)

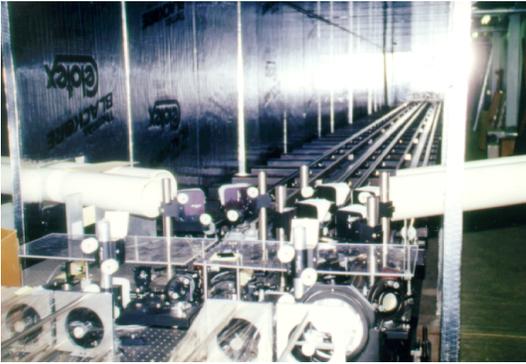


(e)



(f)

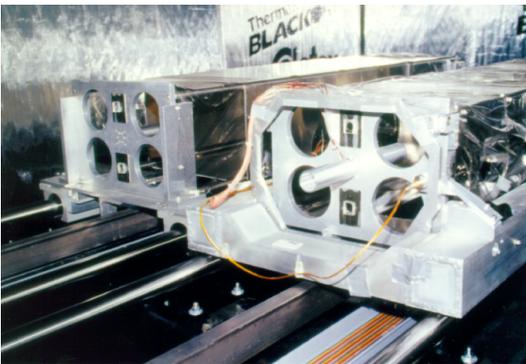
Figure 6.7: External views of interferometric arrays, telescopes, siderostats, and beam transport optics: (a) aerial view of the NPOI, (b) aerial view of SUSI, (c) GI2T telescopes, (d) a COAST siderostat, (e) a SUSI siderostat, (f) PTI evacuated light pipes, with the Palomar 200-in. telescope in the background.



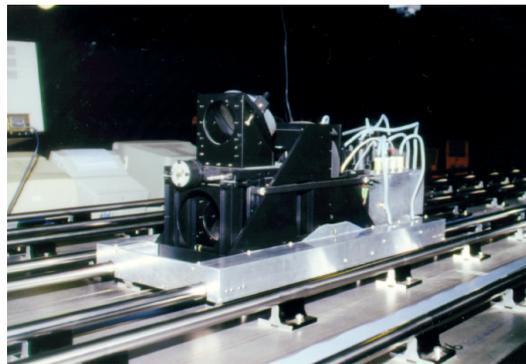
(a)



(b)



(c)



(d)



(e)



(f)

Figure 6.8: Internal views of stellar interferometers, including delay lines and beam combination optics: (a) PTI air delay line area, (b) SUSI air delay line area, (c) PTI delay line carriage, (d) COAST delay line carriage, (e) SUSI beam combiner, (f) COAST beam combiner.

References

- T.A. ten Brummelaar, "Differential path considerations in optical stellar interferometry," *Appl. Opt.* **34**, 2214–2219 (1995).
- T.A. ten Brummelaar and W.J. Tango, "A wavefront tilt correction servo for the Sydney University Stellar Interferometer," *Experimental Astronomy* **4**, 297–315 (1994).
- T.A. ten Brummelaar, W.J. Tango, J. Davis, and R.R. Shobbrook, "Preliminary seeing measurements for SUSI," in *Very High Angular Resolution Imaging*, W.J. Tango and J.G. Robertson, eds., IAU Symp. **158**, 302–304 (Dordrecht, Netherlands: Kluwer Academic, 1994).
- D. Buscher, "Optimizing a ground-based optical interferometer for sensitivity at low light levels," *Mon. Not. R. Astron. Soc.* **235**, 1203–1226 (1988).
- D.F. Buscher, "A thousand and one nights of seeing on Mt Wilson," in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., *Proc. SPIE* **2200**, 260–271 (1994).
- J. Davis, P.R. Lawson, A.J. Booth, W.J. Tango, and E.D. Thorvaldson, "Atmospheric path variations for baselines up to 80 m measured with the Sydney University Stellar Interferometer," *Mon. Not. R. Astron. Soc.* **273**, L53–L58 (1995).
- J. Davis and W.J. Tango, "Measurement of the atmospheric coherence time," *Pub. Astron. Soc. Pac.* **108**, 456–458 (1996).
- J. Davis, W.J. Tango, A.J. Booth, T.A. ten Brummelaar, R.A. Minard, and S.M. Owens, "The Sydney University Stellar Interferometer I: the instrument," *Mon. Not. R. Astron. Soc.* **303**, 773–782 (1999).
- J. Davis, W.J. Tango, and E.D. Thorvaldson, "Dispersion in stellar interferometry: simultaneous optimization for delay tracking and visibility measurements," *Appl. Opt.* **37**, 5132–5136 (1998).
- C.A. Haniff, J.E. Baldwin, P.J. Warner, and T.R. Scott, "Atmospheric phase fluctuation measurement: interferometric results from the WHT and COAST telescopes," in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., *Proc. SPIE* **2200**, 407–417 (1994).
- M. Hrynevych, *Diffraction Effects in Michelson Stellar Interferometry*, Ph.D. thesis, University of Sydney (1992).
- P.R. Lawson and J. Davis, "Dispersion compensation in stellar interferometry," *Appl. Opt.* **35**, 612–620 (1996).
- W.J. Tango, "Dispersion in stellar interferometry," *Appl. Opt.* **29**, 516–521 (1990).
- W.J. Tango and R.Q. Twiss, "Michelson stellar interferometry," *Prog. Opt.* **17**, 239–277 (1980).
- W.A. Traub, "Polarization effects in stellar interferometers," in *High Resolution Imaging by Interferometry*, F. Merkle, ed., *Proc. ESO Conf.* **29**, 1029–1038 (European Southern Observatory: Garching, Germany, 1988).

Part II

Measurement and Calibration of Fringe Parameters

Chapter 7

Visibility Estimation and Calibration

M. MARK COLAVITA

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

7.1 Introduction

Fringe visibility is the fundamental observable in interferometric imaging, as it is related to the object brightness via the van Cittert–Zernike theorem. Visibility is generally complex, and can be expressed as $\Gamma = V \exp(-j\phi)$, where V is the visibility amplitude and ϕ is the fringe phase. We ordinarily work with normalized visibility, i.e., $0 < V < 1$. With a two-element, single-beam (i.e., non-cophased) interferometer, the fringe phase is corrupted by the atmosphere and only the visibility amplitude is useful for imaging (typically, parametric imaging of compact sources; see, for example Boden *et al.* 1999; Mozurkewich *et al.* 1991). While phase and visibility are usually estimated via the same mechanism, the discussion here will be limited to estimation of the visibility amplitude.

This brief review will touch on visibility estimation via fringe scanning, the signal-to-noise ratio of the visibility estimator, estimator and atmospheric biases, and approaches to calibration.

7.2 Fringe Scanning and Matched Filtering

Visibility is just the contrast of the spatial fringe pattern. Most measurement schemes used with Michelson combiners use fringe scanning to convert the spatial fringe pattern to a

temporal one; demodulation of such a temporal signal presents a well-studied measurement problem. This problem occupies only a small niche in the larger field of phase-shifting interferometry for optical testing. The key aspects of the problem for stellar interferometry are accommodating the low intensity of the faint stellar source and the limitations of low-light-level detectors, so some of the more sophisticated phase-shifting algorithms are not readily applicable to our problem.

Fringe scanning to temporally encode the spatial fringe pattern can use step or continuous scanning. Generally, for fast scanning to follow atmospheric fringe motion, continuous scanning is used to avoid settling-time issues. Both sawtooth and triangular waveforms are used, depending on the problem (cf. Shao *et al.* 1988; Colavita *et al.* 1999). This is discussed further by in Section 8.4.4.

Visibility estimation can be analyzed from several perspectives; one is to just consider it a matched filter problem. The temporal fringe pattern for a normalized scan rate can be written as $I = N(1 + V \cos(t + \phi))$, where N is flux, V is visibility, and ϕ is fringe phase. This can also be written as the sum $N + X \cos(t) + Y \sin(t)$, where X and Y are the fringe quadratures. This signal is readily demodulated using the orthogonal basis functions $1, \cos(t), \sin(t)$; from estimates of $N, X,$ and Y we can compute the fringe phase and visibility. With integrating detectors, we must discretize the basis functions to limit the number of required reads (and attendant read noise), leading to, for example, four and eight-bin algorithms. For the common four-bin algorithm, we approximate sin and cos by quadrature squares waves; the basis functions remain orthogonal, but not optimal, and there is a 0.9 dB signal-to-noise penalty on bright stars. Figure 7.1 illustrates fringe estimation at PTI (Colavita *et al.*, 1999).

There is considerable literature on this subject; a number of references are given by Lawson at the end of Chapter 8.

7.3 Visibility Estimation and Signal-to-Noise Ratio

If the fringe phase is known, the fringe visibility can be estimated coherently using a coordinate rotation on the measured fringe quadratures, normalized by the flux. If the phase is unknown, we normally use an incoherent “energy” estimator. Starting with the latter, squared fringe visibility V^2 can be estimated using the four-bin algorithm as (see Colavita 1999)

$$V^2 = \frac{\pi^2}{2} \frac{\langle X^2 + Y^2 - \text{Bias} \rangle}{\langle N \rangle^2} \quad (7.1)$$

where the brackets refer to averaging over a number of samples, with the sample time chosen to freeze the atmospheric fringe motion. The standard deviation of the V^2 estimator, σ_{V^2} , can be calculated from the fourth-order statistics of Gaussian read noise and Poisson photon

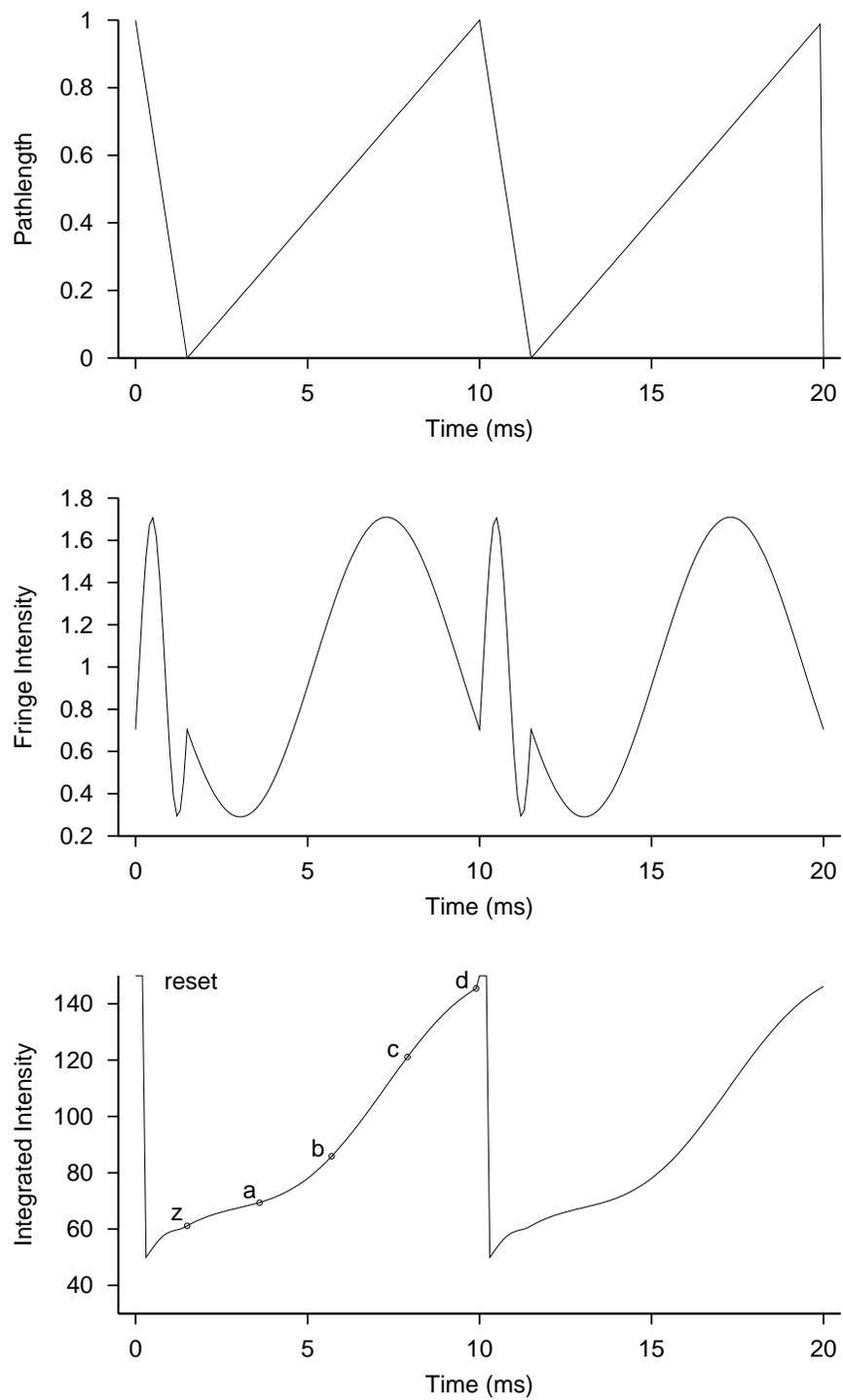


Figure 7.1: Fringe scanning with a sawtooth waveform at PTI.

noise processes. When photon-noise limited, the signal-to-noise ratio ($1/\sigma_{V^2}$) is

$$\frac{1}{\sigma_{V^2}} \propto \begin{cases} \sqrt{N}, & N \gg 1 \\ N, & N \ll 1 \end{cases}. \quad (7.2)$$

When read-noise limited (with read noise σ_{rn}), the signal-to-noise ratio is given by

$$\frac{1}{\sigma_{V^2}} \propto \left(\frac{N}{\sigma_{\text{rn}}} \right)^2, N \ll \sigma_{\text{rn}}^2. \quad (7.3)$$

Thus, as one becomes photon-noise starved or read-noise limited, the signal-to-noise ratio drops precipitously with decreasing light level as shown in Figure 7.2. The figure plots the signal-to-noise ratio for a single sample; the signal-to-noise ratio improves as the square root of the number of samples comprising the averages in Equation 7.1. However, especially when read noise is involved, there is a fairly steep wall, beyond which unreasonable numbers of samples are required to obtain a good final signal-to-noise ratio.

As mentioned above, if we know the fringe phase, we can use a coherent estimator. If we know the fringe phase in real-time, we can use phase referencing to increase the coherent integration time. If we know it a posteriori, we can use it to de-rotate and sum the fringe quadratures before computation of the visibility. Both approaches help move the detection problem away from the photon-starved or detector-noise-limited regimes.

Figure 7.3 illustrates the potential gain of coherently combining the fringe quadratures in groups of 10 samples before computing V^2 with Equation 7.1. Note that for high signal-to-noise ratios, coadding provides no benefit. More detail on this problem, and a number of references, are given by Colavita (1999).

7.4 Estimator Biases

For an accurate estimate of visibility, it is necessary to compensate for biases attributable to background, dark current, and detector imperfections. Also, as the V^2 estimator is an energy estimator, we must correct for biases attributable to squaring quantities which include noise. Bias-corrected visibility estimators are the first step in a data-calibration pipeline, which will also include periodic observation of calibrator objects.

The first set of V^2 biases are the offsets, or zero points, of the estimates of X , Y , and N ; these are typically calibrated through interspersed background measurements, i.e., measurement of the quantities when pointed at dark sky. While in principle X and Y should have no offsets, even in the presence of a finite background, in practice detector nonlinearities and reset tails introduce small offsets requiring calibration.

The second set of biases are the variances of the underlying Poisson photon-noise and Gaussian detector-noise processes which arise when squaring the fringe quadratures to compute the numerator of Equation 7.1. The photon noise bias takes the form $\langle X^2 + Y^2 \rangle =$

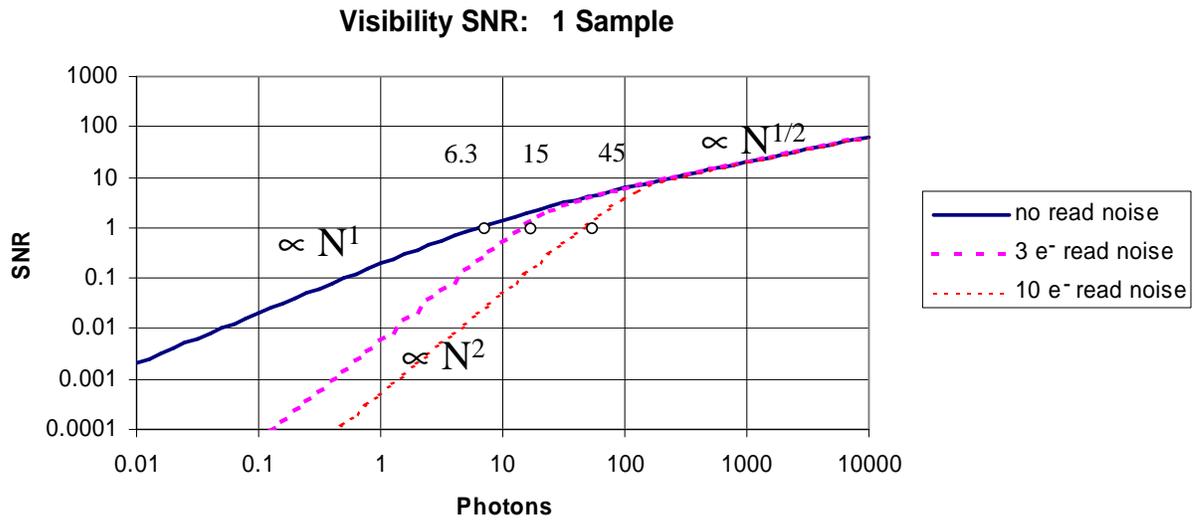


Figure 7.2: Signal-to-noise ratio of the V^2 estimator for 1 sample vs. photons per sample, for different amounts of read noise. 6.3, 15, and 45 electrons per sample are required to achieve an signal-to-noise ratio of 1 for the case of 0, 3, and 10 electrons read noise.

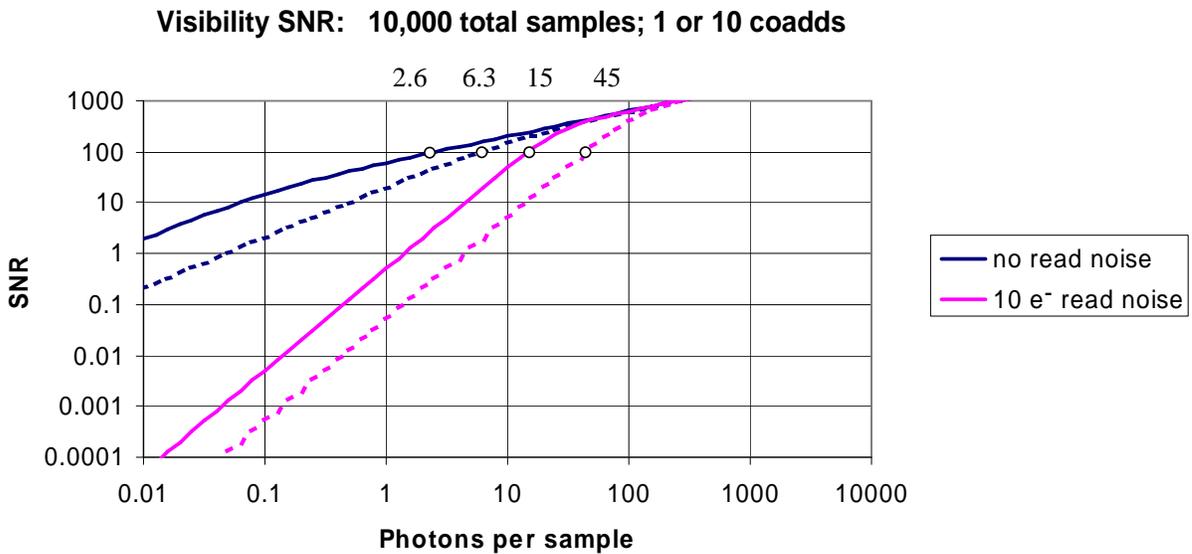


Figure 7.3: Signal-to-noise ratio of the V^2 estimator vs. photons per sample after averaging for 10,000 total samples, with and without coadding the phasors into groups of 10 samples before computing V^2 . With no read noise, a signal-to-noise ratio of 100 is achieved in 10,000 samples with 6.3 and 45 electrons per sample for the case of 0 and 10 electrons read noise, and with 2.6 and 15 electrons per sample when the phasors are coadded to groups of 10 before computing the visibility.

$k\langle N \rangle$, where k is the gain of the signal chain (counts/electron), with X , Y , and N in detector units. The read-noise bias takes the form $\langle X^2 + Y^2 \rangle = 4k^2\sigma_{\text{rn}}^2$, where the read noise variance is for a single double-correlated read (a four-bin algorithm has been assumed). These biases can be calibrated several ways: one approach is to determine the detector-noise term from $\langle X^2 + Y^2 \rangle$ as measured on dark sky, and the photon-noise term from the measured value of $\langle N \rangle$ and an estimate of k . The gain k can be readily determined from measurements of $\langle X^2 + Y^2 \rangle$ at two different light levels (see Colavita 1999).

It's important to note that, especially at low light levels, errors in measuring these variance terms can become the dominant error in estimating visibility, as the signal-to-noise ratio of these calibrations is the same as that of the measurements they are being used to calibrate. Thus observation planning must allocate sufficient integration time for the calibration measurements.

This topic is discussed in more detail in the context of PTI by Colavita (1999).

7.5 Atmospheric Biases

I can do no more than briefly touch on the effects of atmospheric biases; for more detail, see, for example Tango and Twiss (1980).

Atmospheric biases are challenging to calibrate to high precision as the atmospheric statistics are not stationary. The approach used by most groups is to minimize these biases through design or observing strategy; apply modest modeling, when appropriate, to compensate for first-order effects; and use rapid observations of calibrators to track the residual visibility variations.

7.5.1 Spatial Wavefront Errors

For slow guiding, the visibility reduction attributable to wavefront errors can be written

$$\langle V^2 \rangle = \exp[-2.06(d/r_0)^{5/3}], \quad (7.4)$$

where d is the subaperture diameter and r_0 is the atmospheric coherence diameter. This term can be reduced through fast guiding, or through the use of adaptive optics (AO). Except for a factor of 2 in the exponent, the Strehl of an AO system follows a similar expression, and the AO literature addresses this problem in detail (cf. Tyson 1997).

However, for interferometers, which usually process only a single spatial mode per aperture, the option exists to apply modal filters to correct the visibility reduction attributable to spatial wavefront errors. Single-mode optical fibers are often used for this purpose. They can be used either to filter the combined light from the two apertures, or to implement a fully fiber-optic beam combiner. As the fibers select only a single spatial mode, the visibility (attributable to wavefront errors only) is given by the scintillation formula $V^2 =$

$4I_1I_2/(I_1 + I_2)^2$, where I_1 and I_2 are the coupled intensities from each aperture into the fiber(s), which vary with fluctuations in the instantaneous wavefront distortion. For a post-combination fiber, one can show that only the average ratio of $\langle I_1 \rangle / \langle I_2 \rangle$ is needed to calibrate the residual V^2 reduction (Shaklan *et al.*, 1992). While it can be hard to get an accurate simultaneous measurement with this architecture, it is often adequate to use a nearby calibrator, assuming integration times long enough that only the systematic part of the ratio remains. Alternatively, this term can be calibrated through the use of a fully single-mode combiner, which simultaneously samples the coupled intensities from each aperture (Coudé du Foresto, 1994).

As an example, Figures 7.4 and 7.5 present data from PTI plotting V^2 vs. time for a number of sources during a night. Each point represents a 25-sec average. Of note is the difference in visibility between the broadband white-light channel, which is not spatially filtered, and the spectrometer channel (actually, a composite estimator of all of the spectrometer channels), which uses a post-combination single-mode spatial filter. On this night, use of the fiber doubles the measured V^2 .

7.5.2 Temporal Errors

For fringe tracking that is slow compared with the sample time (generally a good approximation), the visibility reduction attributable to fringe motion can be written

$$\langle V^2 \rangle = \exp[-(T/T_{0,2})^{5/3}], \quad (7.5)$$

where T is the coherent integration time and $T_{0,2}$ is the two-aperture variance-definition coherence time. While there are no modal filters applicable to temporal blurring, some compensation for this visibility reduction can be applied using contemporaneous measurements of the fringe phase motion. In particular, a visibility correction of the form $V^2 \rightarrow V^2 \exp(C_0 \sigma_{\Delta\phi}^2)$ can be applied, where $\sigma_{\Delta\phi}^2$ is the measured phase jitter during the observation interval, and C_0 is a scale factor which can be derived from an atmospheric model (see Colavita 1999).

7.5.3 Finite Coherence

The V^2 envelope with respect to delay x is just the magnitude squared of the Fourier transform of the system bandpass. For a rectangular bandpass,

$$V^2 \propto \text{sinc}^2\left(\frac{\pi x}{R\lambda}\right), \quad (7.6)$$

where λ is wavelength and R is the spectrometer resolution $R = \lambda/\Delta\lambda$. Thus for accurate visibility measurements, one must calibrate the shape of the envelope, or work in a narrow band. At PTI, we have found it convenient to combine the visibility estimates from several narrow-band spectrometer channels to synthesize a wide-band channel with a long effective coherence length.

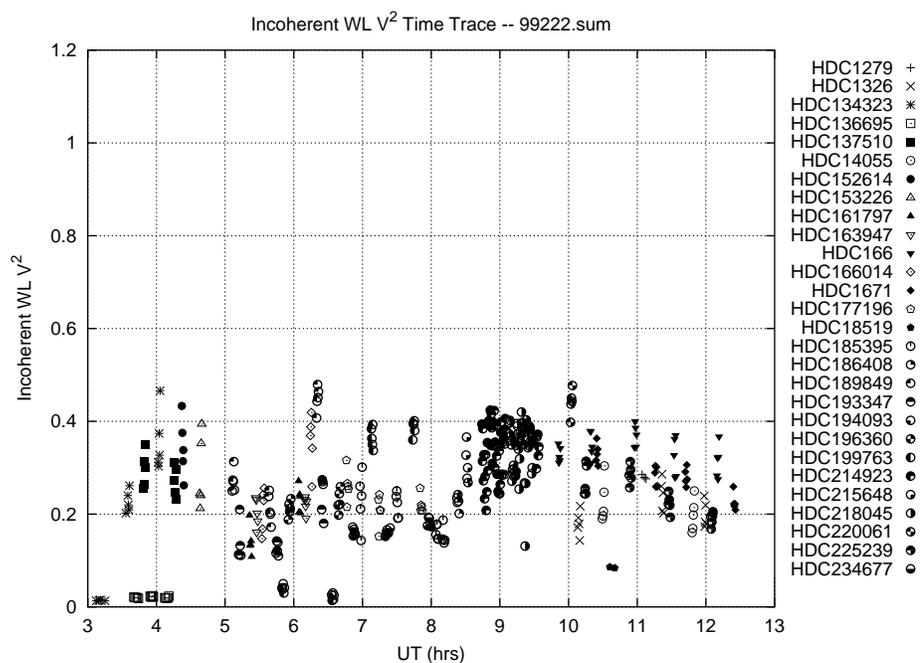


Figure 7.4: V^2 values from the white-light channel on PTI vs. time. Each point represents a 25-sec average.

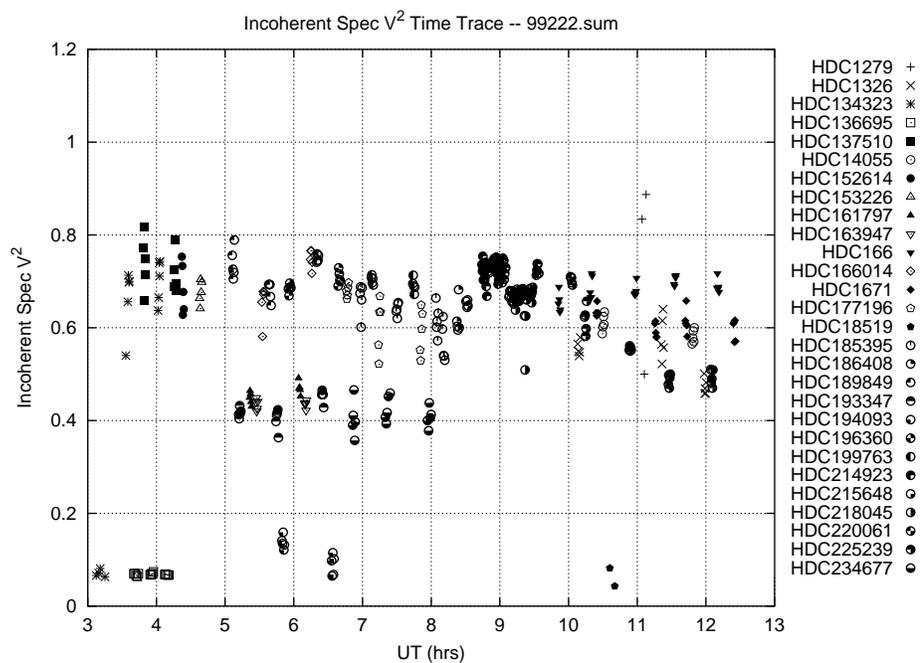


Figure 7.5: V^2 values from the spectrometer channel on PTI vs. time. Each point represents a 25-sec average of the composite spectral value.

7.5.4 Mismatched Stroke

Systematic errors accrue if the length of the pathlength-modulation stroke doesn't match the wavelength. This introduces errors in both the fringe phase and visibility estimates. However, it can be corrected through a transformation applied to the measured fringe quadratures (Colavita *et al.*, 1999).

7.6 Conclusion

With a two-element interferometer, the amplitude of the complex fringe visibility is the primary observable. By converting spatial fringe patterns to temporal ones, we reduce the measurement to a matched-filter problem that is readily solved. While we typically average a number of short exposures to provide a good visibility estimate, at low light levels the estimator statistics change from \sqrt{N} to N or N^2 , establishing an effective limiting magnitude. This magnitude can be increased through larger apertures with AO and longer coherent integration times with phase referencing. For the case of a fringe-tracking interferometer where the signal-to-noise limit is attributable to the use of narrow channels on the spectrometer side of the system, coherent visibility estimators can be used to improve the visibility estimates. No matter what type of estimator is used, accurate correction must be applied for biases in the estimator itself, as well as biases attributable to atmospheric and systematic errors.

Acknowledgments

This work was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

- A.F. Boden, C.D. Koresko, G.T. van Belle, M.M. Colavita, P.J. Dumont, J. Gubler, S.R. Kulkarni, B.F. Lane, D. Mobley, M. Shao, J.K. Wallace, and G.W. Henry, "The visual orbit of ι Pegasi," *Astrophys. J.* **515**, 356–364 (1999).
- M.M. Colavita, "Visibility estimators for the Palomar Testbed Interferometer," *Pub. Astron. Soc. Pac.* **111**, 111–117 (1999).
- M.M. Colavita, J.K. Wallace, B.E. Hines, Y. Gursel, F. Malbet, D.L. Palmer, X.P. Pan, M. Shao, J.W. Yu, A.F. Boden, P.J. Dumont, J. Gubler, C.D. Koresko, S.R. Kulkarni, B.F. Lane, D.W. Mobley, and G.T. van Belle, "The Palomar Testbed Interferometer," *Astrophys. J.* **510**, 505–521 (1999).
- V. Coudé du Foresto, "Integrated optics in astronomical interferometry," in *Very High Angular Resolution Imaging*, J.G. Robertson and W.J. Tango, eds., Proc. IAU Symp. 158, 261–271 (Dordrecht: Kluwer Academic, 1994).

- D. Mozurkewich, K.J. Johnston, R.S. Simon, P.F. Bowers, R. Gaume, D.J. Hutter, M.M. Colavita, M. Shao, and X.P. Pan, “Angular diameter measurements of stars,” *Astron. J.* **101**, 2207–2219 (1991).
- S.B. Shaklan, M.M. Colavita, and M. Shao, “Visibility calibration using single-mode fibers in a long-baseline interferometer,” in *High Resolution Imaging by Interferometry II*, J.M. Beckers and F. Merkle, eds., ESO Conf. and Workshop **39**, 1271–1283 (Garching, Germany: European Southern Observatory, 1992).
- M. Shao, M.M. Colavita, B.E. Hines, D.H. Staelin, D.J. Hutter, K.J. Johnston, D. Mozurkewich, R.S. Simon, J.L. Hershey, J.A. Hughes, and G.H. Kaplan, “The Mark III stellar interferometer,” *Astron. Astrophys* **193**, 357–371 (1988).
- W.J. Tango and R.Q. Twiss, “Michelson stellar interferometry,” *Prog. Opt.* **17**, 239–277 (1980).
- R.K. Tyson, *Principles of Adaptive Optics* (New York: Academic Press, 1997).

Chapter 8

Phase and Group Delay Estimation

PETER R. LAWSON

JET PROPULSION LABORATORY
PASADENA, CALIFORNIA

In this Chapter I would like to describe the methods that are used for phase and group-delay estimation, and to outline the expected performance of each technique. The methods of phase estimation that are described here have been applied (with variations) at the Mark III interferometer (Shao *et al.*, 1988), the Palomar Testbed Interferometer (PTI) and the Navy Prototype Optical Interferometer (NPOI). Likewise, the methods of group-delay tracking to be described here, although different in each case, have been applied at PTI (Colavita *et al.*, 1999), NPOI (Benson *et al.*, 1998), the Sydney University Stellar Interferometer (Davis *et al.*, 1995; Lawson, 1995), and the Grand Interféromètre à 2 Télescopes (Koechlin *et al.*, 1996).

There are other methods of fringe detection and measurement, which do not rely on phase or group delay estimation as described in this chapter. These other approaches are methods of *coherence envelope* tracking and detect the location of the fringe packet by sweeping or scanning the delay line back and forth through the entire fringe envelope, with a throw several times larger than the coherence length. Such approaches have been used successfully at SUSI* (Davis *et al.*, 1999), COAST (Baldwin *et al.*, 1994), and IOTA (Traub, 1998). The advantage of envelope tracking is that it is straightforward to implement. The sweep is generally made much larger than the coherence envelope and therefore the envelope need only be roughly centered in the sweep. If the baseline solution is accurate and the path variations introduced by the atmosphere are small, corrections to the tracking position

*The method used at SUSI, described by Davis *et al.* (1999), does not use a fast sweep but instead uses steps through the coherence envelope and a method of fringe measurement first described by Tango and Twiss (1980).

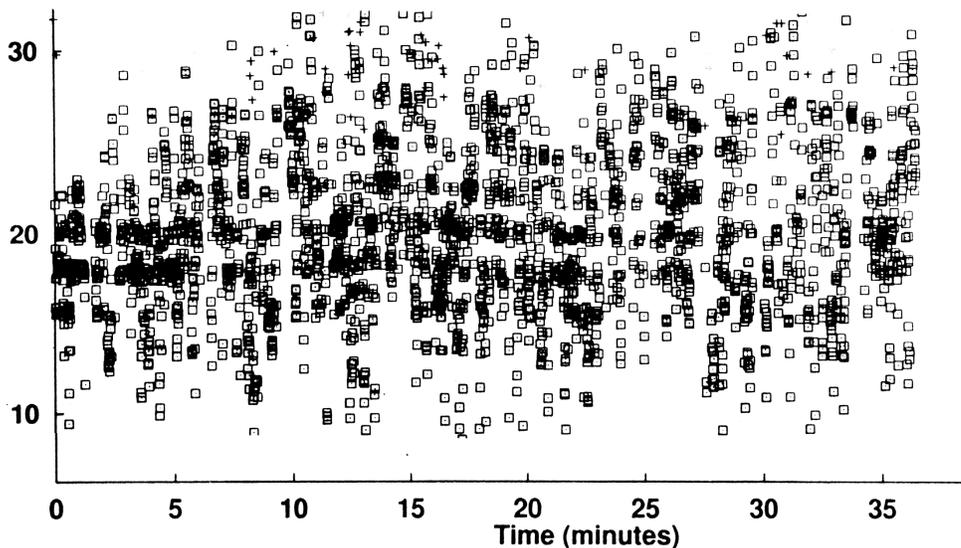


Figure 8.1: Group delay as a function of time observed with PTI in conditions of poor seeing. Several parallel fringes are visible in the plot, illustrating the behavior of the phase-tracking algorithm as it makes mistakes in identifying the central fringe. This information can be used to correctly unwrap the phase during post-processing of the data.

may be made infrequently—perhaps every few minutes. However, the fringes are only sampled for a fraction of the observation time, and the process may be labour intensive, time consuming, and inefficient.

In the following I will describe methods of phase and group delay estimation that allow the fringes to be observed on or near the peak of the coherence envelope. Methods of envelope tracking will not be considered further.

8.1 Motivation

All visibility measurements must be made in such a way that calibration of the data is later possible by observing unresolved reference sources. The visibility loss experienced by fringes observed in source and calibrator should be the same in each case, and fringes should therefore be measured at the same fixed position on the coherence envelope. Without some form of servo control the coherence envelope would move during an observation, because of errors in the astrometric model and random path variations ($\sim 10 \mu\text{m}$ rms per meter of baseline) introduced by the atmosphere. If the observations are to be made always at the same place on the coherence envelope (e.g. at the peak) some form of phase or group-delay tracking must be used.

Astrometric interferometers estimate the relative angular position of celestial objects through accurate measurements of the delay required to obtain fringes on each source. All astrometric interferometers currently use a combination of phase and group delay estimation to provide accurate delay measurements. Phase estimation is used to provide the highest resolution, and group-delay estimation is used so that the phase is unwrapped with reference to the peak of the coherence envelope, as illustrated in Figure 8.1.

Imaging interferometers that estimate closure phase must be capable of simultaneous phase measurements over at least three baselines. In practice this is accomplished by imposing a different spatial or temporal modulation to each baseline and then estimating phase through Fourier techniques or phase-tracking algorithms. Although phase tracking is not always necessary for imaging, as shown for example by the COAST interferometer (Baldwin *et al.*, 1994), phase-estimation methods are nonetheless used to estimate closure phase.

8.2 Phase and Group Delay

The optical path difference between the combined wavefronts in an interferometer can be expressed in terms of the indices of refraction of the different media n_i and the path lengths in each arm of the interferometer that the light traverses, x_{1i} and x_{2i} :

$$x(\kappa) = \sum_{k=0}^K (x_{1i} - x_{2i}) n_i(\kappa) \quad (8.1)$$

where $\kappa = 1/\lambda$ is the spectroscopic wavenumber at a wavelength λ .[†] If we now make the change of variables $x_i = x_{1i} - x_{2i}$, and assume that we have a vacuum delay x_0 and K dispersive media, we have that the phase of the fringes can be given by

$$2\pi\kappa x(\kappa) = 2\pi\kappa \left[x_0 + \sum_{k=1}^K x_i n_i(\kappa) \right] \quad (8.2)$$

In general the optical path-difference x is wavelength dependent, because the light that travels to the beam combiner from each arm of the interferometer may have passed through different paths in vacuum, and different dispersive pathlengths in air and glass: at shorter wavelengths the index of refraction is higher, light travels slower through the media, and the optical path-difference is larger. If we consider an interferometer that only admits a restricted bandwidth of light, there is light of a particular wavelength that arrives “first” and one that arrives “last,” and we can speak of the ensemble of waves as traveling as a

[†]Here I distinguish between the *spectroscopic* wavenumber, $\kappa = 1/\lambda$, and the wavenumber that is customarily used in optics, $k = 2\pi/\lambda$. This distinction is suggested by Born and Wolf (1980) and will be used throughout this chapter.

group or packet whose mean progress is described by a group velocity. The group delay is proportional to the rate-of-change of phase as a function of wavenumber, evaluated at the center of the band.

$$\text{group delay } (\kappa_0) = \left. \frac{d}{d\kappa} \kappa x(\kappa) \right|_{\kappa_0} \quad (8.3)$$

$$= x_0 + \sum_{k=1}^K \left. \frac{d}{d\kappa} \kappa x_i n_i(\kappa) \right|_{\kappa_0} \quad (8.4)$$

If we have simply a vacuum path-difference, $x(\kappa) = x_0$, the group delay is independent of wavelength, and the fringe phase is a linear function of wavenumber.

A phase-tracking algorithm would seek the location of a position of constant phase, where the fringe visibility is highest. A group-delay tracking algorithm would seek a location of constant group delay—where the number of fringes across the bandwidth is maintained constant.

A long-baseline stellar interferometer must typically compensate for a *vacuum* delay of several tens of meters and random variations of air path of several tens of microns rms. In principle, all ground-based interferometers should use vacuum delay lines, but many do not either because of funding restrictions or the perception that longitudinal dispersion is less of a problem at infrared wavelengths. Examples of fringes distorted by longitudinal dispersion have been presented by Lawson (1997) and include an illustration of the wavelength dependence of group delay. If the dispersion is uncompensated then fringes will appear to have a reduced visibility and fringes in different wavelength bands will arrive delayed one from the other, making it impossible to record fringes in two or more bands simultaneously. The effects of dispersion have been studied by numerous authors (Lacasse and Traub, 1988; Tango, 1990; ten Brummelaar, 1995; du Foresto *et al.*, 1995; Lawson and Davis, 1996; Léveque *et al.*, 1996; Davis *et al.*, 1998; Daigne and Lestrade, 1999) and dispersion compensators have been implemented, mostly on an experimental basis, at the I2T, SUSI, and PTI. Dispersion compensators will be used with the GI2T/REGAIN interferometer, the Keck Interferometer, the VLTI, and the CHARA Array—all of which use air delay lines.

It is interesting to note that if dispersion is present the fringe phase at the peak of the coherence envelope (at zero group delay) will not necessarily be zero. As the delay line introduces an increasing dispersive air path, the brightest fringe in the coherence envelope (the central fringe) will move away from the position of zero group delay. When it has moved more than half a wavelength, the fringe that followed it will be closer to the peak of the envelope and will become the new central fringe. The position of zero-phase and zero-group-delay will appear to move with respect to each other in a sawtooth pattern as a function of delay.

8.3 Model of the Fringe

The intensity of a fringe pattern can be expressed (cf. Chapter 7, Born and Wolf 1980) as follows:

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} |\gamma_{12}| \cos(\phi_{12} - \varphi), \quad (8.5)$$

where

$$\varphi = \frac{2\pi}{\lambda}(s_2 - s_1). \quad (8.6)$$

λ is the wavelength of the light, I_1 and I_2 are the intensities of the light in each arm of the interferometer, and γ_{12} is the complex degree of coherence with modulus $|\gamma_{12}|$ and argument $\phi_{12} - \varphi$, where φ arises from the path difference, and ϕ_{12} contains information about the source. The parameters s_1 and s_2 are two optical pathlengths. The contrast, or visibility, of the fringes is the ratio of the fringe amplitude to the total background illumination,

$$\mathcal{V} = \frac{2\sqrt{I_1 I_2} |\gamma_{12}|}{I_1 + I_2}. \quad (8.7)$$

If we make a change of variables introducing the spectroscopic wavenumber $\kappa = 1/\lambda$, letting

$$I_s = 2\sqrt{I_1 I_2}, \quad I_b = I_1 + I_2 - I_s, \quad (8.8)$$

and

$$x = (s_2 - s_1), \quad (8.9)$$

then we have[‡]

$$I(\kappa, x) = I_s \left[1 + |\gamma_{12}| \cos(2\pi\kappa x - \phi_{12}) \right] + I_b. \quad (8.10)$$

The substitution of $x = s_2 - s_1$ in this case is to indicate that the phase offset is entirely piston phase, with no tilt component. Under conditions where $I_1 = I_2$ then the visibility of the fringes is the modulus of the complex degree of coherence,

$$\mathcal{V} = |\gamma_{12}|. \quad (8.11)$$

A source that is *non* quasi-monochromatic may still be treated as such if it is observed with an instrumental bandwidth that is sufficiently small. Bright fringes will occur wherever the path difference x is an integer multiple of 2π at most wavelengths. This will be so at *all* wavelengths only when x is zero and when the dispersion is the same in each arm of the interferometer. The reduction in the fringe visibility at increasing values of x is described by the coherence envelope.

[‡]Although not explicitly indicated here, a phase offset should be introduced to distinguish between fringes that are produced by light traveling single or double-pass through the beam-splitter. The offset phase is $\pi/2$ if the beam-splitter is used in the normal way: two beams of starlight enter the beam-splitter from opposite sides and are combined. However, the phase offset is 0 if the source of light is an artificial star that shines out through the beam-splitter, sending two beams out which are then autocollimated and returned. This is discussed at greater length by Traub in Chapter 3 (Section 3.2.3).

8.3.1 Coherence Envelope

If we observe fringes using a finite bandwidth (ie. not quasi-monochromatic), the recorded intensity is the integral of $I(\kappa, x)$ over wavenumber, weighted by a filter function $W(\kappa)$ that describes the bandpass.

$$I(\bar{\kappa}, x) = \int_{-\infty}^{\infty} W(\kappa - \bar{\kappa}) I(\kappa, x) d\kappa, \quad (8.12)$$

where $\bar{\kappa}$ is the center of the passband.

The filter function includes both the shape of the bandpass and the frequency response of the detector; it has values that are large within the bandwidth, and near zero outside. The result of this averaging is to reduce the sensitivity of the interferometer to fringes of large delay: when the bandwidth $\Delta\kappa$ partially spans a fringe (in the wavenumber domain κ) then the visibility appears to be reduced. This is simple to illustrate.

Let us introduce a change of variables, such that $\kappa' = \kappa - \bar{\kappa}$, and perform the integration in Equation 8.12 with respect to κ' . If we insert Equation 8.10 into Equation 8.12, ignore the background I_b for now, and rearrange the terms we have

$$\begin{aligned} I(\bar{\kappa}, x) = I_s [& 1 + |\gamma| \cos(2\pi\bar{\kappa}x - \phi) \int_{-\infty}^{\infty} W(\kappa') \cos(2\pi\kappa'x) d\kappa' \\ & - |\gamma| \sin(2\pi\bar{\kappa}x - \phi) \int_{-\infty}^{\infty} W(\kappa') \sin(2\pi\kappa'x) d\kappa'] \end{aligned} \quad (8.13)$$

where the subscripts have been dropped from γ_{12} and ϕ_{12} . Now if we define $\Omega(x)$ as the Fourier transform of $W(\kappa)$, then we have:

$$\Omega(x) = |\Omega(x)| e^{j\phi_\Omega} = \int_{-\infty}^{\infty} W(\kappa) e^{j2\pi\kappa x} d\kappa, \quad (8.14)$$

and therefore

$$|\Omega(x)| \cos \phi_\Omega = \int_{-\infty}^{\infty} W(\kappa) \cos(2\pi\kappa x) d\kappa, \quad (8.15)$$

$$|\Omega(x)| \sin \phi_\Omega = \int_{-\infty}^{\infty} W(\kappa) \sin(2\pi\kappa x) d\kappa. \quad (8.16)$$

Equation 8.12 may therefore be written in the form

$$I(\bar{\kappa}, x) = I_s \left[1 + |\gamma_x| \cos(2\pi\bar{\kappa}x - \phi + \phi_\Omega) \right] \quad (8.17)$$

where the apparent visibility $|\gamma_x|$ is the product of the true visibility and the modulus of the Fourier transform of the filter function, evaluated at the current delay:

$$|\gamma_x| = |\gamma| |\Omega(x)|. \quad (8.18)$$

The transfer function $\Omega(x)$ describes the *coherence envelope*. If $W(\kappa)$ is symmetric then $\Omega(x)$ is real valued, $\phi_\Omega = 0$, and only at zero delay, where the envelope is at its peak, is the true visibility observed.

A Rectangular Bandpass

If a detector has a rectangular bandpass then its coherence envelope would resemble a sinc function:

$$W(\kappa) = \begin{cases} 0, & |\kappa| > \Delta\kappa/2 \\ 1, & |\kappa| < \Delta\kappa/2 \end{cases} \quad \text{and} \quad \Omega(x) = |\Delta\kappa| \frac{\sin \pi x \Delta\kappa}{\pi x \Delta\kappa}. \quad (8.19)$$

If a bandwidth of $\Delta\lambda$ is used at a wavelength λ , the same interval expressed in wavenumber is as follows:

$$\Delta\kappa = \frac{1}{(\lambda - \Delta\lambda/2)} - \frac{1}{(\lambda + \Delta\lambda/2)}, \quad \text{therefore} \quad \Delta\kappa = \frac{\Delta\lambda}{\lambda^2 - (\Delta\lambda/2)^2}. \quad (8.20)$$

If we assume that the fractional bandwidth is very small we can ignore the second term in the denominator.

$$\Delta\kappa \simeq \frac{\Delta\lambda}{\lambda^2}. \quad (8.21)$$

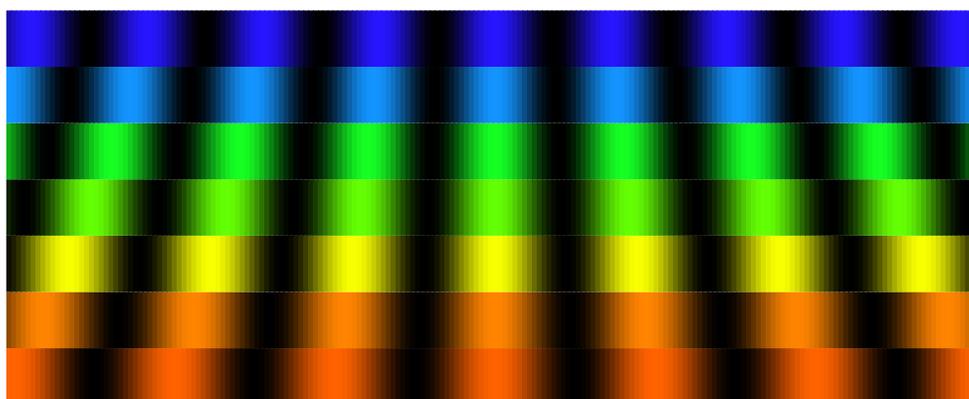
The sinc function is characterized by the location of its first zero crossing, where $x = 1/\Delta\kappa$. This distance can be thought of as the coherence length of the starlight under observation.

The Coherence Envelope and the Color of Fringes

Figure 8.2 illustrates in a more intuitive way how the coherence envelope would become narrower when fringes are observed over an increasingly large bandwidth, which encompasses many different colors (wavelengths). Fringes from a Young's double-slit experiment are shown in Figure 8.2(a), as seen through filters at different wavelengths. Figure 8.2(b) shows the corresponding fringe pattern if all the colors are viewed simultaneously. Note that as you increase the path-difference and move further away from the central fringe, either left or right from the center of the page, the fringes become less and less distinct because the fringes at each color cease to add constructively. The larger the range of colors, the sooner the fringes disappear as the pathlength is increased.

8.3.2 Channeled Spectrum

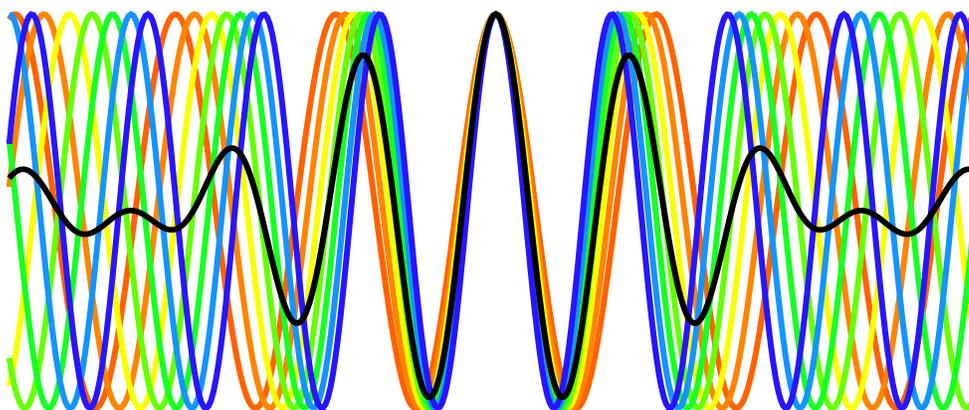
Fizeau and Foucault (1845) were the first to point out that fringes are still observable in the spectrum of interfered light, even when the path differences were so large that the white-light fringes have completely vanished. Suppose what you observed was a pattern similar to the one illustrated in Figure 8.2(b) and you somehow separated the different



(a)



(b)



(c)

Figure 8.2: Narrow band and white-light fringes: (a) Fringes from a Young's double-slit experiment, shown for seven different wavelengths or colors of light. (b) The corresponding "white-light" fringe after summing together the different colored fringes. (c) A cross section of the fringes at each color and the corresponding white-light fringe. Note that the color-wavelength relationship is not to scale. [After A.A. Michelson, *Light Waves and Their Uses* (University of Chicago Press: 1902), Plate II.]

colors to obtain the fringe patterns of Figure 8.2(a). You can now see fringes at large path differences (left and right in the figure) where none were previously visible. Which is to say, if we disperse the light in a spectrometer so that a bandwidth $\Delta\kappa$ is separated into M smaller bands of width $\Delta\kappa/M$, the coherence length for fringe detection changes from $1/\Delta\kappa$ to $M/\Delta\kappa$; it becomes M times larger. Fringes observed in dispersed light have been termed Edser-Butler fringes, or fringes of equal chromatic order, and produce a spectrum that is *channeled* with fringes (Steele, 1987).

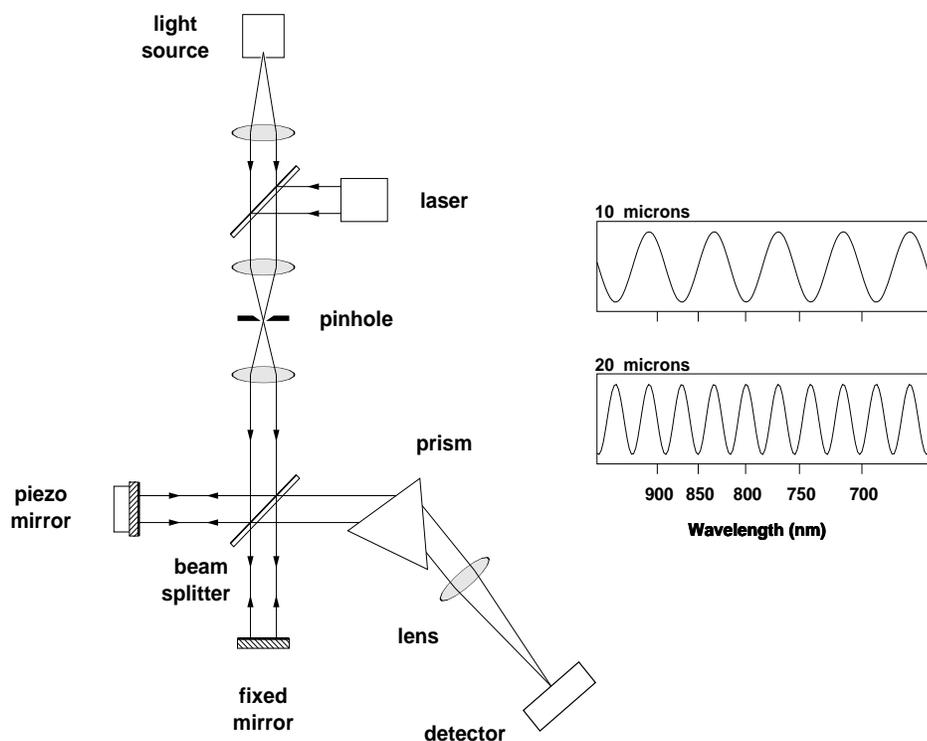


Figure 8.3: Layout of a Michelson interferometer used to produce channeled spectra. Simulations of channeled spectra are shown for path-differences of 10 and 20 μm , using a prism of BK7 glass.

We can re-write Equation 8.17 as a channeled spectrum (here arbitrarily setting ϕ_Ω to zero) in the following way:

$$f_k(x) = I_k \left[1 + |\gamma_k| \cos(2\pi\bar{\kappa}_k x - \phi_k) \right], \quad (8.22)$$

where k is an index number that counts the pixels across the spectrometer; such that at pixel k the wavelength and wavenumber are λ_k and κ_k respectively, I_k is the intensity of the stellar spectrum, $|\gamma_k|$ is the fringe visibility amplitude, and ϕ_k is the fringe phase.

As can be seen in Equation 8.22 the number of fringes in the spectrum across a given bandwidth is directly proportional to the optical path-difference. If p fringes are counted

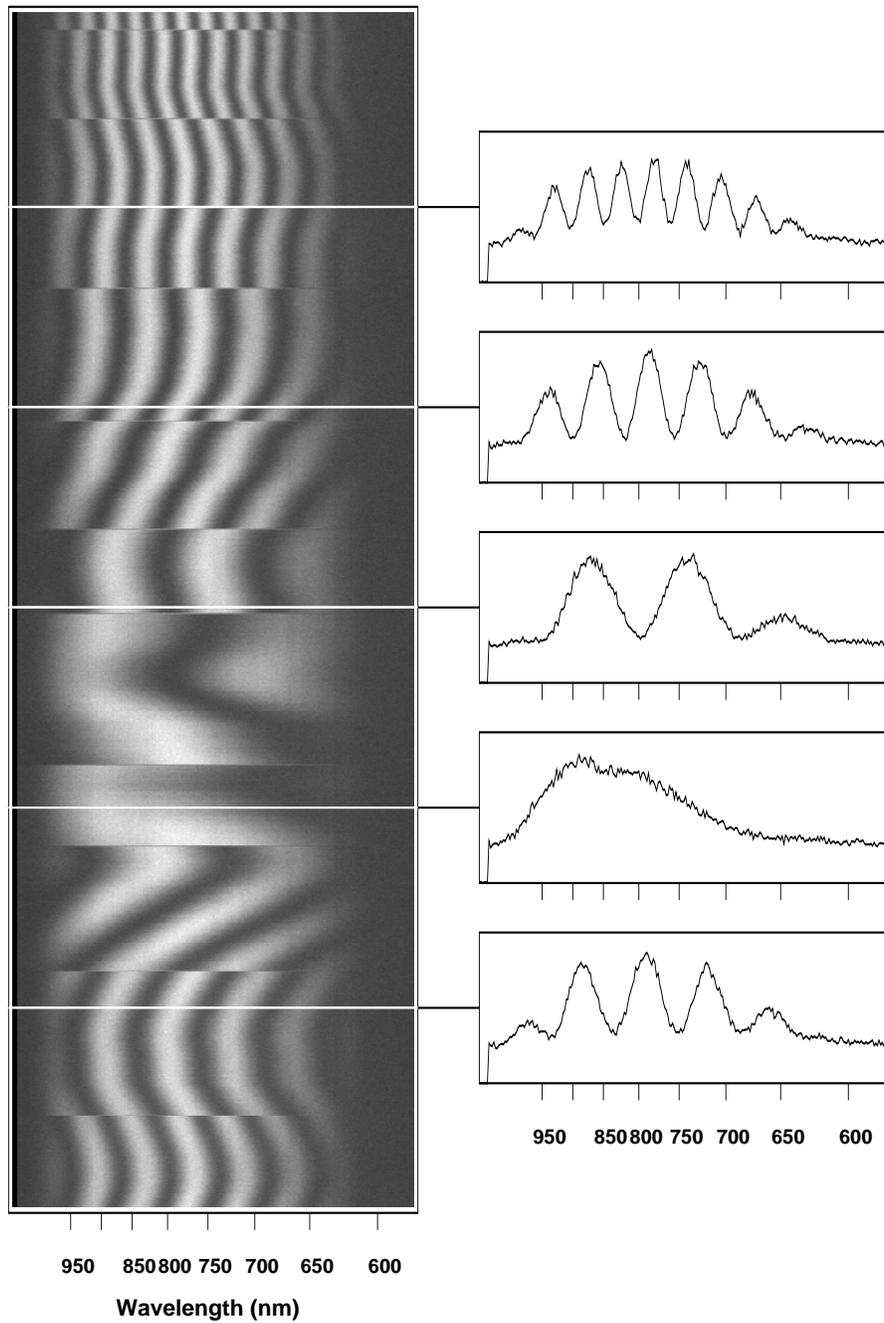


Figure 8.4: Channeled spectra produced at COAST using an artificial star. Various examples are shown as the delay line is stepped from one side of path-equality to the other.

between wavelengths λ_{\min} and λ_{\max} , then we have

$$p = \frac{1}{2\pi} \left[\frac{2\pi x}{\lambda_{\min}} - \frac{2\pi x}{\lambda_{\max}} \right], \quad (8.23)$$

and therefore

$$x = \frac{p}{\Delta\kappa}, \quad (8.24)$$

where

$$\Delta\kappa = \left[\frac{1}{\lambda_{\min}} - \frac{1}{\lambda_{\max}} \right]. \quad (8.25)$$

The optical path-difference can therefore be determined simply by estimating the number or frequency of fringes in the channeled spectrum.[§] Figure 8.3 shows a laboratory Michelson interferometer with a prism on its output. Simulated channeled spectra, as would be seen on the detector, are illustrated. Figure 8.4 shows a time sequence of actual channeled spectra recorded with a CCD camera as the path-difference is stepped from one side of zero path-difference to the other.

Numerous applications for channeled spectra have been described in the literature, including their use for analyzing spectroscopic measurements (Edser and Butler, 1898), for measuring absolute phase shift and dispersion (Sanderman, 1971), and for the analysis of thin films (Feldman, 1984).

8.4 Methods of Tracking Fringe Phase

Let us now return to a description of interference fringes, and look at the approaches that have been used to measure fringe phase with stellar interferometers.

All methods of phase-measurement interferometry involve a modulation of the fringes. Stellar interferometers that measure phase typically use a temporal modulation of the optical path-difference to sample the fringe. The modulation frequency is chosen fast enough so that (for the ground-based instruments) atmospheric path-fluctuations are effectively frozen.

We can rewrite the equation of the fringes, Equation 8.10, using the trigonometric identity

$$\cos(a - b) = \cos(a) \cos(b) + \sin(a) \sin(b), \quad (8.26)$$

and express the fringe pattern as follows:

$$I(\kappa) = \frac{1}{\tau} [N + X \cos(2\pi\kappa x) + Y \sin(2\pi\kappa x)], \quad (8.27)$$

where $\kappa = 1/\lambda$, τ is the measurement period, and

$$N = \tau I_s, \quad (8.28)$$

$$X = \tau I_s |\gamma_{12}| \cos(\phi_{12} - \phi_0), \quad (8.29)$$

$$Y = \tau I_s |\gamma_{12}| \sin(\phi_{12} - \phi_0), \quad (8.30)$$

[§]In this description, the number of fringes in the channeled spectrum only determines the distance from a zero group-delay, but not whether the offset is positive or negative. At least two measurements of channeled fringes at different delays are required to remove the ambiguity in sign.

where we have assumed the background I_b is identically zero. It can now be seen that with at least three measurements at different *known* positions of x we can solve for the three fringe parameters N , X , and Y , and that with many more measurements we can improve the accuracy of the estimate by applying a least-squares algorithm. The numerous different approaches to phase measurement interferometry are described in detail by Creath (1988). In the following it will be assumed that if the total measurement period is τ , and that if M measurements are made of the fringe, they are each made over a time τ/M .

8.4.1 Phase Tracking with Four Quarter-Wavelength Steps

For the purpose of illustration, let us assume the pathlengths in the interferometer are stepped in quarter-wavelength steps. The fringe is given, as above, by

$$I(\kappa, n) = \frac{1}{\tau} [N + X \cos(2\pi\kappa x_n) + Y \sin(2\pi\kappa x_n)], \quad (8.31)$$

and the pathlength modulation is

$$x_n = \frac{n}{4} \frac{1}{\kappa_0}, \quad n = 0, 1, 2, 3 \quad (8.32)$$

where λ_0 is the wavelength of the fringe measurement, ie: $\kappa_0 = 1/\lambda_0$. The four measurements, each integrated for a time $\tau/4$, can be written

$$A = (N + X) / 4, \quad (8.33)$$

$$B = (N + Y) / 4, \quad (8.34)$$

$$C = (N - X) / 4, \quad (8.35)$$

$$D = (N - Y) / 4, \quad (8.36)$$

and these can be reduced to the following three equations

$$A + B + C + D = N, \quad (8.37)$$

$$A - C = X/2, \quad (8.38)$$

$$B - D = Y/2. \quad (8.39)$$

We now arrive at a simple expression for the phase,

$$\phi_{12} - \phi_0 = \tan^{-1} \left(\frac{B - D}{A - C} \right). \quad (8.40)$$

The way that the A , B , C , and D counts are combined to produce phasors depends both on the form of the equation describing the fringe and the initial phase of the sweep—it is therefore not surprising that there are a confusing number of correct but somewhat different descriptions of the four-bin algorithm.

8.4.2 Phase Tracking with a Linear Pathlength Sweep

No stellar interferometer actually uses a phase-*stepping* approach. In practice the pathlength is not stepped, but is varied as a triangle or sawtooth wave with a peak-to-peak amplitude of one wavelength, $\lambda_0 = 1/\kappa_0$, or some integer number of wavelengths. Here we will consider only the case of a single-wavelength sweep. The following derivation is also described in a slightly different way by Wyant (1975, Equation 15ff) and by Colavita (1985, Chapter 4).

During the course of one cycle of duration τ , the photon counts A , B , C , and D are each recorded over a time interval of $\tau/4$. The measurement comprises separate integrations which are each an average across part of the total sweep,

$$\frac{1}{\Delta x} \int_{x_n}^{x_{n+1}} I(\kappa, n) dx \quad \text{where} \quad \Delta x = x_{n+1} - x_n. \quad (8.41)$$

If we have four separate integrations each across a quarter wavelength, then we set $\Delta x = 1/(4\kappa_0)$. We will also set the boundaries of the integration so that the total sweep is one wavelength, symmetric about a zero path-difference:

$$x_n = \frac{n}{4} \frac{1}{\kappa_0} - \frac{3}{4} \frac{1}{\kappa_0}, \quad n = 1, 2, 3, 4, 5. \quad (8.42)$$

It is straightforward to perform the integrations described by Equations 8.41, 8.42, and 8.27. We have that A is bounded by (x_1, x_2) , B is bounded by (x_2, x_3) , C is bounded by (x_3, x_4) , and D is bounded by (x_4, x_5) . We also have that

$$\int_{x_n}^{x_{n+1}} \cos(2\pi\kappa x) dx = \frac{1}{2\pi\kappa} [\sin(2\pi\kappa x_{n+1}) - \sin(2\pi\kappa x_n)], \quad (8.43)$$

and

$$\int_{x_n}^{x_{n+1}} \sin(2\pi\kappa x) dx = \frac{1}{2\pi\kappa} [\cos(2\pi\kappa x_n) - \cos(2\pi\kappa x_{n+1})]. \quad (8.44)$$

We can now write the A , B , C , and D as follows:

$$A = \frac{N}{4} + X \frac{1}{2\pi} [\sin(-\pi/2) - \sin(-\pi)] + Y \frac{1}{2\pi} [\cos(-\pi) - \cos(-\pi/2)], \quad (8.45)$$

$$B = \frac{N}{4} + X \frac{1}{2\pi} [\sin(0) - \sin(-\pi/2)] + Y \frac{1}{2\pi} [\cos(-\pi/2) - \cos(0)], \quad (8.46)$$

$$C = \frac{N}{4} + X \frac{1}{2\pi} [\sin(\pi/2) - \sin(0)] + Y \frac{1}{2\pi} [\cos(0) - \cos(\pi/2)], \quad (8.47)$$

$$D = \frac{N}{4} + X \frac{1}{2\pi} [\sin(\pi) - \sin(\pi/2)] + Y \frac{1}{2\pi} [\cos(\pi/2) - \cos(\pi)], \quad (8.48)$$

which may be simplified to

$$A = \frac{N}{4} + \frac{1}{2\pi} (-X - Y), \quad (8.49)$$

$$B = \frac{N}{4} + \frac{1}{2\pi} (X - Y), \quad (8.50)$$

$$C = \frac{N}{4} + \frac{1}{2\pi}(X + Y), \quad (8.51)$$

$$D = \frac{N}{4} + \frac{1}{2\pi}(-X + Y). \quad (8.52)$$

$$(8.53)$$

We can now write

$$\frac{B - D}{A - C} = \frac{Y - X}{X + Y}. \quad (8.54)$$

Dividing through by X and using the relationship $Y/X = \tan(\phi_{12} - \phi_0)$, we have

$$\frac{B - D}{A - C} = \frac{\tan(\phi_{12} - \phi_0) - 1}{1 + \tan(\phi_{12} - \phi_0)}. \quad (8.55)$$

Using the trigonometric relationship

$$\tan(a - b) = \frac{\tan a - \tan b}{1 + \tan a \tan b}, \quad (8.56)$$

we have finally that

$$\tan^{-1}\left(\frac{B - D}{A - C}\right) = \phi_{12} - \phi_0 - \frac{\pi}{4}, \quad (8.57)$$

$$\phi_{12} - \phi_0 = \tan^{-1}\left(\frac{B - D}{A - C}\right) + \frac{\pi}{4}. \quad (8.58)$$

The phase shift of $\pi/4$ corresponds to a lag of half an integration bin, as would be expected.

Although the minimum number of bins that one could use would be three, algorithms with more than four bins are also in use. For example, the Navy Prototype Optical Interferometer uses an eight-bin algorithm where the real and imaginary components of the phase are calculated as follows (Benson, 1998):

$$\begin{aligned} \cos(\phi) &\propto (A - E) + 0.5\sqrt{2}(B - D - F + H), \\ \sin(\phi) &\propto (C - G) + 0.5\sqrt{2}(B + D - F - H), \end{aligned}$$

with the integration bins now extending from A through H . The phase offset in this case would be $\pi/8$.

Cassaing *et al.* (2000) have suggested that a servo that only seeks to find the zero-phase position, and does not share data with the science instrument, need only estimate the sine phasor, as that would suffice for a zero-seeking servo.

8.4.3 Simultaneous Phase Measurements at Several Wavelengths

Astrometric interferometers often make phase measurements at several different wavelengths simultaneously. The actual modulation is chosen to be one wavelength at the longest wavelength where data is measured. For data at shorter wavelengths, the data acquisition is halted briefly when the modulation exceeds one wavelength, and then resumed when it returns. The boundaries of the A , B , C , and D bins are re-defined according to the timing appropriate at each wavelength.

8.4.4 Triangle vs Sawtooth Modulations

A sawtooth waveform is preferable when the detector being used to record the fringe has appreciable readout noise.

A triangle wave produced by a moving piezo will have non-linearities that are different in its up and down-strokes. These non-linearities will produce a bias in the fringe phase measurement—a phase measured using an up-stroke will have a different bias than a phase measured using a down-stroke. The problem arises that all phase estimates should have the *same* bias, so that phase-difference measurements (used for the astrometry) are unbiased.

At the Mark III interferometer, each phase estimate was made as an average between a measurement made on the up-stroke and a measurement made on the down-stroke. All phase estimates therefore had the same bias. In practice several cycles of up and down-strokes were averaged if the atmospheric conditions permitted. Photomultiplier tubes were used that photon-counted without incurring read noise.

At the Palomar Testbed Interferometer, the detector that is used is a NICMOS III detector. It has very high read noise, and even with multiple non-destructive reads, the read noise is about 12 electrons rms. Each phase estimate should use the absolute minimum number of reads necessary, so that the highest signal-to-noise is achieved in each cycle. A sawtooth waveform is therefore used. It has only an up-stroke, and so every phase measurement has the same bias—and only four reads (one measurement of A , B , C , D) are used rather than eight reads per phase estimate.

8.5 Methods of Tracking Group-Delay

Group-delay tracking has a very long history of use in stellar interferometry. Michelson and Pease (1921) applied this technique to acquire fringes by eye with the 20-ft interferometer using a direct-view prism. Labeyrie (1975) used an identical approach when he demonstrated that fringes could be acquired with two separated telescopes. The I2T interferometer acquired fringes this way up until about 1984, and the GI2T also routinely used a direct-view prism up until about 1995.

The idea of applying this technique using photon counting detectors was no doubt obvious to Labeyrie and was also suggested by Tango and Twiss (1980). The probable limitations of group-delay tracking have been described by numerous authors since then. These include simulations for the IOTA interferometer performed by Nisenson and Traub (1987) and Traub *et al.* (1990); simulations for the COAST interferometer performed by Buscher (1988); signal-to-noise predictions for “photon-starved” operation with the Mark III interferometer (Shao *et al.*, 1988); simulations for SUSI considering photon noise only (Lawson, 1995); and further simulations taking into account detector read noise (ten Brummelaar, 1997) and visibility fluctuations (Lawson *et al.*, 1999). Simulations have also shown that

with the use of *a priori* knowledge and Bayesian analysis methods that performance limits could be further extended (Meisner, 1996; Padilla *et al.*, 1998; Morel and Koechlin, 1998).

Channeled spectra were recorded by Kim (1989) using a PAPA camera at the Mark III interferometer, although without implementing a servo loop. Similar observations were carried out by the author (Lawson, 1994) at the Sydney University Stellar Interferometer, along with observations of atmospheric path fluctuations seen in channeled spectra (Davis *et al.*, 1995). In 1994 the GI2T began to automate a low-bandwidth servo for pathlength control using dispersed fringes (Koechlin *et al.*, 1996). When the Palomar Testbed Interferometer was commissioned in 1995, it automated a method of group-delay tracking using phasor measurements (Colavita *et al.*, 1999), as did the NPOI at about the same time (see for example Benson *et al.* 1998 and Hummel 2000). Future applications of group-delay tracking with space-borne interferometers have been reviewed by Shao and Colavita (1992).

Approaches to Group-Delay Estimation

The group delay can be measured if the combined beams from an interferometer are dispersed in a spectrometer. The detected spectrum of the star will be channeled with fringes whose number is proportional to the optical path-difference.

Approaches to group-delay tracking can be broadly classified according to the form of the measurement and the type of data processing that is used. The measurement will be either of a single channeled spectrum or of fringe phasors.*

1. *Channeled spectrum*: single snap-shot of the stellar spectrum channeled with fringes, but no pathlength modulation (e.g. Lawson 1995, with delay estimation from a transform of real-valued data).
2. *Multi-wavelength phasor measurements*: fringe phasors recorded at multiple wavelengths, with pathlength modulation and methods of phase measurement interferometry (e.g. Colavita *et al.* 1999; Armstrong *et al.* 1998, with delay estimation from a transform of a series of complex numbers).

Approaches to Data Reduction

The approaches to data processing for group-delay tracking could include a cross-correlation with an optimal filter, a least-squares modeling of the data, or a method of power spectrum analysis. Each of these approaches would use a simplified model of the fringe, based on a small set of free parameters, such as fringe frequency, amplitude, phase, and some

*This distinction is perhaps artificial. What I have called the *fringe phasor* approach, would be an n -bin method of phase measurement applied at multiple wavelengths, for which n snap-shots of channeled spectra are required. Although the sign of the delay cannot be determined from a single channeled spectrum, with two or more channeled spectra recorded with a known shift in delay, as is the case with measuring the fringe phasors, the sign is straightforward to determine.

assumptions—including the relationship between detector pixels and wavelength on the spectrometer. The utility of a particular technique depends on how well its assumptions model the data. For instance, the fast Fourier transform (FFT) assumes that the data represent a series of harmonically related sinusoids, but if there exists one sinusoid that does not coincide with any of the harmonics, then the corresponding power spectrum will be poorly reconstructed. Likewise, if the data were sampled at irregular intervals then the sampling will also bias the group delay.

Advances in spectrum analysis have come from deriving power spectra from more accurate assumptions, as is illustrated in the review paper by Kay and Marple Jr. (1981). Spectacular improvements are possible if the proper model is chosen and a strong signal is present. Unfortunately, all of these methods will fail when they are asked to derive power spectra from processes that deviate from their model. This can occur in some cases simply by adding observation noise to the data: at low signal-to-noise levels the resolution is often no better than an FFT approach, and consequently many of the methods are ill adapted for real-time processing. While it may be possible to determine the parameters that describe the power spectrum, one must then recalculate the power spectrum numerous times to locate the fringe peak, performing lengthy calculations. This has meant that only the relatively prosaic FFT and several simple variations of least-squares methods have been used in stellar interferometry. These will now be described.

8.5.1 Channeled Spectrum: Fast Fourier Transform

The most straightforward method is to use, despite its limitations, is the fast Fourier transform. If we choose the FFT to process the data, we can cast the problem in terms of an estimate of visibility using the Discrete Fourier Transform (DFT) with the same model for the fringes used by Walkup and Goodman (1973).

If we can assume that wavenumber is mapped linearly onto the detector then an FFT could be used. If the mapping is non-linear then a DFT would be used with the actual wavenumbers corresponding to each sample in the spectrometer. Let us look at how an FFT would be implemented.

Linear Mapping

If wavenumber κ is mapped linearly onto the detector coordinates ξ ,

$$\kappa = c_0 \xi \tag{8.59}$$

then it is straightforward to describe the sampled and transformed data, where we ignore DFT artifacts in the following discussion. If the detector has M pixels of width $\Delta\xi$, and wavelengths from λ_{\min} to λ_{\max} mapped onto it, then we have

$$\Delta\xi = \frac{1}{Mc_0} \left[\frac{1}{\lambda_{\min}} - \frac{1}{\lambda_{\max}} \right], \quad \text{or} \quad \Delta\xi = \frac{\Delta\kappa}{Mc_0}, \tag{8.60}$$

where $\Delta\kappa$ is the corresponding interval in wavenumber between λ_{\min} and λ_{\max} .

$$\Delta\kappa = \frac{1}{\lambda_{\min}} - \frac{1}{\lambda_{\max}}. \quad (8.61)$$

The samples therefore lie at intervals of wavenumber given by

$$\kappa_m = \kappa_{\min} + mc_0\Delta\xi, \quad m = 0, 1, \dots, M. \quad (8.62)$$

The transform determines the spatial frequency of the fringes detected across the array, that is to say p fringes per M pixels. We have therefore $x = p/\Delta\kappa$,

$$x_p = \frac{p}{Mc_0\Delta\xi}, \quad p = 0, 1, \dots, M/2. \quad (8.63)$$

and p is an index of spatial frequency. Using the expressions for κ_m and x_p we have

$$x\kappa = \frac{mp}{M} + \frac{p\kappa_{\min}}{\Delta\kappa}. \quad (8.64)$$

If we can describe the fringes as in Equation 8.10:

$$I(m) = I_s [1 + |\gamma| \cos(2\pi\kappa x + \phi_\gamma)] + I_b, \quad (8.65)$$

then inserting Equation 8.64 into 8.65 yields

$$I(m) = I_s \left[1 + |\gamma| \cos \left(\frac{2\pi mp_0}{M} + \phi \right) \right] + I_b, \quad (8.66)$$

where $I(m)$ is the average intensity at the m^{th} pixel of the detector,

$$\phi = \frac{2\pi p\kappa_{\min}}{\Delta\kappa} + \phi_\gamma. \quad (8.67)$$

The average total number of photons in each frame of data can be written

$$M_t = M(I_s + I_b), \quad (8.68)$$

where I_s and I_b are the average stellar spectrum and background per pixel.

The Discrete Fourier Transform of the detected channeled spectrum would be

$$\mathcal{I}(p) = \sum_{m=0}^{M-1} I(m) \exp \left[j \frac{2\pi pm}{M} \right], \quad (8.69)$$

whose real and imaginary parts are

$$\text{Re}[\mathcal{I}(p)] = \sum_{m=0}^{M-1} I(m) \cos \left(\frac{2\pi pm}{M} \right) = \begin{cases} M(I_s + I_b) & p = 0 \\ M(I_s|\gamma|/2) \cos \phi & p = p_0 \\ 0 & \text{otherwise} \end{cases} \quad (8.70)$$

and

$$\text{Im}[\mathcal{I}(p)] = \sum_{m=0}^{M-1} I(m) \sin \left(\frac{2\pi pm}{M} \right) = \begin{cases} M(I_s|\gamma|/2) \sin \phi & p = p_0 \\ 0 & \text{otherwise} \end{cases} \quad (8.71)$$

where the factor M arises in performing the sum of the DFT, and the fringe amplitude contains a factor of $1/2$ because the DFT calculates both positive and negative frequency components. These are complex conjugates, this being the transform of real-valued data, and no information would be lost if we discarded the negative frequency half.

The features at $\pm p_0$ are not truly delta functions, but sinc functions whose nulls lie at the locations of the other samples in the spatial frequency domain. For example, if the fringe frequency was in fact somewhere part-way between the frequencies sampled by the FFT, the convolution of the sinc function with the sampling would be more obvious.

Power Spectrum and Periodogram

It would be normal in most approaches of group-delay tracking to form a power spectrum from the complex transform described by Equations 8.70 and 8.71 and afterwards integrate the power spectra to improve the signal-to-noise ratio. From Equations 8.66 and 8.68 we have that the amplitude of the power spectrum would be

$$|\mathcal{I}(p)| = \begin{cases} M_t & p = 0 \\ M(I_s|\gamma|/2) & p = \pm p_0 \\ 0 & \text{otherwise} \end{cases}, \quad (8.72)$$

Examples of fringe signals detected in channeled spectra are shown in Figure 8.5. These are from internal fringes formed with SUSI used in autocollimation. Note the peak at the fringe frequency (located between 30 and 70 cycles) and the large peak at the zero frequency.

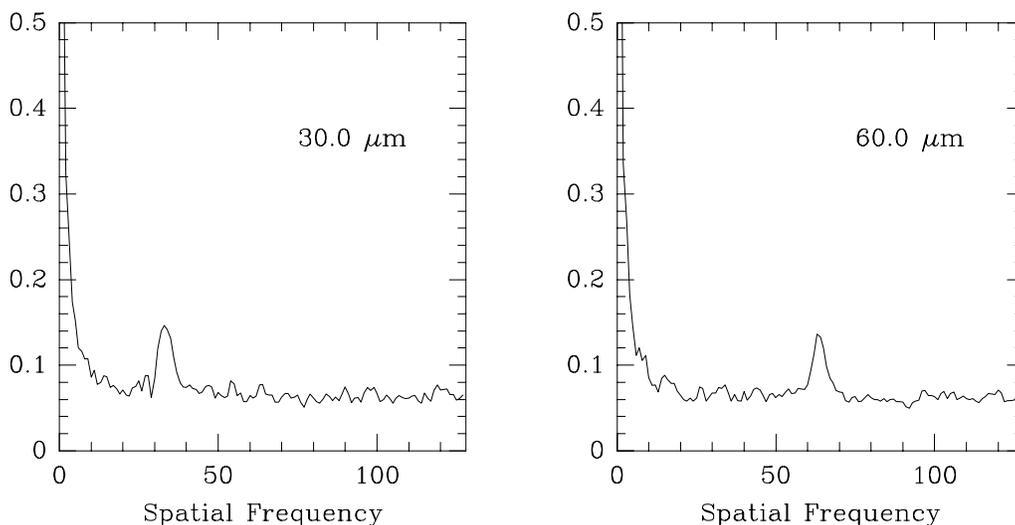


Figure 8.5: Examples of the fringe signal from an FFT processing of channeled spectra. Note the large signal at a spatial frequency of 0. The distances indicated are relative positions of a piezo actuator and are not with respect to the location of zero path-difference. [From Figure 10.7 of Lawson (1994).]

Our ability to locate the spatial frequency of the peak is somewhat better than the “resolution” of the power spectrum. The resolution in the power spectrum is the smallest separation in spatial frequency for which *two* unresolved line-features can be distinguished as separate.[†] If the spectrum only has a single feature (the fringe) the problem reduces to finding the best fit to the data of a known transfer function (i.e. a sinc function in the case of a square bandpass). The peak can be located by padding the data with its mean value before applying the DFT to produce samples of the power spectrum at shorter intervals. The peak can be approximately located this way (it isn’t practical to infinitely pad the data) and can be further determined by applying a three-point parabolic interpolation around those samples nearest the peak.

An advantage of using power spectrum analysis is that it allows incoherent integration of the fringe signal. It is the poor resolution in delay, when compared with phase-tracking methods, that makes this approach attractive. Small changes in delay may be unresolved in the power spectrum, making it possible to integrate numerous noisy power spectra and to thereby improve the sensitivity. Group-delay tracking with the FFT has been seen as particularly suited to low-light-level conditions in which methods of phase tracking would fail.[‡]

The methods of phase measurement require a modulation, and consequently the fringes are very slightly blurred in each sample—the fringes move by $\lambda/4$ per sample, reducing the fringe visibility by $\sim 10\%$. If we are processing a single channeled spectrum as described here without modulating the delay line, the sensitivity for fringe detection is slightly better in comparison.

A significant drawback of this approach is that without some subtle changes to the data processing, it is difficult to track fringes at zero delay. Because in each cycle we only measure a single channeled spectrum we cannot determine the sign of the delay, making this method a poor candidate for implementation as a zero-seeking servo. Furthermore, to avoid the fringe signal being buried in the low spatial frequency profile of the stellar spectrum, the zero-frequency signal must be subtracted in each frame. This is somewhat complicated because of variations in the intensity of the stellar spectrum due to scintillation.

Non-Linear Mapping

A problem that is common to all implementations of group-delay tracking is that the mapping from wavenumber to pixel number will most likely not be linear. Most spectrometers

[†]The resolution is inversely proportional to the total bandwidth detected by the array, and the DFT produces estimates at intervals in delay of $\Delta x = 1/\Delta\kappa$ up until a cut-off of $x = \pm M/(2\Delta\kappa)$, corresponding to a distance of half the coherence length.

[‡]This comparison is only valid if we assume that the sensitivity and noise characteristics (dark current and read noise for example) are the same for the detectors used in each approach.

use either a prism or a grating, neither of which have dispersions that are constant in κ .[§] Consequently the distance between fringes will change throughout the detected spectrum, and the fringes will be partly stretched or compressed. This “chirp” means that although the number of fringes would be the same, the associated frequency is more difficult to identify. The Fourier transform yields a fringe *frequency*, not a number-of-fringes. The transform of the fringes is therefore not a delta function and the peak is broadened with its height reduced. For instance, if the spacing between fringes doubles from one edge of the detector to the other, then the peak would be spread between these two frequencies. This effect is more severe the more fringes are present: at larger path differences the peak becomes progressively broader and lower in height. It becomes more difficult to detect the peak in the presence of noise, and the broadening means that the peak is less well defined. Careful calibration of the wavelength scale of the detector along with the use of a discrete Fourier transform is required to overcome these losses. The effects of longitudinal dispersion within the interferometer must also be understood and accounted for, otherwise the sensitivity of tracking will be degraded, sometimes in unexpected ways (Lawson and Davis, 1996).

8.5.2 Channeled Spectrum: Lomb-Scargle Periodogram

One of the disadvantages of the FFT approach is that the periodogram is not normalized in a way that allows thresholding against noise. It is therefore difficult to judge the significance of a peak in the power spectrum relative to the noise. A better method would be a power spectrum derived from a least-squares fit to the data. This would then give us some measure of the residuals and goodness of the fit.

The Lomb-Scargle periodogram, discussed by Press *et al.* (1992), is such an approach. The advantage of this method is that it normalizes the periodogram (power spectrum) so that it is possible to ignore noise peaks below a set threshold. Although it appears to be a computer intensive approach, W.J. Tango at the University of Sydney has implemented it for real-time fringe tracking with a CCD detector.

Because the Lomb-Scargle approach requires that the data sets be real-valued only, I have not yet seen how it could be adapted to process fringe phasors.

8.5.3 Channeled Spectrum: Least-Squares Fit

Traub *et al.* (1990) describe a method of group-delay tracking which uses a cross-correlation of the data with model functions. It is assumed that a family of functions exist which will closely fit the data providing certain parameters are adjusted. It follows that if these parameters are chosen correctly then it will minimize the least-squared difference between the data and the model. If we were to consider the delay by itself then we would perform

[§]An exception to this is a 60° prism of BK7 glass, which is closely linear over the wavelength range of 600–1000 nm.

the minimization by taking the partial derivative with respect to the delay, x , of the mean squared difference, equating it to zero, and solving for the delay. If the data are represented by the set g_k and the model is $f_k(x)$ then we have

$$\frac{\partial}{\partial x} \left[\sum_{k=1}^K [g_k - f_k(x)]^2 \right] = 0, \quad (8.73)$$

which can be written in full as

$$\frac{\partial}{\partial x} \left[\sum_{k=1}^K [g_k^2 - 2g_k f_k(x) + f_k^2(x)] \right] = 0. \quad (8.74)$$

In this equation only the cross term is of interest. The sum of the g_k^2 terms is a constant and contributes nothing to the minimization. Furthermore, if the model $f_k(x)$ was normalized correctly then the sum of the $f_k^2(x)$ terms would be independent of x , and therefore would also be a constant. We can now express the minimization of the mean square difference as

$$\frac{\partial}{\partial x} \left[\sum_{k=1}^K g_k f_k(x) \right] = 0, \quad (8.75)$$

where the function $f_k(x)$ maximizes the sum of the cross terms. The quantity in brackets is simply the cross-correlation between the model and the data, calculated at zero lag. Traub *et al.* (1990) presented simulations of pathlength motions with peak-to-valley excursions of $1.3 \mu\text{m}$ over 1 second with $|\gamma| = 1.0$. He concluded that delay tracking should be possible at count rates as low as 10 photons per coherence time, with a position uncertainty of $\sim 0.2\lambda$.

8.5.4 Multi-Wavelength Phasor Measurements: Fast or Direct Fourier Transform

Multiple-wavelength phase measurements are complicated somewhat because you can never modulate a *pathlength* to produce the same phase shift at all wavelengths. However, if you can control the timing of your detector then it is possible to bin the data separately at each wavelength and at the same time ignore data at wavelengths where the phase introduced by the phase-shifter (piezo) has already changed by 1λ or more. This is the approach that was used at the Mark III interferometer and which is currently used at PTI and NPOI.

We have then that the four bins A , B , C , and D that characterize ϕ as a function of wavenumber κ are recorded in each cycle of modulation. Let us then assume that the phase ϕ arises from a vacuum path-difference x , such that

$$\phi(\kappa) = 2\pi\kappa x, \quad (8.76)$$

If we measure the quantities A , B , C , D , at M wavenumbers we can calculate

$$h_c(\kappa_m) = A(\kappa_m) - C(\kappa_m) \quad m = 0, \dots, M - 1 \quad (8.77)$$

$$h_s(\kappa_m) = B(\kappa_m) - D(\kappa_m) \quad m = 0, \dots, M - 1 \quad (8.78)$$

so that we now have

$$h_c(\kappa_m) \propto \cos(2\pi\kappa_m x), \quad (8.79)$$

$$h_s(\kappa_m) \propto \sin(2\pi\kappa_m x). \quad (8.80)$$

We can now define the complex number series

$$h(\kappa_m) = h_c(\kappa_m) + jh_s(\kappa_m), \quad m = 0, \dots, M - 1 \quad (8.81)$$

and take its discrete Fourier transform

$$H(x) = \sum_{m=0}^{M-1} h(\kappa_m) \exp(j2\pi\kappa_m x). \quad (8.82)$$

The value of x that locates the peak in the power spectrum $|H(x)|^2$ corresponds to the group delay of the fringes.

The advantage of this method is that the modulation allows the sign of the delay to be unambiguously determined, and because phasors are processed there is no zero-frequency term in the power spectrum: it is straightforward to track at zero group delay. This allows the tracking to be implemented as a zero-seeking servo, and because the mean tracking position can be zero the astrometric error introduced by incorrectly scaling spatial frequencies to delays is of less consequence.

The disadvantage of this brute-force approach is that you cannot weigh the data to distinguish between good and bad estimates of the sine and cosine of the phase. This ability to weight the data is important if we know beforehand that certain pixels in our array are noisier than others. If for example we know that all the sine and cosine measurements estimates are noisy, we would like to have some figure-of-merit to allow us to judge the usefulness of the derived delay estimate.[†]

8.5.5 Multi-Wavelength Phasor Measurements: “Optimal” Estimator

With the conventional approach to group-delay estimation, the delay is inferred from the *spatial frequency* of the fringes in a channeled spectrum. It is assumed that the source is so faint and the atmosphere so unstable that coherent integration is limited to time-scales less than $\sim 2t_0$, and thus *incoherent* integration (the integration of power spectra) is used. The phase information in the complex Fourier transform of the fringes is simply thrown away, because it is assumed to be so corrupted by noise that it is unrecoverable.

At high light levels where a sufficient signal-to-noise is achievable in a time less than t_0 , the phase of the channeled spectra can indeed be extracted. It is then possible to formulate a group-delay estimate using this phase and thereby greatly improve the resolution of the estimates (cf. Equations 8.70 and 8.71 yield the phase ϕ at the fringe frequency, p_0).

[†]Dave Mozurkewich has pointed out that a simple but perhaps heavy-handed way of weighting the data is to ignore data points that are suspect.

This approach is a two-step procedure: one must first correctly identify the spatial frequency of fringes in the channeled spectra (normal group-delay estimate); and secondly extract the phase of those fringes and interpret it in terms of a delay. This approach is *optimal* in the sense that it provides a group-delay estimate with the variance of a phase estimator. This will only work well at high light levels and under circumstances where the dispersion is well understood. Lawson *et al.* (2000) have described an implementation using phasors for use at PTI. The approach has also been independently considered by Mozurkewich, Hummel, and Benson for use at the NPOI (Mozurkewich, 2000). Although the method is not yet in routine use at either interferometer—in part because of the difficulty in modeling the changing atmospheric dispersion—it may ultimately allow noise in group-delay estimates to be greatly reduced.

8.6 Variance of Phase and Group Delay Estimates

The derivations that follow have been previously described by Lawson *et al.* (2000).

8.6.1 Variance of Phase

The expected signal-to-noise ratio (SNR) and rms phase error σ_ϕ for a four-bin phase estimate has been derived by Wyant (1975):

$$\text{SNR} = \frac{2}{\pi} \sqrt{NV^2}, \quad \sigma_\phi = \frac{\pi}{2} \frac{1}{\sqrt{NV^2}}, \quad (8.83)$$

where V is the fringe visibility, N is the number of photons per frame, and the effects of background and detector read noise have been ignored.

8.6.2 Variance of Group Delay

Variance of Phase-Slope Derivation

The group delay is proportional to the slope of the phase as a function of wavenumber. The group delay is defined as

$$\frac{1}{2\pi} \frac{\partial \phi}{\partial \kappa}, \quad (8.84)$$

where $\kappa = 1/\lambda$. If we have several noisy samples of phase at independent wavenumbers, the variance of the group delay is proportional to the variance of the slope of a straight line fit to that data.

If there are M data points and the i^{th} data point has for its coordinate x_i and a variance of σ_i^2 , then the variance of the slope σ_b^2 of a least-squares fit of a line to that set of data is given as (Press *et al.* 1992, Section 15.2 “Fitting data to a straight line,” Equations 15.2.4 to 15.2.9):

$$\sigma_b^2 = \frac{S}{\Delta}, \quad (8.85)$$

where

$$S = \sum_{i=1}^M \frac{1}{\sigma_i^2}, \quad \Delta = S S_{xx} - (S_x)^2, \quad (8.86)$$

and

$$S_x = \sum_{i=1}^M \frac{x_i}{\sigma_i^2}, \quad S_{xx} = \sum_{i=1}^M \frac{x_i^2}{\sigma_i^2}. \quad (8.87)$$

The group delay is $(1/2\pi)$ times the slope of phase with respect to wavenumber, $\kappa = 1/\lambda$. The variance of the group delay is therefore proportional to the variance of the phase—as measured at M data points across a band $\Delta\kappa$. If the variance of the central-fringe phase is σ_ϕ^2 and we assume that the light from the broadband channel is divided equally amongst M pixels, the variance of the phase in each pixel will be $M\sigma_\phi^2$. We can therefore write that

$$S = \frac{1}{\sigma_\phi^2}, \quad (8.88)$$

$$S_x = \frac{1}{\sigma_\phi^2} \left[\frac{1}{M} \sum_{i=1}^M \kappa_i \right] \rightarrow \frac{1}{\sigma_\phi^2} \frac{1}{\Delta\kappa} \int_{\bar{\kappa}-\Delta\kappa/2}^{\bar{\kappa}+\Delta\kappa/2} \kappa d\kappa = \frac{1}{\sigma_\phi^2} \bar{\kappa}, \quad (8.89)$$

$$S_{xx} = \frac{1}{\sigma_\phi^2} \left[\frac{1}{M} \sum_{i=1}^M \kappa_i^2 \right] \rightarrow \frac{1}{\sigma_\phi^2} \frac{1}{\Delta\kappa} \int_{\bar{\kappa}-\Delta\kappa/2}^{\bar{\kappa}+\Delta\kappa/2} \kappa^2 d\kappa = \frac{1}{\sigma_\phi^2} \left[\bar{\kappa}^2 + \frac{\Delta\kappa^2}{12} \right]. \quad (8.90)$$

We have therefore that

$$\Delta = \frac{1}{\sigma_\phi^4} \frac{\Delta\kappa^2}{12}, \quad (8.91)$$

and it follows from Equation 8.85 that the variance of the slope of phase with respect to wavenumber is

$$\sigma_b^2 = 12 \frac{\sigma_\phi^2}{\Delta\kappa^2}, \quad (8.92)$$

independent of the number of pixels M . The rms variations in group delay can therefore be written

$$\sigma_{gd} = \frac{\sqrt{12}}{2\pi} \frac{\sigma_\phi}{\Delta\kappa}. \quad (8.93)$$

Matched Filter Derivation

One can also derive the signal-to-noise ratio for the amplitude group-delay estimator from a simple matched-filter argument. Assume the input data are the phases as a function of wavenumber

$$\phi = \phi_0 + (\kappa - \kappa_0) \frac{\partial\phi}{\partial\kappa}. \quad (8.94)$$

The orthonormal basis functions over the bandwidth $\Delta\kappa$ are respectively,

$$\frac{1}{\sqrt{\Delta\kappa}} \quad \text{and} \quad \sqrt{12} \frac{(\kappa - \kappa_0)}{(\Delta\kappa)^{3/2}}. \quad (8.95)$$

Thus

$$(\Delta\kappa)^{1/2}\phi_0 \quad \text{and} \quad \frac{1}{\sqrt{12}}(\Delta\kappa)^{3/2}\frac{\partial\phi}{\partial\kappa} \quad (8.96)$$

are each estimated with the same error.

8.6.3 Phase and Group Delay Variance Compared

From Equation 8.93 we now have that the ratio of standard deviations of the phase and group delay estimates is

$$\frac{\sigma_{\phi d}}{\sigma_{gd}} = \left[\frac{1}{2\pi} \frac{\sigma_{\phi}}{\kappa} \right] \left[\frac{\sqrt{12}}{2\pi} \frac{\sigma_{\phi}}{\Delta\kappa} \right]^{-1} = \frac{1}{\sqrt{12}} \frac{\Delta\kappa}{\kappa}, \quad (8.97)$$

where $\sigma_{\phi d}$ is the rms path fluctuation corresponding to phase variations σ_{ϕ} .

As an example of the difference between phase and group-delay variations, with PTI and an observation bandwidth of 2.0–2.4 μm and a mean observing wavelength of $\lambda = 2.2 \mu\text{m}$, we have $1/\Delta\kappa = 12 \mu\text{m}$, and can conclude that delay estimates derived from phase estimates will have rms variations 19 times smaller those derived from group-delay estimates.

8.7 Conclusion

The methods of phase and group delay estimation are routinely used in modern stellar interferometers to locate fringes and maintain the observations at a fixed location on the coherence envelope. Other methods of fringe measurement, in particular coherence envelope tracking, may also be used to estimate fringe parameters but are generally less efficient and more labour intensive.

In this Chapter we have reviewed the various methods of phase and group delay estimation that are currently being used in stellar interferometers, with emphasis on the techniques used at SUSI, PTI, and NPOI.

Acknowledgments

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. The author would like to thank Mark Colavita for discussions leading to the derivations of Section 8.6.2.

References

- J.T. Armstrong, D. Mozurkewich, L.J. Rickard, D.J. Hutter, J.A. Benson, P.F. Bowers, N.M. Elias II, C.A. Hummel, K.J. Johnston, D.F. Buscher, J.H. Clark III, L. Ha, L.-C. Ling, N.M. White, and R.S. Simon, “The Navy Prototype Optical Interferometer (NPOI),” *Astrophys. J.* **496**, 550–571 (1998).
- J.E. Baldwin, R.C. Boysen, G.C. Cox, C.A. Haniff, J. Rogers, P.J. Warner, D.M.A. Wilson, and C.D. Mackay, “Design and performance of COAST,” in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., Proc. SPIE **2200**, 118–128 (1994).
- J.A. Benson (1998), private communication.
- J.A. Benson, D. Mozurkewich, and S.M. Jefferies, “Active optical fringe tracking at the NPOI,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 493–496 (1998).
- G.D. Bergland, “A guided tour of the fast Fourier transform,” *IEEE Spectrum* **6**, 41–52 (1969).
- M. Born and E. Wolf, *Principles of Optics*, 6 edn. (Oxford, UK: Pergamon Press, 1980), Section 10.3.1.
- T.A. ten Brummelaar, “Differential path considerations in optical stellar interferometry,” *Appl. Opt.* **34**, 2214–2219 (1995).
- T.A. ten Brummelaar, “Correlation measurement and group delay tracking in optical stellar interferometry with a noisy detector,” *Mon. Not. R. Astron. Soc.* **285**, 135–150 (1997).
- D.F. Buscher, *Getting the most out of COAST*, Ph.D. thesis, University of Cambridge (1988).
- F. Cassaing, B. Fleury, C. Coudrain, P.-Y. Madec, E. Di Folco, A. Glindemann, and S. Lévêque, “Optimized fringe tracker for the VLTI/PRIMA instrument,” in *Interferometry in Optical Astronomy*, P.J. Léna and A. Quirrenbach, eds. Proc. SPIE **4006**, 152–163 (2000).
- M.M. Colavita, *Atmospheric limitations of a two-color astrometric interferometer*, Ph.D. thesis, Massachusetts Institute of Technology (1985).
- M.M. Colavita, J.K. Wallace, B.E. Hines, Y. Gursel, F. Malbet, D.L. Palmer, X.P. Pan, M. Shao, J.W. Yu, A.F. Boden, P.J. Dumont, J. Gubler, C.D. Koresko, S.R. Kulkarni, B.F. Lane, D.W. Mobley, and G.T. van Belle, “The Palomar Testbed Interferometer,” *Astrophys. J.* **510**, 505–521 (1999).
- K. Creath, “Phase-measurement interferometry techniques,” *Prog. Opt.* **26**, 349–393 (1988).
- G. Daigne and J.F. Lestrade, “Astrometric optical interferometry with non-evacuated delay lines,” *Astron. Astrophys. Supp. Ser.* **138**, 355–363 (1999).
- J. Davis, P.R. Lawson, A.J. Booth, W.J. Tango, and E.D. Thorvaldson, “Atmospheric path variations for baselines up to 80 m measured with the Sydney University Stellar Interferometer,” *Mon. Not. R. Astron. Soc.* **273**, L53–L58 (1995).
- J. Davis, W.J. Tango, A.J. Booth, E.D. Thorvaldson, and J. Giovannis, “The Sydney University Stellar Interferometer – II. commissioning observations and results,” *Mon. Not. R. Astron. Soc.* **303**, 783–791 (1999).

- J. Davis, W.J. Tango, and E.D. Thorvaldson, “Dispersion in stellar interferometry: simultaneous optimization for delay tracking and visibility measurements,” *Appl. Opt.* **37**, 5132–5136 (1998).
- E. Edser and C.P. Butler, “A simple method of reducing prismatic spectra,” *Phil. Mag.* **46**, 207–216 (1898).
- A. Feldman, “Determination of fringe order in the channel spectra of thin films,” *Appl. Opt.* **23**, 1193–1196 (1984).
- H. Fizeau and L. Foucault, “Sur le phénomène des interférences entre deux rayons de lumière dans le cas de grandes différence de marche,” *C. R. Acad. Sci.* **30**, 146–159 (1845).
- V. Coudé du Foresto, G. Perrin, and M. Boccas, “Minimization of fibre dispersion effects in double Fourier stellar interferometers,” *Astron. Astrophys.* **293**, 278–286 (1995).
- C.A. Hummel, “Practice of interferometry at NPOI,” in *Interferometry in Optical Astronomy*, P.J. Léna and A. Quirrenbach, eds. *Proc. SPIE* **4006**, 459–464 (2000).
- S.M. Kay and S.L. Marple Jr., “Spectrum analysis—a modern perspective,” *Proc. IEEE* **69**, 1380–1418 (1981).
- E.J. Kim, *Dispersed Fringe Group Delay Astrometry using the Mark III Stellar Interferometer*, M.S.E.E, Mass. Inst. Technology (1989).
- L. Koechlin, P.R. Lawson, D. Mourard, A. Blazit, D. Bonneau, F. Morand, Ph. Stee, I. Tallon-Bosc, and F. Vakili, “Dispersed fringe tracking with the multi- r_0 apertures of the Grand Interféromètre à 2 Télescopes,” *Appl. Opt.* **35**, 3002–3009 (1996).
- A. Labeyrie, “Interference fringes obtained on Vega with two optical telescopes,” *Astrophys. J.* **196**, L71–L74 (1975).
- M.G. Lacasse and W.A. Traub, “Glass compensation for an air filled delay line,” in *High-Resolution Imaging by Interferometry*, F. Merkle, ed., ESO Conf. Workshop, **29**, 959–970 (Garching, Germany: European Southern Observatory, 1988).
- P.R. Lawson, *Group Delay Tracking with the Sydney University Stellar Interferometer*, Ph.D. thesis, University of Sydney (1994).
- P.R. Lawson, “Group-delay tracking in optical stellar interferometry with the fast Fourier transform,” *J. Opt. Soc. Am. A* **12**, 366–374 (1995).
- P.R. Lawson, “Dispersion compensation and fringe tracking,” in *High Angular Resolution in Astrophysics*, A.-M. Lagrange, P. Léna, and D. Mourard, eds. *NATO ASI*, **3350**, 349–353 (Dordrecht: Kluwer Academic, 1997).
- P.R. Lawson and J. Davis, “Dispersion compensation in stellar interferometry,” *Appl. Opt.* **35**, 612–620 (1996).
- P.R. Lawson, T.R. Scott, and C.A. Haniff, “Group-delay tracking and visibility fluctuations in long-baseline stellar interferometry,” *Mon. Not. R. Astron. Soc.* **304**, 218–224 (1999).
- P.R. Lawson, M.M. Colavita, P.J. Dumont, and B.F. Lane, “Least-squares estimation and group delay in astrometric interferometers,” in *Interferometry in Optical Astronomy*, P.J. Léna and A. Quirrenbach, eds. *Proc. SPIE* **4006**, 397–406 (2000).

- S. Léveque, B. Koehler, and O. von der Luhe, “Longitudinal dispersion compensation for the very large telescope interferometer,” *Astrophys. Space. Sci.* **239**, 305–314 (1996).
- J.A. Meisner, “Atmospheric delay tracking in a long-baseline optical stellar interferometer,” *Opt. Eng.* **35**, 1927–1935 (1996).
- A.A. Michelson and F.G. Pease, “Measurement of the diameter of α Orionis with the interferometer,” *Astrophys. J.* **53**, 249–259 (1921).
- S. Morel and L. Koechlin, “Fringe tracking using a priori information on the optical path difference drift,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 1057–1064 (1998).
- D. Mozurkewich (2000), private communication.
- P. Nisenson and W. Traub, “Magnitude limit of the group delay tracking method for long baseline interferometry,” in *Interferometric Imaging in Astronomy*, J.W. Goad, ed., 129–133 (Tucson, AZ: National Optical Astronomy Observatories, 1987).
- C.E. Padilla, V.I. Karlov, L.K. Matson, K. Soosaar, and T. ten Brummelaar, “High-performance fringe tracking algorithms utilizing statistical models at atmospheric turbulence,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 1045–1056 (1998).
- W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C* (Cambridge, UK: Cambridge University Press, 1992).
- R.J. Sanderman, “Use of channeled spectra to measure absolute phase shift and dispersion in two beam interferometry,” *Appl. Opt.* **10**, 1087–1091 (1971).
- M. Shao and M.M. Colavita, “Long-baseline optical and infrared stellar interferometry,” *Ann. Rev. Astron. Astrophys.* **30**, 457–498 (1992).
- M. Shao, M.M. Colavita, B.E. Hines, D.H. Staelin, D.J. Hutter, K.J. Johnston, D. Mozurkewich, R.S. Simon, J.L. Hershey, J.A. Hughes, and G.H. Kaplan, “The Mark III stellar interferometer,” *Astron. Astrophys.* **193**, 357–371 (1988).
- W.H. Steele, *Interferometry* (Cambridge, UK: Cambridge University Press, 1987).
- W.J. Tango, “Dispersion in stellar interferometry,” *Appl. Opt.* **29**, 516–521 (1990).
- W.J. Tango and R.Q. Twiss, “Michelson stellar interferometry,” *Prog. Opt.* **17**, 239–277 (1980).
- W.A. Traub, “Recent results from the IOTA interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 848–855 (1998).
- W.A. Traub, M.G. Lacasse, and N.P. Carlton, “Spectral dispersion and fringe detection in IOTA,” in *Amplitude and Intensity Spatial Interferometry*, J.B. Breckinridge, ed., Proc. SPIE **1237**, 145–152 (1990).
- J.F. Walkup and J.W. Goodman, “Limitations of fringe parameter estimation at low light levels,” *J. Opt. Soc. Am.* **63**, 399–407 (1973).
- J. Wyant, “Use of an ac heterodyne lateral shear interferometer with real-time wavefront correction systems,” *Appl. Opt.* **14**, 2622–2626 (1975).

Chapter 9

Phase Referencing

ANDREAS QUIRRENBACH

UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA

9.1 Introduction

The use of phases in ground-based astronomical interferometry is severely limited by the pathlength fluctuations of the Earth’s atmosphere (see for example the review by Quirrenbach in Chapter 5). Two different approaches are widely used to deal with the problem of atmospheric and instrumental phase corruption: closure phase methods (or phase self-calibration), and phase-referencing. In the latter technique, the phase information from a reference object is used to determine the atmospheric phase, and to correct the phase of the target source accordingly. Both methods have been used extensively at radio wavelengths, so one could hope to apply the same techniques in the visible and near-infrared. However, while in radio astronomy the atmospheric coherence time τ_0 is typically several minutes, and the isoplanatic angle θ_0 several degrees, the corresponding values in the optical regime are only of order ten milliseconds and a few arcseconds.

These limitations have important consequences for phase-referencing in the visible and near-infrared. They preclude the use of source-switching strategies and require the simultaneous observation of target and reference object. While this might appear to be a very restrictive requirement, there are several important applications of phase-referencing to optical long-baseline interferometry. First, the phase difference can be used as the primary observable in “astrometric” applications, e.g. to determine the positional offset of a circumstellar envelope from the central star, or to search for the reflex motion of stars orbited by planets. (In the latter case, a suitable reference object is needed within the isoplanatic patch.) Second, the

reference phase can be used to increase the effective atmospheric coherence time, allowing longer coherent integrations on the target source. As we shall see, phase-referencing can improve the sensitivity of large interferometers by many magnitudes; it is therefore a key technique for imaging faint objects. There are different variants of phase referencing: the reference phase can come from simultaneous observations of a separate object (dual-star observations), from observations of the target source with a second wavelength channel (wavelength bootstrapping), or from a more sensitive baseline in an interferometer array (baseline bootstrapping).

9.2 Principles of Phase Referencing

9.2.1 The Mark III Interferometer

We will use a relatively simple instrument, the Mark III interferometer (Shao *et al.* 1988, see Figure 9.1), to explain some the principles of phase referencing. This means that we can concentrate on one specific fringe-tracking technique (explained below), and that we can ignore detector and background noise compared to the photon noise. Our qualitative conclusions remain valid in many more general situations, however, and our quantitative results can easily be generalized for more complex fringe-tracking schemes and more general sources of noise (see e.g. Shao and Colavita 1992).

The Mark III optical interferometer was operational on Mt. Wilson, CA, from 1986 to 1990. Its basic optical layout was that of a single-baseline Michelson interferometer, with two siderostats feeding the two arms, and vacuum delay lines to compensate for the pathlength difference. The baseline of the instrument could be configured to lengths ranging from 3.0 to 31.5 m, giving some flexibility for measurements of stellar diameters and observations of binary stars. The maximum aperture size of the Mark III was 5 cm. The images of stars from both arms of the interferometer were centered by an angle tracker, which worked in the wavelength range $450 \text{ nm} \lesssim \lambda \lesssim 600 \text{ nm}$.

The delay was modulated with a 500-Hz triangle wave of amplitude 800 nm. If the pathlength difference between the two interferometer arms was within the coherence length, the intensity at the output of the beam combiner would thus vary sinusoidally with time. The phase of this signal was computed in real time and used to track the movement of the fringes due to atmospheric pathlength fluctuations. The closed-loop bandwidth of the fringe tracker was $\sim 20 \text{ Hz}$. Under favorable seeing conditions, the fringes stayed locked for several seconds, sometimes up to a few minutes. Once per second, the fringe amplitude was compared to a preset value. If it was lower than this threshold, the fringe tracker assumed that it was not locked on the central fringe and jumped one fringe; the direction was determined by the visibility gradient.

Dichroic beam splitters in the two output arms of the beam combiner provided four spectral channels. A broad-band channel ($600 \text{ nm} \lesssim \lambda \lesssim 900 \text{ nm}$, giving an effective fringe-tracking

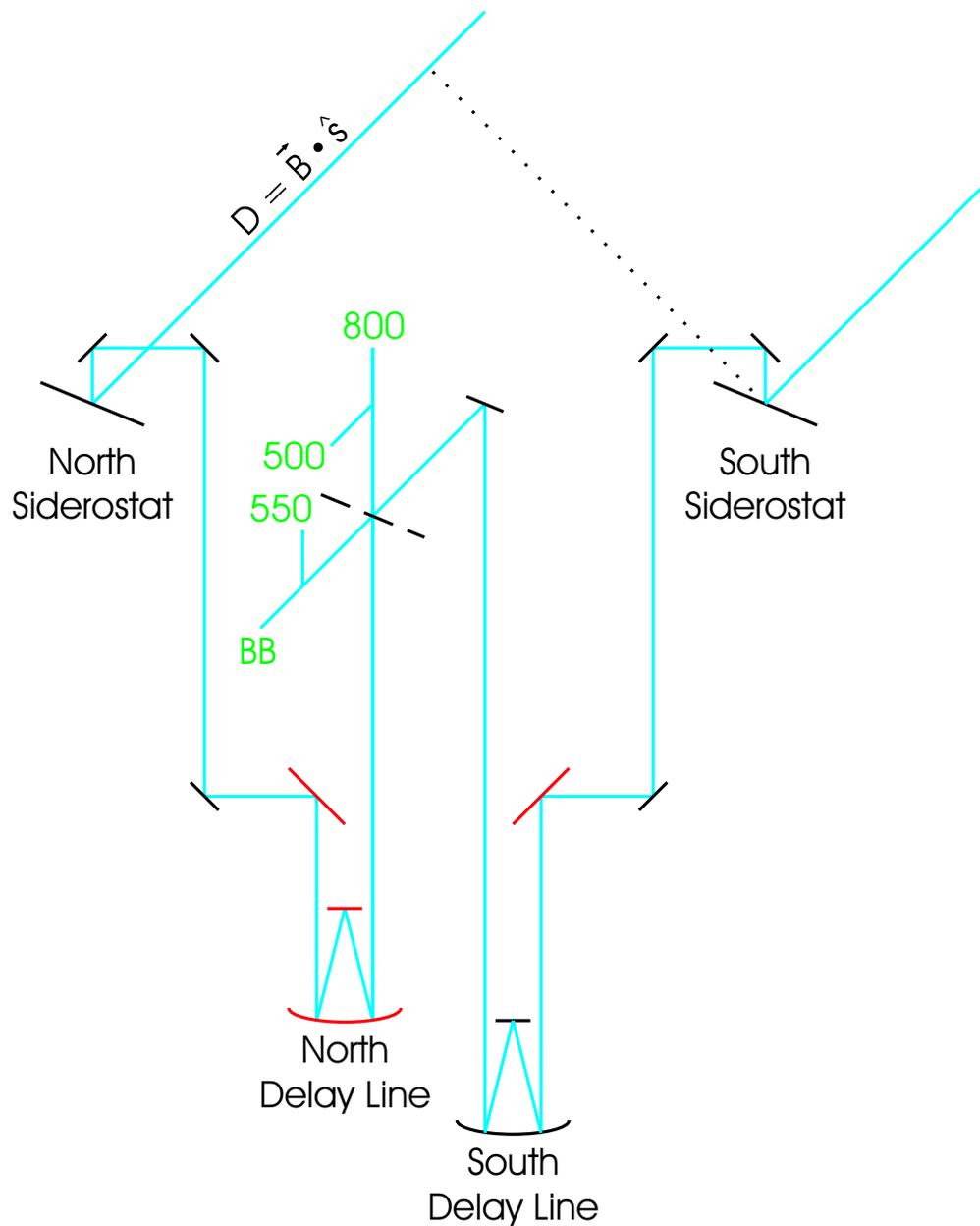


Figure 9.1: Schematic drawing of the Mark III interferometer. The two siderostats feed light into a vacuum system. The two mirrors feeding light into the delay lines are mounted on piezo-electric actuators and are part of the angle-tracking servo loop. The positions of the two optical delay lines are continuously monitored with a laser interferometer. They are optically equivalent, but the cart and the small mirror in one of them are actively controlled and are part of the fringe-tracking servo loop. The beams from the two arms are combined with a 50% reflective mirror. The light in each of the two outputs is divided with a dichroic beam splitter, so that four wavelength channels (broad band for fringe tracking, 500 nm, 550 nm, and 800 nm in the standard setup) are available.

wavelength $\lambda_t \simeq 700$ nm) was used to track the fringes, while substantially narrower band-passes defined by interference filters were used in the other channels to take the scientific data. This arrangement ensured that the fringe packet was much wider in the data channels than in the tracking channel, so that errors in the central fringe identification did not lead to a noticeable visibility reduction.

9.2.2 Visibility Estimation and Signal-to-Noise Ratio

For each of the four spectral channels, arriving photons are counted synchronously with the delay modulation in bins corresponding to $\lambda/4$. (Since the physical stroke is equal to λ only in the channel with the longest wavelength, dead time is added in the electronics at the end of the stroke in the other three channels.) From the four bin counts A , B , C , and D , the square of the visibility V^2 can be estimated using

$$V^2 = \frac{\pi^2}{2} \cdot \frac{\langle X^2 + Y^2 - N \rangle}{\langle N - N_{\text{dark}} \rangle^2} , \quad (9.1)$$

where $X = C - A$ and $Y = D - B$ are the real and imaginary parts of the visibility, $N = A + B + C + D$ is the total number of photons counted, and N_{dark} is the background count rate determined separately on blank sky. This estimator for V^2 is not biased by photon noise (Shao *et al.*, 1988). The visibility phase is estimated using

$$\phi = \arctan \left(\frac{Y}{X} \right) - \frac{\pi}{4} . \quad (9.2)$$

The data are averaged using a combination of coherent and incoherent integrations.* By choosing a coherent integration time T , an observation of total duration $M \cdot T$ is divided into M intervals, which are averaged incoherently. The variance of the V^2 -estimator (Equation 9.1) is then given by

$$\sigma^2 = \frac{\pi^4}{4MN^2} + \frac{\pi^2 V^2}{MN} , \quad (9.3)$$

where N is the number of photons detected per coherent integration time (Colavita, 1985). The signal-to-noise ratio (SNR) of V^2 is therefore

$$\text{SNR}(V^2) = \frac{2}{\pi^2} \cdot \frac{\sqrt{MN} V^2}{\sqrt{1 + \frac{4}{\pi^2} NV^2}} . \quad (9.4)$$

If $NV^2 \gg 1$, the second term in Equation 9.3 dominates, and the variance depends only on the total number of photons detected, MN . If however $NV^2 \ll 1$, the first term is the dominant one, and the variance for a given total duration of the observation (i.e.,

*Coherent integration means that we sort each photon arriving during the integration time in one of the bins A , B , C , D , and use Equation 9.1 to get an estimate of V^2 . Incoherent integration means that we average over many estimates of V^2 . The intuitive meaning is that the coherent integration is used to estimate both amplitude *and phase* of the visibility, whereas the incoherent integration averages over the *modulus* of the visibility.

constant total number of photons MN) decreases with increasing coherent integration time, $\sigma^2 \propto N^{-1} \propto T^{-1}$; this implies that the signal-to-noise ratio of V^2 is $\propto T^{1/2}$ (for constant $M \cdot T$). We will call the two cases the “photon-rich” and “photon-starved” regimes, although NV^2 , and not N , is the critical quantity.

The extremely important results captured in Equations 9.3 and 9.4 have a simple intuitive interpretation. If the coherent integration time is sufficiently long, we get a good estimate of the amplitude *and phase* of the complex visibility. We can then stop the coherent integration, write out V^2 for a data sample, and average over these samples later without losing sensitivity. This is the photon-rich regime. If we are forced to stop the coherent integration (e.g., because of variations in the atmospheric or instrumental phase) before we get a meaningful phase measurement, we can still estimate V^2 for each data sample, but averaging over these estimates gives the poorer signal-to-noise characteristic of the photon-starved regime.

While these considerations show that it is advantageous to choose T large enough to get into the photon-rich regime, values larger than a fraction of the atmospheric coherence time will lead to serious phase changes and therefore to unacceptable degradation of the visibility. In the Mark III “standard” data reduction for measurements of stellar diameters and binary stars, $T = 4$ ms is adopted, which gives a coherence loss of a few per cent for seeing conditions typical for Mt. Wilson.

Several calibrator stars are normally included in the observing list for each night. They are used to determine the “system visibility” V_{sys}^2 , i.e., the value of V^2 observed for unresolved stars, as a function of seeing, zenith angle, time, and angle of incidence on the siderostat mirrors. For the seeing calibration, a seeing index S is calculated for each observation from the residual delay (Mozurkewich *et al.*, 1991). After removing the relatively strong dependence of V^2 on S , calibration with respect to the other variables normally leads to only a slight further improvement. (This situation is changed for phase-referenced data, where an additional strong decrease of V^2 with zenith angle has to be taken into account, see Section 9.3.8). The raw values of V^2 determined from Equation 9.1 are then divided by V_{sys}^2 to obtain calibrated data V_{cal}^2 for further analysis. Both the internal noise, with contributions from photon noise and from short-term fluctuations, and the calibration uncertainty contribute to the error of V_{cal}^2 . The two terms are added in quadrature to obtain formal error bars.

9.2.3 Phase-Referenced Visibility Averaging

The wide-band tracking channel in the Mark III interferometer provides a phase reference, which can be used to extend the coherent integration time T beyond the limit imposed by the atmospheric turbulence. This method provides a means of obtaining substantially

better signal-to-noise in the photon-starved regime, or even to make a transition into the photon-rich regime. The phase-referenced quantities X_r , Y_r , V_r , and ϕ_r are defined by

$$X_r + iY_r = V_r e^{i\phi_r} = V_s e^{i(\phi_s - \frac{\lambda_t}{\lambda_s}\phi_t)} , \quad (9.5)$$

where λ_s , V_s , ϕ_s are the wavelength, visibility, and phase in the signal channel, and λ_t , ϕ_t the wavelength and phase in the tracking channel. In practice, V_r^2 is computed from Equation 9.1 using X_r and Y_r instead of X and Y ; this procedure retains the advantage of using an unbiased estimator.

Equation 9.5 assumes that the atmospheric phase at λ_s is given by $(\lambda_t/\lambda_s)\phi_t$. If this were the case exactly, there would be no coherence losses, and the integration time could be arbitrarily long. A number of systematic effects (discussed in more detail in Section 9.3, see also Quirrenbach *et al.* 1994) can lead to a decorrelation of the phases between the signal and tracking channels, however. They introduce additional phase noise, which reduces the system visibility and limits the maximum integration time. The dependence of the system visibility on seeing and zenith angle is also made steeper, which increases the uncertainty of the calibration. In practice, therefore, phase-referenced averaging involves trading off some calibration accuracy for the gain in signal-to-noise.

9.2.4 Limb Darkening of Arcturus

A good example for the use of phase-referenced visibility averaging are the Mark III observations of limb darkening in Arcturus (Quirrenbach *et al.*, 1996). The main challenge of limb-darkening measurements is the need to collect data in the vicinity of and beyond the first zero of the visibility function.[†] The signal-to-noise ratio (Equation 9.4) of these measurements is normally very small. Since $V^2 \ll 1$, even observations of extremely bright stars like Arcturus may be in the “photon-starved” regime, and phase-referencing may lead to a substantial improvement. Fortunately the fringe-tracking channel of the Mark III interferometer provides a convenient phase reference. The signal-to-noise ratio in this channel is much higher than in the 550 nm signal channel, first because the tracking channel has a much larger bandwidth, and second because the tracking wavelength is longer and therefore has a higher visibility (see also Figure 9.4).

Figure 9.2 shows Mark III visibility measurements for Arcturus at 550 nm. The three baselines were chosen to bracket the first zero of the visibility function at that wavelength. The data were processed with phase-referencing (Equation 9.5), and a coherent integration time of 256 ms was chosen. The 3σ upper limit for the smallest visibilities plotted in Figure 9.2 is $V^2 \leq 10^{-4}$. This means that the data close to the zero would be in the photon-starved regime for the “standard” Mark III coherent integration time of 4 ms; the error bars would be much larger with the standard processing. The effect of varying the coherent integration time is illustrated in Figure 9.3. In this figure, the formal error of V^2

[†]On short baselines the visibility function of a limb-darkened disk is virtually indistinguishable from that of a somewhat smaller uniform disk.

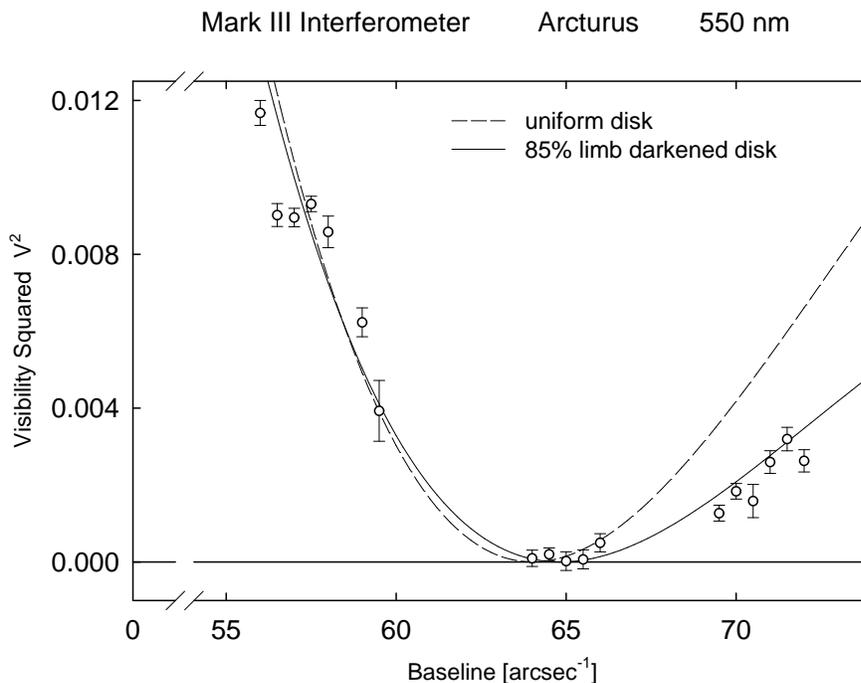


Figure 9.2: Mark III visibility data on α Boo (Arcturus) at $\lambda_s = 550$ nm on three different baselines near the first zero of the visibility function. The data were processed with the phase-referenced averaging algorithm, using a coherent integration time of 256 ms.

is plotted versus V^2 for a number of observations very close to the zero of the visibility function. Each observation was processed with four different coherent integration times (4 ms, 32 ms, 256 ms, and 1024 ms). We see that with increasing integration time both the formal errors (vertical position of the data points in Figure 9.3), as well as the scatter between them (horizontal spread of the points) get smaller. The dashed lines correspond to a -1σ -deviation from $V_{\text{cal}}^2 = 0$, and a $+1\sigma$ -deviation from $V_{\text{cal}}^2 = 2 \cdot 10^{-4}$. It can be seen from the figure that almost all data points are consistent with $0 \leq V_{\text{cal}}^2 \leq 2 \cdot 10^{-4}$ on the 1σ level, but the uncertainty of the V^2_{cal} estimate gets much smaller with increasing coherent integration time.

9.2.5 Further Applications of Same-Source Phase Referencing

The Mark III observations of Arcturus described in the previous section are an example of *wavelength bootstrapping*. This technique uses the fact that the signal-to-noise ratio may be high at a certain wavelength λ_1 , but low at another wavelength λ_2 . An important case, illustrated in Figure 9.4, is the situation where the difference in signal-to-noise is due to V^2 being high at λ_1 , but low at λ_2 . It is then possible to observe at λ_2 , while the fringe tracker is working at λ_1 . For example, wavelength bootstrapping is useful for imaging stellar photospheres, where λ_1 can be in the IR and λ_2 in the visible, or for imaging circumstellar

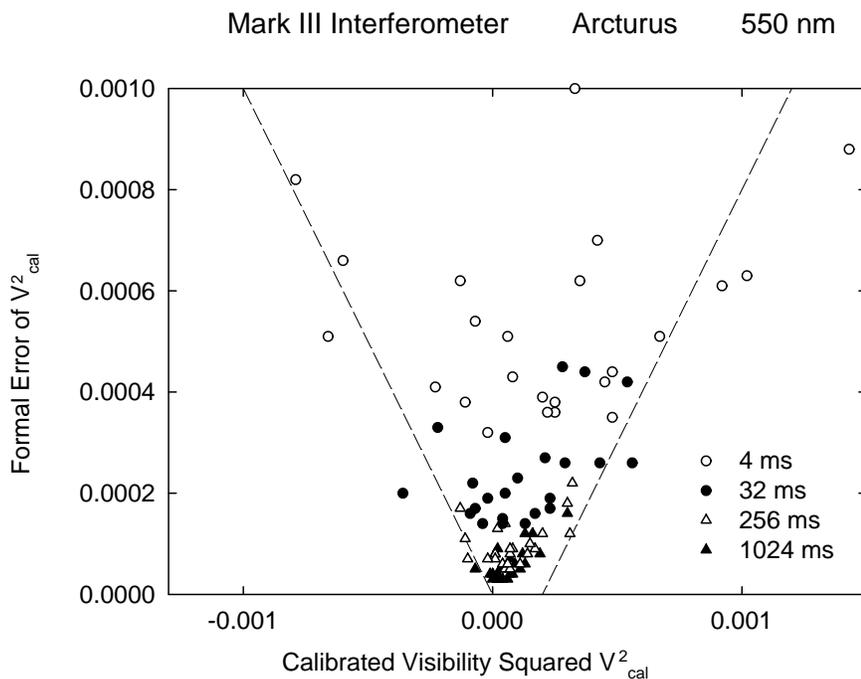


Figure 9.3: Mark III data on α Boo (Arcturus) at 550 nm, very close to the first zero of the visibility function (corresponding to the group of points near 65 arcsec^{-1} in Figure 9.2). The plot shows the formal error of V_{cal}^2 as a function of V_{cal}^2 , for four different coherent integration times. All points that are compatible with $0 \leq V_{\text{cal}}^2 \leq 2 \cdot 10^{-4}$ to within 1σ lie in the wedge-shaped region between the two dashed lines.

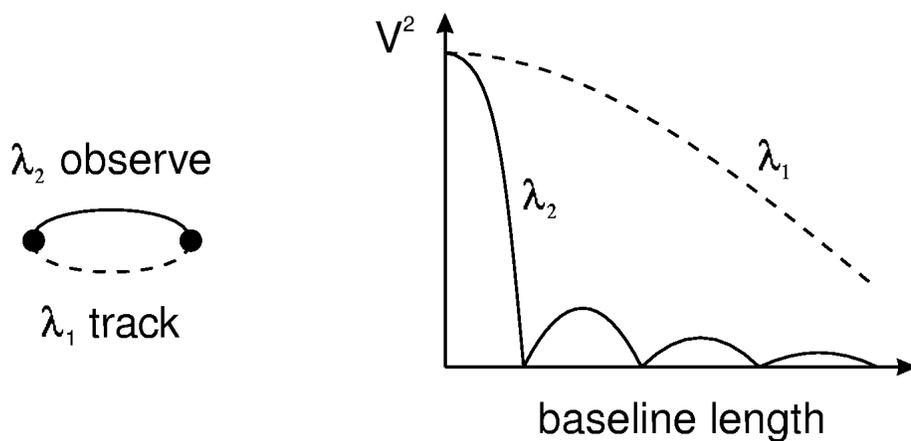


Figure 9.4: Wavelength bootstrapping. The fringes are tracked at a long wavelength. The observations are done on the same baseline, but at a shorter wavelength, where the resolution is higher, but V^2 much smaller.

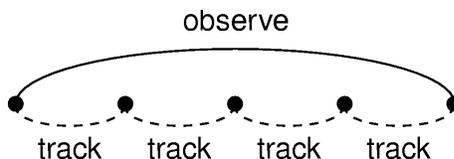


Figure 9.5: Baseline bootstrapping. The fringes are tracked on the short baselines of the array (in this case a five-element linear configuration), where V^2 is high. The observations can then be done on the long baselines.

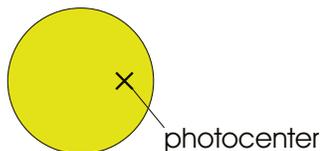
matter, where λ_1 can be in the continuum (where the small stellar photosphere dominates) and λ_2 in a line emitted by the extended material.

In interferometer arrays with more than two telescopes, a different variant of phase referencing is possible: baseline bootstrapping (see Figure 9.5) uses the signal on the short baselines of the array for the fringe-tracking servo, while data are taken on the long baselines, where V^2 can be much lower.[‡] The configuration of the Navy Prototype Optical Interferometer, which is optimized for observations of stellar surface structure, has been laid out specifically to make use of baseline bootstrapping (Mozurkewich and Armstrong, 1992). A related idea has been developed for arrays with telescopes of different sizes, such as the VLTI and the Keck Interferometer. Fringe tracking is required only on the more sensitive baselines which involve at least one large telescope, while bootstrapping enables observations on the baselines between two small telescopes. It is also possible, of course, to combine baseline bootstrapping and wavelength bootstrapping.

So far we have discussed applications of phase-referencing that use the phase relation between the reference channel and the signal channel only implicitly, to remove the atmospheric phase and to increase the interferometric sensitivity. One can also make explicit use of the referenced phases and use them for *phase-referenced imaging* or *phase-referenced spectroscopy*. If the reference star can be considered a point source (or if its structure phase can be computed and subtracted from the reference phase), the referenced phase can be used directly as the Fourier phase in an image reconstruction algorithm. An example is emission-line observations of circumstellar matter. In many cases the continuum emission of the stellar photosphere provides a nearly point-like reference for the much more extended line emission. The phase difference between line and continuum is then an observable that can be used for imaging the line emission. If sufficient spectral resolution is available, this can even be done separately for a number of radial velocity channels. By referencing to the continuum phase, these channel maps can be registered with respect to each other and with respect to the continuum. One should note that true imaging with full phase information is possible in this way even with a single-baseline instrument (if data are collected succes-

[‡]Clearly, when the pathlength from telescope 1 is kept equal to the pathlength from telescope 2, and the pathlength from telescope 2 equal to that from telescope 3, the paths from telescopes 1 and 3 are also equal. However, one has to keep in mind that the phase errors accumulate along the chain of baselines that are co-phased in this manner.

wavelength outside molecular band



wavelength inside molecular band

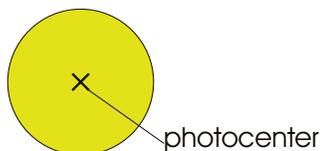


Figure 9.6: The shift of the star – planet photocenter with wavelength gives rise to an interferometric phase shift that can be exploited to obtain a spectrum of the planet.

sively on a sufficient number of points in the (u, v) plane to satisfy the Nyquist sampling theorem).

Another potential application of phase-referencing is the spectroscopy of faint stellar companions (Quirrenbach, 2000). For example, the near-infrared spectra of extrasolar planets should be characterized by extremely deep absorption bands of water and methane. The photocenter of a star-planet system is therefore slightly different outside the molecular bands, where the planet is relatively bright, and within the bands, where it is much dimmer (see Figure 9.6). The shift of the photocenter is proportional to the planet / star brightness ratio and can thus be used as a proxy for the planet spectrum. The shift of the photocenter gives rise to a corresponding wavelength dependence of the interferometer phase, which can be measured if the signal-to-noise ratio is sufficient and systematic effects are kept small. In the case of “hot Jupiters,” which are quite favorable because the planets are close to their parent stars and therefore hot and bright, the expected effect on the interferometer phase is ~ 0.5 mrad on the longest baselines of the Keck Interferometer or VLTI. This could be measured with a signal-to-noise ratio of ~ 3000 , but reducing the systematic instrumental and atmospheric effects to that level will be a very challenging task.

9.2.6 Off-Source Phase Referencing

A common characteristic of the techniques discussed so far is that the reference phase is measured on the target object itself, either at a different wavelength or on a different baseline. This helps for observations of bright objects in the low-visibility (and therefore “photon-starved”) regime, and for specific spectroscopic applications. For faint objects, however, one would clearly like to emulate the phase calibration procedure widely used in radio astronomy in which the atmospheric phase is determined from a bright source near the

target. In radio interferometry one can slew the telescope between target and reference in intervals of several minutes, but because of the short atmospheric coherence time at visible and near-infrared wavelengths, here the target and the reference have to be observed truly simultaneously. Off-source fringe tracking is therefore possible only in interferometers with a field much wider than feasible in a Michelson instrument; either a wide-field (e.g., Fizeau) setup or a *dual-star system* is required. In a dual-star interferometer, each telescope accepts two small fields and sends two separate beams through the delay lines. The delay difference between the two fields is taken out with an additional short-stroke differential delay line; an internal laser metrology system is used to monitor the delay difference (which is equal to the phase difference multiplied with $\lambda/2\pi$, of course). Dual-star interferometry has been demonstrated by the Palomar Testbed Interferometer (Colavita *et al.*, 1999); it is a vital component of the plans for the Keck Interferometer (Colavita *et al.*, 1998) and the VLT Interferometer (Quirrenbach *et al.*, 1998).

The dual-star technique has been developed mainly for interferometric astrometry (another application of phase referencing that makes explicit use of the phase difference), but it can also be used for phase-referenced visibility averaging or phase-referenced imaging. The most important problem encountered by all off-source phase-referencing techniques is anisoplanatism, i.e., the fact that atmospheric fluctuations are only partly correlated in different sky directions (see Section 9.3.7 below). The phase noise associated with anisoplanatism causes astrometric errors, and reduces the phase-referenced visibility dramatically if the distance to the reference source exceeds the isoplanatic angle. The need to find a reference object within the isoplanatic patch is a severe limitation for off-source phase-referencing; the chances to find a suitably bright star for a randomly chosen target are typically one in a hundred or worse. Still, there are a number of important astrophysical applications for this technique: astrometric searches for unseen companions (e.g., planets)[§], observations in clusters (e.g., near the Galactic Center), and programs in which a few suitable targets can be drawn from comparatively long lists (e.g., observations of extragalactic sources that happen to be close in the sky to a bright star).

The reference source can also be used for adaptive optics wavefront sensing, if such a system is available. In this case the whole entrance pupil of the interferometer is made fully cophased and the sensitivity of the interferometer is essentially identical to the sensitivity of a single telescope with the same diameter. It is thus important to realize that *bright objects are needed to cophase an interferometer, but very faint sources can be observed in a limited field around these reference sources.*

[§]In this case the target is normally a nearby and therefore bright star, which can be used for fringe tracking. It is still necessary to find nearby astrometric reference stars, but they can be much fainter, because phase-referenced fringe tracking can be applied to them: the *astrometric target* is the *interferometric reference* for the astrometric reference stars.

9.3 Phase Decorrelation Mechanisms

9.3.1 Phase Errors and Coherence Losses

We will now discuss a number of mechanisms that lead to phase errors and therefore to coherence losses and to a reduction of the phase-referenced visibility. These effects can be broadly divided into two classes, namely those mechanisms that are due to errors in the determination of the phase in the reference channel (Sections 9.3.2–9.3.5), and those that are due to atmospheric propagation effects (Sections 9.3.6–9.3.9). While some of the former processes are instrument-dependent and can be reduced (or even avoided) by improved interferometer and fringe-detector designs, the latter class sets fundamental limits to the application of phase-referencing methods from the ground. We will again use phase-referenced visibility averaging with the Mark III interferometer to give some specific numerical examples (see also Quirrenbach *et al.* 1994).

If the variance of the referenced phase ϕ_r associated with a decorrelation mechanism is $\sigma_{\phi,r}^2$, it will reduce V_r^2 by a factor η , which can be computed from

$$\eta = e^{-\sigma_{\phi,r}^2} . \quad (9.6)$$

For assessing the individual mechanisms, it is not only important to compare the numerical values of the associated phase variances, but also to note their dependencies on observing conditions (e.g. seeing, zenith angle) and particularly on stellar parameters (e.g. colors). While the standard calibration procedure will correct for a uniform reduction of V^2 , and to some extent for variations with observing conditions, effects that differ from star to star can introduce systematic errors that are difficult to detect. A priori limits on these effects are therefore necessary for practical applications of phase-referenced visibility averaging.

9.3.2 Photon Noise in the Tracking Channel

The finite number of photons detected during each coherent integration interval (4 ms in the Mark III case) sets a fundamental limit to the precision of the reference phase determination. The variance of ϕ_r due to photon noise in the tracking channel is

$$\sigma_{\phi,r}^2 = \left(\frac{\lambda_t}{\lambda_s}\right)^2 \sigma_{\phi,t,\text{phot}}^2 = \left(\frac{\lambda_t}{\lambda_s}\right)^2 \cdot \frac{2}{N_t V_t^2} , \quad (9.7)$$

where N_t and V_t are the number of the photons counted and the visibility in the tracking channel. $\sigma_{\phi,r}^2$ depends on the brightness and color of the star, and even on the baseline length (through V_t^2). However, for the fringe tracker to work reliably under average seeing conditions, $N_t V_t^2 \simeq 70$ is needed for the 4 ms sampling interval, giving $\eta \simeq 0.98$ for $\lambda_t = 700$ nm, $\lambda_s = 800$ nm, and $\eta \simeq 0.95$ for $\lambda_t = 700$ nm, $\lambda_s = 500$ nm. Thus the visibility reduction is slight even for stars that are close to the sensitivity limit of the fringe tracker, and negligible for stars that are substantially brighter. It is also possible to introduce the signal-to-noise in the tracking channel as an additional independent variable in the calibration process, if very high accuracy is required.

9.3.3 Color and Visibility Dependence of the Effective Tracking Wavelength

To achieve high sensitivity (and to keep the errors due to photon noise small), the bandpass in the fringe-tracking channel should be made as wide as possible. The effective wavelength to be used in Equation 9.5 is then given by

$$\lambda_t = \frac{\int d\lambda \lambda W_t(\lambda) N(\lambda) V(\lambda)}{\int d\lambda W_t(\lambda) N(\lambda) V(\lambda)} , \quad (9.8)$$

where $N(\lambda)$ is the number of photons emitted by a star as a function of wavelength, $V(\lambda)$ the visibility, and $W_t(\lambda)$ the combined response of atmosphere, instrument, and detector. If the wavelength used in Equation 9.5 differs from the true effective wavelength by $\delta\lambda_t$, the resultant variance of the reference phase is

$$\sigma_{\phi,r}^2 = \left(\frac{\delta\lambda_t}{\lambda_s} \right)^2 \cdot \langle \phi_t^2 \rangle . \quad (9.9)$$

As evident from Equation 9.8, the true effective wavelength depends on stellar colors and diameters, and on the baseline length. If for simplicity one uses $\lambda_t = 700$ nm for all stars, $\delta\lambda_t \lesssim 25$ nm for the parameters of the Mark III interferometer. With the additional assumption that the residual atmospheric phase rms not tracked by the fringe tracker $\sqrt{\langle \phi_t^2 \rangle} \lesssim 2$ rad, $\eta \gtrsim 0.99$ is derived from Equation 9.9.

9.3.4 Stroke Mismatch

In pathlength modulation schemes like that used by the Mark III, any difference between the stroke of the 500 Hz pathlength modulation and the wavelength λ will also lead to errors in the phase estimation, since then the bins A , B , C , and D do not correspond exactly to $\lambda/4$. (This correspondence is assumed implicitly in Equation 9.2.) For each channel, the gating of the electronic counters for A , B , C , and D has to be set by the on-line control system to match one quarter of the nominal wavelength. In this way, an effective stroke s is created for each channel. Defining

$$\varepsilon = \frac{2\pi}{\lambda} \cdot (s - \lambda) \quad \text{and} \quad \delta = \frac{\cos \varepsilon/4}{1 + \sin \varepsilon/4} , \quad (9.10)$$

it has been shown by Colavita (1985) that

$$\tan \phi_{\text{est}} = \delta \cdot \tan \phi_{\text{true}} , \quad (9.11)$$

where ϕ_{est} is the phase estimated from Equation 9.2, and ϕ_{true} is the true phase. For a complete treatment of the effect of the stroke mismatch, these equations have to be integrated over λ , with a suitable weighting function representing the bandpass of the tracking channel. To first order, however, it can be assumed that the phase error is given by Equations 9.10 and 9.11, evaluated at $\lambda = \lambda_t$. For $s_t - \lambda_t \leq 25$ nm, a phase error $\phi_{\text{est}} - \phi_{\text{true}} \leq 2^\circ$ is then obtained. Errors of this order can be safely ignored for most visibility averaging applications, but may be important for phase-referenced imaging and spectroscopy.

9.3.5 Fringe Jumps

An ideal fringe tracker would follow the atmospheric pathlength fluctuations to a fraction of λ_t , and ϕ_t would always be well within the interval $(-\pi, \pi)$. In practice, however, temporary excursions from the central fringe that are larger than $\lambda/2$ may occur, and the phase has to be “unwrapped” by the phase-referencing algorithm. This is done by imposing the requirement that the phase in successive data segments (4 ms intervals for the Mark III) should be continuous. While this process normally works well, occasional misidentifications are possible. It is obvious from Equation 9.5 that a 360° error in ϕ_t will lead to a phase jump in ϕ_r .

If the average number of these jumps during the coherent integration time T is small, the coherence loss is not dramatic. This requirement sets an upper limit to T . Since the probability of unwrapping errors depends only on the seeing and on the signal-to-noise in the tracking channel, it can be accounted for in the calibration procedure. In a series of tests with the Mark III, it turned out that the degradation of the phase-referenced visibility V_r due to fringe jumps was not serious for integration times up to 2 s, for average seeing conditions on Mt. Wilson.

9.3.6 Dispersion

While Equation 9.5 assumes that the atmospheric pathlength fluctuations are independent of wavelength, they are actually larger in the blue spectral range than in the red, because of dispersion. The two-color dispersion coefficient D is defined by

$$D = \frac{n(\lambda_t) - 1}{n(\lambda_s) - n(\lambda_t)} \quad , \quad (9.12)$$

where $n(\lambda)$ is the refractive index of air at λ . Typical values for $\lambda_t = 700$ nm and $\lambda_s = 450, 500, 550,$ and 800 nm are $D = 59, 87, 137,$ and -364 , respectively. If the total “unwrapped” phase in the tracking channel is denoted Φ_t , a phase error $(\lambda_t/\lambda_s)(\Phi_t/D)$ is introduced by the dispersion. Since the largest phase excursions occur on long time scales, this sets a limit to the coherence time. For Kolmogorov turbulence, the coherence time $t_{0,r}$ of ϕ_r is given by

$$t_{0,r} = |D|^{6/5} t_{0,s} \quad , \quad (9.13)$$

where $t_{0,s}$ is the atmospheric coherence time in the data channel (Colavita, 1992). Under average conditions on Mt. Wilson, $t_{0,s}$ is of order 6 to 8 ms at 500 nm. For integration times up to about 2 s, the coherence losses due to dispersion are therefore tolerable for visibility averaging, and they can be taken into account by the calibration procedure.

It is obviously possible to deal with dispersion explicitly by using

$$\tilde{\phi}_r = \phi_s - \frac{\lambda_t}{\lambda_s} \phi_t - \frac{\lambda_t}{\lambda_s} \cdot \frac{\Phi_t}{D} \quad (9.14)$$

instead of ϕ_r as defined in Equation 9.5. While this approach can reduce the phase errors by a factor ~ 10 , a residual effect due to water vapor fluctuations remains, because their dispersion is different from the values applicable to dry air.

9.3.7 Anisoplanatism

If the reference phase is measured on a star at an angular separation θ from the target object, there will be some decorrelation because the light from the two sources passes through different turbulence cells. The angle θ_i for which the variance of the relative phase is 1 rad^2 is called the isoplanatic angle. In interferometric applications, the independent contributions from the two arms of the interferometer have to be taken into account, giving a somewhat smaller value for θ_i . Under the assumption of a Kolmogorov turbulence spectrum with refractive index structure constant $C_n^2(h)$ at height h , the interferometric isoplanatic angle is

$$\theta_i = \left[5.82 k^2 (\sec z)^{8/3} \int_0^\infty dh C_n^2(h) h^{5/3} \right]^{-3/5}, \quad (9.15)$$

where $k = 2\pi/\lambda$ is the wavenumber (assumed here to be equal for the target and reference channels), and z the zenith angle. While this expression holds for small apertures, a somewhat more optimistic estimate is obtained for larger apertures (Colavita, 1992). Typical values for θ_i are of order a few arcseconds, much larger than the interferometric field of view of a Michelson interferometer. In applications where the reference phase is measured on the object of scientific interest itself, anisoplanatism does not occur at all. However, it is the most severe limitation for dual-star interferometry. We see from Equation 9.15 that $\theta_i \propto k^{-6/5} \propto \lambda^{6/5}$; this means that finding reference stars for dual-star interferometry is much easier at longer wavelengths.

9.3.8 Differential Refraction

An effect somewhat similar to anisoplanatism occurs even when the angular separation between the target and the reference is zero. If $\lambda_s \neq \lambda_t$, the beams at the two wavelengths follow different paths through the atmosphere at non-zero zenith angles, due to differential refraction. For a Kolmogorov turbulence spectrum, the corresponding phase variance is

$$\sigma_{\phi,r}^2 = 5.82 k_s^2 \left[\frac{h_0 (n(\lambda_t) - 1) e^{-h_1/h_0}}{D} \right]^{5/3} \tan^{5/3} z \sec^{8/3} z \int_0^\infty dh C_n^2(h) \left(1 - e^{-h/h_0} \right)^{5/3}, \quad (9.16)$$

where $k_s = 2\pi/\lambda_s$ is the wavenumber in the signal channel, $n(\lambda_t)$ is the atmospheric index of refraction at λ_t , D is the atmospheric dispersion between λ_s and λ_t defined by Equation 9.12, h_0 is the scale height of the atmospheric density, h_1 is the elevation of the observatory site above sea level, z is the zenith angle, and $C_n^2(h)$ is the refractive index structure constant. Again, this estimate might be somewhat pessimistic, since averaging over the aperture has not been taken into account.

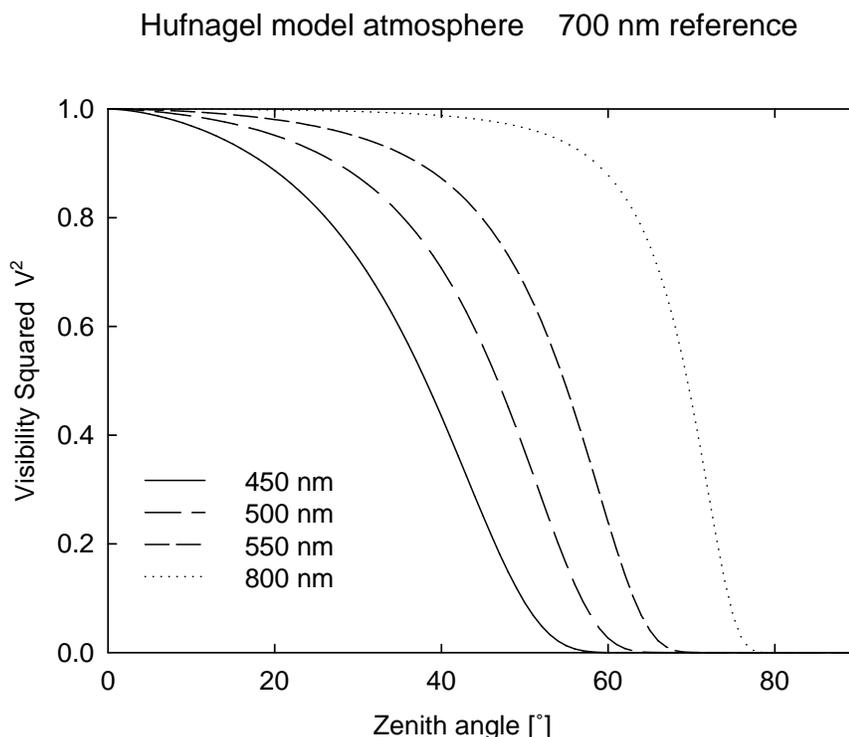


Figure 9.7: Reduction of V^2 due to differential refraction as a function of zenith angle z , predicted from a Hufnagel (1974) model atmosphere. The reference wavelength $\lambda_t = 700$ nm; the wavelength in the data channels $\lambda_s = 450, 500, 550,$ and 800 nm.

The phase variance due to differential refraction depends very strongly on z ; while it is negligible close to the zenith, it is the dominant decorrelation mechanism at intermediate to large zenith angles for the parameters of the Mark III phase-referenced visibility averaging experiments. From Equation 9.16 it is obvious that differential refraction—like anisoplanatism—is more strongly affected by high-altitude turbulence than by disturbances close to the ground. This is expected, since the beams from target and reference coincide at the telescope aperture; their separation increases with height when they are traced back through the atmosphere. To carry out quantitative calculations of differential refraction, it is therefore necessary to know the turbulence profile; in the absence of better measurements we use the model for the atmospheric turbulence as a function of height h (in m) by Hufnagel (1974),

$$C_n^2(h) = 2.7 \cdot \left(2.2 \cdot 10^{-53} h^{10} e^{-h/1000} + 10^{-16} e^{-h/1500} \right) . \quad (9.17)$$

Figure 9.7 shows the reduction of V_r^2 derived from a numerical integration of Equation 9.16, with the Hufnagel turbulence profile. The values $h_0 = 8300$ m, $h_1 = 1700$ m (applicable to Mt. Wilson), $\lambda_t = 700$ nm, and $\lambda_s = 450, 500, 550,$ and 800 nm were used. This figure demonstrates that differential refraction leads to a much steeper dependence of the system visibility with zenith angle in the phase-referenced data than in incoherent averages. This

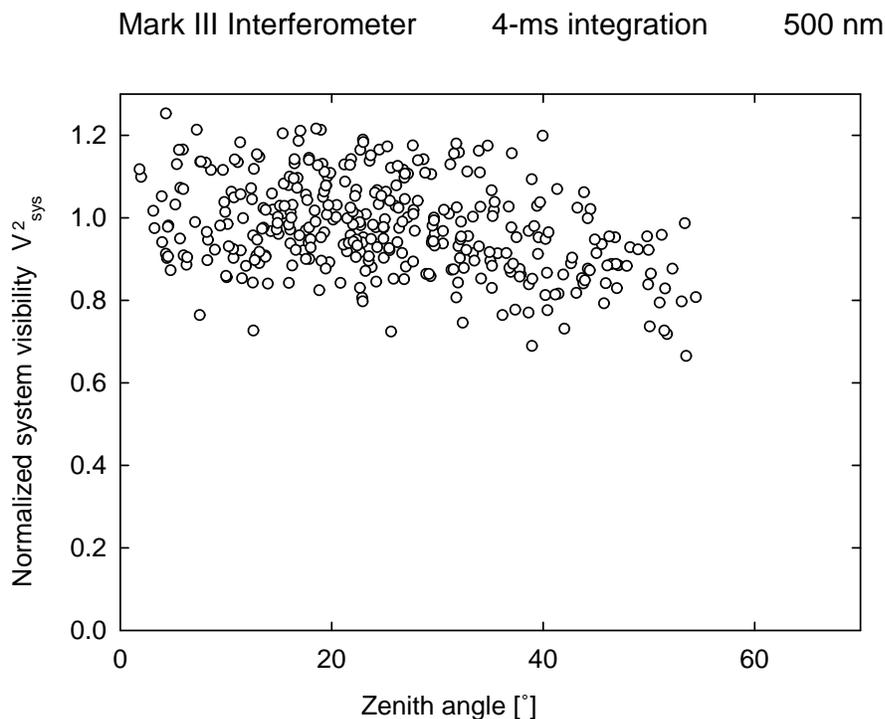


Figure 9.8: Observed V^2 divided by an estimate V_{est}^2 from photometric data, for 16 stars at 500 nm. The data were obtained on the nights July 29 and July 31, 1989; they are plotted a function of zenith angle z . Each night was normalized to 1 at $z = 0$. Each measurement corresponds to one 75-s observation. The standard data reduction procedure was used, which averages the 4-ms samples incoherently.

effect is particularly important in the blue spectral range, where the dispersion is large (small values of D). Differential refraction therefore restricts the application of phase-referenced visibility averaging to moderate zenith angles, depending on the wavelength λ_s and on the seeing.

Figure 9.8 shows the Mark III system visibility for two nights (July 29 and 31, 1989) as a function of zenith angle z , for the data integrated incoherently with the standard method; it has been normalized to $V_{\text{sys}}^2 = 1$ at $z = 0$. It is obvious that V_{sys}^2 varies only slightly with z ; this variation is mostly due to the degradation of the seeing for longer pathlengths through the atmosphere. Figure 9.9 shows the same data, but processed with the phase-referencing algorithm, using an integration time of 1024 ms. A strong reduction of the system visibility is now apparent at $z \gtrsim 40^\circ$. The solid line indicates the visibility reduction due to differential refraction predicted by the Hufnagel (1974) atmosphere model. The qualitative agreement between the observations and this model demonstrates that differential refraction is indeed the dominant reason for coherence losses at intermediate to large zenith angles.

Mark III Interferometer 1024-ms integration 500 nm Hufnagel model

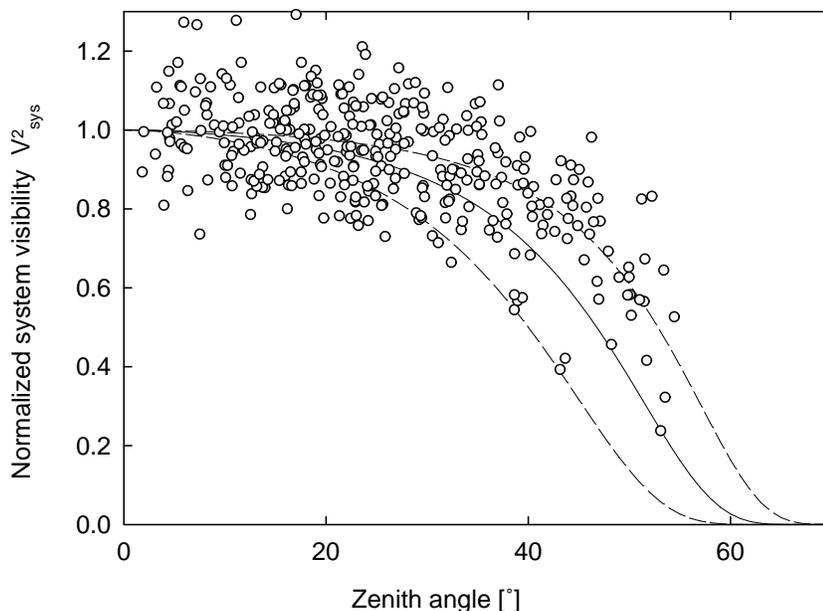


Figure 9.9: The same data as in Figure 9.8, but processed with the phase-referenced averaging algorithm. The coherent (phase-referenced) integration time is 1024 ms. The solid curve is the visibility reduction due to differential refraction predicted by the Hufnagel model atmosphere; the dashed curves correspond to atmospheres that have 0.5 and 2 times the C_n^2 of the Hufnagel model at all heights.

9.3.9 Diffraction

Finally, if $\lambda_s \neq \lambda_t$, there will be some decorrelation because of diffraction. The phase variance due to diffraction is related to the intensity scintillation variance $\sigma_{\ln I}^2$ by

$$\sigma_{\phi,r}^2 = G(\lambda_t/\lambda_s) \sigma_{\ln I}^2(\lambda_t) \quad , \quad (9.18)$$

with a function $G(r)$, which can be approximated by

$$G(r) \simeq \left(r^{1/2} (r - 1) / 2 \right)^{4/3} \quad (9.19)$$

for $1 \leq r \lesssim 1.5$ (Colavita, 1992). Observed values for $\sigma_{\ln I}^2$ on Mt. Wilson range from 0.005 to 0.05. The larger of these values gives $\sigma_{\phi,r}^2 = 0.0073$, or $\eta = 0.99$ for $\lambda_t = 700$ nm, $\lambda_s = 500$ nm. Since all stars are affected equally, the calibration procedure takes into account the small coherence loss due to diffraction.

References

- M.M. Colavita, *Atmospheric Limitations of a Two-Color Astrometric Interferometer*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA (1985).
- M.M. Colavita, “Phase-referencing for stellar interferometry,” in *High-Resolution Imaging by Interferometry II*, J.M. Beckers and F. Merkle, eds. ESO Conference and Workshop Proceedings, **39**, 845–851 (Garching, Germany: European Southern Observatory, 1992).
- M.M. Colavita, A.F. Boden, S.L. Crawford, A.B. Meinel, M. Shao, P.N. Swanson, G.T. van Belle, G. Vasisht, J.M. Walker, J.K. Wallace, and P.L. Wizinowich, “The Keck Interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 776–784 (1998).
- M.M. Colavita, J.K. Wallace, B.E. Hines, Y. Gursel, F. Malbet, D.L. Palmer, X.P. Pan, M. Shao, J.W. Yu, A.F. Boden, P.J. Dumont, J. Gubler, C.D. Koresko, S.R. Kulkarni, B.F. Lane, D.W. Mobley, and G.T. van Belle, “The Palomar Testbed Interferometer,” *Astrophys. J.* **510**, 505–521 (1999).
- R.E. Hufnagel, “Variations of atmospheric turbulence,” in Digest of Technical Papers Presented at the Topical Meeting on *Optical Propagation through Turbulence*, Wa1–1 – Wa1–4 (Washington, DC: Optical Society of America, 1974).
- D. Mozurkewich and J.T. Armstrong, “Array layout and beam combination considerations necessary for imaging with a long baseline optical interferometer,” in *High-Resolution Imaging by Interferometry II*, J.M. Beckers and F. Merkle, eds. ESO Conference and Workshop Proceedings, **39**, 801–804 (Garching, Germany: European Southern Observatory, 1992).
- D. Mozurkewich, K.J. Johnston, R.S. Simon, P.F. Bowers, R. Gaume, D.J. Hutter, M.M. Colavita, M. Shao, and X.P. Pan, “Angular diameter measurements of stars,” *Astron. J.* **101**, 2207–2219 (1991).
- A. Quirrenbach, “Astrometry with the VLT interferometer,” in *From Extrasolar Planets to Cosmology: The VLT Opening Symposium*, J. Bergeron and A. Renzini, eds., in press (Berlin, Germany: Springer-Verlag, 2000).
- A. Quirrenbach, V. Coudé du Foresto, G. Daigne, K.-H. Hofmann, R. Hofmann, M. Lattanzi, R. Osterbart, R. le Poole, D. Queloz, and F. Vakili, “PRIMA — study for a dual-beam instrument for the VLT Interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 807–817 (1998).
- A. Quirrenbach, D. Mozurkewich, D.F. Buscher, C.A. Hummel, and J.T. Armstrong, “Phase-referenced visibility averaging in optical long-baseline interferometry,” *Astron. Astrophys.* **286**, 1019–1027 (1994).
- A. Quirrenbach, D. Mozurkewich, D.F. Buscher, C.A. Hummel, and J.T. Armstrong, “Angular diameter and limb darkening of Arcturus,” *Astron. Astrophys.* **312**, 160–166 (1996).
- M. Shao and M.M. Colavita, “Long-baseline optical and infrared stellar interferometry,” *Ann. Rev. Astron. Astrophys.* **30**, 457–498 (1992).
- M. Shao, M.M. Colavita, B.E. Hines, D.H. Staelin, D.J. Hutter, K.J. Johnston, D. Mozurkewich, R.S. Simon, J.L. Hershey, J.A. Hughes, and G.H. Kaplan, “The Mark III stellar interferometer,” *Astron. Astrophys.* **193**, 357–371 (1988).

Part III

Astrometry

Chapter 10

Wide-Angle Astrometry

DONALD J. HUTTER

ASTROMETRY DEPARTMENT, U.S. NAVAL OBSERVATORY
WASHINGTON, DC 20392

The goal of wide-angle astrometry is to determine accurate relative-positions of stars that are widely separated on the sky. The problem via interferometry is to recover the two coordinates for each star from the observed delays. In principle, sufficient delay measurements would allow the baseline vectors of the interferometer to be determined along with the delay constants and the positions of the stars. However, the actual situation is greatly complicated by the presence of the atmosphere and the fact that neither the delay “constants” nor the baseline vectors are stable over time. The design of, and the analysis of the data from, any ground-based optical interferometer must overcome all three of these effects. The design, operation, and the analysis of data from the Navy Prototype Optical Interferometer (NPOI) are here presented as examples of how to overcome the effects of the atmosphere and the instrumental instabilities in order to achieve accurate wide-angle astrometry. The status of the implementation of these techniques at the NPOI is presented.

10.1 Introduction

The Navy Prototype Optical Interferometer (Armstrong *et al.*, 1998a), located on Anderson Mesa, AZ (Figure 10.1), is a joint project of the U.S. Naval Observatory and the Naval Research Laboratory in cooperation with the Lowell Observatory. The NPOI includes arrays for imaging and for astrometry. The imaging array consists of six movable 50-cm siderostats feeding 12-cm apertures, with baseline lengths from 2.0 m to 437 m. The astrometric array consists of four fixed 50-cm siderostats feeding 12-cm apertures (soon to be increased to 35 cm), with baseline lengths from 19 m to 38 m. The arrays share vacuum



Figure 10.1: An aerial view of the NPOI showing the array as viewed from the north-east. The road surrounding the array can be seen along with parts of the north (right) west (center) and east (left) arms of the imaging array (station piers and vacuum feed lines visible). The astrometric siderostat shelters are the white structures at the array center. The beam-combining laboratory is the long structure to the right of the array center, with the ‘long’ delay line vacuum tanks (under construction) extending towards the north. The interferometer control building appears at the far right-center.

feed and delay-line systems. The NPOI features rapid tip-tilt star tracking, active group-delay fringe tracking over a wide band (450–850 nm in 32 channels), and a high degree of automation. The astrometric array includes an extensive baseline metrology system to measure the motions of the siderostats with respect to the local bedrock to 100 nm accuracy. Additional details of the NPOI design can be found in Armstrong *et al.* (1998b), Clark *et al.* (1998), Mozurkewich (1994), and White *et al.* (1998).

The initial goal of wide-angle astrometric observations with the NPOI will be to produce a catalog of positions for ~ 1000 of the brighter Hipparcos stars with an internal accuracy of 1–3 mas. Astrometric observations of radio stars will be used to orient the NPOI catalog with respect to the fundamental reference frame defined by extragalactic radio sources. With an anticipated operational lifetime of more than a decade, the NPOI will significantly improve the measured proper motions of these (and additional) stars. (Hipparcos positional accuracies will have already degraded to ~ 10 mas by 2001 due to proper motion uncertainties.) Position measurements repeated at regular intervals will also allow unambiguous separation of binary motion from proper motion, an accomplishment that might be difficult to achieve from space-based observations that are likely to be repeated only at intervals of decades.

The NPOI catalog will also be used to check for and correct any systematic rotation in the Hipparcos reference system (estimated at up to 0.25 mas/yr), and will be used to check for, and possibly correct, proper-motion induced zonal systematics in the Hipparcos system.

Thus, the NPOI, with a capability of milliarcsecond astrometry and a long, continuous operational lifetime, will be capable of maintaining the optical reference frame by improving the proper motions of thousands of the brighter Hipparcos stars through repeated observations.

10.2 Wide-Angle Astrometry

The problem of astrometry via interferometry is to recover the two coordinates for each star from the observed delays. In the absence of atmospheric effects, the geometrical delay can be defined as

$$d_{G,ij}(t) \equiv d_j(t) - d_i(t) = \mathbf{B}_{ij}(t) \cdot \hat{\mathbf{s}}_0 - C_{ij}, \quad (10.1)$$

where the geometrical delay is the difference between the delay line lengths d_i and d_j that is required to equalize the effective optical paths from the star to the point of beam combination via each of two apertures i and j , $\mathbf{B}_{ij}(t)$ is the baseline between the apertures, $\hat{\mathbf{s}}_0$ is the star position, and C_{ij} is the difference between the “fixed” internal optical path lengths C_i and C_j within the instrument (which are independent of wavenumber in the case of the NPOI since the optical paths are in vacuum). In principle, sufficient delay measurements would allow the baseline vectors to be determined along with the delay constants, and the positions of the stars (e.g., Hummel *et al.* 1994). However, the actual situation is greatly complicated by the presence of the atmosphere, and the fact that neither the delay “constants” (C_{ij}), nor the baseline vectors are stable over time. Thermal drifts in the positions of siderostats, and all subsequent elements in the optical paths prior to the point of beam combination, typically produce drifts in the delay constants of up to tens of $\mu\text{m}/\text{hour}$. Mechanical imperfections in the siderostats can produce $\sim 10 \mu\text{m}$ changes in the baseline vectors even between successive observations of widely spaced stars! Finally, the atmosphere also induces large ($\geq \mu\text{m}$), delay fluctuations on timescales as short as milliseconds. The design of the instrument and data analysis for the NPOI attempts to overcome all three of these effects.

10.3 Baseline Metrology

The design goal of the astrometric array of the NPOI is to measure stellar positions to 1–3 mas precision over the entire sky accessible from Anderson Mesa. This precision depends on knowing the 19–38 m astrometric baselines to $\sim 100 \text{ nm}$. Since the baselines are not stable to that degree, we must measure changes due to such effects as thermal changes and mechanical imprecisions in the siderostat mounts and correct for them in the astrometric

solution (Equation 10.1). The baseline metrology system (Hutter 1992, Elias 1994), described in Figure 10.2, is designed to monitor the motions of the siderostats with respect to local bedrock and to one another, to the required level of precision. This system consists of a number of laser interferometers tied to four temperature-stabilized super-Invar reference plates, one next to each siderostat. Five laser interferometers measure the position of a “cat’s-eye” retroreflector (Danchi *et al.*, 1986) near the intersection of the rotation axes of each siderostat relative to the adjacent reference plate*. Typical results for the observed motion of the siderostat retroreflector are shown in Figure 10.3. The translation and tilt of the reference plates are, in turn, measured by two other metrology subsystems. In the first case, six laser interferometers monitor changes in the distances of three points on each plate from retroreflectors embedded in a deep subsurface layer. The second subsystem consists of interferometers along lines of sight between the reference plates that detect motions of the plates in the horizontal plane. Most of the length of each laser interferometer line of sight is in vacuum. Corrections are applied to the laser metrology data for changes in the optical path length due to variations of the temperature of the transmissive optics (and the variations in temperature, pressure, and relative humidity of the small remaining air paths).

Together, these various subsystems contain 56 laser interferometers that will allow a continuous measurement of the time evolution of the baselines with respect to an Earth-fixed reference system with sub-micron error.

10.4 Constant Term

Several techniques can be used to measure the temporal variations in the delay zero-point offset (“constant term”) on each interferometer baseline. If rapid, and/or discontinuous path length variations are present, then an internal laser metrology system must be used to continuously monitor the optical paths through the instrument. A single color (preferably infrared) metrology system could be employed to measure the relative changes in the paths, in combination with occasional fringe tracking observations of an internal white-light source, with the siderostats in autocollimation. Alternatively, a two-color, absolute metrology system could be employed. Both these systems pose significant technical challenges, including, for example, providing a retroreflector near the siderostat pivot that can be used *simultaneously* for internal path monitoring *and* baseline metrology. Fortunately, in the case of the NPOI, the temporal variations in the constant terms are relatively slow and continuous. In this case, it is practical to determine the differential optical path variations in the vacuum feed system and beam combining optics by periodic (~ 30 min) white-light observations, and determine variations due to siderostat pivot motion using baseline metrol-

*Since it can be shown (Hines *et al.*, 1990) that any point on the siderostat mirror surface can, if consistently used, be defined as one end of the baseline vector, the motion of a retroreflector perpendicular to the mirror surface can be monitored to determine the temporal variations of the baseline vector. In practice, the retroreflector is placed as close to the siderostat pivot as possible (within a few μm) to allow monitoring of its position with laser beams of fixed orientation.

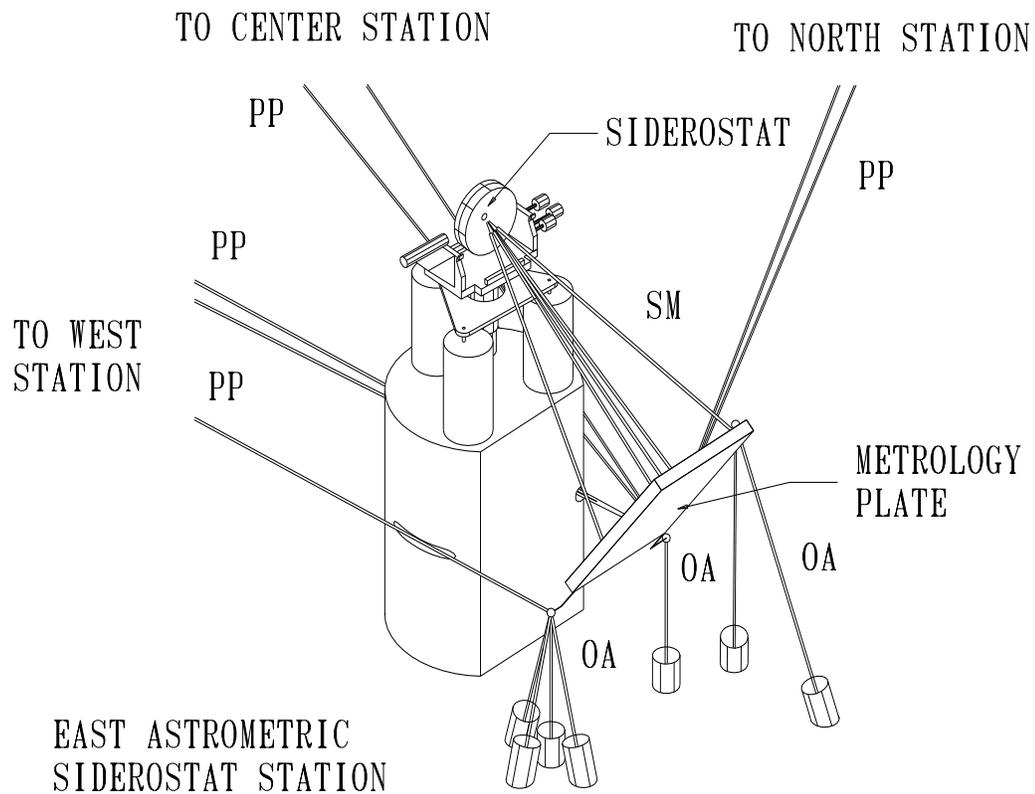


Figure 10.2: The baseline metrology system of the NPOI consists of 56 laser interferometers configured (in three subsystems) to monitor changes in the locations of the siderostats, in three dimensions, with respect to the local bedrock. This elevation view of the east astrometric siderostat station shows parts of the three metrology subsystems: the siderostat metrology (SM; five beams), which monitors the distance from the metrology plate to the siderostat pivot point; the optical anchor metrology (OA; seven beams), which monitors the motion of the metrology plate with respect to bedrock 7 m below the surface; and the pier-to-pier metrology (PP; seven beams), which monitors the horizontal motions of the metrology plates with respect to one another.

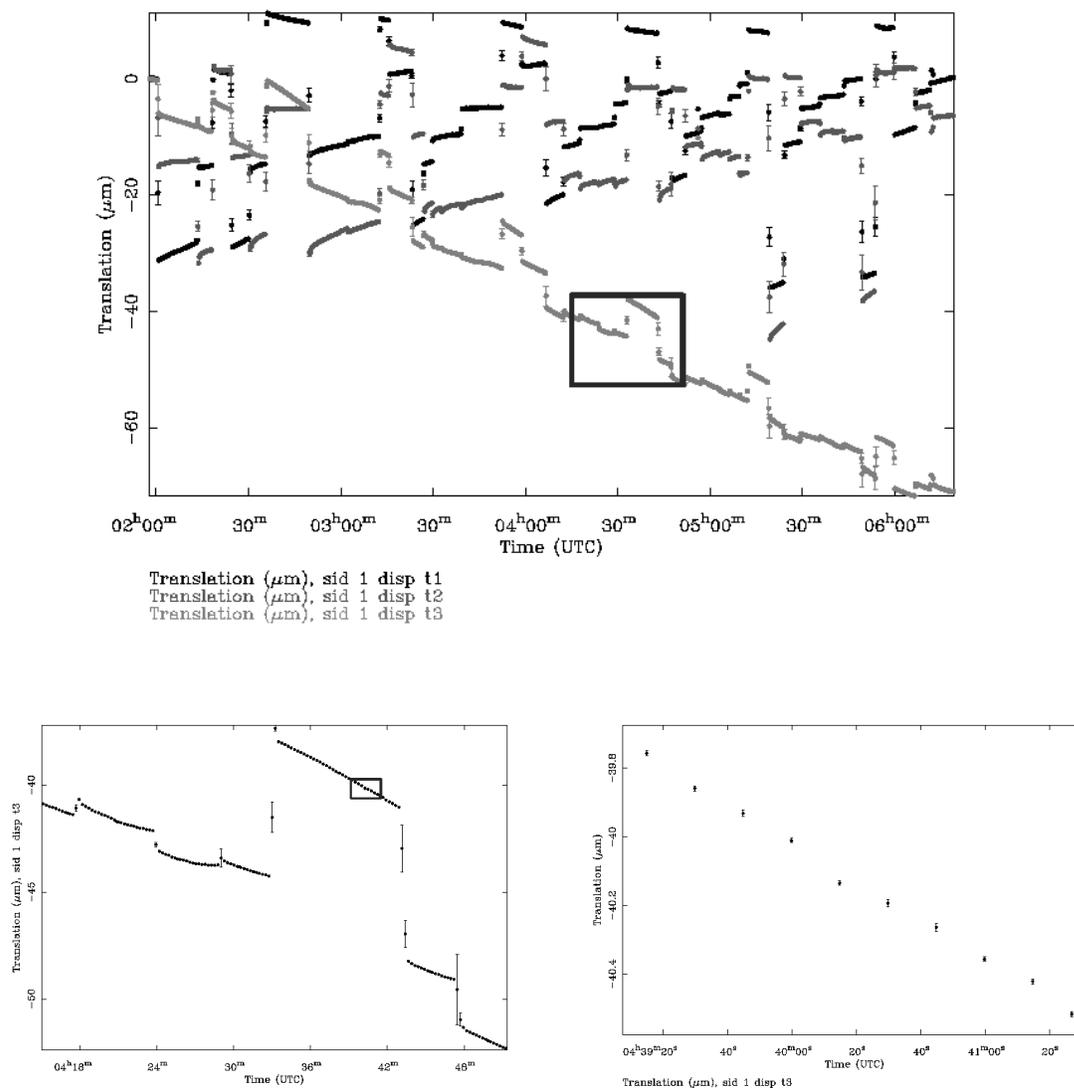


Figure 10.3: The motion of one siderostat pivot versus time, derived from siderostat metrology. The points plotted in the top panel represent the components of the pivot-point motion in the east, north, and vertical directions (upper, middle, and lower traces, respectively). The data show both the characteristic long-term drifts due to thermal contraction of the siderostat and its pier over the course of the night, and the shorter-term step-like changes (observed when the siderostat slews between stars) due to mechanical imperfections of the siderostat that affect the location of the siderostat pivot. The bottom left and right panels show successive blow-ups of a section of the top panel to illustrate the typically small formal errors of the derived siderostat motion (typically 5–10 nm).

ogy measurements.[†] The data from the white-light observations can be interpolated to the time of each stellar observation, then corrected for the siderostat pivot position at the time of that observation.

10.5 Compensation of Atmospheric Effects

The correction of the delay variations caused by atmospheric turbulence relies on the fact that, in the optical, atmospheric dispersion varies in a significantly nonlinear manner with wavelength. The method outlined below is a generalization of the “two-color” method (e.g., Colavita *et al.* 1987) that was successfully applied to astrometric data from the Mark III interferometer (Hummel *et al.*, 1994).

The path change in the atmosphere above an array element i is equivalent, to good approximation, to replacing a length P_i of the vacuum path with air. The change in equivalent path length is given by $A_i(\sigma, t) = P_i(t)[n(\sigma) - 1]$, where $n(\sigma)$ is the refractive index of air and σ is wavenumber. The observed delay, the delay required for effective path-length equality, is then given by

$$d_{ij}(\sigma, t) = \mathbf{B}_{ij}(t) \cdot \hat{\mathbf{s}}_0 - C_{ij} - A_{ij}(\sigma, t) = d_{G,ij}(t) + d_{A,ij}(\sigma, t), \quad (10.2)$$

where $A_{ij} \equiv A_j - A_i$, and the atmospheric delay $d_{A,ij}$ is the delay required to compensate for A_{ij} .

The observed phase ϕ_{ij} , measured with respect to the direction $\hat{\mathbf{s}}_0$ is then given by

$$\phi_{ij}(\sigma, t) = \phi_{S,ij}(\sigma) + \phi_{A,ij}(\sigma, t), \quad (10.3)$$

where the atmospheric phase $\phi_{A,ij}$ is defined by

$$\phi_{A,ij}(\sigma, t) \equiv 2\pi\sigma d_{A,ij}(\sigma, t) = -2\pi\sigma[A_j(\sigma, t) - A_i(\sigma, t)]. \quad (10.4)$$

The variation of $\phi_{A,ij}(\sigma, t)$ with σ can then be used to estimate A_{ij} , because $n(\sigma) - 1$ has a significant σ^2 dependence at optical wavelengths. If we expand $\phi_{A,ij}$ as a Taylor series around some σ_0 , we obtain

$$\phi_A(\sigma) = \phi_A(\sigma_0)[D_1 + \sigma D_2 + \Phi(\sigma)], \quad (10.5)$$

where the D coefficients depend on σ_0 , $n(\sigma_0) - 1$, $dn/d\sigma$, and $d^2n/d\sigma^2$. The first term gives a constant offset in phase. The second term produces a phase term that is indistinguishable from an error in the position of the star. The $\Phi(\sigma)$ term makes it possible to determine the air-path mismatch. With the air path mismatch determined, we can calculate $d_{A,ij}(\sigma)$ and subtract it from $d_{ij}(\sigma)$, resulting in an estimate of $d_{G,ij}$, which is the desired datum for determining the stellar positions (Equation 10.1).

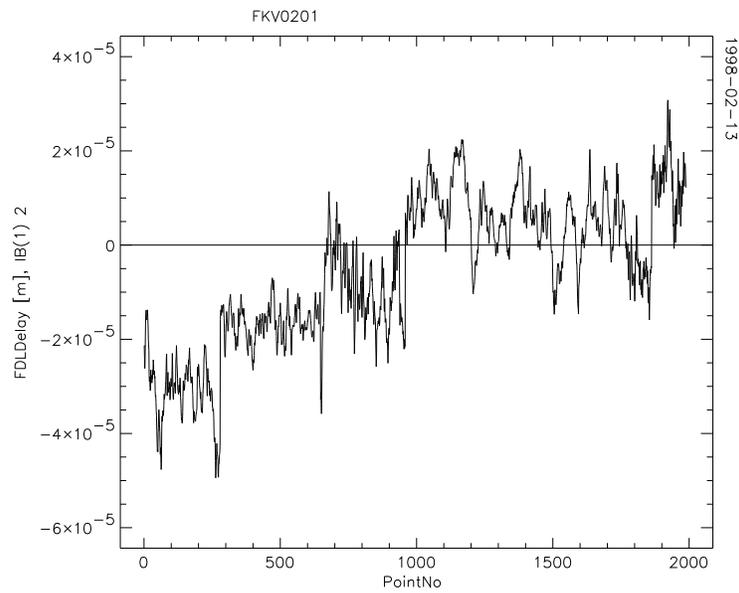
[†]The component of the motion of the siderostat pivot in the direction of the siderostat feed contributes to the variations in the “constant” terms.

The NPOI, with 32 spectral channels, is well adapted for the application of this technique for the correction of atmospheric effects. The use of vacuum delay lines renders the interferometer insensitive to plane-parallel atmospheric refraction and allows simultaneous fringe tracking in spectral channels over the entire optical bandpass. Fringe tracking on the NPOI is implemented via a group delay tracking technique, which makes use of the fringes across a spectrum of the combined light, along with delay modulation. The delay on each baseline is modulated by a small number of wavelengths while, for each spectral channel and baseline, synchronously measuring the photon count rates in eight bins per wavelength of modulation. (Data are collected in this way for 2 ms, a period short compared to the temporal coherence time of the atmosphere.) The complex Fourier transform of each set of bins provides the complex fringe visibility for that channel. A second Fourier transform over all the channels yields the group delay used for fringe tracking. Knowledge of the group delay for each 2-ms period allows one to rotate the complex visibility phasors by $e^{(2\pi id\sigma)}$, where d is the group delay and σ is the wavenumber of the channel. This allows the coherent addition (averaging) of the data to provide sufficient signal-to-noise to determine the variation of the fringe phase with wavenumber and thus determine the dispersion correction. All the data reported here were coherently averaged to 200 ms. Application of the dispersion corrections typically results in a four to eight-fold reduction in the rms variations of the delays within an individual scan for bright stars. Results of dispersion correction on typical stellar data are shown in Figure 10.4.

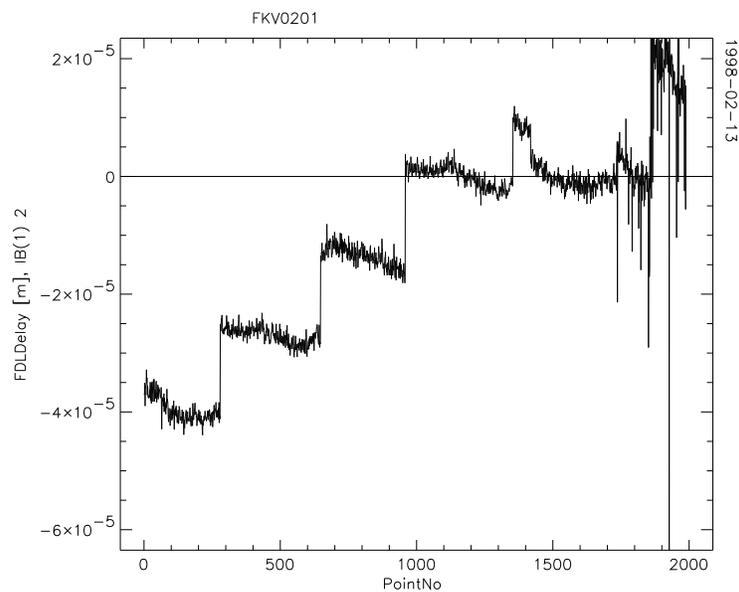
10.6 Current Status

The implementation of baseline metrology, constant-term measurement, and correction of atmospherically induced delay fluctuations has advanced to the point that accurate wide-angle observations with the NPOI are near at hand.

Dispersion compensation of the internal optical path observations of the white-light source, which are effected by delay fluctuations due to air paths between the siderostats and the vacuum-feed system (and to a much lesser extent the air paths in the vicinity of the beam combiner), can now be corrected to ~ 300 nm. This correction results in white-light fringe tracking data that varies sufficiently smoothly to allow accurate interpolation of the instrumental delay offset to the time of the stellar observations from white-light observations spaced at practical intervals (~ 30 min).



(a)



(b)

Figure 10.4: An example, for observations of FK5 201 on 02-13-98 UT, of the application of dispersion corrections to the observed delays (on the center-east baseline). Figure 10.4a (top) displays the residuals, relative to the calculated delay (estimated from the nominal star position, baseline, and time), for all of the 200-ms coherently averaged delays. (These data were collected during seven scans, spaced over 4 hours). Figure 10.4b (bottom) displays the result of the application of the dispersion corrections to the uncorrected delays of Figure 10.4a. Note the much reduced, and more random, distribution of the points within each scan. (Discontinuities between scans are likely due to baseline motion, and uncertainties in the star's position.)

Dispersion compensation of stellar observations resulting in a reduction in the rms fluctuations of observed delays to ~ 500 nm per (typically 90-second) observation has been achieved, but only for bright (third magnitude) stars. However, improvements in the techniques for coherently averaging the 2-ms fringe data promise similar accuracies in the atmospheric compensation of observations of much fainter stars in the near future. To date, measurements of siderostat pivot motion have demonstrated repeatability at the ~ 200 -nm level, while known, externally applied baseline motions (and resulting delay changes) can be reproduced to the level of ~ 1 μm , even before dispersion compensation is applied to the measured delays.

Recent advances in all these techniques for the correction of stellar observations to the corresponding “geometrical” delays promise achievement of accurate wide-angle astrometric solutions for stellar positions in the very near future.

Acknowledgments

The NPOI project is funded by the Oceanographer of the Navy and the Office of Naval Research.

References

- J.T. Armstrong, D. Mozurkewich, L.J. Rickard, D.J. Hutter, J.A. Benson, P.F. Bowers, N.M. Elias II, C.A. Hummel, K.J. Johnston, D.F. Buscher, J.H. Clark III, L. Ha, L.-C. Ling, N.M. White, and R.S. Simon, “The Navy Prototype Optical Interferometer (NPOI),” *Astrophys. J.* **496**, 550–571 (1998a).
- J.T. Armstrong, D. Mozurkewich, T.A. Pauls, and A.R. Hajian, “Bootstrapping the NPOI: keeping long baselines in phase by tracking on short baselines,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 461–466 (1998b).
- J.H. Clark III, L. Ha, D. Mozurkewich, and J.T. Armstrong, “Design of the long delay lines for the Navy Prototype Optical Interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 497–504 (1998).
- M.M. Colavita, M. Shao, and D.H. Staelin, “Two-color method for optical astrometry: theory and preliminary measurements with the Mark III stellar interferometer,” *Appl. Opt.* **26**, 4113–4122 (1987).
- W.C. Danchi, A. Arthur, R. Fulton, M. Peck, B. Sadoulet, E.C. Sutton, C.H. Townes, and R.H. Weitzmann, “A high precision telescope pointing system,” in *Advanced Technology Optical Telescopes III*, L.D. Barr, ed., Proc. SPIE **628**, 422–428 (1986).
- N.M. Elias II, “Baseline metrology system of the U.S. Naval Observatory Astrometric Interferometer,” in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., Proc. SPIE **2200**, 71–75 (1994).

- B.E. Hines, M. Shao, M.M. Colavita, and D.H. Staelin, "Use of laser metrology optical truss for monitoring baseline motion," in *Amplitude and Intensity Spatial Interferometry*, J.B. Breckinridge, ed., Proc. SPIE **1237**, 87–100 (1990).
- C.A. Hummel, D. Mozurkewich, N.M. Elias II, A. Quirrenbach, D.F. Buscher, J.T. Armstrong, K.J. Johnston, R.S. Simon, and D.J. Hutter, "Four years of astrometric measurements with the Mark III optical interferometer," *Astron. J.* **108**, 326–336 (1994).
- D.J. Hutter, "Siderostat/site metrology system for the USNO Astrometric Optical Interferometer," in *High-Resolution Imaging by Interferometry II*, J.M. Beckers and F. Merkle, eds., 1187–1194 (Garching, Germany: European Southern Observatory, 1992).
- D. Mozurkewich, "A hybrid design for a six way beam combiner," in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., Proc. SPIE **2200**, 76–80 (1994).
- N.M. White, D. Mozurkewich, J.T. Armstrong, D.J. Hutter, K.J. Johnston, J.H. Clark III, and L. Ha, "Infrastructure and layout of the Navy Prototype Optical Interferometer at the Lowell Observatory," in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 471–477 (1998).

Chapter 11

Ground-Based Narrow-Angle Astrometry

M. MARK COLAVITA

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

11.1 Introduction

While atmospheric turbulence imposes severe limitations on the accuracy of wide-angle astrometric measurements performed from the ground, the limitations are far less severe for differential measurements over small fields. Performance in this regime is useful for a number of problems, including the search for extrasolar planets. As an indirect technique, astrometry measures the transverse reflex motion of the parent star for evidence of an unseen companion, analogous to radial-velocity measurements, which sense the velocity of the longitudinal reflex motion.

The astrometric signature of a Jupiter-Sun system, seen from a distance of 10 pc, has an amplitude of 500 μas (1 mas peak-to-peak), and establishes an upper limit on the accuracy of astrometric techniques to perform a useful search around nearby stars; single-measurement accuracies of $< 100 \mu\text{as}$ are needed to search for lower-mass planets as well as to provide high-confidence detections. The amplitude of the astrometric signature can be written

$$\theta = \frac{m}{M} \frac{r}{L}, \quad (11.1)$$

where m and M are the planet and star masses, r is the orbital radius, and L is the distance of the system from the Earth. The signature can also be written

$$\theta = \frac{m}{M^{2/3}} \frac{P^{2/3}}{L}, \quad (11.2)$$

where P is the period of the system. Thus astrometry is most sensitive to planets with large orbital radii and long periods, and is complementary to the search space for radial-velocity measurements (Marcy and Butler, 1998).

11.2 Atmospheric Effects

The turbulent atmosphere introduces well-known spatial, temporal, and angular coherence losses parameterized by the coherence diameter r_0 , coherence time τ_0 , and the isoplanatic angle θ_0 . For astrometry, we are most interested in how the astrometric error integrates down with time, which cannot be derived simply from the isoplanatic angle. It is intuitive that the astrometric error for a differential measurement should decrease with decreasing field as the atmosphere becomes common mode. The “sweet spot” for such measurement occurs with a long-baseline interferometer when the star separation is made smaller than the isokinetic angle B/h , where B is the interferometer baseline and h is an effective atmospheric height. In this regime, the error behavior is given by (Shao and Colavita, 1992)

$$\sigma_{\delta\theta} = 300B^{-2/3}\theta t^{-1/2} \text{ arcsec}, \quad (11.3)$$

where we adopt a particular Mauna Kea atmospheric model. In this equation the error is given for integration time t in seconds and star separation θ in radians. This result assumes a strict infinite-outer-scale Kolmogorov atmosphere. Expected deviations from this behavior generally produce better performance. Thus, for a 20-arcsec star separation and a 100-m interferometer, the atmospheric error in one hour of integration time should be less than about 20 μas .

Much more detail on narrow-angle interferometric astrometry is presented by Shao and Colavita (1992).

11.3 Other Errors

For astrometry, an optical interferometer can be looked at geometrically; the problem is identical to the case of a radio interferometer (see Thompson *et al.* 1986). The delay x measured with the interferometer can be related to the interferometer baseline B and the star unit vector s as $x = B \cdot s$. Thus, measurements of delay in conjunction with knowledge of the baseline gives the angle of the star with respect to the baseline vector. The measured delay can be written

$$x = l + k^{-1}\phi, \quad (11.4)$$

where l is the laser-monitored internal delay, ϕ is the fringe phase, and k is the wavenumber of the interfering light.

We can capture most aspects of the measurement problem by reducing it to two dimensions and doing a small-angle approximation for sources near normal to the instrument, viz. $\theta \simeq x/B$. A trivial sensitivity analysis yields the error in the astrometric measurement to be

$$\delta\theta = \frac{\delta l}{B} + k^{-1} \frac{\delta\phi}{B} - \frac{\delta B}{B} \theta. \quad (11.5)$$

The first term incorporates systematic errors in measuring the internal delay; the second term incorporates errors in measuring the fringe phase, including photon and detector noise; the third term incorporates errors in measurement or knowledge of the interferometer baseline.

The long baselines achievable on the ground help reduce the requirements on systematic error control, which are challenging, but within the state of practice. For example, with a 100-m baseline, 10- μ as systematic accuracy requires a 5-nm total length error. With differential measurements, certain systematic errors become common mode and do not affect accuracy. In addition, the astrometric measurement can be performed in a switching mode, reducing requirements on long-term thermal stability.

The dependence on θ in the third term of Equation 11.5 illustrates the difference in the requirements on baseline knowledge between wide- and narrow-angle astrometry. For wide-angle astrometry, $\theta \simeq 1$, leading to the intuitive result that the required fractional accuracy on the baseline is equal to the desired astrometric accuracy. However, for small fields, the requirement on baseline accuracy decreases: essentially, the baseline becomes more common mode to the differential measurement. For example, for a narrow-angle field of 20 arcsec, the requirements on the baseline are reduced by a factor of 10^4 compared with a wide-angle measurement.

The ability to measure the fringe phase places a limit on the achievable accuracy in a given integration time. The error $\delta\phi$ in a phase measurement can be written in terms of the signal-to-noise ratio SNR_ϕ ,

$$\delta\phi = (\text{SNR}_\phi)^{-1}, \quad (11.6)$$

where

$$\text{SNR}_\phi^2 \simeq \frac{1}{2} \frac{N^2 V^2}{N + B + M\sigma^2}, \quad (11.7)$$

where N is the total photon count, B is the total background and dark count, σ^2 is the read-noise variance, and M is the number of reads needed to make the phase measurement. The detection error shows up in the error expression, Equation 11.5, reduced by the baseline. Thus, long baselines help by reducing astrometric error for a given source brightness, or by improving sensitivity for a given accuracy.

11.4 Implementing a Narrow-Angle Measurement

Exploiting the tens-of-microarcsec astrometric accuracy possible with a ground-based narrow-angle astrometric measurement requires the ability to utilize nearby reference stars. One

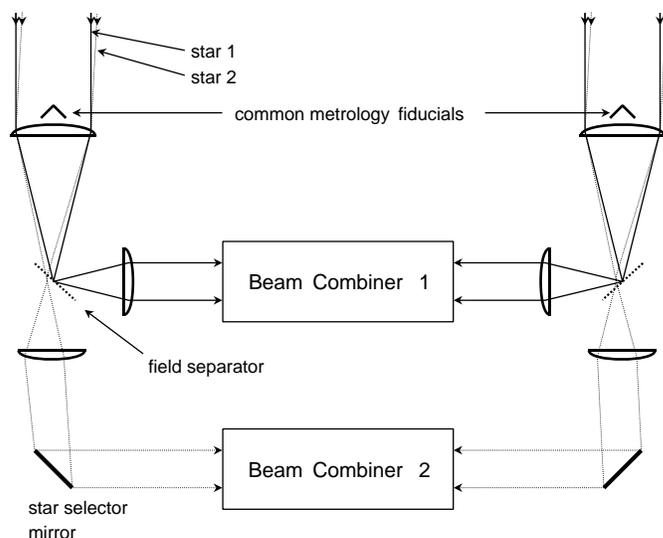


Figure 11.1: Dual-star architecture.

approach to this problem uses a dual-star architecture (Shao and Colavita, 1992), as shown in Figure 11.1. It consists of a long-baseline interferometer with dual beam trains. The light at each aperture forms an image of the field containing the target star and the astrometric reference. A dual-star feed separates the light from the two stars into separate beams which feed separate interferometer beam combiners. These beam combiners are referenced with laser metrology to a common fiducial at each collector. The two beam combiners make simultaneous measurements of the delays for the two stars.

Over a small field, reference stars will invariably be faint, and ordinarily would not be usable by the interferometer. However, searching for exoplanets is a unique problem in that the target star is nearby, and hence bright, and can serve as a phase reference. With phase referencing, the bright target star is used as a probe of the atmospheric turbulence within the isoplanatic patch of the target star. By compensating for the fringe motion of the target star with an optical delay line, the fringe motion of the faint astrometric reference star is frozen, allowing for long integration times which greatly increase sensitivity.

The radius of the isoplanatic patch increases with wavelength, and is 20–30 arcsec at $2.2 \mu\text{m}$. With phase referencing and 1.5–2.0-m telescopes, astrometric references can be detected around most potential planetary targets.

Conducting a narrow-angle measurement with an architecture like that of Figure 11.1 involves two steps. The first step is wide-angle astrometry using known reference stars to solve for the interferometer baseline. As discussed above, the required baseline precision for a narrow-angle measurement is much less than for a wide-angle measurement, and the accuracies available from these wide-angle measurements provide sufficient accuracy. There are some subtleties regarding the wide-angle baseline as thus solved and the narrow-angle

baseline applicable to the science measurement, and an auxiliary system may be required to tie these two baselines together.

The second step is to implement the measurement through chopping. In this approach, one interferometer beam combiner always tracks the target star. The other beam combiner switches repeatedly between the target star and the reference star. This “chopping” approach requires instrument stability only over the chop cycle. The use of even a low-resolution spectrometer in the fringe detector makes the ground-based measurements relatively insensitive to differential chromatic refraction.

In general, measurements on two orthogonal baselines are needed to detect systems with arbitrary inclinations. Measurements with respect to two reference stars are also desirable; with redundant measurements, astrometric noise in a reference star is, in most cases, separable from the desired (planetary) signature.

The Palomar Testbed Interferometer (Colavita *et al.*, 1999) was designed to demonstrate most aspects of narrow-angle astrometry for application to one of the key science modes of the Keck Interferometer (Colavita *et al.*, 1998; van Belle *et al.*, 1998). Recent results from PTI demonstrate a night-to-night repeatability of 100 μas on a bright visual binary (Boden *et al.*, 2000).

11.5 Conclusion

Long-baseline interferometers can exploit the behavior of the atmosphere over a small field to conduct high-accuracy measurements for applications such as exoplanet detection. The particular nature of this problem, i.e., that the target is bright and serves as a phase reference, allows cophasing the interferometer to obtain high sensitivity within the isoplanatic patch. While astrometry at the full accuracy allowed by the atmosphere is challenging, the long baselines achievable on the ground help moderate the effects of fringe-detection noise and systematic errors attributable to metrology and baseline knowledge.

Acknowledgments

This work was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

- A.F. Boden et al. “Long-baseline differential astrometry of the 61 Cygni binary system,” *Astrophys. J.*, in preparation (2000).
- G.T. van Belle, A.F. Boden, M.M. Colavita, M. Shao, G. Vasisht, and J.K. Wallace, “Astrometry with the Keck interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 362–372 (1998).
- M.M. Colavita, A.F. Boden, S.L. Crawford, A.B. Meinel, M. Shao, P.N. Swanson, G.T. van Belle, G. Vasisht, J.M. Walker, J.K. Wallace, and P.L. Wizinowich, “The Keck interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 776–784 (1998).
- M.M. Colavita, J.K. Wallace, B.E. Hines, Y. Gursel, F. Malbet, D.L. Palmer, X.P. Pan, M. Shao, J.W. Yu, A.F. Boden, P.J. Dumont, J. Gubler, C.D. Koresko, S.R. Kulkarni, B.F. Lane, D.W. Mobley, and G.T. van Belle, “The Palomar Testbed Interferometer,” *Astrophys. J.* **510**, 505–521 (1999).
- G.W. Marcy and R.P. Butler, “Detection of extrasolar giant planets,” *Ann. Rev. Astron. Astrophys.* **36**, 57–98 (1998).
- M. Shao and M.M. Colavita, “Potential of long-baseline infrared interferometry for narrow angle astrometry,” *Astron. Astrophys.* **262**, 353–358 (1992).
- A.R. Thompson, J.M. Moran, and G.W. Swenson Jr., *Interferometry and Synthesis in Radio Astronomy* (New York: Wiley, 1986).

Part IV

Optical Synthesis Imaging

Chapter 12

Interferometry with Two Telescopes

H. MELVIN DYCK

U.S. NAVAL OBSERVATORY
FLAGSTAFF, ARIZONA

12.1 Introduction

As was discussed by Boden in Chapter 2, faithfully reconstructing complex astronomical images requires the use of many interferometric baselines and closure phases. Owing to phase smearing in the atmosphere, fewer than three telescopes are not capable of recovering phase information in a general way (cf. Chapter 13). However, two telescopes can recover the amplitude of the complex visibility function (Born and Wolf, 1999) and are useful for investigating a limited range of scientific problems. These problems belong to a class for which the structure is known *a priori* and only a few global parameters need to be determined from observations. As we increase the number of parameters to be extracted from models fitted to the observed visibility amplitude, we also increase the risk of misinterpreting the nature of the source. In this chapter, we concentrate on the more limited class of sources where one or two global parameters suffice to yield interesting astrophysical information. We shall also try to stress limitations to the interpretation that may arise from expected astrophysical violations to the *a priori* knowledge of the source.

12.2 Simple Concepts of Aperture Synthesis

A single-mirror telescope is able to construct accurate images because it provides a continuum of interferometric baselines ranging in size from zero up to the diameter of the mirror. The resolution of such an instrument, called the point-spread function (Schroeder,

2000), may be determined by diffraction or by aberrations either intrinsic to the mirror or introduced by an unstable atmosphere. Let's consider this in more detail. In Figure 12.1(a), we show a solid circle meant to represent the edge of a mirror of arbitrary diameter D . We refer to this part of Figure 12.1 as the “aperture plane.” Let's suppose that the mirror is placed on the ground with directions to north and east as shown. If we consider any two infinitesimal sub-apertures in the mirror, represented by the small gray dots in the figure, then we may envision a single baseline, B , between them. We know from the theory of interferometry (see Chapter 2) that we may sample an image with a resolution of approximately λ/B (Schroeder, 2000) where λ is the wavelength of the observation. Inspection of the figure shows us that we may resolve all angular scales $\theta \geq \lambda/D$ for all possible pairs of elementary sub-apertures in the mirror. This is the fundamental principle of aperture synthesis. We have synthesized the full aperture D by taking all possible pairs of elementary sub-apertures in the mirror. In this case we are dealing with a filled aperture system.

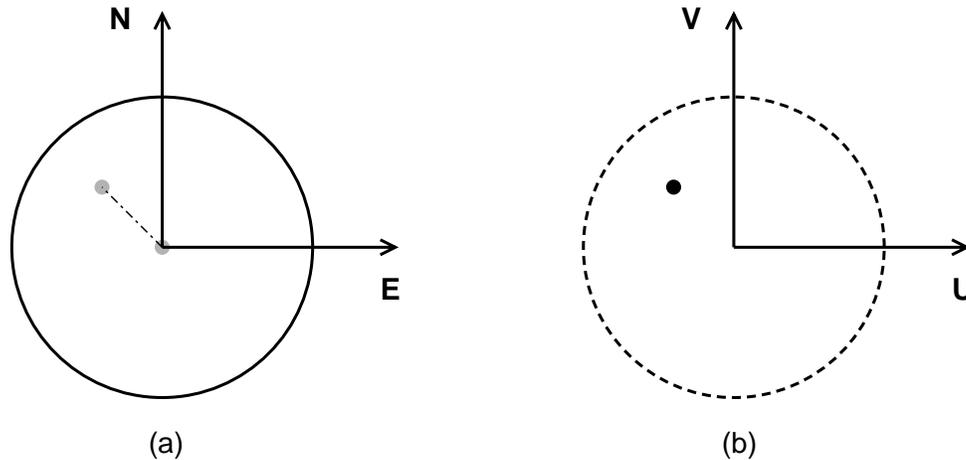


Figure 12.1: The relationship between baseline in the aperture plane and spatial frequency in the Fourier plane.

There is another plane of interest in imaging, namely the Fourier plane or (u, v) plane, shown in Figure 12.1(b). The coordinate complement to spatial resolution in the aperture plane (λ/B) is spatial frequency, B/λ , in the Fourier plane. Note that a single baseline in the aperture plane maps into a single point in the Fourier plane. All possible baselines filling the aperture plane will map into all possible points in the Fourier plane. The filled aperture system will completely sample the Fourier plane from zero spatial frequency out to a maximum spatial frequency D/λ , equal to the radius of the dashed circle in the figure. The Fourier coordinate axes (u, v) correspond to the spatial axes E, N in the aperture plane. Note that the Fourier representation of the point-spread function is the optical transfer function while its modulus is the modulation transfer function (Schroeder, 2000).

Now suppose the solid circle in the figure is composed of a few discrete, separated sub-apertures of finite size. This is an example of an unfilled or sparse aperture system. We

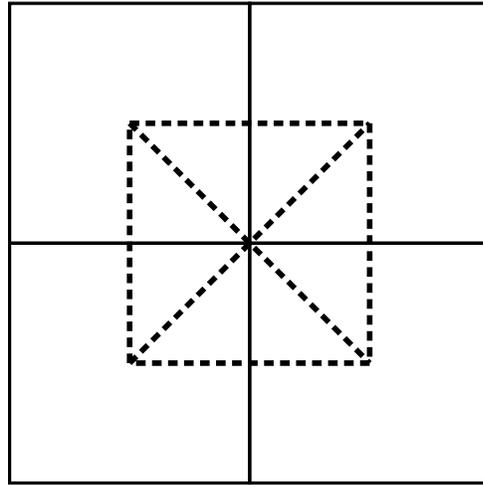


Figure 12.2: All possible baselines for a 2×2 telescope array.

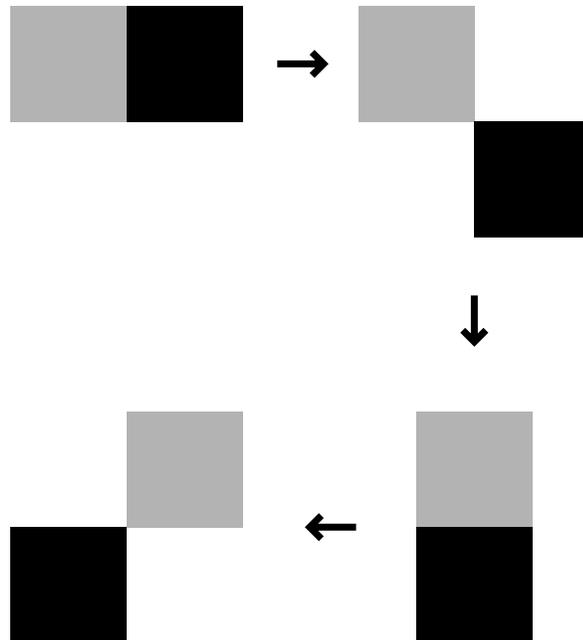


Figure 12.3: Synthesizing a 2×2 telescope by re-positioning 2 elements.

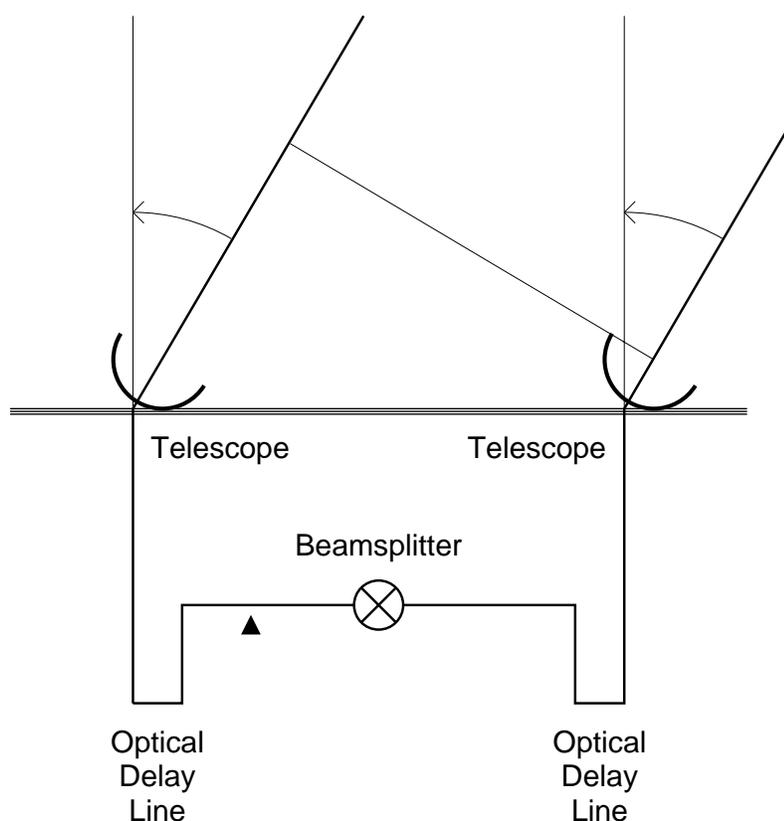


Figure 12.4: A simple schematic picture of a two-telescope interferometer, showing the optical delay lines. With the delay lines balanced, as shown in the figure, the position of the fringe occurs at the filled triangle for the source observed to the right of the zenith.

may draw baselines between all possible pairs. For a snapshot observation (i.e. infinitesimal observation time compared to the rotation of the Earth), there will be a finite number of baselines in the aperture plane, corresponding to a finite number of spatial frequencies in the Fourier plane. As an example, we have shown in Figure 12.2 an interferometer formed from four contiguous square apertures, arranged in a 2×2 array. The array is shown as the solid squares in the figure. Also shown as dashed lines are the six possible baselines obtainable from the array. The point-spread and modulation-transfer functions for this array will have only discrete values. Such sparse aperture functions contribute to the difficulty of interpretation of complex images. Note also that of the six baselines shown in Figure 12.2, two are redundant, giving the same image information. The redundant pairs are the horizontal and vertical pairs in the figure.

We may synthesize the 2×2 interferometric array from only a single pair of telescopes, moving one or both telescopes to achieve all possible baselines. This concept is shown in Figure 12.3 in a sequence of four different configurations. So, we have illustrated how we may synthesize a larger aperture from a number of sub-apertures by physically moving the

sub-apertures. These concepts apply to the snapshot mode of observation. Of course the actual resolution of any interferometric pair is not the physical separation between the pairs but the projected separation of the pair as viewed from the source. This is illustrated in Figure 12.4, where we show the basic elements of a two-telescope interferometer including the optical delay lines and the optical correlator (labeled the beam splitter) where the interference occurs. In this figure the source is supposed to be to the right of the direction to the zenith. In order to obtain interference at the beam splitter the optical delay line for the telescope on the right would need to be adjusted to add in a path length equal to the extra distance traveled by the light rays coming to the telescope on the left. As the Earth rotates, the amount of delay changes from a maximum at the horizon to zero at the zenith.

The spatial resolution on the source is the projected distance between the two apertures as seen from the source. Clearly, this resolution will change as the telescopes track the source, ranging from zero when the source is on the horizon to some maximum value (equal to the physical separation in the figure) when the source is at the zenith. Thus, we may also synthesize an aperture by allowing the Earth to rotate for a fixed physical configuration of the array. This has been referred to as Earth rotation aperture synthesis.

12.3 Optical Delay and (u, v) Plane Tracks

For any given array the aperture synthesis process involves a combination of physical movement of the telescopes in the array and Earth's rotation. An interferometer with many telescopes will sweep out much of the (u, v) plane corresponding to all the possible baselines. A two-telescope interferometer has much sparser coverage in the (u, v) plane. A nice discussion of the equations for the u and v coordinates for a pair of telescopes is developed in the paper by Fomalont and Wright (1974), for example. We follow their formalism in this chapter. In Figure 12.5 we sketch out the coordinate frame for the development. Let the origin of the coordinate system be centered on one of the mirrors of the pair. One axis points north, one points east, and one points to the zenith; these axes are shown as solid lines and are labeled N, E, and L in the figure. We may draw a vector \mathbf{B} , shown as a dashed line in the figure, from one mirror to the other that has components B_E , B_N , and B_L . Let b = mean latitude of the interferometer (and neglecting curvature of the Earth), δ = declination of the source, and h = hour angle of the source. The (u, v) coordinates corresponding to a snapshot projection of the baseline, in spatial frequency units (cycles/arcsec), are

$$u = (B_E \cos h - B_N \sin b \sin h + B_L \cos b \sin h) / 206265 \lambda, \quad (12.1)$$

and

$$v = (B_E \sin \delta \sin h + B_N(\sin b \sin \delta \cos h + \cos b \cos \delta) \\ - B_L(\cos b \sin \delta \cos h - \sin b \cos \delta)) / 206265 \lambda.$$

We have computed the (u, v) plane tracks for a simple two-telescope interferometer situated at a latitude $b = 33^\circ$, having baseline components $B_E = B_N = 100$ m and operated at

$\lambda = 2.2 \mu\text{m}$. Tracks are shown for sources at four different declinations assuming the interferometer tracks the source over the range of hour angles $-4 \leq h \leq +4$. These tracks are shown plotted in Figure 12.6 which illustrates how little of the (u, v) plane is filled by this simple interferometer. One could laboriously move the telescopes and repeat the observations but this is not so important for the simple sources we will consider in this chapter. We may also see the dependence of the tracks upon source position. Note that not only does the projected separation change but also the orientation of the baseline with respect to the source changes as the Earth rotates.

To finish this section we give below the relation for the optical delay (in the same units as the baseline vector components) in terms of telescope and source parameters.

$$\Delta = - B_E \cos \delta \sin h \quad (12.2)$$

$$- B_N (\sin b \cos \delta \cos h - \cos b \sin \delta) \quad (12.3)$$

$$+ B_L (\cos b \cos \delta \cos h + \sin b \sin \delta). \quad (12.4)$$

12.4 Simple Models of Astrophysical Sources

In this section we apply the principles discussed in the previous section but limit them to a two-telescope interferometer. A simplification adopted for the models is that the visibilities are observed along a single, arbitrary, radial cut through the source. For this case we will plot visibility amplitude $V(s)$ as a function of spatial frequency s .

One may think of several astrophysical systems where observations at a few baseline separations and azimuths, projected on the source, will serve to characterize the system. Some of these simple systems and the relevant astrophysical parameters are listed below:

- Binary stars (one may obtain stellar masses)
- Single-star angular diameters (one may obtain linear radii and effective temperatures)
- Limb-darkened diameters (one may obtain the atmospheric temperature structure)
- Circumstellar shells (one may characterize mass-loss phenomena)
- Departures from circular symmetry in stars (one may detect non-radial pulsation and rapid rotation).

Of course, it is understood that other pieces of information are needed to derive the parameters mentioned above. All of the above systems present relatively simple visibility curves that may be adequately sampled by a two-telescope interferometer.

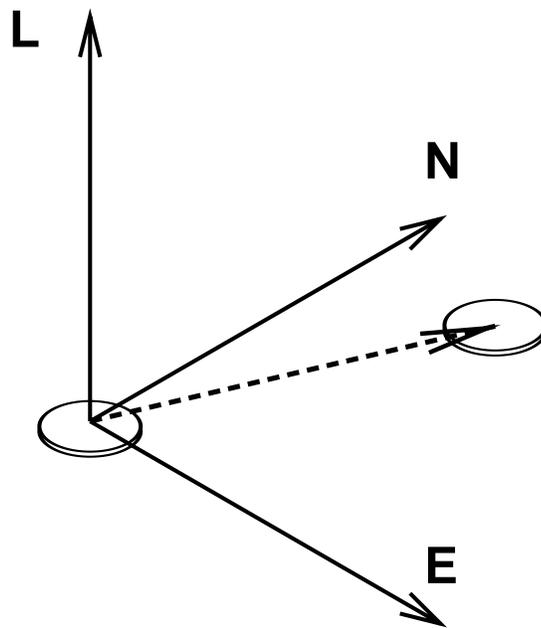


Figure 12.5: The three-axis coordinate system for a two-telescope interferometer.

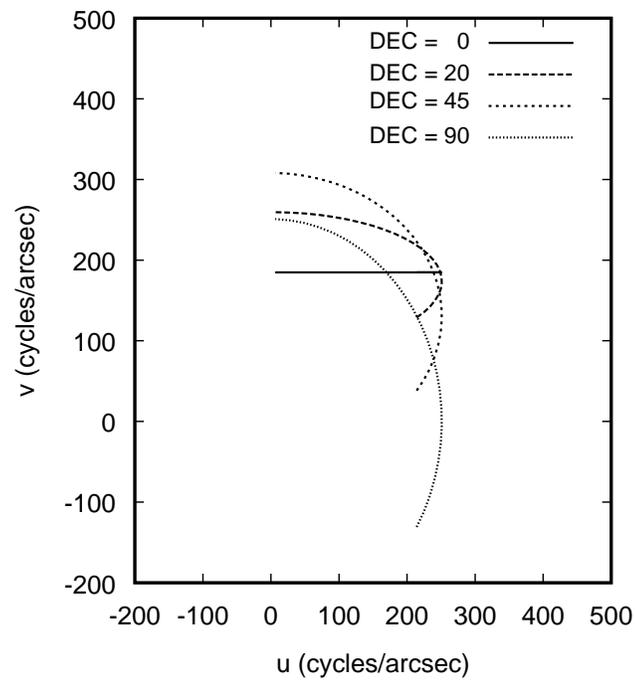


Figure 12.6: (u, v) tracks for a two-telescope interferometer for sources at different declinations and observed over a range of hour angles 4 hours either side of the meridian.

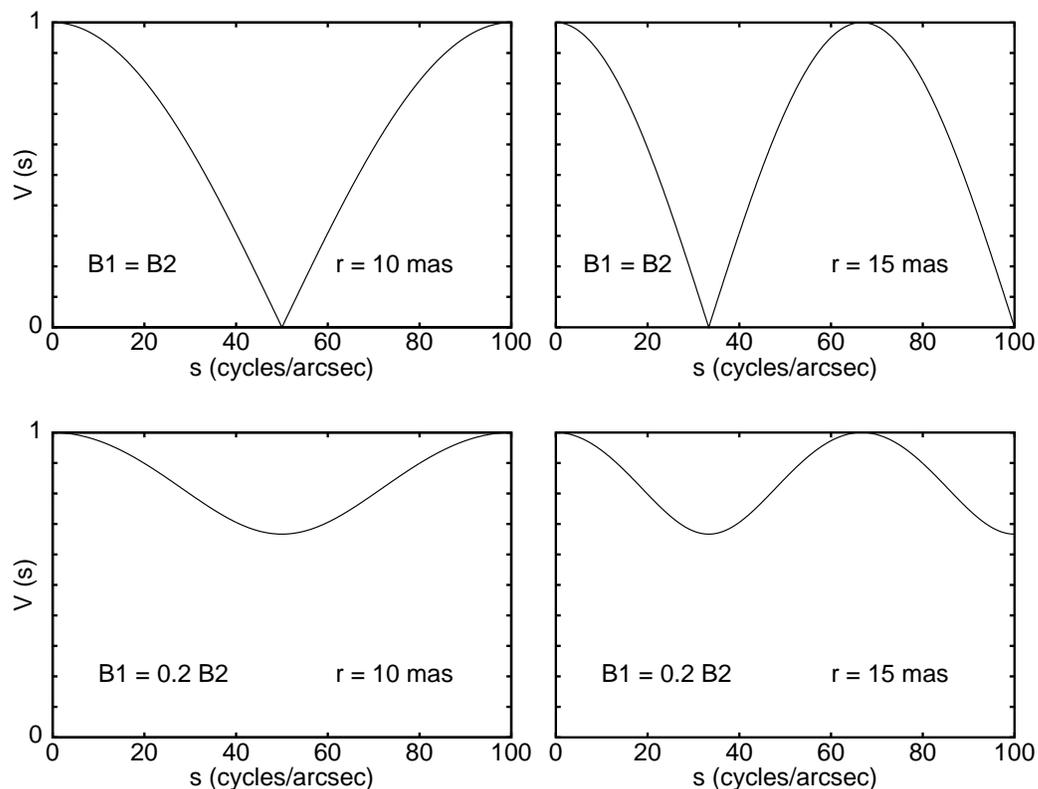


Figure 12.7: Double-star visibility curves. Note how the position and depth of the minima change with separation and brightness ratio, respectively.

12.4.1 Binary Stars

Let's begin with a binary star system. We suppose that both stars are unresolved and define $B_1 =$ brightness of source 1, $B_2 =$ brightness of source 2 and $r =$ separation between sources (measured in arcsec). The visibility amplitude may be shown to be

$$V(s) = [P_0 + (1 - P_0) \cos^2(\pi s r)]^{1/2}, \quad (12.5)$$

where the parameter P_0 is defined as

$$P_0 = \left(\frac{B_1 - B_2}{B_1 + B_2} \right)^2. \quad (12.6)$$

In Figure 12.7 we have plotted some examples of two unresolved points, where we have varied the spacing and brightness ratio for the two components. One may see that the closer together the stars become, the more cycles occur in the visibility amplitude function. Also, when the stars are equally bright, the first minimum of the visibility amplitude goes to zero. Note that the first minimum becomes shallower as the brightness ratio changes

from unity. The properties of the first minimum may be used to determine the separation and brightness ratio for the stars. If we let V_{\min} = amplitude of the first minimum in the visibility function and s_{\min} = spatial frequency corresponding to that first minimum then we find that

$$r = \frac{1}{2s_{\min}}, \quad (12.7)$$

and

$$\frac{B_1}{B_2} = \frac{1 + V_{\min}}{1 - V_{\min}}. \quad (12.8)$$

Note that the separation obtained in this way will be the component of the binary separation in one direction, the direction given by the interferometer baseline. To determine the true separation, this measurement must be repeated in at least one other direction. Normally, by observing a binary system for an extended period during the night, one would obtain snapshot values of $V(s)$ for a variety of projected telescope separations and orientations with respect to the source. These observations may be used to solve for the binary source parameters. Obvious possible problems with such observations would arise if the source had additional components or if one or both stars were resolved.

Fruitful binary star investigations have been carried out at the Mark III (Hummel *et al.*, 1995), IOTA (Dyck *et al.*, 1995), PTI (Boden *et al.*, 1999) and NPOI (Hummel *et al.*, 1998) among others.

12.4.2 Stellar Angular Diameters

Next we consider single-star angular diameters and suppose that the stars appear as simple, circular, uniformly-bright disks projected on the sky. If a = diameter of the source (in arcsec) then the visibility function may be shown to be

$$V(s) = \left| \frac{2J_1(\pi as)}{\pi as} \right|, \quad (12.9)$$

where $J_1(\pi as)$ is the first-order Bessel function of the first kind. In Figure 12.8 we have plotted visibility functions for sources of three different diameters. Note that successive secondary maxima decrease in amplitude with increasing spatial frequency. The frequency of the first zero of the function, s_0 , scales inversely with the angular diameter of the source such that $a \simeq 1.22/s_0$. In principle, if we know that the source really is a uniformly-bright circular disk, the angular diameter may be determined from the measurement of the visibility amplitude at a single spatial frequency. For example, when $V(s) \geq 0.13$, approximately, then we know that the measurement lies at spatial frequencies smaller than s_0 . A single measurement may be used to obtain the diameter uniquely.

Interferometric investigations of single star diameters have been carried out by Michelson and Pease (1921), Di Benedetto and Rabbia (1987), Mozurkewich *et al.* (1991), Dyck *et al.* (1998), van Belle *et al.* (1999), and Nordgren *et al.* (1999).

Obvious difficulties with the interpretation will arise if there are spots on the surface or in the presence of limb-darkening. Surface structure, in general, will be too complex to be treated adequately by observations with two telescopes. Limb darkening may be assumed to be circularly symmetric and primarily affects the part of the visibility curve at frequencies $s \geq s_0$, where it depresses the visibility curve below the value derived from uniformly-bright circular disks. At the present time model atmosphere calculations exist that allow one to estimate more realistic center-to-limb brightness variations than can be obtained from the assumption of a uniformly bright circular disk. One should consult, for example, recent work by Claret (1998) for stars hotter than about 4000 K and Hofmann and Scholz (1998) for cooler, more extended stars. These papers will lead the reader to other investigations. A convenient analytic expression for the center to limb brightness variations was developed by Michelson and Pease (1921) and suggested by Hestroffer (1997) as a useful approximation to the models. In image space this has the form

$$\frac{I(\mu)}{I(0)} = \mu^\alpha, \quad (12.10)$$

where $I(\mu)$ = disk brightness at angle μ , $\mu = \cos \theta$, and α = an exponent best describing the model atmosphere. The angle θ is the angle between the normal to the stellar surface and the direction to the observer, at the point observed on the star. Hestroffer (1997) has obtained an analytic form for the visibility from this center to limb brightness function:

$$V(s) = \Gamma(n + 1) \frac{|J_n(\pi as)|}{(\pi as/2)^n}, \quad (12.11)$$

where $n = (\alpha + 2)/2$.

Direct determination of limb darkening has been carried out by Hajian *et al.* (1998).

12.4.3 Circumstellar Shells

Circumstellar shell structures may be detected if they are optically thin, for example, as superpositions on the underlying star. Suppose we modeled such a system by a coaxial uniform disk and point source. Here, the uniform disk represents the shell while the point source represents the central star. Let a = diameter of the shell and V_P = ratio of power radiated by the star to total power radiated by the system. The visibility function for this combination may be shown to be

$$V(s) = V_P + (1 - V_P) \left[\frac{2 J_1(\pi as)}{\pi as} \right]. \quad (12.12)$$

An example of this kind of visibility function has been shown in Figure 12.9.

Observations of circumstellar shells superposed on central stars have been made with the ISI interferometer. Reports of these observations may be found, for example, in Danchi *et al.* (1994). One may see examples of composite visibility functions such as the one described above in this paper.

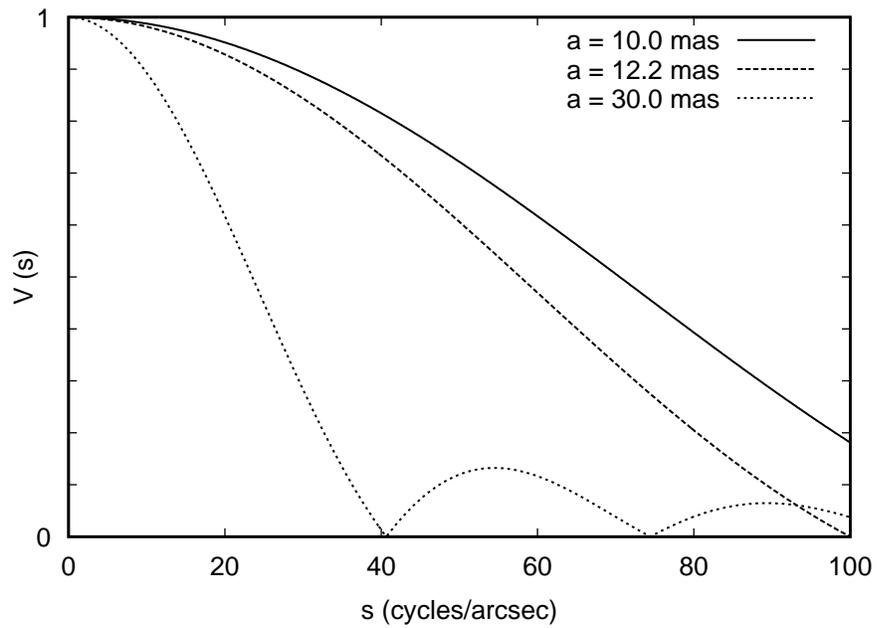


Figure 12.8: Visibility curves for uniformly-bright circular disks. Note the decreasing amplitude of successive maxima beyond the first null.

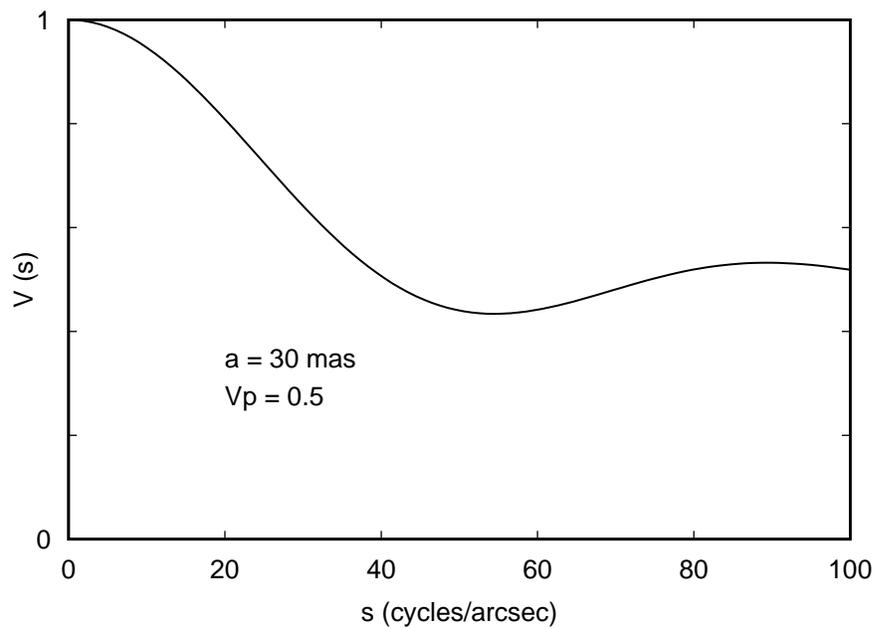


Figure 12.9: The visibility curve for a uniform-disk surrounding a point. The disk size is 30 mas and there is 50% in each of the two components.

We finish this section with a brief mention of departures from circular symmetry. For simple stars, one may detect this departure by observing at more than one angle projected on the star. Fitting a uniform disk or limb darkened disk visibility function to the observations will show a different scale size in different directions.

12.5 Cautions and Caveats

The most important limitation imposed upon the choice of model is set by the resolution. It is stating the obvious that very poor angular resolution may render a source unresolved. As the resolution is improved there will be a regime where the source is seen to be different from a point but for which there will be nearly complete uncertainty about its nature. It is an interesting exercise to take the model formulations presented in the previous section, choose variable size scales for each model and see how closely the visibilities can be made to agree at the very lowest spatial frequencies.

Apart from these sampling considerations, the choice of model for any set of observations will depend upon the signal-to-noise ratio of the observations, the confidence in the level of systematic errors in the system, the number and orientation of the baselines used and the expected complexity of the source brightness distribution.

12.6 Comparisons

There are now enough interferometers operating that we may reasonably ask about the quality of the measurements. There are two easy ways to do this, in the absence of ground truth observations. First, one may inter-compare interferometers measuring the same objects. Second, one may compare the measurements of interferometers with the most current predictions of models. Neither of these methods defines the absolute accuracy of interferometric observations.

Let's first compare measurements of the same objects made with different interferometers. In Figure 12.10 the fractional difference between 2.2 μm measurements made at IOTA (Dyck *et al.*, 1998) and those made at CERGA (Di Benedetto and Rabbia, 1987) or FLUOR (Perrin 1996; Perrin *et al.* 1998a,b) is plotted versus the IOTA diameter. Here, diameters obtained from uniform disk fits to the visibility data are used, with no corrections made for the effects of limb darkening. The important points of the figure are that (1) there is no large systematic difference among the measurements and (2) the scatter is larger for the smaller diameter stars. Quantitatively, the mean systematic difference between the IOTA measurements and the other measurements is $(\text{Other IR} - \text{IOTA})/\text{IOTA} = 0.017 \pm 0.107$ (rms). That is, the IOTA uniform disk diameters are less than 2% smaller than the other diameters, with a typical error of about 11% for a single measurement. Most of the error in the measurements is attributable to the measurement error of the IOTA diameters.

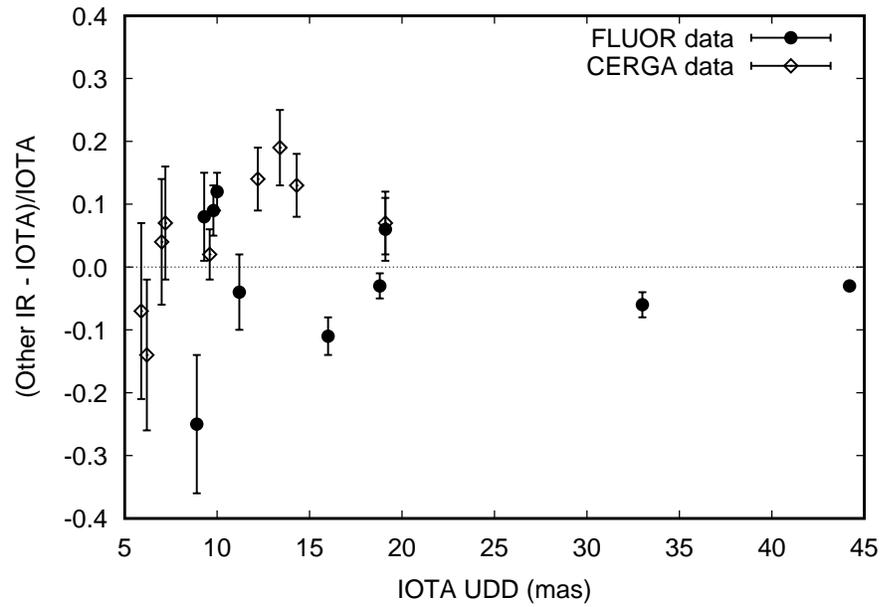


Figure 12.10: A comparison of 2.2- μm uniform disk diameters measured with IOTA to those measured with CERGA and FLUOR.

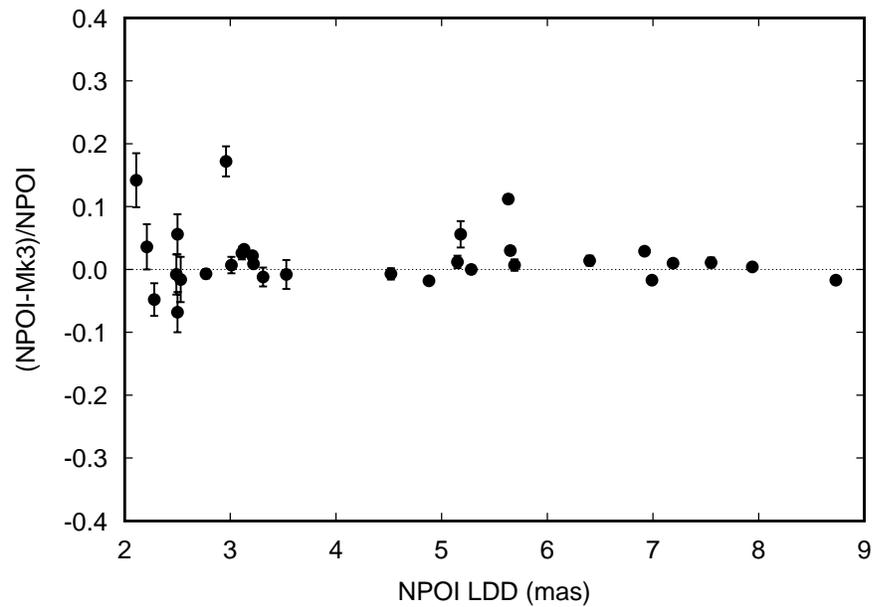


Figure 12.11: A comparison of limb-darkened angular diameters measured at NPOI and the Mark III at red wavelengths.

In Figure 12.11, the same comparison is made between the NPOI and the Mark III for stars measured in common at wavelengths near 800 nm. These results have been tabulated by Nordgren (1999). Here, the comparison is between limb-darkened angular diameters, since the observations at the two interferometers were made at somewhat different wavelengths. Note again the relative increase in scatter for the smaller diameter stars. The quantitative systematic difference is $(\text{NPOI} - \text{Mark III})/\text{NPOI} = 0.018 \pm 0.052$. That is, the NPOI diameters are less than 2% larger than the Mark III diameters with a typical error of the diameter of about 5%.

Thus, optical and infrared interferometers working at the same or nearly the same wavelengths are capable of reproducing each others results to levels better than about 2%.

Next, we may compare interferometric observations to current model predictions. Blackwell *et al.* (1990) and Blackwell and Lynas-Gray (1994), using the infrared flux method (IRFM), have computed and tabulated diameters for stars in the temperature range 4000–8500 K. Bell and Gustafsson (1989) have also tabulated diameters, but that are not based principally upon infrared observations. The largest overlap between observed and computed diameters exists for the near infrared, so we have used the 2.2- μm observations made at IOTA by Dyck *et al.* (1998), at CERGA by Di Benedetto and Rabbia (1987) and at PTI by van Belle *et al.* (1999). These comparisons are shown plotted in Figure 12.12, where IRFM refers to the diameters from Blackwell and collaborators and BG refers to the diameters from Bell and Gustafsson. In this figure, the comparison is made between the fractional difference in diameter and the intrinsic V–K color of the star. One may see that the general level of the difference is less than about 20% with the exception of 41 Cyg, the very discrepant point in the upper left part of the figure. Excluding this star, the comparisons are $(\text{Observed} - \text{IRFM})/\text{IRFM} = -0.019 \pm 0.061$ and $(\text{Observed} - \text{BG})/\text{BG} = -0.009 \pm 0.071$. That is, the observations and the models agree to the level 1–2%. If the 41 Cyg diameter is left in the plot, there is a suggestion that there may be a difference between theory and observation that depends upon color. This will have to be verified with additional observations.

A final assessment of the capability of interferometers may be obtained from double star observations. In Figure 12.13 there is a plot of the orbital position of the secondary about the primary for Mizar A, taken from Hummel *et al.* (1998). Data are shown from both the Mark III and the NPOI at optical wavelengths. The best fit orbit yields an $(\text{observed} - \text{computed})_{rms} \simeq 60 \mu\text{as}$!

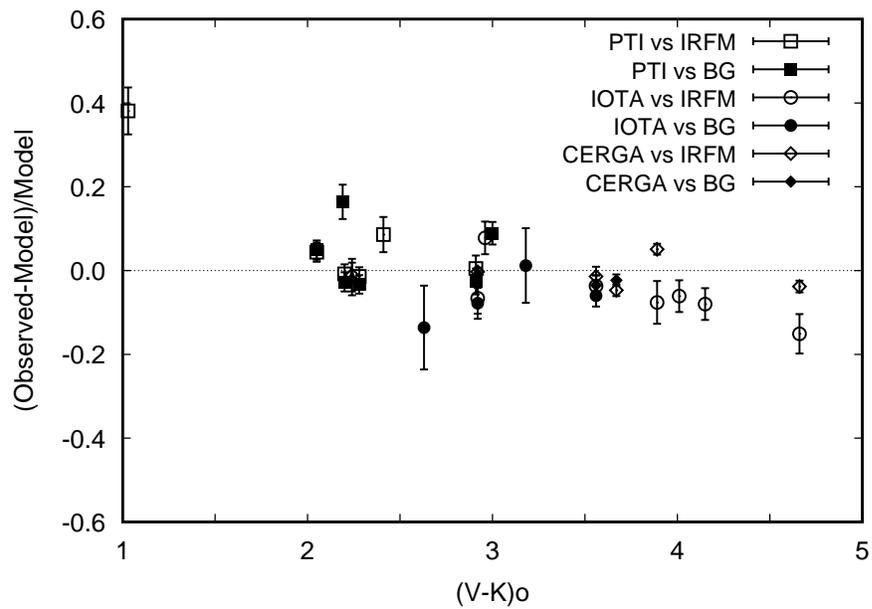


Figure 12.12: A comparison of angular diameters computed from model stellar atmospheres and limb-darkened diameters obtained at $2.2 \mu\text{m}$. The codes are explained in the text.

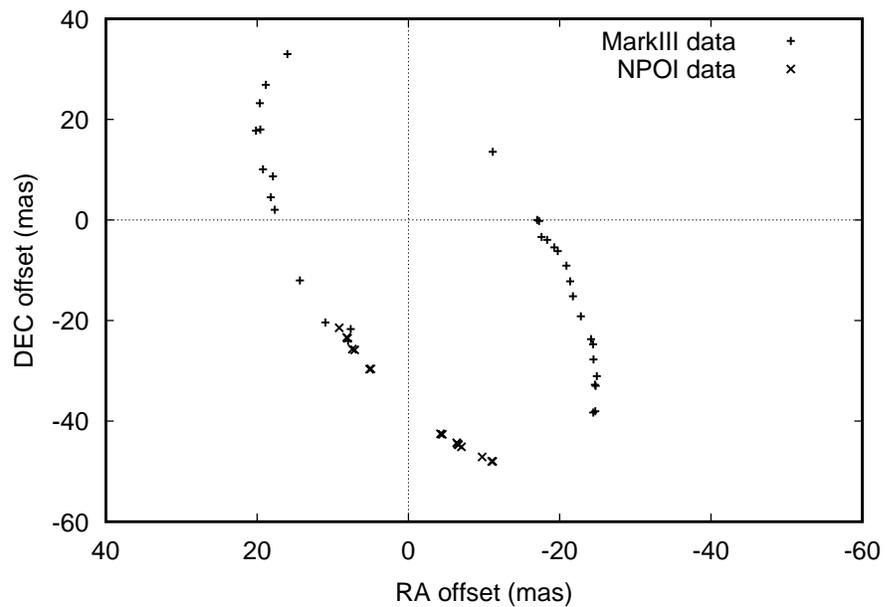


Figure 12.13: Observations of the apparent orbit for the binary star Mizar A.

References

- R.A. Bell and B. Gustafsson, “The effective temperatures and colours of G and K stars,” *Mon. Not. R. Astron. Soc.* **236**, 653–708 (1989).
- G.T. van Belle, B.F. Lane, R.R. Thompson, A.F. Boden, M.M. Colavita, P.J. Dumont, D.W. Mobley, D. Palmer, M. Shao, G.X. Vasisht, J.K. Wallace, M.J. Creech-Eakman, C.D. Koresko, S.R. Kulkarni, X.P. Pan, and J. Gubler, “Radii and effective temperatures for G, K, and M giants and supergiants,” *Astron. J.* **117**, 521–533 (1999).
- G.P. Di Benedetto and Y. Rabbia, “Accurate angular diameters and effective temperatures for eleven giants cooler than K0 by Michelson interferometry,” *Astron. Astrophys.* **188**, 114–124 (1987).
- D.E. Blackwell and A.E. Lynas-Gray, “Stellar effective temperatures and angular diameters determined by the infrared flux method (IRFM): revisions using improved Kurucz LTE atmospheres,” *Astron. Astrophys.* **282**, 899–910 (1994).
- D.E. Blackwell, A.D. Petford, S. Arribas, D.J. Haddock, and M.J. Selby, “Determination of temperatures and angular diameters of 114 F-M stars using the infrared flux method (IRFM),” *Astron. Astrophys.* **232**, 396–410 (1990).
- A.F. Boden, C.D. Koresko, G.T. van Belle, M.M. Colavita, P.J. Dumont, J. Gubler, S.R. Kulkarni, B.F. Lane, D. Mobley, M. Shao, J.K. Wallace, and G.W. Henry, “The visual orbit of ι Pegasi,” *Astrophys. J.* **515**, 356–364 (1999).
- M. Born and E. Wolf, *Principles of Optics*, 7 (expanded) edn. (Cambridge, UK: Cambridge University Press, 1999).
- A. Claret, “Comprehensive tables for the interpretation and modeling of the light curves of eclipsing binaries,” *Astron. Astrophys. Supp. Ser.* **131**, 395–400 (1998).
- W.C. Danchi, M. Bester, C.G. Degiacomi, L.J. Greenhill, and C.H. Townes, “Characteristics of dust shells around 13 late-type stars,” *Astron. J.* **107**, 1469–1513 (1994).
- H.M. Dyck, G.T. van Belle, and R.R. Thompson, “Radii and effective temperatures for K and M giants and supergiants. II,” *Astron. J.* **116**, 981–986 (1998).
- H.M. Dyck, J.A. Benson, and F.P. Schloerb, “Imaging a binary star with a two-telescope Michelson interferometer,” *Astron. J.* **110**, 1433–1439 (1995).
- E.B. Fomalont and M.C.H. Wright, “Interferometry and aperture synthesis,” in *Galactic and Extra-galactic Radio Astronomy*, G.L. Verschuur and K.I. Kellerman, eds., 256–290 (New York: Springer-Verlag, 1974).
- A.R. Hajian, J.T. Armstrong, C.A. Hummel, J.A. Benson, D. Mozurkewich, T.A. Pauls, N.M. Elias II, D.J. Hutter, K.J. Johnston, L.J. Rickard, and N.M. White, “Direct confirmation of stellar limb darkening with the Navy Prototype Optical Interferometer,” *Astrophys. J.* **496**, 484–489 (1998).
- D. Hestroffer, “Centre to limb darkening of stars,” *Astron. Astrophys.* **327**, 199–206 (1997).
- K.-H. Hofmann and M. Scholz, “Limb-darkening and radii of non-Mira M giant models,” *Astron. Astrophys.* **335**, 637–646 (1998).

- C.A. Hummel, J.T. Armstrong, D.F. Buscher, D. Mozurkewich, A. Quirrenbach, and M. Vivekanand, "Orbits of small angular scale binaries resolved with the Mark III interferometer," *Astron. J.* **110**, 376–390 (1995).
- C.A. Hummel, D. Mozurkewich, J.T. Armstrong, A.R. Hajian, N.M. Elias II, and D.J. Hutter, "Navy Prototype Optical Interferometer observations of the double stars Mizar A and Matar," *Astron. J.* **116**, 2536–2548 (1998).
- A.A. Michelson and F.G. Pease, "Measurement of the diameter of α Orionis with the interferometer," *Astrophys. J.* **53**, 249–259 (1921).
- D. Mozurkewich, K.J. Johnston, R.S. Simon, P.F. Bowers, R. Gaume, D.J. Hutter, M.M. Colavita, M. Shao, and X.P. Pan, "Angular diameter measurements of stars," *Astron. J.* **101**, 2207–2219 (1991).
- T.E. Nordgren, "Stellar angular diameters from the NPOI and Mark III: A comparison," *Bull. Am. Astron. Soc.* **194**, 8205 (1999).
- T.E. Nordgren, M.E. Germain, J.A. Benson, D. Mozurkewich, J.J. Sudol, N.M. Elias II, A.R. Hajian, N.M. White, D.J. Hutter, K.J. Johnston, F.S. Gauss, J.T. Armstrong, T.A. Pauls, and L.J. Rickard, "Stellar angular diameters of late-type giants and supergiants measured with the Navy Prototype Optical Interferometer," *Astron. J.* **118**, 3032–3038 (1999).
- G. Perrin, *A fiber recombination unit for the IOTA interferometer: Application to the study of late-type stars*, Ph.D. thesis, Université de Paris VII (1996).
- G. Perrin, V. Coudé du Foresto, S.T. Ridgway, J.-M. Mariotti, N.P. Carleton, and W.T. Traub, "Stellar parameters with high-accuracy visibilities," in *Fundamental Stellar Parameters: The Interaction Between Observation and Theory*, T. R. Bedding, A.J. Booth, and J. Davis, eds., 18–21 (Dordrecht: Kluwer Academic, 1998a).
- G. Perrin, V. Coudé du Foresto, S.T. Ridgway, J.-M. Mariotti, W.A. Traub, N.P. Carleton, and M.G. Lacasse, "Extension of the effective temperature scale of giants to types later than M6," *Astron. Astrophys.* **331**, 619–626 (1998b).
- D.J. Schroeder, *Astronomical Optics*, 2 edn. (San Diego, CA: Academic Press, 2000).

Chapter 13

An Introduction to Closure Phases

JOHN D. MONNIER

SMITHSONIAN ASTROPHYSICAL OBSERVATORY
CAMBRIDGE, MASSACHUSETTS

13.1 Introduction

Phase distortions due to atmospheric turbulence, as emphasized by Quirrenbach in Chapter 5, cause a variety of ill effects. They both limit the maximum useful size of a collecting aperture and the longest allowed integration times, setting severe limitations on sensitivity. However even when a fringe is detected, the random and unknown atmospheric path delays cause phase shifts which erase information about the *intrinsic* phase arising from source structure. This chapter will discuss the use of *closure phases*, first invented for use in radio interferometry to recover most, if not essentially all, of this lost phase information for interferometric arrays with three or more telescopes.

13.1.1 Telescope Errors: Complex Gain

In an interferometric array, amplitude and phase errors associated with telescope i can be conceptualized in terms of a complex gain, \tilde{G}_i , where the tilde is used to indicate a complex number endowed with both an amplitude and phase. At radio wavelengths, the electric field of the incoming radiation can be directly measured at each telescope; in the visible/infrared, the field is not measured before interference, but rather is modified by the atmosphere and optics in each telescope before beam combination. In either case, the “measured” electric field can be represented as follows:

$$\tilde{E}_i^{\text{measured}} = \tilde{G}_i \tilde{E}_i^{\text{true}} \quad (13.1)$$

$$= |G_i| e^{i\Phi_i^G} \tilde{E}_i^{\text{true}}. \quad (13.2)$$

The amplitude of \tilde{G} corresponds to the overall scale factor, collectively representing all telescope-specific effects which modify the intensity of the received stellar radiation, e.g. mirror reflectivity, detector sensitivity, local scintillation. The phase Φ_i^G encodes all telescope-specific phase shifts, such as those due changing optical pathlengths from thermal expansion/contraction or *atmospheric turbulence* above the telescope.

How do such errors effect the measurement of the complex visibility? When light from two telescopes i and j are interfered, the visibility $\tilde{\mathcal{V}}_{ij}$ is derived from the contrast of the resulting fringes. Using Equation 13.2, we can see how the telescope-specific errors, represented by complex gains \tilde{G} , affect the measured visibility:

$$\text{Since } \tilde{\mathcal{V}}_{ij} \propto \tilde{E}_i \cdot \tilde{E}_j^*, \quad (13.3)$$

$$\tilde{\mathcal{V}}_{ij}^{\text{measured}} = \tilde{G}_i \tilde{G}_j^* \tilde{\mathcal{V}}_{ij}^{\text{true}} \quad (13.4)$$

$$= |G_i| |G_j| e^{i(\Phi_i^G - \Phi_j^G)} \tilde{\mathcal{V}}_{ij}^{\text{true}}. \quad (13.5)$$

Atmospheric Phase Errors

From Equation 13.5, we can see that the measured phase of a detected fringe is shifted by the phase difference of the phase offsets at the individual telescopes. This can be easily seen in the idealized interferometer sketched in Figure 13.1. In this figure, an optical interferometer is represented by a Young's two-slit experiment (Born and Wolf 1999). Flat wavefronts from a distant source impinge on the slits and produce an interference pattern on an illuminated screen; this interference pattern drawn corresponds to the field *intensity*, not the electric field strength.

The spatial frequency of these (intensity) fringes is determined by the distance between the slits (in units of the wavelength of the illuminating radiation). However if the pathlength above one slit is changed (due to a pocket of warm air moving across the aperture, for example), the interference pattern will be shifted by an amount depending on the difference in pathlength of the two legs in this simple interferometer. If the extra pathlength is half the wavelength, the fringe pattern will shift by half a fringe, or π radians. The phase shift is completely independent of the slit separation, and only depends on slit-specific phase delays (as in Equation 13.5).

Why Not Average Phase?

One might think that the intrinsic phase of the interference pattern could be recovered by averaging over many realizations of the atmosphere. Even when the average atmosphere-

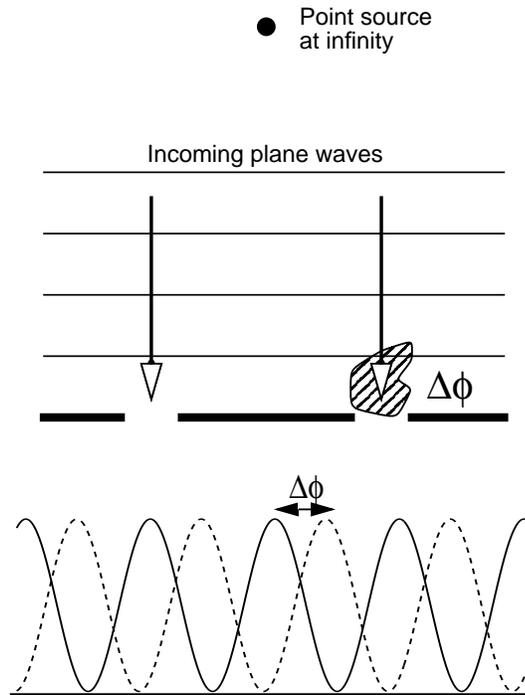


Figure 13.1: Phase errors at telescopes cause fringe shifts, as can be seen through analogy with Young's two-slit experiment.

induced phase shift is zero, the intrinsic phase can not be deduced if the rms phase shift $\Delta\phi_{\text{atm}}$ is greater than about 1 radian.

This is illustrated in Figure 13.2. The top panel (case 1) shows the distribution of measured phases when $\Delta\phi_{\text{atm}}$ is less than one radian, while the bottom panel (case 2) shows what happens when $\Delta\phi_{\text{atm}} \gg 1$ radian. Clearly, unless one knows *which* fringe one is measuring, the phase wrapping induced by large atmospheric fluctuations completely scrambles the phase information on any given baseline. Under typical seeing conditions, one can expect pathlength fluctuations up to 5–10 μm , hence most optical and infrared optical interferometers operate under conditions similar to case 2, and baseline phase information is destroyed by the turbulent atmosphere.

The loss of this phase information has serious consequences. Imaging of non-centrosymmetric objects rely on the Fourier phase information encoded in this intrinsic phase of interferometer fringes. Without this information, imaging can not be done except for simple objects such as disks or round stars. Fortunately, a number of strategies have evolved to circumvent these difficulties.

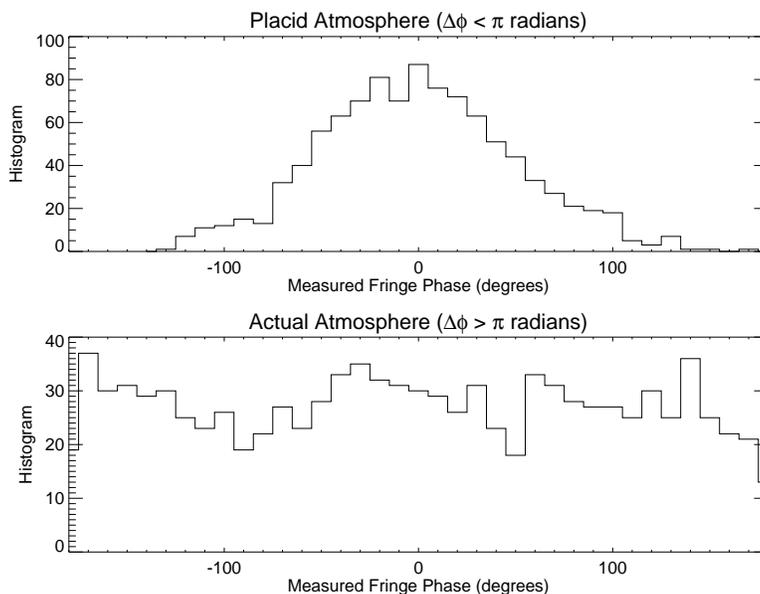


Figure 13.2: *Top Panel:* A simulation of measured fringe phases when the atmosphere induces small phase shifts. The intrinsic phase (0 degrees) could be extracted from such a dataset. *Bottom Panel:* A simulation of measured fringe phases for a turbulent atmosphere which causes phase shifts greater than π radians. All phase information is lost.

13.1.2 Phase Referencing

In Chapter 9 of this volume, possible techniques for recovering this phase information using *phase referencing* are described. I will briefly mention only three of these.

1. **Nearby Sources.** If a bright point source (or source with well-known structure) lies within an isoplanatic patch (see Chapter 5), then its fringes will act as a probe of the atmospheric conditions. By measuring the instantaneous phases of fringes from the bright source, one can correct the corrupted phases on the science target. This has been applied to narrow-angle astrometry where fringe-phase information is used for determining precise relative positions of nearby stars (Shao and Colavita 1992; Colavita *et al.* 1999). While it would be very valuable to use an artificial guide star for phase referencing a long baseline interferometer, current laser beacons are too spatially extended; schemes exist to circumvent this problem (Gavel *et al.* 1998).
2. **Measuring $\Delta\Phi$.** In the millimeter and sub-millimeter bands, phase shifts caused by fluctuations in the column density of atmospheric water-vapour can be monitored by observing its line emission. This information can be used to phase-compensate the interferometer, allowing longer coherent integrations and accurate fringe-phase determination on the target (Wiedner 1998, and references therein). In the mid-infrared, strategies to actively monitor ground-level turbulence using temperature

sensors are being explored by the Infrared Spatial Interferometry group at Mt. Wilson motivated by recent atmospheric studies (e.g., Bester *et al.* 1992).

3. **Spectral Lines.** Another possibility is to observe a target at multiple wavelengths and to use data from one part of the spectrum to calibrate another. For example, one might use fringes formed by the continuum emission to phase reference a spectral line (e.g., Vakili *et al.* 1997).

Phase referencing is not feasible for most sources, and one must make use of *closure quantities*.

13.2 The Closure Quantities

13.2.1 Closure Phase and the Bispectrum

Consider Figure 13.3 in which a phase delay is introduced above telescope 2. This causes a phase shift in the fringe detected between telescopes 1–2, as discussed in the last section. Note that a phase shift is also induced for fringes between telescopes 2–3; however, this phase shift is equal but *opposite* to the one for telescopes 1–2.

$$\begin{array}{lcl} \text{Observed} & \text{Intrinsic} & \text{Atmosphere} \\ \Phi(1-2) & = & \Phi_0(1-2) + [\phi(2) - \phi(1)], \end{array} \quad (13.6)$$

$$\Phi(2-3) = \Phi_0(2-3) + [\phi(3) - \phi(2)], \quad (13.7)$$

$$\Phi(3-1) = \Phi_0(3-1) + [\phi(1) - \phi(3)]. \quad (13.8)$$

Hence, the sum of three fringe phases, between 1–2, 2–3, and 3–1, is insensitive to the phase delay above telescope 2.

$$\begin{aligned} \text{Closure Phase (1-2-3)} &= \Phi(1-2) + \Phi(2-3) + \Phi(3-1) \\ &= \Phi_0(1-2) + \Phi_0(2-3) + \Phi_0(3-1). \end{aligned} \quad (13.9)$$

This argument holds for arbitrary phase delays above any of the three telescopes. In general, the sum of three phases around a closed triangle of baselines, the *closure phase*, is a good interferometric observable; that is, it is independent of telescope-specific phase shifts induced by the atmosphere or optics.

The idea of closure phase was first introduced by Jennison to compensate for poor phase stability in early radio VLBI work (Jennison 1958). Although Jennison also described an optical counterpart (Jennison, 1961), Rogstad (1968) is usually credited with first suggesting closure phase techniques at optical wavelengths. The first experiments at optical telescopes, using aperture masks, were only carried out some twenty years later (Baldwin *et al.* 1986; Haniff *et al.* 1987; Readhead *et al.* 1988; Haniff *et al.* 1989). Currently only two separate-element interferometers have succeeded in obtaining closure-phase measurements, in the

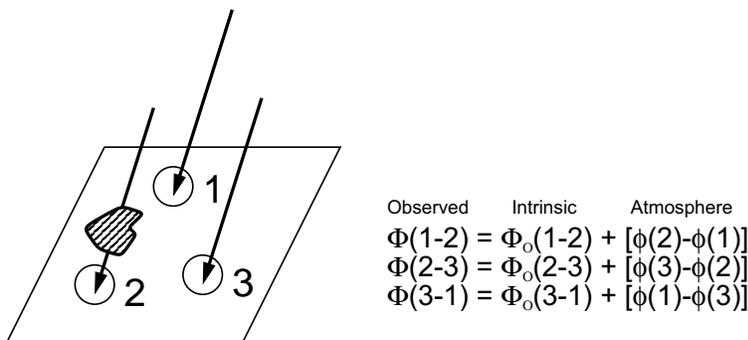


Figure 13.3: Phase errors introduced at any telescope in an array causes equal but opposite phase shifts, canceling out in the *closure phase*, which is the sum of $\Phi(1-2)$, $\Phi(2-3)$, and $\Phi(3-1)$ (see Readhead *et al.* 1988).

optical/infrared, first at COAST (Baldwin *et al.* 1996) and soon after at NPOI (Benson *et al.* 1997).

Another way to derive the invariance of the closure phase to telescope-specific phase shifts is through the *bispectrum*. The bispectrum $\tilde{B}_{ijk} = \tilde{\mathcal{V}}_{ij}\tilde{\mathcal{V}}_{jk}\tilde{\mathcal{V}}_{ki}$ is formed through triple products of the complex visibilities around a closed triangle, where ijk specifies the three telescopes. Using Equation 13.5, we can see how the telescope-specific errors affect the measured bispectrum:

$$\tilde{B}_{ijk} = \tilde{\mathcal{V}}_{ij}^{\text{measured}} \tilde{\mathcal{V}}_{jk}^{\text{measured}} \tilde{\mathcal{V}}_{ki}^{\text{measured}} \quad (13.10)$$

$$= |G_i||G_j| e^{i(\Phi_i^G - \Phi_j^G)} \tilde{\mathcal{V}}_{ij}^{\text{true}} \cdot |G_j||G_k| e^{i(\Phi_j^G - \Phi_k^G)} \tilde{\mathcal{V}}_{jk}^{\text{true}} \cdot |G_k||G_i| e^{i(\Phi_k^G - \Phi_i^G)} \tilde{\mathcal{V}}_{ki}^{\text{true}} \quad (13.11)$$

$$= |G_i|^2 |G_j|^2 |G_k|^2 \tilde{\mathcal{V}}_{ij}^{\text{true}} \cdot \tilde{\mathcal{V}}_{jk}^{\text{true}} \cdot \tilde{\mathcal{V}}_{ki}^{\text{true}}. \quad (13.12)$$

From the above derivation, one can see the bispectrum is a complex quantity, and that the phase is identical to the closure phase. The use of the bispectrum for reconstructing diffraction-limited images was developed independently (Weigelt 1977) of the closure-phase techniques, and the connection between the approaches realized only later (Roddiier 1986).

13.2.2 Closure Amplitudes

When one has four or more telescopes, another important closure quantity can be formed, the *closure amplitude*. The closure amplitude is constructed to be independent of the telescope-specific gain amplitudes. The closure amplitude A_{ijkl} can be defined in a variety of ways, but here it is defined in terms of four telescopes $ijkl$:

$$A_{ijkl} = \frac{|\tilde{\mathcal{V}}_{ij}^{\text{measured}}| |\tilde{\mathcal{V}}_{kl}^{\text{measured}}|}{|\tilde{\mathcal{V}}_{ik}^{\text{measured}}| |\tilde{\mathcal{V}}_{jl}^{\text{measured}}|} \quad (13.13)$$

$$= \frac{|\tilde{G}_i| |\tilde{G}_j| |\tilde{\mathcal{V}}_{ij}^{\text{true}}| |\tilde{G}_k| |\tilde{G}_l| |\tilde{\mathcal{V}}_{kl}^{\text{true}}|}{|\tilde{G}_i| |\tilde{G}_k| |\tilde{\mathcal{V}}_{ik}^{\text{true}}| |\tilde{G}_j| |\tilde{G}_l| |\tilde{\mathcal{V}}_{jl}^{\text{true}}|} \quad (13.14)$$

$$= \frac{|\tilde{\mathcal{V}}_{ij}^{\text{true}}| |\tilde{\mathcal{V}}_{kl}^{\text{true}}|}{|\tilde{\mathcal{V}}_{ik}^{\text{true}}| |\tilde{\mathcal{V}}_{jl}^{\text{true}}|} \quad (13.15)$$

In radio interferometry, this is an important quantity and can be used to compensate for detector gain fluctuations and changing antenna efficiencies. However in the optical and infrared regime, varying fringe amplitudes are not caused by telescope-specific gain changes (such as scintillation), but rather arise from baseline-dependent decorrelation effects related to atmospheric turbulence. Hence, the closure amplitude is not very important to consider for optical/IR interferometers, although recent advances, such as the use of high-order adaptive optics systems and fast fringe tracking, may make this quantity more interesting in the future.

13.2.3 Measuring Closure Phases

Before we discuss how to interpret the closure phases, let us first discuss the appropriate way of measuring them.

Averaging in the Complex Plane

Under noisy conditions, each individual closure-phase measurement may vary from -180° to 180° . Obviously, averaging the scalar phase in such a case will be useless as shown in Section 13.1.1. However, there is a crucial difference between averaging the closure phases and averaging the fringe phases, and this is illustrated in Figure 13.4. The source of the phase variations in the latter case was atmospheric phase shifts, or “phase noise.” However, the closure phase can be thought of as the phase of the bispectrum, and the noise (from detector or Poisson statistics) is an “additive noise.” In the left panel of Figure 13.4, the measured bispectrum (thick, solid arrow) is shown in the complex plane and is the sum of a (small) “true” signal (thin, solid arrow) and a larger “noise” vector (dashed arrow). The resulting vector (bispectrum measurement) could lie anywhere on the dashed circle; since the closure phase is the phase of the complex bispectrum, one can see that the closure phase is hardly constrained. However, the right panel shows why averaging of the bispectrum is effective, unlike averaging of the fringe phasors in Section 13.1.1. The “true signal” component of the vector-averaged bispectrum increases linearly with the number of samples N , while the “noise signal” is undergoing a random walk whose rms amplitude grows like \sqrt{N} . Simply speaking, the signal-to-noise ratio of the bispectrum measurement should grow like \sqrt{N} . Methods of averaging noisy closure phases are discussed in Woan and Duffett-Smith (1988). More detailed (and accurate) analyses of the bispectrum under a wider range of signal-to-

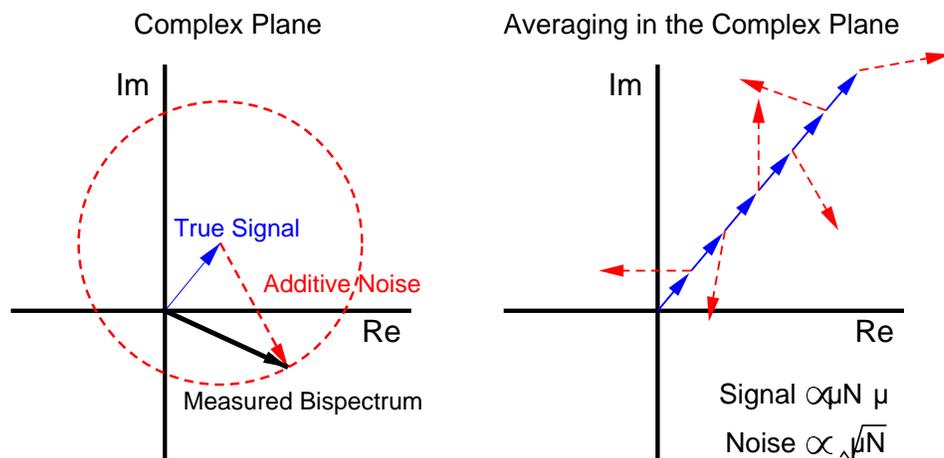


Figure 13.4: The bispectrum is a good observable for optical interferometry. The measured bispectrum is a complex quantity composed of the *true signal* and noise. Vector averaging allows the signal to grow linearly with the number of independent data points N , while the amplitude of the noise term will only grow by \sqrt{N} .

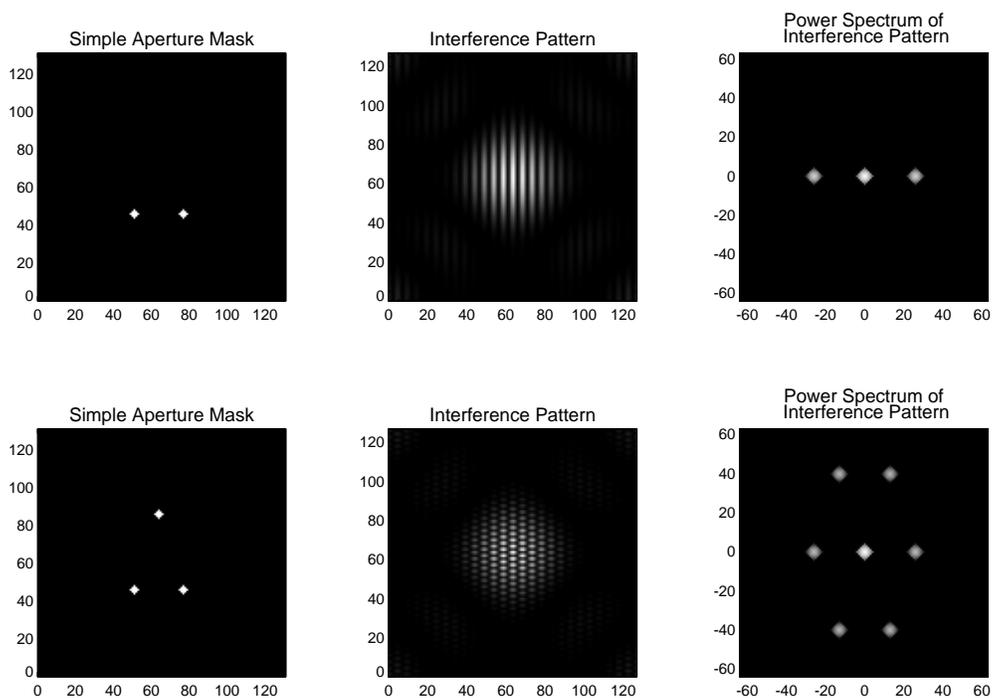


Figure 13.5: The principles of aperture masking. The left panels show the aperture masks being used, while the middle panels show the resulting interference patterns. The right panels result from taking the 2-D power spectra (modulus-squared of 2-D Fourier transform) of the image interference patterns. Each individual fringe pattern becomes a pair of isolated spots in the Fourier plane.

noise conditions are discussed in Readhead *et al.* (1988) and Kulkarni (1989). Sources of bias in the bispectrum are evaluated in Pehlemann *et al.* (1992).

Aperture Masking Example

Let us consider an example from aperture masking. Two simple implementations of aperture masking are illustrated in Figure 13.5. The top series of panels document the use of the simplest mask possible, one with only two holes. The first panel shows the aperture mask being used, while the middle panel displays the corresponding “image” at the focal plane of the telescope. One sees a fringe pattern as expected from a simple Young’s experiment, with fringes running in the same direction as the hole-separation vector. The overall size of the fringe envelope is determined by the diffraction patterns of the individual holes. The two-dimensional power spectrum of the image fringes can be used to determine the exact fringe frequency and fringe amplitude, as can be seen in the right-most panel. The spot in the center is the DC-term, and is proportional to the square of the total flux in the image frame. The spots to the right and left correspond to the fringe frequency and the spatial frequency value for them is proportional to the hole-separation. Since the power spectrum of a real function is inversion-symmetric, there is equal fringe power at both positive and negative spatial frequencies. Note the finite sizes of the spots in the power spectrum; this results from the finite size of the holes, reflecting the multiple baselines that stretch from one hole to the other. Note that the geometry of the power spectrum is directly related to the autocorrelation of the pupil mask via the Convolution Theorem (see Chapter 2, Section 2.5).

The bottom series of panels shows the next most complicated arrangement of holes for an aperture mask; one hole has been added to the first mask. One can clearly see in the middle pattern the effect of the additional hole. Now three baselines exist and thus three intersecting fringe patterns form, adding constructively and destructively in a pattern of spots. One can extract the amplitudes and exact spatial frequencies of the three fringe patterns by taking the power spectrum, which appears to the right. The original left-right baseline is still there, but now one can see the two new diagonal baselines. The closure phase can be found by Fourier transforming each image created through such a mask, and the Fourier phases summed from the appropriate spatial frequency bins. The location of the origin (phase center) during the Fourier transform does affect the individual phases determined for each baselines. However, its straightforward to prove that the closure phase is not sensitive to the choice of origin. This is equivalent to saying that the closure phase is independent of image translation. Indeed, for any set of three telescopes, it would be impossible to distinguish between an image translation or atmospheric phase disturbances.

An example of closure-phase analysis has been included in Figure 13.6 from actual data using an aperture mask on the Keck Telescope (see Section 13.3.5). The top panel shows the phases measured on three separate baselines for 100 short-exposure images on a point-source calibrator. Except for the shortest baseline (filled circles) which somewhat cluster

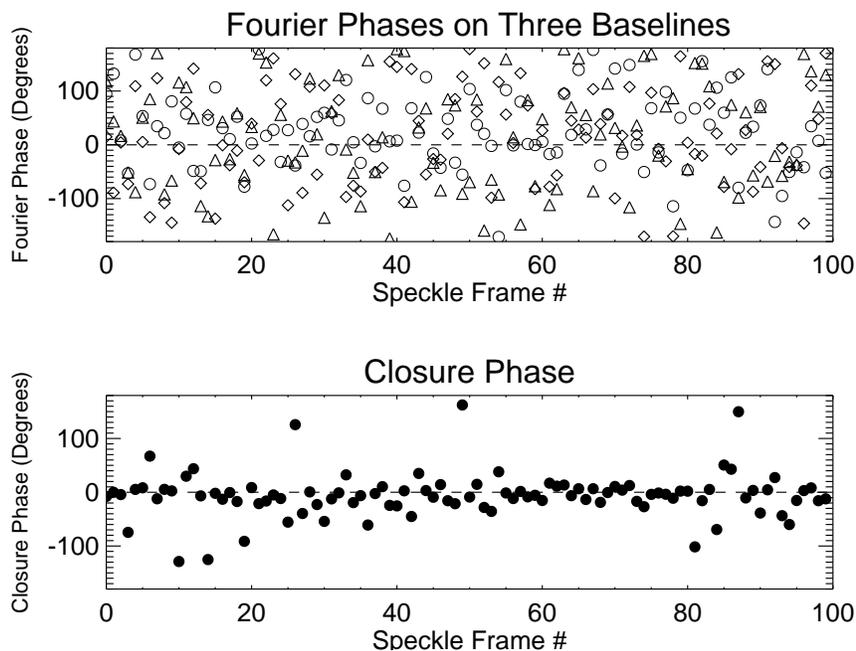


Figure 13.6: This figure illustrates the utility of closure phase averaging from real data. The top panel shows the Fourier phases of three fringe patterns, corresponding to a closed triangle of three baselines. The observed phases are marked by three different plot symbols for 100 separate frames. The bottom panel shows the closure-phase signal resulting from summing the above phases around the baseline triangle.

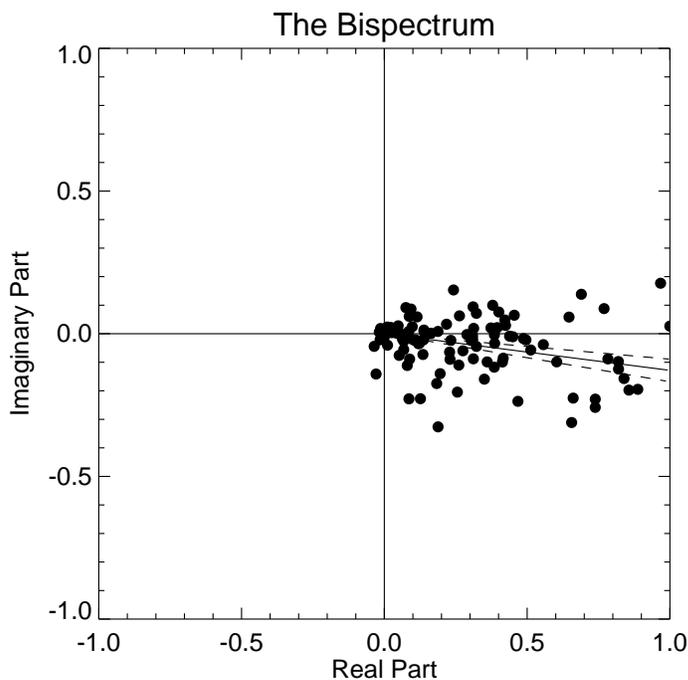


Figure 13.7: This figure shows the 100 bispectrum measurements and the vector average.

around 0° , the individual Fourier phases are pretty evenly spread between -180° and $+180^\circ$. However, when the phases are summed in the appropriate way, this closure-phase quantity has much less dispersion than the individual phase measurements and is equal to about zero degrees. Using proper weighting, the uncertainty in the above measurement of the closure phase can be reduced to only $\sim 2^\circ$ for just 100 frames; such accurate closure phases are essential for reliable image reconstructions of complicated sources. As discussed above, the best way to determine the closure phase is to average the bispectrum in the complex plane, as shown in Figure 13.7.

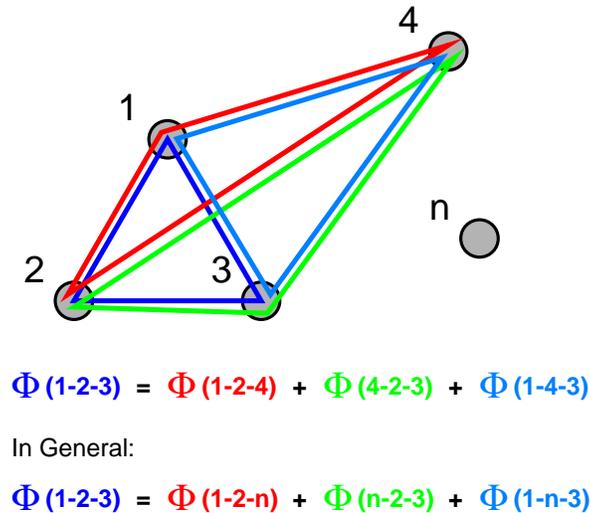


Figure 13.8: This figure illustrates important closure-phase relations.

13.2.4 Closure Phase Relations

For N telescopes, there are “ N choose 3,”

$$\binom{N}{3} = \frac{(N)(N-1)(N-2)}{(3)(2)},$$

possible closing triangles. However, there are only

$$\binom{N}{2} = \frac{(N)(N-1)}{2}$$

independent Fourier phases; clearly not all the closure phases can be independent. Figure 13.8 illustrates how any given closure phase can be expressed as a sum of three others. The number of *independent* closure phases is only

$$\binom{N-1}{2} = \frac{(N-1)(N-2)}{2},$$

equivalent to holding one telescope fixed and forming all possible triangles with that telescope. The number of independent closure phases is always less than the number of phases

Table 13.1: Phase information contained in the closure phases alone

Number of Telescopes	Number of Fourier Phases	Number of Closing Triangles	Number of Independent Closure Phases	Percentage of Phase Information
3	3	1	1	33%
7	21	35	15	71%
21	210	1330	190	90%
27	351	2925	325	93%
50	1225	19600	1176	96%

one would like to determine, but the *percent* of phase information retained by the closure phases improves as the number of telescopes in the array increases. Table 13.1 lists the number of Fourier phases, closing triangles, independent closure phases, and recovered percentage of phase information for telescope arrays of 3 to 50 elements. For example, approximately 90% of the phase information is recovered with a 21-telescope interferometric array (e.g., Readhead *et al.* 1988). This phase information can be coupled with other image constraints (e.g., finite size and positivity) to reconstruct the source brightness distribution (see Section 13.3).

13.2.5 Simple Cases

Equal Binary

Since the closure phases are independent of the phase center, one can strategically place the origin in order to more easily determine the Fourier phases for a given brightness distribution. For example, consider the equal binary system depicted in Figure 13.9. The complex visibility can be easily written by choosing the origin midway between the two components. If \mathbf{u} is the baseline vector and ρ is the separation vector of the binary, we have

$$\begin{aligned}
 V &= 0.5 \left[\exp \left(-2\pi \mathbf{u} \cdot \frac{\rho}{2} \right) + \exp \left(2\pi \mathbf{u} \cdot \frac{\rho}{2} \right) \right] \\
 &= \cos \left(2\pi \mathbf{u} \cdot \frac{\rho}{2} \right)
 \end{aligned} \tag{13.16}$$

Note the abrupt phase jump when the visibility amplitude goes through a null. These discontinuities are smoothed out when the two components are not precisely equal.

But what about the closure phases? Since a closure phase is simply a sum of three phases, we can immediately see that all the closure phases must be either 0° or 180° . In fact, this is true not just for equal binaries, but *any point-symmetric brightness distribution*. This is easily proven: by placing the origin (phase center) at the location of point-symmetry, then we can make the imaginary part of the Fourier transform disappear (i.e., all odd basis functions must be zero). Hence, the phases of *all* Fourier components must be either 0° or 180° .

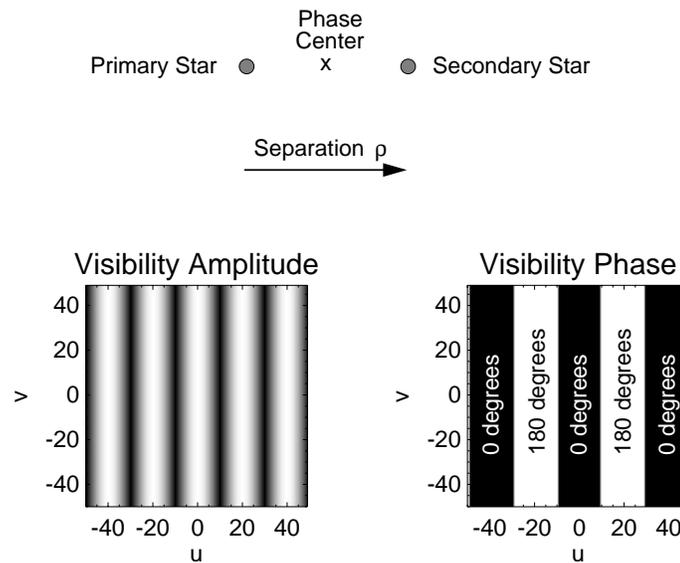


Figure 13.9: This figure shows the complex visibility for an equal binary system. With the above choice for the phase center, the Fourier phases can be represented simply. Notice the abrupt phase jumps when visibility amplitude goes through a null.

For an equal binary then, we would expect to see abrupt closure-phase jumps between 0° and 180° if one of the baselines traverses a null in the visibility pattern. This indeed has been observed with COAST (Baldwin *et al.*, 1996) and with the NPOI (Benson *et al.*, 1997). One can determine the binary separation (and brightness ratio) from the closure-phase information alone.

Faint Hotspot on Stellar Surface

In Figure 13.10, a more complicated example is illustrated, a star with a hotspot. This can be thought of as unequal binary with one of the components being resolved. For closing triangles with all short baselines (compared to that needed to resolve the star itself), the flux from the star itself dominates the visibility measurement. Hence, the system looks mostly centro-symmetric (like a round star), and we expect closure phases to be small. For triangles containing all long baselines, the star itself is mostly resolved and the hotspot dominates the appearance. In this limiting case as well, the closure phases should all be zero. This illustrates how an interferometer acts as a spatial filter, allowing only image details of certain spatial scales to be detected.

It is only at intermediate baselines, when the star itself is partially resolved, that non-zero (and non- 180°) values of the closure phases are expected. Hence, we see the importance of having a variety of closing triangles available, since only some of them will contain useful information about the source structure. For instance, without intermediate-sized closing

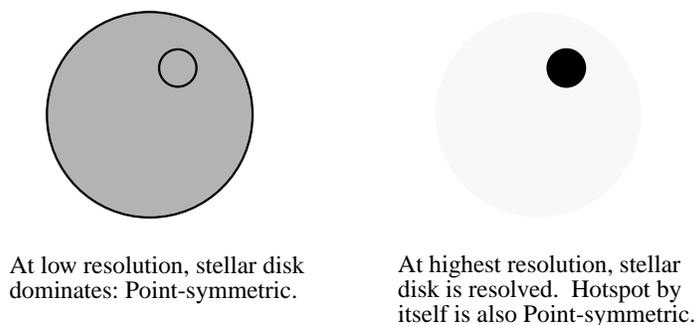


Figure 13.10: This illustrates the behavior of closure phases for a source consisting of a star with a hotspot. At both high and low spatial resolutions, the closure phases should be near zero. Only at intermediate baselines will the closure phases have significant non-zero values.

triangles, we would know the size of the star and reveal the presence of a hotspot, but we would only weakly constrain the hotspots position on the stellar surface. Very convincing measurements of non-zero/ 180° closure phases (or more simply, imaginary bispectral components) can be found in Tuthill (1994) and associated papers, establishing the presence of hotspots on a number of evolved giants and supergiants.

13.2.6 Summary

There are a few important points to remember from this section.

- The bispectrum is always real for sources with *point symmetry*. That is, the closure phases are all 0° or 180° .
- Closure phases are not sensitive to an overall translation of image. A translation is indistinguishable from atmospheric phase delays for any given closing triangle.
- The closure phases, or bispectrum, are independent of *telescope-specific* phase errors. Non-zero closure phases from a point source can result from having non-closing triangles and phase delays after beam combination.

13.3 Imaging

While modeling visibility and closure-phase data with simple models is useful, one would like to make an image unbiased by theoretical expectations. With a large number of measurements, images of arbitrary complexity should be attainable using optical/infrared interferometers and reliable closure-phase measurements. This section will discuss strategies currently employed, based on the techniques of Very Long Baseline Interferometry (VLBI) in the radio.

13.3.1 Converting Bispectrum to Complex Visibilities

In order to make an image from an interferometer, one needs estimates of the complex visibilities over a large portion of the (u, v) plane, both the amplitudes *and* phases. Unfortunately, there are not as many independent closure phases as there are Fourier phases (for non-redundant arrays), hence there is not a unique mapping from closure phases to Fourier phases. This would seem to make inverting the closure phases into the original Fourier phases a mathematical impossibility.

Baseline Redundancy—Baseline Bootstrapping

One way to get around this problem is to introduce baseline redundancy into the interferometric array. Figure 13.11 shows a simple four-element, linear, redundant array. Baselines connecting telescopes 1–2, 2–3 and 3–4 are all identical. Hence, the number of independent Fourier Phases is reduced from 6 to only 3. We can always assign the shortest baseline Fourier phase to be zero, since this encodes the position information which can not be extracted from the closure-phase data anyway. Consider the arrows drawn on Figure 13.11; these represent some of the baselines and closure triangles possible to have in this geometry. The Fourier phase of baseline 1–3 can be determined from the Fourier phases on the identical baselines 1–2 and 2–3 (which are equal and set to zero) and the closure phase 1–2–3. Now that we have the phase for baseline 1–3, we can determine the Fourier phase for baseline 1–4, by using the closure phase 1–3–4 and the (now) known phases for baselines 1–3 (just deduced) and 3–4 (same as 1–2 because of baseline redundancy). This process can be repeated to directly solve for all the Fourier phases given only the closure phases and setting the short baseline to zero phase.

While this process has found application for direct image reconstruction based on speckle interferometry (e.g., Cruzalebes *et al.* 1996; Koresko 1993), there are a number of problems with this approach. Because the short baseline phases must be determined first in order to “spiral-out” and solve for longer baselines, the long-baseline phases have much higher

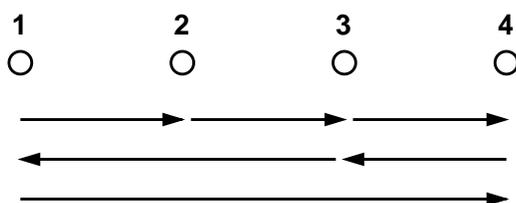


Figure 13.11: This is an example of a redundant four-element array geometry. Notice how baselines 1–2, 2–3, and 3–4 are identical; this allows the phase on baseline 1–3 to be determined using the closure phase 1–2–3 and only a single Fourier phase 1–2.

noise. This poor noise propagation can be improved by regularizing the inversion process, taking into account the noise and solving for a best fit. Simple regularization methods fail to properly handle phase wrapping and make this process vulnerable to significant errors unless handled with relatively high numerical sophistication.

Redundant arrays result in significantly poorer sampling in the Fourier plane, (u, v) coverage, than non-redundant alternatives when there are only a few telescopes. This would argue against deploying interferometric elements in such a manner. Interestingly however, baseline bootstrapping methods for allowing fringe tracking on long baselines with low-visibility amplitudes may necessitate partially redundant arrays for future sensitive optical/infrared interferometers (see Chapter 14). Under these conditions, the ability to directly solve for the Fourier phases from the closure phases may be quite valuable.

Applying Image Constraints

Another way of determining the Fourier phases from a limited number of closure phases is to apply image constraints. For instance, brightness distributions are always non-negative and generally have a finite extent. This is especially true in the infrared where thermal emission is almost always quite compact. These constraints introduce correlations in the Fourier amplitudes and phases, and essentially remove degrees of freedom from our inversion problem. The rest of this section will explore this strategy for image reconstruction using closure phases.

13.3.2 Imaging Goals

The goals of an image-reconstruction procedure can be stated quite simply: find an image which fits both the visibility amplitudes and closure phases within experimental uncertainties. However in practice, there are an infinite number of candidate images which satisfy this criterion, because interferometric data are always incomplete and noisy. Furthermore, the closure phases can not be used to unambiguously arrive at Fourier phase estimates as stated above, even under ideal noise-free conditions.

Additional constraints are imposed to “select” an image as the best-estimate of the true brightness distribution. Some of the most common ones are described below.

- **Limited Field-of-View.** This constraint is always imposed in aperture synthesis imaging, even for a fully-phased array (e.g., VLA). Limiting the field-of-view introduces correlations in the complex visibility in the (u, v) plane. This is a consequence of the Convolution Theorem (Chapter 2, Section 2.5), where a multiplication in image-space is equivalent to a convolution in the corresponding Fourier-space.
- **Positive-Definite.** Since brightness distributions can not be negative, this is a sensible constraint. While clearly limiting the range of “allowed” complex visibilities, there are

few obvious, intuitive effects in the Fourier-plane; one is that the visibility amplitude is maximum at zero spatial frequency. The Maximum Entropy Method (see Section 13.3.3) naturally incorporates this constraint.

- “Smoothness.” The Maximum Entropy Method, for instance, selects the “smoothest” image consistent with the data.
- *A Priori* Information. One can incorporate previously known information to constrain the possible image reconstructions. For instance, a low-resolution image may be available from a single-dish telescope. Another commonly encountered example is a point source embedded in a nebulosity; one might want the reconstruction algorithm to take into account that the source at the center is point-like from theoretical arguments.

13.3.3 “Standard” Aperture Synthesis Imaging

For a phased interferometric array (e.g., the VLA), one can use a number of aperture synthesis techniques to produce an estimate of an image based on sparsely sampled Fourier components. These procedures basically remove artifacts, i.e. sidelobes, of the interferometer’s point-source response arising from uneven sampling of the (u, v) plane. These procedures do not incorporate closure phases, but work by inverting the Fourier amplitudes and phases to make an image. A brief explanation of the most popular algorithms CLEAN and MEM follow with additional references for the interested reader. See Perley *et al.* (1986) for essays on these topics aimed at radio astronomers.

CLEAN

Originally described by Högbom (1974), CLEAN has been traditionally the most popular algorithm for image reconstruction in radio interferometry because it is both computationally efficient and intuitively understandable. Given a set of visibility amplitudes and phases over a finite region of the Fourier plane, the “true” image can be estimated by simply setting all other spatial frequencies to zero and taking the Fourier Transform. As one might expect, this process leads to a whole host of image artifacts, most damaging being positive and negative “sidelobes” resulting from non-complete coverage of the Fourier plane; we call this the “dirty map.” The unevenly-filled Fourier plane can be thought of as a product of a completely-sampled Fourier plane (which we desire to determine) and a spatial frequency mask which is equal to 1 where we have data and 0 elsewhere. Since multiplication in Fourier space is identical to convolution in image space, we can take the Fourier transform of the spatial frequency mask to find this convolving function; we call this the “dirty beam.” Now the image reconstruction problem can be recast as a “deconvolution” of the dirty map with the dirty beam.

The dirty map is CLEANed by subtracting the dirty beam (scaled to some fraction of the map peak) from the brightest spot in the dirty map. This removes sidelobe structure and artifacts from the dirty map. Repeating this process with dirty beams of ever decreasing amplitudes leads to a series of delta-functions which, when combined, fit the interferometric data. For visualization, this map of point sources is convolved with a Gaussian function whose full-width at half-maximum (FWHM) values are the same as the dirty beam; this removes high spatial resolution information beyond the classic “Rayleigh” criterion cutoff. One major weakness with CLEAN is that this smoothing changes the visibility amplitudes, hence the CLEANed image no longer strictly fits the interferometric data, especially the spatial frequency information near the diffraction limit. Another weakness is that CLEAN does not directly use the known uncertainties in the visibility data, and hence there is no natural method to weight the high signal-to-noise data more than the low signal-to-noise data during image reconstruction. Further discussion of various implementations of CLEAN can be found in Clark (1980), Schwab (1984), Cornwell (1983), and Chapter 7 of Perley *et al.* (1986) by T. Cornwell.

MEM

The maximum entropy method (MEM) makes better use of the highest spatial frequency information by finding the *smoothest* image consistent with the interferometric data. While enforcing positivity and conserving the total flux in the frame, “smoothness” is estimated here by a global scalar quantity S , the “entropy.” If f_i is the fraction of the total flux in pixel i , then

$$S = - \sum_i f_i \ln \frac{f_i}{I_i} \quad (13.17)$$

after the thermodynamic quantity; I_i is known as the *image prior* and must be specified by the user. The MEM map f_i will tend toward I_i when there are little (or noisy) data to constrain the fit. Often I_i is assumed to be a uniformly bright background, however one can use other image priors if additional information is available, such as the overall size of the source which may be known from previous observations.

Mathematically, MEM solves the multi-dimensional (N =number of pixels) constrained minimization problem which only recently has become computationally realizable on desktop computers. Maintaining an adequate fit to the data ($\chi^2 \sim$ number of degrees of freedom), MEM reconstructs an image with maximum S . MEM image reconstructions always contain some spatial frequency information beyond the diffraction limit in order to keep the image as “smooth” as possible consistent with the data. Because of this, images typically have maximum spatial resolution a few times smaller than the typical Rayleigh-type resolution encountered with CLEAN (“super-resolution”). Further discussions of MEM and related Bayesian methods can be found in Pina and Puetter (1992), Narayan and Nityananda (1986), Skilling and Bryan (1984), Gull and Skilling (1984), and Sivia (1987).

Unfortunately, MEM images also suffer from some characteristic artifacts and biases. Photometry of MEM-deconvolved images is necessarily biased because of the positivity constraint; any noise or uncertainty in the imaging appears in the background of the reconstruction instead of the source, systematically lowering the estimated fluxes of compact sources. Also, fields containing a point source embedded in extended emission often show structure reminiscent of Airy rings, the location of the rings being influenced by the wavelength of the observation and not inherent to the astrophysical source. Fortunately, these imaging artifacts are greatly alleviated for asymmetric structures, when closure phases and not the visibility amplitudes play a dominant role in shaping the reconstructed morphology.

13.3.4 Including Closure Phase Information

The above algorithms were designed to use Fourier amplitudes and phases, not closure phases. Early image reconstruction algorithms incorporated closure-phase information by using an iterative scheme (Thompson *et al.* 1986; Readhead and Wilkinson 1978). The following steps summarize this process:

1. Start with a “phase model” based on either prior information or setting all phases to zero.
2. Determine candidate phases by using some values from the “phase model” and enforcing all the (self-consistent) closure-phase relations (see Section 13.2.4).
3. Using CLEAN or MEM, perform aperture synthesis mapping on the given visibilities and candidate phases. At this stage, image constraints such a positivity and/or finite support are applied.
4. Use this image as a basis for a new “phase model.”
5. Go to step 2 and repeat until the process converges to a stable image solution.

“Self-Calibration”

Cornwell and Wilkinson (1981) introduced a modification of the above scheme by explicitly solving for the telescope-specific errors as part of the reconstruction step. Hence the measured (corrupted) Fourier phases are fit using a combination of intrinsic phases (which are used for imaging using CLEAN/MEM) plus telescope phase errors. In this scheme, the closure phases are not explicitly fit, but rather are conserved in the procedure since varying telescope-specific errors can not change any of the closure phases. Figure 13.12 shows a flow diagram for this procedure, and requires thoughtful consideration in order to fully understand the power and elegance of “self-cal.”

Self-calibration works remarkably well for large number of telescopes, but requires reasonably high signal-to-noise ratio ($\text{SNR} \gtrsim 5$) in the measured complex visibilities. Once the

Self-Calibration

models intrinsic Fourier phases
plus telescope errors

$$\Phi_{ij}^{intrinsic} = \Phi_{ij}^{measured} - \underbrace{(\phi_i - \phi_j)}_{\text{telescope errors}}$$

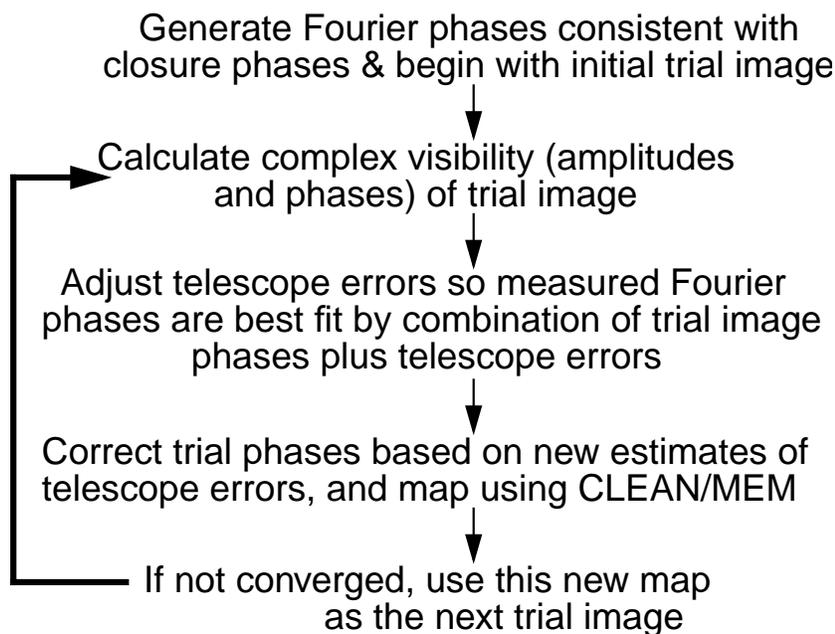


Figure 13.12: This is a flow diagram for a incorporating closure-phase information into CLEAN/MEM aperture synthesis imaging algorithms based on the “self-calibration” procedure of Cornwell and Wilkinson (1981).

signal-to-noise ratio decreases below this point, the method completely fails. This conceptualization, while useful for radio interferometry, fails for optical/infrared interferometry where the good observables are the closure phases themselves, not corrupted Fourier phases. This is because the time-scale for phase variations in the optical/infrared is much less than a second, as opposed to minutes/hours in the radio.

13.3.5 Imaging Results from Keck Aperture Masking

Detailed images of dusty circumstellar envelopes have been made by performing aperture masking interferometry on the Keck-I telescope at near-IR wavelengths. Both non-redundant and partially-redundant array geometries have been used with success. The methodology was as follows. A set of visibility amplitudes and closure phases was obtained from a series of 100 speckle frames. These data were calibrated by using an equal number of frames taken on a nearby point-source calibrator. A set of Fourier phases consistent with the measured closure phases (to within noise) was then generated using an iterative relaxation technique (Monnier 1999), avoiding phase-wrapping problems of Singular Value Decomposition methods. This allowed the data to be represented in terms of visibility amplitudes and “corrupted” Fourier phases. After data conversion to a VLBI data format, the self-calibration/MEM routines of Sivia (1987) were used to produce diffraction-limited images. Results from this experiment as well as details regarding observing methodology and mask geometries can be found in various publications (e.g., Tuthill *et al.* 1998; Monnier 1999; Monnier *et al.* 1999; Tuthill *et al.* 1999; Tuthill *et al.* 2000).

IRC +10216 and WR 104

In this section I present images of dust shells around IRC +10216 and WR 104. To illustrate the importance of phase information to accurately reconstruct images of complicated objects, I have included MEM reconstructions using no phase information as well as reconstructions taking advantage of the measured closure phases. Figure 13.13 contains these results.

Spurious symmetrization is observed in the images with no phase information; this is because the phases were set equal to zero for definiteness. This enforced point-symmetry and explains the symmetry observed in the upper panels. Note that the dust shells around these stars do not conform to the spherically-symmetric, uniform-outflowing dust density distributions which were theoretically expected. These dust shells have been fit for many years using spherically-symmetric dust shell models, because insufficient data existed to make maps free of theoretical bias. Imaging of astrophysical objects is quite important for confirming theoretical ideas, and these examples showcase the potential of long-baseline interferometry to produce images of complicated environments using closure-phase imaging techniques when arrays with sufficient numbers of telescopes come into service.

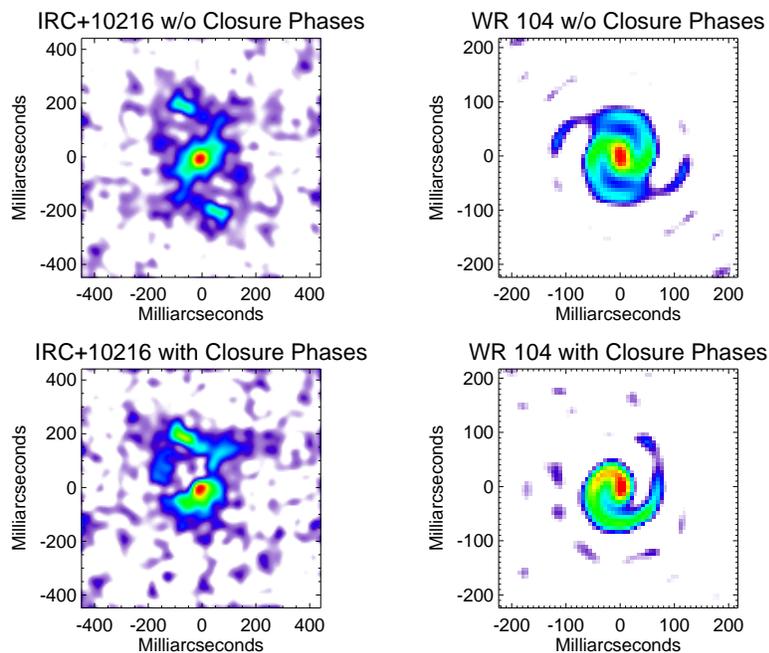


Figure 13.13: (*top panels*) These panels show K-band reconstructions of IRC +10216 and WR 104 with no phase information. (*bottom panels*) These panels show the maps when closure-phase information is taken into account.

13.4 Outstanding Issues

There are a number of problems with existing software being used for image reconstruction by long-baseline optical/infrared interferometry groups. The adoption of phase-referencing techniques by the VLA and VLBA has allowed direct Fourier inversion of the interferometry data, using CLEAN or MEM, without requiring a self-calibration step. For this reason (I believe), scant software or theoretical progress has been made in the last 10 years. Disappointingly, the immense potential of the exponential increase in computing power has remained unrealized.

13.4.1 Overview of Problems

Here is an incomplete list of some of the outstanding problems with the current status of image reconstruction software for optical/infrared long-baseline interferometry data:

- There is no agreed data storage format for optical/infrared interferometry data. Standard radio formats (UVFITS) record complex visibilities rather than the bispectrum or closure phases. Some have suggested a NASA-derived standard be adopted, since IPAC will presumably need to archive data from upcoming interferometry projects (e.g., SIM and the Keck Interferometer). Others believe an extension of FITS, within the AIPS++ paradigm, may be better.

- Algorithms should fit directly to the good observables, i.e. the closure phases (bispectrum) and power spectra. This is straightforward, although more computationally expensive. The “self-cal” concept of fitting to the telescope errors breaks down, since the atmosphere is cycling the phase many times every second.
- Fits based on the scalar χ^2 have baseline-dependent residuals when used with most regularization schemes. Use of multi-resolution approaches or more sophisticated “goodness-of-fit” criteria should be attempted.
- A point source embedded in extended nebulosity is often encountered, yet common image reconstruction algorithms (e.g., MEM) introduce bad artifacts under these conditions (“ringing”). Use of this type of *a priori* information should be included in new algorithms.
- New generation of interferometers have from three to seven elements (usually three). This will result in quite uneven Fourier coverage, and new strategies, both for planning observations and reconstructing images, will likely be required if imaging of moderately complex sources are desired.

Note that solving the above problems will also extend the usefulness of speckle interferometry, for most users of these techniques are not optimally using their data. In addition, there may be advantages of combining aperture-masking with new adaptive optics systems (e.g., Haniff and Wilson 1994).

13.4.2 New Possibilities

Before concluding this section, I want to mention some interesting new techniques which show promise for solving some of the above problems. While efforts to reconstruct images based on the bispectrum alone have a relatively long history (e.g., Weigelt 1977), only recently have these techniques seen success imaging complicated environments (Hofmann and Weigelt 1993; Weigelt *et al.* 1998). Recent work applied to mm-wave interferometry, which has similar atmospheric phase issues, can be found in Katagiri *et al.* (1997). Some of the artifacts of MEM and CLEAN could be avoided by developing new aperture synthesis approaches, based on wavelets (Starck *et al.* 1994), pixon-based methods (Pina and Puetter 1993; Yahil 1999), or WIPE (Lannes *et al.* 1997). For the serious reader, I recommend the series of difficult but interesting papers relating closure-phase imaging to algebraic graph theory (Lannes 1990; Lannes *et al.* 1997; Lannes 1998b; Lannes 1998a); the power of these ideas are surely underutilized.

13.5 Summary of Important Points

- Closure phases and closure amplitudes are insensitive to *telescope-specific* errors such as atmospheric phase delays.
- Closure phases and closure amplitudes *can not* be used to calibrate *baseline-dependent* problems.
- All point-symmetric objects (including disks) yield closure phases of either 0° or 180° , i.e. the bispectrum is entirely *real*. Hence, closure phases act as strong probes of asymmetric structure, even when source structure can not be fully resolved.
- Software and theoretical work is absolutely necessary for optimal imaging with the current generation of long baseline interferometers.
- Despite the total scrambling of Fourier phases by a turbulent atmosphere, the use of closure phases and image constraints allow nearly complete recovery of all phase information.

Acknowledgments

I would like to thank Chris Haniff for helpful discussions.

References

- J.E. Baldwin, C.A. Haniff, C.D. Mackay, and P.J. Warner, “Closure phase in high-resolution optical imaging,” *Nature* **320**, 595–597 (1986).
- J.E. Baldwin, M.G. Beckett, R.C. Boyesen, D. Burns, D.F. Buscher, G.C. Cox, C.A. Haniff, C.D. Mackay, N.S. Nightingale, J. Rogers, P.A.G. Scheuer, T.R. Scott, P.G. Tuthill, P.J. Warner, D.M.A. Wilson, and R.W. Wilson, “The first images from an optical aperture synthesis array: mapping of Capella with COAST at two epochs.” *Astron. Astrophys.* **306**, L13–L16 (1996).
- J.A. Benson, D.J. Hutter, N.M. Elias II, P.F. Bowers, K.J. Johnston, A.R. Hajian, J.T. Armstrong, D. Mozurkewich, T.A. Pauls, L.J. Rickard, C.A. Hummel, N.M. White, D. Black, and C.S. Denison, “Multichannel optical aperture synthesis imaging of Zeta 1 Ursae Majoris with the navy prototype optical interferometer.” *Astron. J.* **114**, 1221–1226 (1997).
- M. Bester, W.C. Danchi, C.G. Degiacomi, L.J. Greenhill, and C.H. Townes, “Atmospheric fluctuations - empirical structure functions and projected performance of future instruments,” *Astrophys. J.* **392**, 357–374 (1992).
- M. Born and E. Wolf, *Principles of Optics*, 7 (expanded) edn. (Cambridge, UK: Cambridge University Press, 1999).
- B.G. Clark, “An efficient implementation of the algorithm ‘CLEAN’,” *Astron. Astrophys.* **89**, 377–378 (1980).

- M.M. Colavita, J.K. Wallace, B.E. Hines, Y. Gursel, F. Malbet, D.L. Palmer, X.P. Pan, M. Shao, J.W. Yu, A.F. Boden, P.J. Dumont, J. Gubler, C.D. Koresko, S.R. Kulkarni, B.F. Lane, D.W. Mobley, and G.T. Van Belle, “The Palomar Testbed Interferometer,” *Astrophys. J.* **510**, 505–521 (1999).
- T.J. Cornwell, “A method of stabilizing the clean algorithm,” *Astron. Astrophys.* **121**, 281–285 (1983).
- T.J. Cornwell and P.N. Wilkinson, “A new method for making maps with unstable radio interferometers,” *Mon. Not. R. Astron. Soc.* **196**, 1067–1086 (1981).
- P. Cruzalebes, E. Tessier, B. Lopez, A. Eckart, and D. Tiphene, “Diffraction limited near-infrared imaging of the Red Rectangle by bispectral analysis,” *Astron. Astrophys. Sup. Ser.* **116**, 597–610 (1996).
- D.T. Gavel, H.W. Friedman, and S.S. Olivier, “Wide-baseline optical interferometry with laser guide stars,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 793–799 (1998).
- S.F. Gull and J. Skilling, “The maximum entropy method,” in *Indirect Imaging*, J.A. Roberts, ed., (Cambridge University Press: Cambridge, UK, 1984), pp. 267–279.
- C.A. Haniff, D.F. Buscher, J.C. Christou, and S.T. Ridgway, “Synthetic aperture imaging at infrared wavelengths,” *Mon. Not. R. Astron. Soc.* **241**, 51P–56P (1989).
- C.A. Haniff, C.D. Mackay, D.J. Titterton, D. Sivia, and J.E. Baldwin, “The first images from optical aperture synthesis,” *Nature* **328**, 694–696 (1987).
- C.A. Haniff and R.W. Wilson, “Closure-phase imaging with partial adaptive correction,” *Pub. Astron. Soc. Pac.* **106**, 1003–1014 (1994).
- K.-H. Hofmann and G. Weigelt, “Iterative image reconstruction from the bispectrum,” *Astron. Astrophys.* **278**, 328–339 (1993).
- J.A. Högbom, “Aperture synthesis with a non-regular distribution of interferometer baselines,” *Astron. Astrophys. Sup. Ser.* **15**, 417–426 (1974).
- R.C. Jennison, “A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent,” *Mon. Not. R. Astron. Soc.* **118**, 276–284 (1958).
- R.C. Jennison, “The Michelson stellar interferometer: a phase sensitive variation of the optical instrument,” *Proc. Phys. Soc.* **78**, 596–599 (1961).
- S. Katagiri, K.-I. Morita, N. Kawaguchi, and M. Hayakawa, “An imaging algorithm using the bispectrum in radio interferometry,” *Pub. Astron. Soc. Japan* **49**, 123–129 (1997).
- C. Koresko, *The Early Evolution of Binary Stars*, Ph.D. thesis, Cornell University (1993).
- S.R. Kulkarni, “Self-noise in interferometers – radio and infrared,” *Astron. J.* **98**, 1112–1130 (1989).
- A. Lannes, “Remarkable algebraic structures of phase-closure imaging and their algorithmic implications in aperture synthesis,” *J. Opt. Soc. Am. A* **7**, 500–512 (1990).
- A. Lannes, “Imaging capabilities of weak-phase interferometric devices,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 996–1003 (1998a).

- A. Lannes, “Weak-phase imaging in optical interferometry,” *J. Opt. Soc. Am. A* **15**, 811–824 (1998b).
- A. Lannes, E. Anterrieu, and P. Marechal, “CLEAN and WIPE,” *Astron. Astrophys. Sup. Ser.* **123**, 183–198 (1997).
- J.D. Monnier, P.G. Tuthill, B. Lopez, P. Cruzalebes, W.C. Danchi, and C.A. Haniff, “The last gasps of VY CMa: Aperture synthesis and adaptive optics imaging,” *Astrophys. J.* **512**, 351–361 (1999).
- J.D. Monnier, *Infrared Interferometry and Spectroscopy of Circumstellar Envelopes*, Ph.D. thesis, University of California at Berkeley (1999).
- R. Narayan and R. Nityananda, “Maximum entropy image restoration in astronomy,” *Ann. Rev. Astron. Astrop.* **24**, 127–170 (1986).
- E. Pehlemann, K.-H. Hofmann, and G. Weigelt, “Photon bias compensation in triple correlation imaging and observation of R 136,” *Astron. Astrophys.* **256**, 701–714 (1992).
- R.A. Perley, F.R. Schwab, and A.H. Bridle, in *Synthesis Imaging: Course Notes from an NRAO Summer School*, NRAO Workshop **13** (Green Bank, WV: National Radio Astronomy Observatory, 1986).
- R.K. Pina and R.C. Puetter, “Incorporation of spatial information in Bayesian image reconstruction: The maximum residual likelihood criterion,” *Pub. Astron. Soc. Pac.* **104**, 1096–1103 (1992).
- R.K. Pina and R.C. Puetter, “Bayesian image reconstruction - the pixion and optimal image modeling,” *Pub. Astron. Soc. Pac.* **105**, 630–637 (1993).
- A.C.S. Readhead and P.N. Wilkinson, “The mapping of compact radio sources from VLBI data,” *Astrophys. J.* **223**, 25–36 (1978).
- A.C.S. Readhead, T.S. Nakajima, T.J. Pearson, G. Neugebauer, J.B. Oke, and W.L.W. Sargent, “Diffraction-limited imaging with ground-based optical telescopes,” *Astron. J.* **95**, 1278–1296 (1988).
- F. Roddier, “Triple correlation as a phase closure technique,” *Opt. Comm.* **60**, 145–148 (1986).
- D.H. Rogstad, “A technique for measuring visibility phase with an optical interferometer in the presence of atmospheric seeing,” *Appl. Opt.* **7**, 585–588 (1968).
- F.R. Schwab, “Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry,” *Astron. J.* **89**, 1076–1081 (1984).
- M. Shao and M.M. Colavita, “Potential of long-baseline infrared interferometry for narrow-angle astrometry,” *Astron. Astrophys.* **262**, 353–358 (1992).
- D. Sivia, *Phase Extension Methods*, Ph.D. thesis, Cambridge University, UK (1987).
- J. Skilling and R.K. Bryan, “Maximum entropy image reconstruction: General algorithm,” *Mon. Not. R. Astron. Soc.* **211**, 111–124 (1984).
- Jean-Luc Starck, Albert Bijaoui, Bruno Lopez, and Christian Perrier, “Image reconstruction by the wavelet transform applied to aperture synthesis,” *Astron. Astrophys.* **283**, 349–360 (1994).

- A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., *Interferometry and synthesis in radio astronomy* (New York: Wiley-Interscience, 1986).
- P. G. Tuthill, *Imaging Stars Through the Atmosphere*, Ph.D. thesis, University of Cambridge, England (1994).
- P. G. Tuthill, J. D. Monnier, and W. C. Danchi, "Imaging of stellar disks and mass-loss envelopes in evolved stars," in *Asymptotic Giant Branch Stars*, A. Lebre, T. Le Bertre and C. Waelkens, eds., IAU Symposia **191**, 331–336 (1998).
- P. G. Tuthill, J. D. Monnier, and W. C. Danchi, "A dusty pinwheel nebula around the massive star WR 104." *Nature* **398**, 487–489 (1999).
- P. G. Tuthill, J. D. Monnier, W. C. Danchi, E. Wishnow, and C. A. Haniff, "Michelson interferometry with the Keck-I telescope," *Pub. Astron. Soc. Pac.* **112**, 555–565 (2000).
- F. Vakili, D. Mourard, D. Bonneau, F. Morand, and P. Stee, "Subtle structures in the wind of P Cygni." *Astron. Astrophys.* **323**, 183–188 (1997).
- G. Weigelt, Y. Balega, T. Bloeker, A. J. Fleischer, R. Osterbart, and J. M. Winters, "76 mas speckle-masking interferometry of IRC+10216 with the SAO 6m telescope: Evidence for a clumpy shell structure," *Astron. Astrophys.* **333**, L51–L54 (1998).
- G. P. Weigelt, "Modified astronomical speckle interferometry 'speckle masking'," *Opt. Comm.* **21**, 55–59 (1977).
- M. C. Wiedner, *Atmospheric water vapour and astronomical millimetre interferometry*, Ph.D. thesis, University of Cambridge, United Kingdom (1998).
- G. Woan and P.J. Duffett-Smith, "Determination of closure phase in noisy conditions," *Astron. Astrophys.* **198**, 375–378 (1988).
- A. Yahil, private communication (1999).

Chapter 14

Interferometer Design for Synthesis Imaging

DAVID MOZURKEWICH

NAVAL RESEARCH LABORATORY
WASHINGTON, DC

This Chapter reviews the problem of synthesis imaging at optical and infrared wavelengths, how to design interferometric arrays, and how to build the beam combination optics.

14.1 Good Fringes vs Good Science

The present generation of interferometers has a sensitivity limitation that restricts us to observing bright stars. In the future, the Keck interferometer and other arrays may exceed this limit, but until then we are restricted to observing bright stars.

Since we are using instruments that can only observe stars, we should admit up front that we are observing stars and we must assert with great confidence that stars are interesting. For this to be true, we need to make images of stellar surfaces. With these images we can see spots form and evolve and we can watch the stars rotate. Perhaps we will see the spots form at preferred latitudes and perhaps those preferred latitudes will change in a cyclic way as they do on our Sun. Stars are interesting with these kinds of observations, and so we need to make images.

A large portion of this talk will cover imaging strategies. That may seem like an odd topic. After all radio astronomers taught us how to make images 40 years ago. But their images are at longer wavelengths. For visible light observations, they only solved half the problem. To see what is missing, look at Figure 14.1. It shows the visibility amplitude

versus projected baseline length for two stars, one with a featureless uniform disk and one that is almost as featureless but with limb-darkening. What is most obvious is that the two visibility curves are nearly identical on short baselines.

The situation is similar in Figure 14.2, which shows one of the interesting stars I want to image along with one of the boring stars from Figure 14.1. The interesting star has limb-darkening *and* a spot. Again, on short baselines the curves are very similar. If I were to change the diameter of this spotted star a little bit, the position of the first minimum would shift and we would have virtually indistinguishable curves out to the first zero. On longer baselines there is a factor of two (or more) difference in the visibility amplitude between the stars, but these amplitudes are low—less than 0.1.

To summarize, if we want to make high dynamic range images of stars, or even separate stars with any surface structure from those without, we have to measure low visibility fringes.

Fringes are easiest to detect on short baselines, but the interesting science requires the long baselines. It can be extremely difficult even to find fringes on long baselines. Another complication is that we have to track the fringes. The atmosphere moves the fringes around by a lot more than a few wavelengths, and on the long baselines where fringes are weak there is not enough signal-to-noise in a short integration* time to find them and position the delay lines properly. A long integration time does not work because the fringes are washed out by the atmosphere.

There is the problem. Stated somewhat differently, you have a choice of either recording useful data *or* looking at interesting sources. This is not the kind of choice I like to make.

The existing optical interferometers do not have a lot of dynamic range and their sensitivity drops dramatically with fringe visibility. The signal-to-noise is a monotonically increasing function of NV^2 , where N is the number of photons detected in an integration and V is the fringe visibility amplitude. To observe a star on a baseline where its visibility amplitude is 0.1 requires 100 times as much sensitivity as an observation of the same star on a baseline where it is unresolved. The low visibility amplitude observation needs 100 times as many photons in an integration to give the same signal-to-noise. This is a loss of 5 magnitudes, and 5 magnitudes is a lot of sensitivity to lose when the limiting magnitude for unresolved stars is in the range of 5 to 8. We absolutely need have a scheme to get around this sensitivity limit if we are going make useful images. Consider the following five possibilities.

*In this Chapter, I use integration to refer to data that goes into a single estimate of the fringe parameters. An observation is a collection of integrations that are combined to increase the signal-to-noise to a useful level.

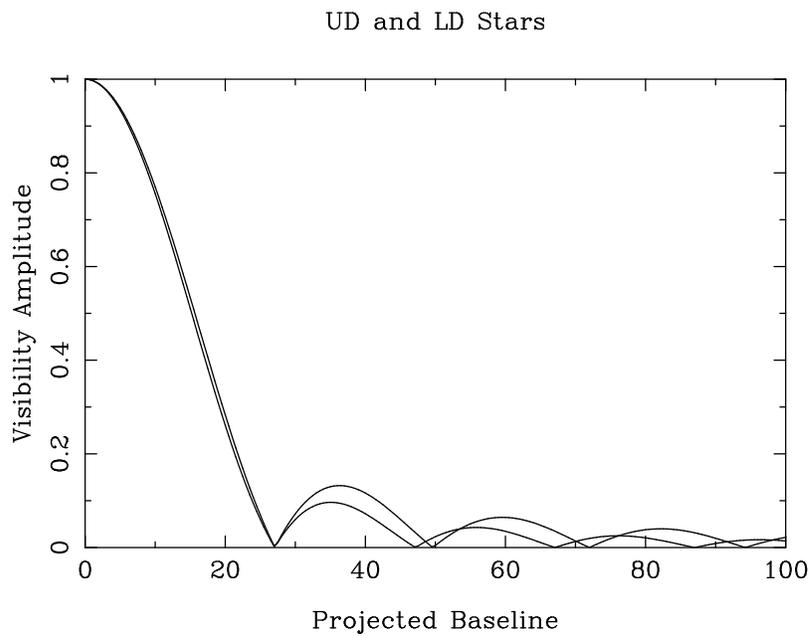


Figure 14.1: Visibility amplitude as a function of projected baseline for uniform and limb-darkened stars.

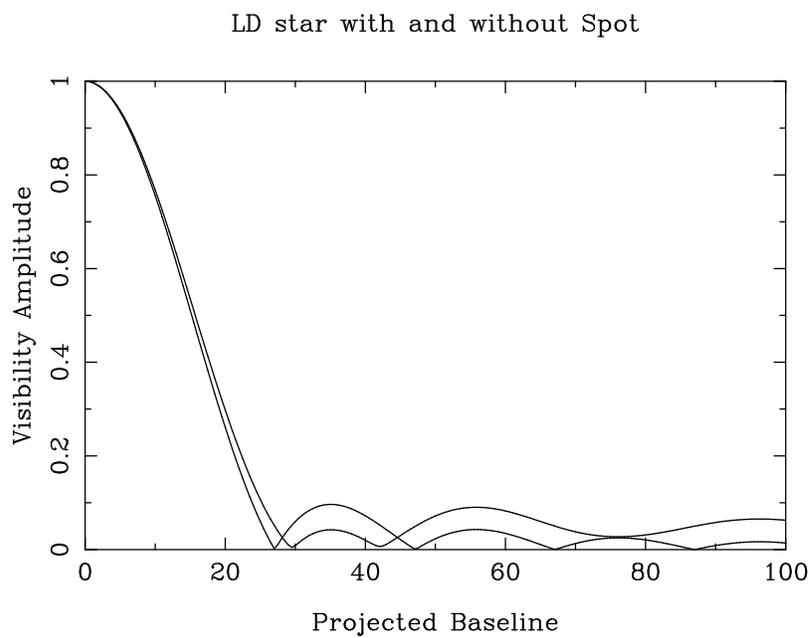


Figure 14.2: Visibility amplitude as a function of projected baseline for the same diameter: a limb-darkened star with and without a central spot.

14.1.1 Alternative 1: Integrate Forever

The simplest scheme to try is to make up for the lower signal-to-noise by observing longer. This does not work. A star with a visibility amplitude of 0.1 requires 100 times as many photons per integration. But we cannot integrate 100 times as long to get those photons because the maximum integration time is limited by phase fluctuations in the atmosphere. We have to use the data from short integrations to estimate the visibility amplitude (or rather the square of the amplitude). This will be a noisy estimate. The signal-to-noise can be improved by averaging a large number of estimates, and will increase by the square root of the number of integrations. The signal-to-noise ratio is given by

$$\text{SNR} = \frac{V^2}{\sigma(V^2)} = \frac{2}{\pi^2} \frac{NV^2}{\sqrt{1 + \frac{4}{\pi^2}NV^2}}. \quad (14.1)$$

This is a monotonically increasing function of NV^2 . In the high signal-to-noise regime, $NV^2 > 3$ per integration, the required observing time increases as V^{-2} . A star whose visibility amplitude is 0.1 requires 100 times the integration time. This converts tens of seconds into twenties of minutes. A reasonable observation time. But most of the objects we want to image will be in the photon-starved regime, $NV^2 < 2$ per integration. Here, the integration time increases as V^{-4} ; a 1-second observation is converted into 3 hours. However bad this may seem, the actual situation will be worse, since this is an upper bound to how well we can do. In the photon-starved regime, we cannot determine the phase of the fringe—or even if it is present—during the atmospheric coherence time. Hence, we cannot fringe-track and we have to blindly point the delay lines. We are required to build an interferometer that allows us to control the delay *open loop* to better than the coherence length of the starlight. But the atmosphere is moving the fringe delay around by typically 50 μm peak-to-peak. This requires a coherence length longer than 50 μm and we need to restrict the fractional bandpass to something on the order of 1%. To observe a low visibility amplitude fringe, we need to restrict the bandpass and hence the number of photons detected per integration. NV^2 is reduced yet again and the single integration signal-to-noise drops even more.

Another way of blind pointing—running an interferometer without fringe tracking—is to use a broad-band channel and scan in delay across the entire region where the fringe could occur. The power spectrum of these data will show a peak whose shape is proportional to the shape of the optical bandpass and whose amplitude is proportional to N^2V^2 . In this approach, we do not lose due to a narrow bandpass, but a large fraction of the integration time is spent at delays where there are no fringes. This adds noise to the data and is little better than the narrow-band approach.[†] In addition, the broad-band visibility is harder to interpret. My conclusion is that there is no really good way to just integrate forever in the hope of seeing a fringe.

[†]A third variant is to combine these methods using a spectrograph to simultaneously record 100 channels of narrow-band data. Here, detector read noise is the limitation.

It is true that for very short baselines, say a few meters, the atmospheric fluctuations are small enough that integrating forever can produce results. However, long baselines are needed for imaging stars, so this is not an option for the problem which I am interested in.

14.1.2 Alternative 2: Sources with High Visibility

A second alternative is to restrict ourselves to looking only at sources that provide plenty of signal for fringe tracking. There are indeed sources that have high visibility at the same baselines where there's structure, as shown by the following examples:

Figure 14.3 shows a binary star. The visibility varies wildly as function of baseline length. Even if you can only observe on the baselines where the visibility amplitude is high, you can measure the period and amplitude. When you measure how fast the envelope falls off with projected baseline length, you estimate the diameter of the primary star, and if you measure how the mean amplitude drops, you also get the diameter of the secondary. A lot of information can be obtained from this type of observation. There are people, such as Christian Hummel at the US Naval Observatory, who are making a little cottage industry out of observing double stars with optical interferometry, and they are doing some great science.

Figure 14.4 shows a star with a uniformly-bright, circumstellar disk. This is a very nice source to observe. What we see here is the fringe pattern of the disk superimposed on the fringe pattern of the star. The drop in visibility amplitude at short baselines occurs when we resolve the size of the disk. The long plateau is formed by fringes from the star. The wiggles superimposed on the plateau are from the disk. Their size and location tell us about structure on the disk for spatial scales larger than the star. The star embedded in the disk has converted our problem from one of detecting low visibility fringes to one of measuring small variations in higher-amplitude fringes.

There are sources like this. At infrared wavelengths, from 2 to 10 μm , a number of young stars have disks and old stars have shells. There are enough of these sources to do a fair amount of science. However at visible wavelengths, you have only Be stars, and perhaps μ Cep, but I'm not sure there's anything else that fits into this category and is also bright enough for the current generation of interferometers.

Unfortunately, with this alternative you are letting the instrument tell you what science you are allowed to find interesting, whereas, if possible, you should be building an instrument that will let you observe the sources that interest you. For this reason, I do not like Alternative 2—even though it is a valid imaging strategy.

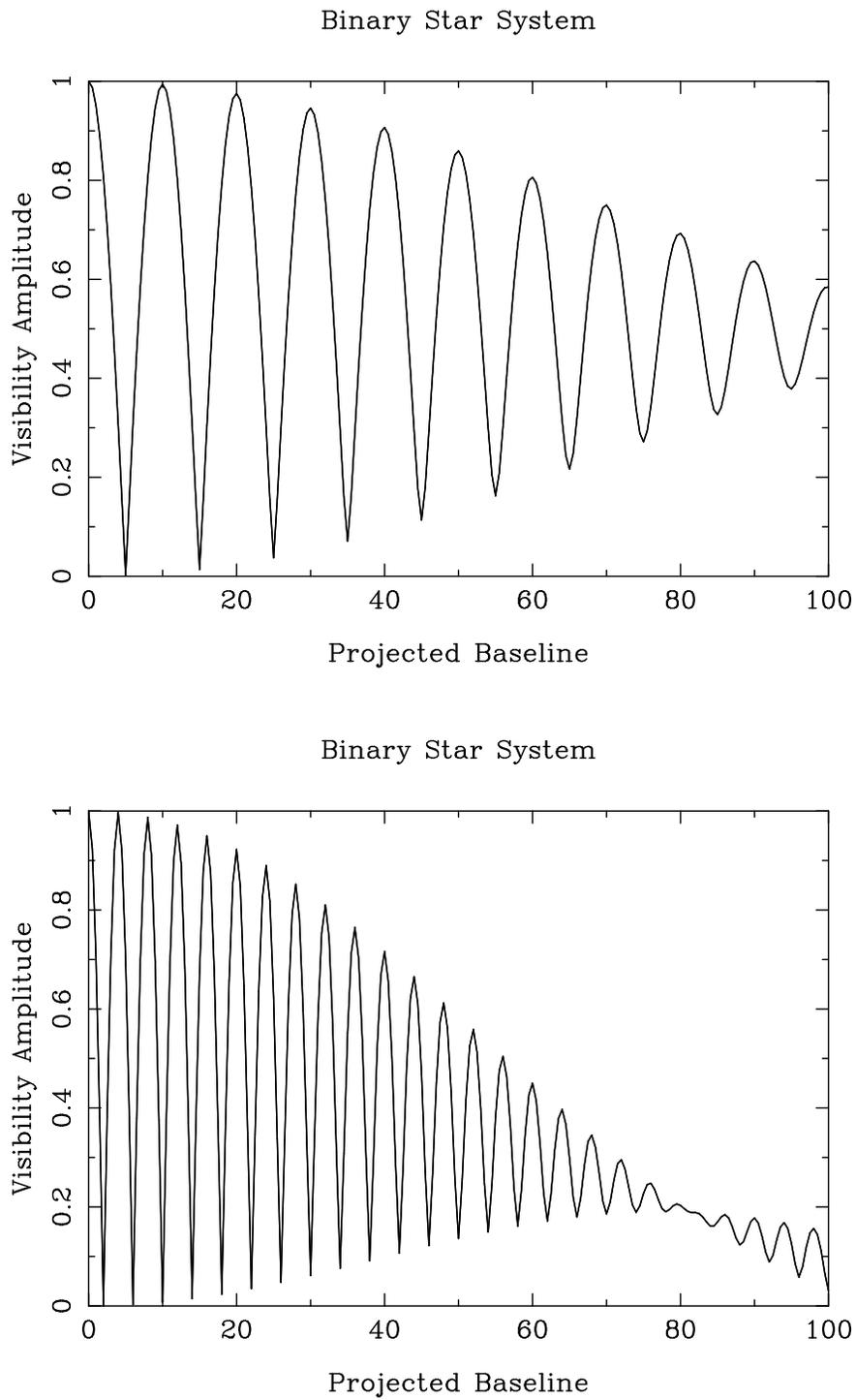


Figure 14.3: Visibility amplitude as a function of projected baseline for two binary star systems with different stellar diameters. The system shown in the upper plot has equal brightness components.

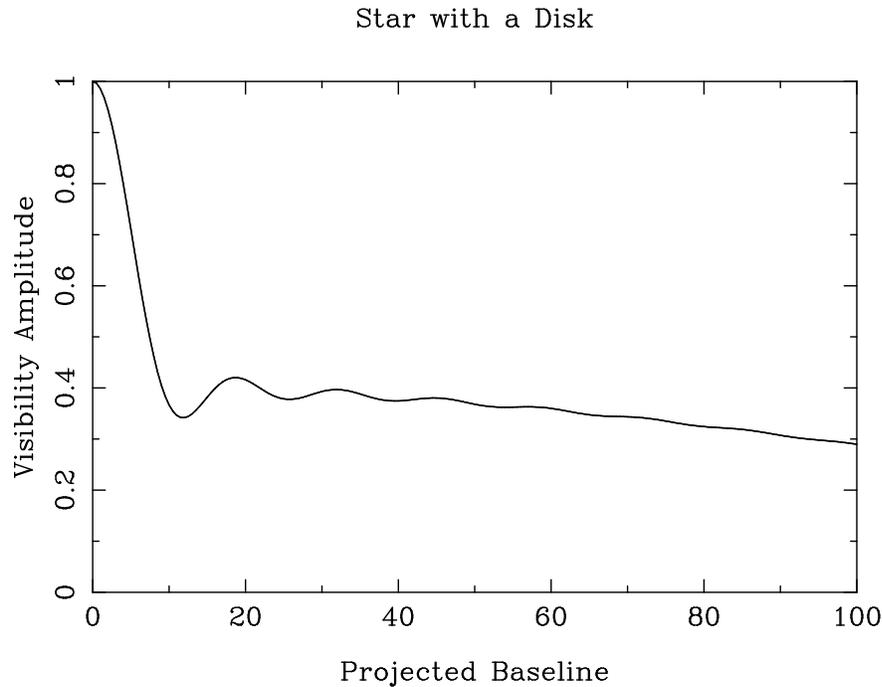


Figure 14.4: Visibility amplitude as a function of projected baseline for a compact star surrounded by an extended disk.

14.1.3 Alternative 3: Wavelength Bootstrapping

We have been looking at plots of visibility versus projected baseline. This is not the whole story. The ordinate should be spatial frequency. The spatial frequency is the projected baseline divided by the observation wavelength. So far we have only discussed what happens when we change the baseline between the telescopes. We can also change the wavelength. A broad-band (multi-wavelength) system can simultaneously span a significant range of spatial frequencies on one of these curves. The longest wavelengths will usually have the highest visibility amplitudes. The shortest wavelengths always have the most resolution. The fringe tracker uses the highest-visibility wavelengths while the science detectors record the low visibilities. This is referred to as “wavelength bootstrapping.” Figure 14.5 shows an example from the Mark III. The star is Arcturus. Fringe tracking was done in the 700–800 nm range while simultaneously, observing at 550 nm. Because the visibility amplitude is a very steep function of baseline length, there is enough visibility at the longer wavelength (i.e., shorter “baseline”) to determine the fringe position and stabilize it. There are measured squared visibility amplitudes in this figure almost as low as 10^{-4} . That’s low, corresponding to 5 magnitudes of sensitivity loss in the photon-rich regime and up to 10 magnitudes in the photon-starved regime.

The main advantage of wavelength bootstrapping is that the fringe tracking data can be used to stabilize the fringe—either in hardware or software. With this accomplished, it

is possible to overcome the atmospherically induced limit on the integration time. This allows us to push the data from the photon-starved regime into the photon-rich regime and can shorten the observation time needed by a factor of 100 or more. Note that it is not necessary for the star to be the same shape at all the wavelengths. It is only necessary for it to be compact at some wavelengths and resolved at others.

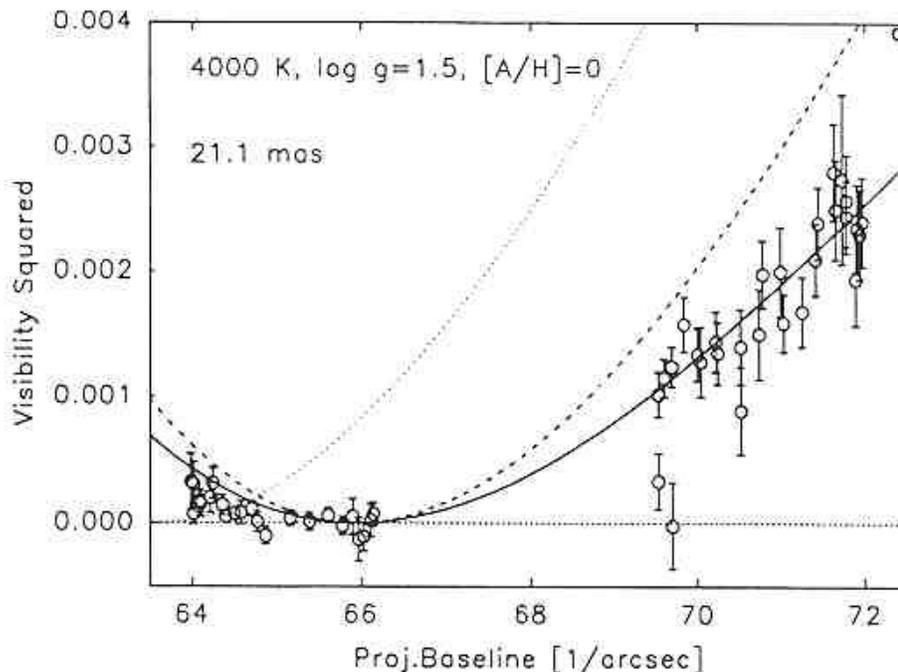


Figure 14.5: Squared visibility amplitude versus projected baseline for Arcturus showing the application of wavelength bootstrapping. This 550 nm data was taken while fringe tracking in the 700 to 800 nm band (Quirrenbach *et al.*, 1996).

Although wavelength bootstrapping is useful, it has limitations. The fringe tracking is performed at the longest wavelength, but it must be done with enough accuracy and speed to stabilize fringes at the shortest wavelength. Also, unless we have good adaptive optics, we must limit the size of the aperture to that which is allowed at the shortest wavelength since the atmosphere introduces a relative piston between two apertures that are co-located but of different diameters. These limitations on the aperture size and integration time severely limit the sensitivity of the long wavelength measurement. And even if we have enough sensitivity for the observations, there is still the problem of knowing the index of refraction of air with sufficient accuracy (due to the changing and unknown water vapor content) to predict the fringe position from a measurement at a greatly different wavelength. Because of these types of effects, it will be difficult to apply wavelength bootstrapping over more than a factor of two or three in wavelength. But a factor of two or three in wavelength corresponds to a factor of two or three in resolution and that can make a big difference in the type of science that can be done. I dislike interferometer designs where the blue light is siphoned off for use in angle tracking and adaptive optics leaving only the long wavelength

portion of the bandpass for fringe measurements. I think wavelength bootstrapping *must* be an integral part of any imaging interferometer.

14.1.4 Alternative 4: Baseline Bootstrapping

It is also possible to stabilize the fringes on a long baseline through “baseline bootstrapping.” Figure 14.6 shows a 5 telescope linear array. The observing procedure is to simultaneously fringe track on baseline A–B, and on baseline B–C, and on baseline C–D, etc. Although only five are shown, you may string together as many telescopes as you want. The change in phase on each of the short baselines is determined, and those phases are combined in various ways to obtain the change in phase on each of the long baselines.

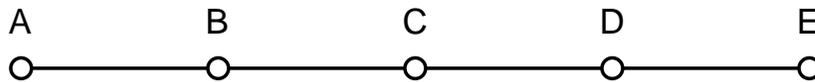


Figure 14.6: A redundant linear array useful for baseline bootstrapping.

This procedure works because of the closure-phase relationships. The change in delay due to the atmosphere and the instrument affects all the light intercepted by one telescope in the same way. Thus the phase changes induced on the path from the star through telescope “B” to the beam combiner adds exactly as much phase to baseline A–B as it subtracts from baseline B–C. Changes in the baseline phase due to source structure do not cancel out in this procedure and that is good because those are the phases we want to measure. But these phases vary slowly—at least a lot more slowly than the atmosphere. Although source phases changing with time limit the maximum integration time in principal, they do not impact the usefulness of this method.

The goal here is the same as it is with wavelength bootstrapping; estimate the phase on the longest baseline with enough accuracy to stabilize it. Then the integration time can be extended beyond the atmospheric limit, and in the process a couple factors of V^2 are removed from the required observing time. But this method is not perfect: When N short baselines are strung together to phase a long baseline, the phase noise on all the short baselines contributes to the uncertainty in the long baseline phase. The phase noise on a long baseline will be \sqrt{N} times the phase noise on a short baselines. As a result, the system visibility is lower on the longer baselines. But as long as we are aware of this effect and calibrate, it should not be a limitation.

Figure 14.7 shows an example of baseline bootstrapping data from three baselines of the NPOI. The squared visibility amplitudes for the two short baselines are shown in the upper panel. The long baseline data is shown in the lower panel. The signal-to-noise on the longest baseline is too low for the fringes to be detected or tracked during the observations.

Nevertheless, the data was obtained because the fringes were detected and tracked on the two short baselines. This technique is just starting to be exploited and is a very good way to make an imaging interferometer work.

14.1.5 Alternative 5: Guide Star Methods

The fifth scheme is the use of a guide star. If the fringes are too faint to allow fringe tracking on the star itself, you need a method to estimate the delay, or at least the change in delay, other than by observing the fringes on that star. The obvious technique is to use the delay measured on a nearby star. In radio interferometry, this is standard operating procedure and is used to remove phase variations induced by the ionosphere; the telescopes switch back and forth between two sources. The phase variations observed on the bright source are applied to the faint source, allowing longer integration times.

At optical wavelengths, the atmosphere varies with a coherence time on the order of 10 ms. The two stars must be observed simultaneously since this is too fast to allow switching between the stars. Fringes are only tracked on the guide star but the delay corrections deduced from that tracking are applied to both stars. The phase, however, decorrelates with increasing angle as well as time. At visible wavelengths, the maximum allowed angle at the best astronomical sites is less than ten arc-seconds. 10 arcsec regions surrounding the brightest stars only covers a very small fraction of the sky. This does not make for a practical instrument.

In the infrared the atmosphere is more benign; the pathlength errors it introduces are a smaller fraction of a wavelength and do not hurt the observations as much as they do at shorter wavelengths. This technique is potentially very exciting for observations at 10 μm . At shorter wavelengths around 2 μm the system visibility amplitudes will be low, but the technique will certainly produce some good science. Calibration at these wavelengths is going to be interesting: The visibility amplitude has already been significantly reduced by decorrelation between the two paths, and so small changes in the seeing will cause small changes in that decorrelation which can result in large changes in the calibration. The system visibility amplitude may be such a strong function of the seeing that it will be difficult or impossible to calibrate the observations. Recently, I heard a rumor that measurements from the Palomar Testbed Interferometer suggest I am wrong. I hope the rumour is true, and I am anxiously awaiting the publication of these results.

I should point out that guide star methods are probably restricted to natural guide stars. Laser guide stars that light up the atmosphere are intrinsically incoherent, and as a result, you can use them to correct the wavefront flatness but not absolute phase.

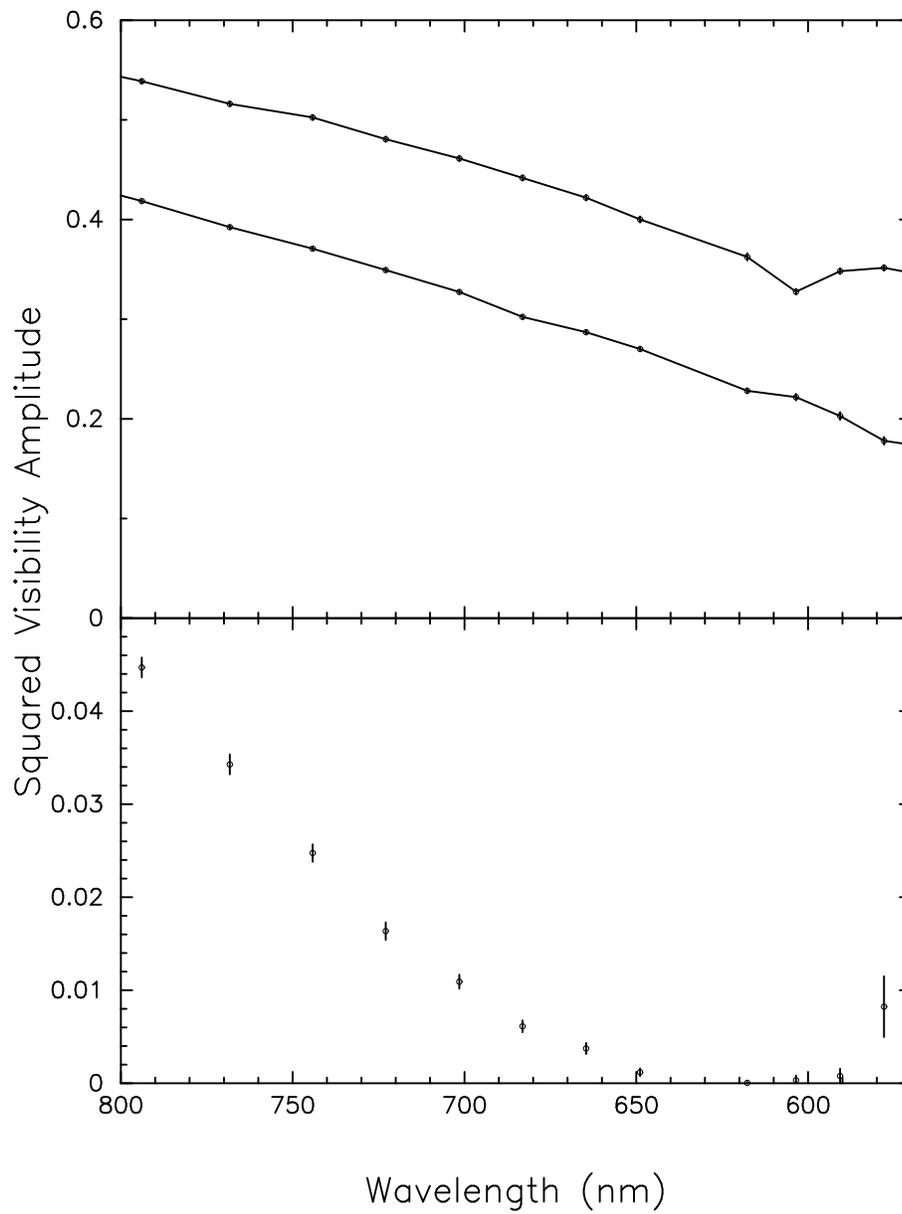


Figure 14.7: Fringe data as a function of wavelength from the NPOI. The data was taken simultaneously on all three baselines. The longest baseline (lower curve) was only obtained because the fringes were stabilized on that baseline using the two shorter baselines (upper curves).

14.2 Design of Interferometric Arrays

14.2.1 Optimal Design of Interferometric Arrays

A lot of work has gone into how to design arrays and every few years someone will come by and say “This is the right way to do it.” There are plenty of examples. I am not going to spend much time discussing them because most of this work is not relevant to our problem.

- **VLA Y.** The VLA was laid out in a Y-configuration with a power law spacing of the elements along the arms. This gives good images using Earth-rotation synthesis and is often quoted as the correct way to layout an array (see for example Mathur 1969 and Chow 1972).
- **Random Arrays.** When the telescopes are arranged in a geometric pattern such as a **Y** or a circle, some of that pattern bleeds into the (u, v) coverage with the sample points falling along lines and arcs. This patterning can be removed by laying out an array with a more random distribution. The actual configurations are determined by trial and error or exhaustive searches and give more uniform (u, v) distributions that lack the patterning seen in more conventional arrays.
- **Special Purpose Arrays.** There is an interesting masters thesis from MIT recently (see Kong *et al.* 1998) where the author looked at how to arrange the apertures to create regions in the image close to the central peak that are devoid of sidelobes. His application was to design an array that is optimized for a planet search. Other optimizations should be possible for whatever applications are of interest.

There are a lot of possibilities (Mathur, 1969; Golay, 1970; Chow, 1972; Cornwell, 1988; Keto, 1997). I doubt very much if the “optimum” designs are truly applicable to *synthesis imaging at optical wavelengths from the ground with Michelson interferometers*. This is because, although we can argue about the existence of a *best* array configuration, there already exist a lot of *pretty good* configurations. And it turns out the pretty good configurations are not that much worse than the best configuration, whatever that is. As a result, practical considerations will win out over the optimization processes involved in these array designs.

14.2.2 Partially Redundant Arrays

In my view, imaging requires redundant arrays. If we are going to image fine-scale structure on the surface of stars, we need to measure visibilities on baselines where the fringe visibility is almost unmeasurable. This requires baseline bootstrapping. The best arrays for baseline bootstrapping are redundant arrays. Arrays that are redundant enough to allow baseline bootstrapping can have remarkably good (u, v) coverage. I say partially redundant because a completely redundant array has all the elements evenly spaced along a line. You do not want a completely redundant array.

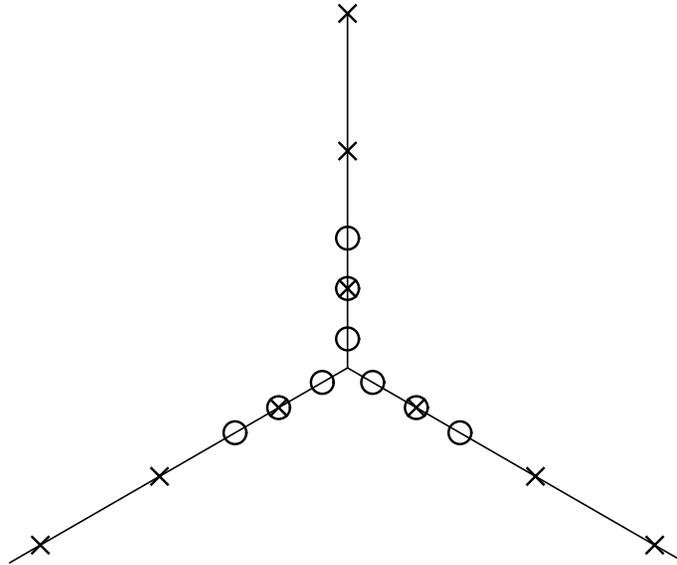


Figure 14.8: Layout of the NPOI showing its partially redundant configuration.

14.2.3 Array Design with Vacuum Feed Systems

It seems obvious to me that imaging at optical wavelengths requires a vacuum feed system. The star light has to be transferred from the telescopes to the optics lab and there is just too much air along that path. That path is horizontal and close to the ground. There is always turbulence along the ground. If the path is not evacuated, there will be problems with local seeing along that path. If the path is not evacuated, there will be problems with dispersion because the telescopes are not all at the same distance from the lab and after adding an air delay there will be more air in one path than in the others. There can be problems with refraction because simple measures to control seeing will cause the air to stratify. Just to get rid of these problems the feed system should be in a vacuum pipe. If the feed system is in a vacuum, then a random array, or even a circle, will have pipes strewn all over the site. You will want to collect the pipes together into lines or a Y for convenience. This is particularly true when you realize the telescopes have to be moved to match the resolution of the array to the size of the source being imaged. It would be nice to reuse the same pipes and telescope stations for different array layouts. This leaves us with a partially redundant Y. For purely practical reasons this seems to be the only reasonable choice and is probably the way we are going to build arrays.

14.2.4 NPOI Configuration

The NPOI is a good example of what I mean by a partially redundant array. The layout of the inner part of the NPOI is shown in Figure 14.8. Two array configurations are shown; one marked with circles, the other with an \times . In the complete system, all stations are part of two configurations. The nearest neighbors within a configuration are all equally spaced

allowing baseline bootstrapping. For a six-telescope configuration with three telescopes per arm, the longest baseline is $(2\sqrt{3} + 1) = 4.464$ times longer than the short baselines. The six-element completely redundant (linear) array has a maximum baseline five times the shortest, roughly 10% longer than the NPOI design. A non-redundant six-element array consists of 15 unique baselines. In the partially redundant NPOI design only two baselines are redundant, 13 are unique. By contrast, the fully redundant, linear array has only five unique baselines. This partially redundant type of design seems to be an excellent compromise between (u, v) plane coverage and baseline bootstrapping.

The (u, v) plane coverage obtained from this six-telescope design is asymmetric with highest resolution perpendicular to the unpopulated baseline. For best imaging, we need to observe the same target three times, once with each arm unpopulated. Or we need to have nine telescopes.

14.2.5 Polarization and Beam Rotation

When designing an interferometer, you must remember that the optics are not idealized entities. It is important to specify, or at least understand, all optical properties of each element in the system. Arguably the hardest part of building an interferometer is providing specifications for all the components. For example, consider the simplest optical component in the system, a mirror. For building a telescope, all we have to do is know the surface quality and reflectivity.

For an interferometer, the specifications of a mirror are more complicated. A root-mean-squared surface quality is not enough since the micro-roughness can dominate the far-field reflectivity. The reflectivity is usually given for unpolarized light at normal incidence. This is good enough for a telescope where most of the reflections are near normal incidence, but for an interferometer, the polarization dependence is important. Using the unpolarized reflectivity will provide estimates of the system throughput that can either greatly underestimate or overestimate performance. The sign of the error depends on the orientation of the mirrors in the optical train.

Mirror coatings not only absorb and polarize the light, they also act as waveplates. The orientation of the waveplate depends on the orientation of the plane of reflection. The retardance of the waveplate depends on the type of coating and on the angle of incidence. Data can be frightfully difficult to come by. It is generally assumed that dielectric coatings are worse than metal coatings. Silver is considerably better than aluminum.

The effect of these waveplates is to produce an optical path length through the system which depends on the orientation of the electric-field vector. If this polarization-dependent phase shift is different in the optical trains from the different telescopes to the beam combiner, the result is a loss of visibility. And the effect is large; even for silver coatings the phase shift can build up to a radian after only three or four reflections. Fortunately, the visibility loss depends on the difference in phase shift between the two beams being interfered. The

resulting design rule is to insist that the optical trains from each telescope to the beam combiner consist of the same number of reflections, each with the same angle of incidence and orientation and each with the same coating. If the coatings cannot be the same—for example with a reflection off a single beam splitter—that angle of incidence should be made as small as possible.

There is another reason to make the optical paths identical: beam rotation. Consider a 100% linearly polarized source and a feed system made from perfect mirrors. The light from each telescope at the beam combiner will also be linearly polarized. If the polarization vectors are not parallel, the visibility amplitude will be decreased by the cosine of that misalignment angle. Although it is harder to visualize, the same visibility loss occurs for unpolarized light.

The visibility loss due to the polarization-dependent phase shift can be solved by inserting a polarizer in the beam immediately before the detector. Because this *throws away half the light*, it may seem like a bad idea. But we need star light for different tasks—angle tracking and perhaps adaptive optics, fringe tracking, and science. Dividing the light by polarization for these different tasks may be an efficient way to handle the phase shift problems. However, the visibility loss due to beam rotation cannot be solved with a single polarizer.

14.3 Beam Combination and Modulation

Now we need to talk about beam combination and modulation. A single-baseline interferometer needs only to measure visibility amplitude, because the usefulness of a baseline phase is destroyed by the atmosphere. However, closure phases are needed to make an image and closure phases require three telescopes. To get the data needed for imaging, a beam combiner is needed that combines light from three or more telescopes simultaneously.

Given a number of telescopes, we need some means of combining the light from those telescopes and determining whether or not a fringe for each baseline is present in the data. If some or all of the fringes are missing, we need to move the delay lines to find them. Once a fringe is present, we have to calculate the amplitude and phase of that fringe. We will now discuss fringe-detection schemes and beam-combination schemes—topics that are closely related to each other.

14.3.1 Pupil-Plane Combination: Passive Detection

Tango and Twiss (1980) described what I will call *passive* detection. A simple two-way beam combiner for this method is illustrated in Figure 14.9. Light from two telescopes enter the beam combiner from the two inputs. They are combined at the surface of the beam splitter. The two combined beams are collected by two detectors. We label the intensity of the light detected by the two detectors A and B . If there is no fringe present

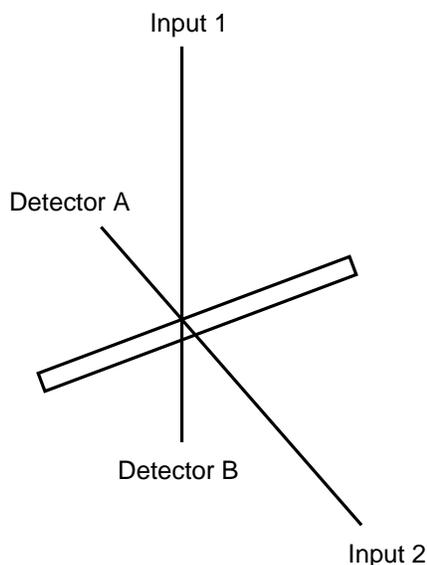


Figure 14.9: Beam combiner with two inputs and two outputs.

then $(A - B)$ will average to zero and have the same noise as $(A + B)$. If a fringe is present, then one side of the beam splitter will have a higher signal than the other side. Which side is higher depends on the phase of the fringe. As the fringe-phase varies, the fluctuations in $(A - B)$ will increase as the visibility amplitude increases but the fluctuations on $(A + B)$ will remain unchanged. The visibility amplitude can be determined by comparing the statistics of the sum and difference signals:

$$\langle V^2 \rangle = 2 \frac{\langle (A - B)^2 \rangle}{(\langle A \rangle + \langle B \rangle)^2}. \quad (14.2)$$

This is a very simple technique for measuring a fringe amplitude. A hidden assumption is that there has to be enough fringe motion during the observation for the fringe phase to be uniformly distributed. This should not be a problem most of the time but may become an issue for short observations, for high-precision measurements, or if a fringe tracker is used to partially stabilize the fringe packet. This is not a particularly useful method for imaging since phase information is discarded.

14.3.2 Image-Plane Combination: Spatial Modulation

As was discussed by Traub in Chapter 3, an image-plane beam combiner takes the beams of light from each of the telescopes and first lines them up so that they are parallel to each other, but separated by different spacings. This set of beams is passed through a single lens to form an image. For an unresolved star it is an Airy disk with fringes superimposed on it. The size of the Airy disk corresponds to the diameter of a single beam. The fringe frequencies depend on the spacing of beams. The two beams closest together will have the

widest fringes. The two beams furthest apart will have the highest frequency fringes. The phase of the fringe is the offset of the fringe peak from the center of the Airy disk. The easiest way to analyze the data is to take a Fourier transform. The transform will have a peak at each fringe frequency corresponding to a beam spacing. The amplitude and phase of the Fourier transform at a peak is the visibility amplitude and phase of the baseline formed from the two beams whose spacing at the lens corresponds to the fringe frequency. Needless to say, it is important to make sure no two beams have the same spacing.

14.3.3 Pupil-Plane Combination: Temporal Modulation

Pupil-plane beam combination brings the beams together on a beam splitter with zero beam spacing. Therefore the fringe frequency is also zero. As the phase changes, the image of the combined beam does not change shape as it does for the image-plane combination. The entire image changes brightness. Energy can still be conserved since images form on both sides of the beam splitter. As one image increases in brightness, the other becomes fainter.

To detect the fringe, the delay is modulated at a rate faster than the fringe motion induced by the atmosphere and the detectors are read out synchronously with that modulation. The signal varies sinusoidally with time. The phase of the signal relative to the phase of the modulation is the fringe phase.

Compared to image-plane detection, pupil-plane detection required fewer detectors but they have to be sampled faster.

Biases

The instrument and the atmosphere can bias the fringe measurement. The amplitude of the Fourier transform of the fringe data is not a delta function, it is a sinc function because of the finite length of the data stream. The peak occurs at the fringe frequency and it is the value at the peak that corresponds to the fringe visibility. If we do not know the exact ratio of the wavelength to the modulation length, the measured fringe amplitude will be wrong. And it will always be lower than the true value. You may think that this is not a problem; simply build an instrument with a stable modulation and measure it. It is not that simple. The effective wavelength is affected not only by the instrumental bandpass, but also by the star's spectrum and visibility variation across the bandpass. Fortunately, these effects are only important for the widest bandpass observations and the scientific interpretation of the result usually requires a narrower bandpass.

The atmosphere is a more important limitation since it is moving the fringe while the instrument is modulating the delay. As a result, the actual fringe frequency is a little higher or lower than the instrumental fringe frequency. Near the peak, the sinc function is quadratic. The error increases as the square of the atmospheric motion. This change

in fringe frequency is one of the few atmospheric calculations that is relatively easy. The atmospheric coherence time, t_0 , is defined as the amount of time it takes the atmosphere to change the phase of the fringe by 1 radian. Therefore, the change in fringe frequency is

$$\delta f = \frac{1}{2\pi t_0} \quad (14.3)$$

On average, the fringe frequency differs from the instrumental fringe frequency by δf . The visibility amplitude error depends on the fractional fringe modulation error and can be decreased by increasing the modulation frequency. For typical atmosphere and observations in the visible, the modulation frequency should be on the order of 500 Hz for 1 percent measurements.

14.3.4 Demodulating Multiple Baselines

When light from multiple telescopes is combined on a single detector, the signals corresponding to the different baselines have to be separated. This is best done by imposing a different modulation on each baseline. The delay modulation has to be imposed on the light from each telescope before combination. Interpretation of the data is easiest and the resulting signal-to-noise is highest if the modulation uses a piecewise linear waveform: linear so that the modulation on the detector varies sinusoidally and can be analyzed using Fourier transforms, and piecewise so that the maximum delay can remain bounded. All of the delay modulations should be of the same frequency but of different amplitudes. The fringe frequency on a baseline is the difference in the modulation *amplitudes* of the two stations forming that baseline. If the modulation time is T and the difference in modulation amplitudes between the two stations forming the baseline is k wavelengths, then the fringe frequency is $f = k/T$. The modulation amplitudes must be chosen so that all of the fringe frequencies are different.

An example is shown in Table 14.1. This example uses six stations. There are 15 baselines and with modulation amplitudes of 0, 1, 4, 10, 12, and 17. There are no duplicate fringe frequencies. The fringe frequencies range from 1 to 17 with 14 and 15 missing. For more than four stations, it is always necessary to leave some intermediate frequencies unpopulated. These unpopulated frequencies can be useful for calibration.

The data from each linear modulation segment is Fourier transformed. The amplitude will have a peak at each modulation frequency. The amplitude and phase of that peak is the amplitude and phase of the fringe on the baseline corresponding to that fringe frequency.

Cross-Talk

An example of a multi-baseline data set can be shown in Figure 14.10. The upper panel shows the raw fringe data for a six-telescope, all-on-one combiner. It is a time series consisting of 15 sine waves of various amplitudes and phases. In this example, only the

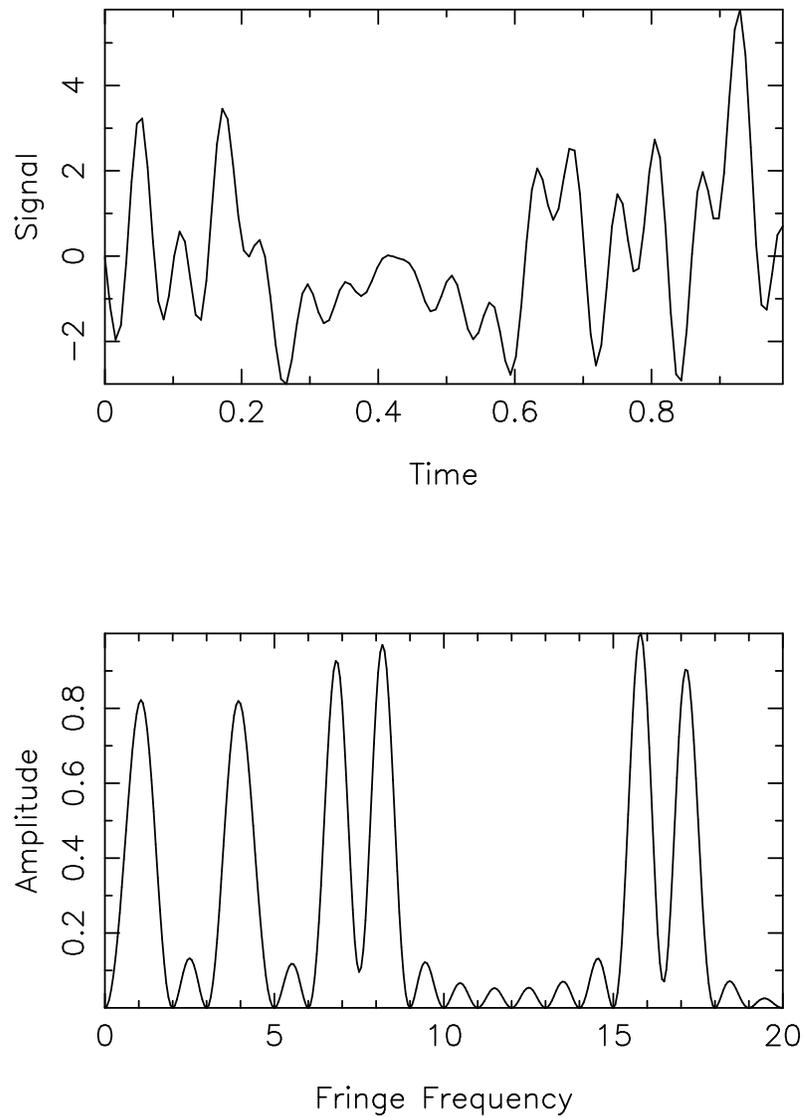


Figure 14.10: An example of multi-baseline data for six telescopes and 15 baselines. The upper panel shows the raw data as a time series. The lower panel shows the power spectrum. The height of a peak is proportional to the square of the visibility amplitude on that baseline.

Table 14.1: Demodulation schemes for multiple baselines.

Telescope Number	Modulation Amplitude	Baseline	Fringe Frequency
A	0		
B	1	A-B	1
C	4	A-C	4
D	10	A-D	10
E	12	A-E	12
F	17	A-F	17
		B-C	3
		B-D	9
		B-E	11
		B-F	16
		C-D	6
		C-E	8
		C-F	13
		D-E	2
		D-F	7
		E-F	5

six shortest baselines have high enough visibility amplitude for the fringes to be seen. The phases are random. The lower panel shows the power spectrum. The height of a peak is proportional to the square of the visibility amplitude on that baseline. Each of the peaks is a sinc function. The wiggles at frequencies between the six main peaks are sidelobes of those peaks. If the fringe frequencies are all integral multiples of the lowest fringe frequency, then each peak falls at a zero of all the other peaks and the measurements are independent. But the atmosphere moves the fringe frequencies around so we cannot maintain this condition and there is cross-talk between the baselines.

Even though this cross-talk is small, it is important. Remember that to make an image of a stellar surface, we need to measure visibility amplitudes on some baselines that are factors of 10 or more smaller than the tracking baselines. The cross-talk increases linearly with δf . As with the bias, the cross-talk can be decreased by increasing the modulation frequency, but for similar fringe parameter fidelity, it constrains the fringe frequency to be an order of magnitude higher than the constraint imposed by the bias.

One solution to this problem is to multiply the data by a tapered window function before Fourier transforming. This apodization reduced the height of the sidelobes at the price of increasing the width of the central peak. The result is a decrease in modulation frequency, but an increase in modulation amplitude—and a nearly complete elimination of cross-talk.

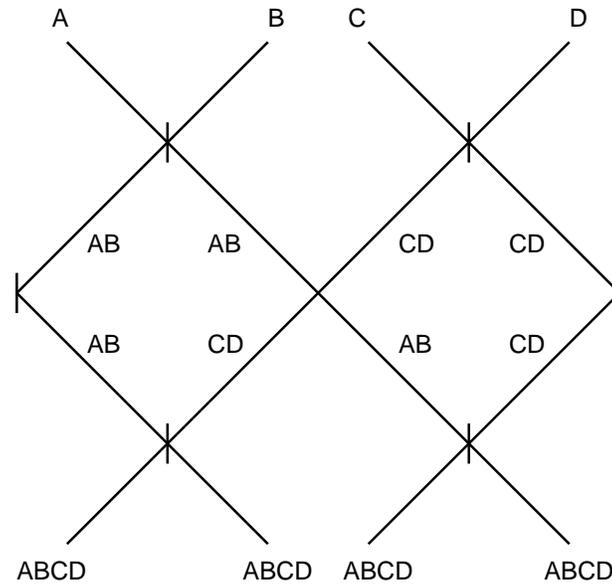


Figure 14.11: Beam combiner with four inputs and four outputs, similar to the one used at the Cambridge Optical Aperture Synthesis Telescope (COAST).

14.4 Beam Combination Techniques

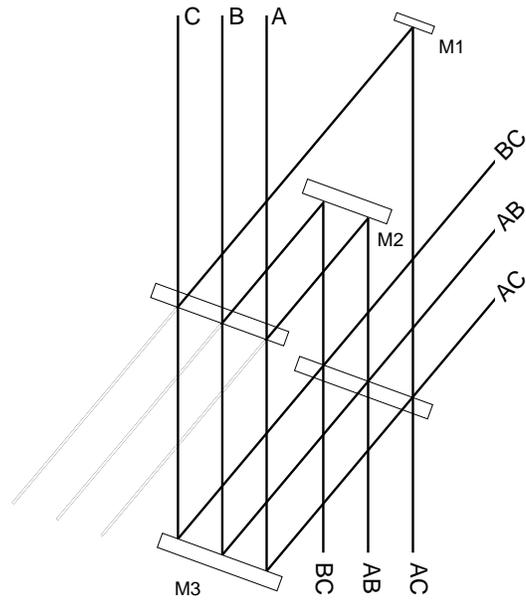
There is more to designing a beam combiner than deciding whether it will operate in the pupil plane or in the image plane. Beam topology plays an important role both in determining performance and in imposing performance requirements on the rest of the interferometer. Four possible topologies will be described and compared in this section.

14.4.1 All-On-One Combination

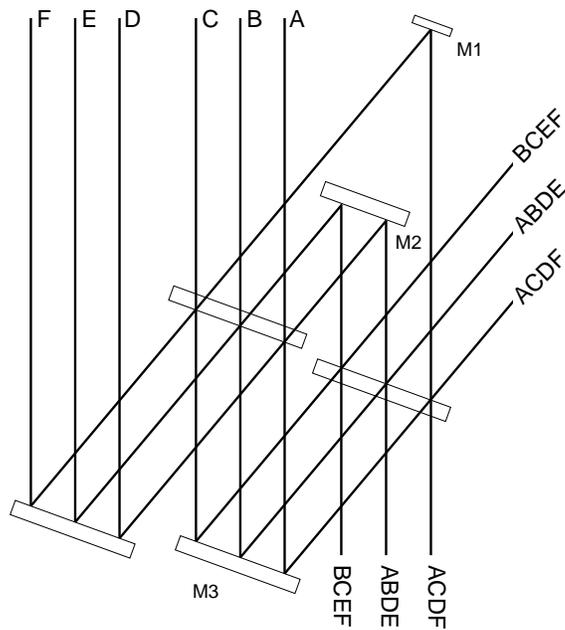
The beam combiner shown in Figure 14.11 is modeled after the one used at COAST, which is a multi-stage four-way beam combiner. Light from the telescopes is first combined in pairs. Then these pairs are recombined with other pairs. The result is that each detector sees light from *all* of the telescopes. This topology is easily extended to any power of two; combining light from 2^E telescopes requires 2^E detectors. A combiner designed for 2^E beams can be used for fewer beams by simply not using some of the inputs. It will still need 2^E detectors to collect all of the light.

14.4.2 Pairwise Combination

All of the beams do not have to be combined on a single detector. We could design a beam combiner that uses a different detector for each baseline. The example shown in the upper portion of Figure 14.12 was built at the NPOI. One advantage of this design over an all-on-



(a)



(b)

Figure 14.12: The Navy Prototype Optical Interferometer (NPOI) beam combiners. (a) a pairwise combiner for three beams. (b) An extension of the combiner in (a) for use with six beams. This hybrid design is intermediate in performance between a pairwise and an all-on-one combiner.

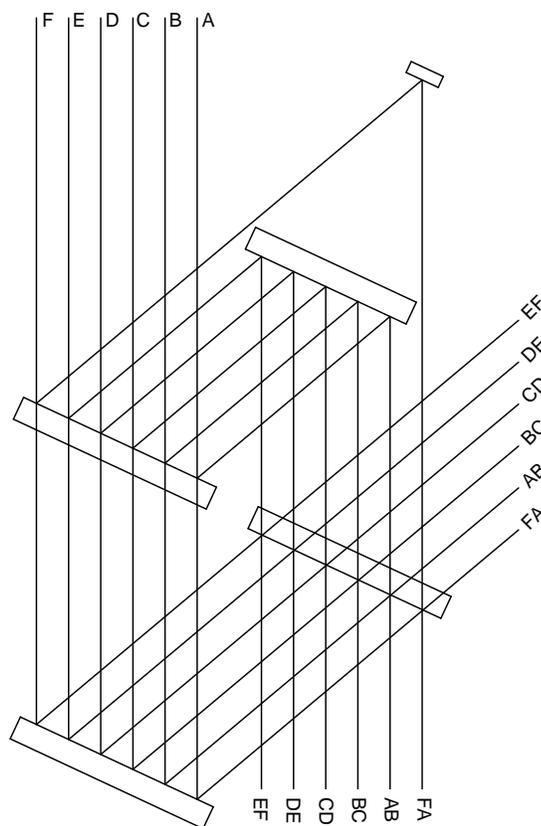


Figure 14.13: A partial pairwise beam combiner for six telescopes.

one design is that there is no chance of cross-talk. A disadvantage is that motion of any of the optics shown introduces a closure-phase error. A disadvantage of a pairwise combiner is that its complexity, and the number of detectors required, increases as the square of the number of beams, not linearly as in an all-on-one combiner.

The lower panel of Figure 14.12 shows an intermediate scheme that was adopted for six-way combination at the NPOI. The first beam splitter produces pairs which are passed through the three-way combiner to form quadruples. This topology requires fewer detectors and smaller modulation amplitudes than either pairwise or all-on-one combination.

14.4.3 Partial Pairwise Combination

A topology I am starting to like is the partial pairwise beam combiner. An example is shown in Figure 14.13 for six telescopes. There are six pairwise outputs corresponding to baselines A–B, B–C, C–D, D–E, E–F, and F–A. This is only 6 of the 15 possible baselines. As a result, this is not a particularly good science beam combiner, but it has enough functionality for a co-phasing beam combiner since fringe tracking on only five baselines is needed to stabilize the fringes on all of the baselines. Separating instrument functionality

into independent subsystems is a good thing to do from an engineering standpoint. If we are going to dedicate some of the star light to co-phasing the instrument, then this is a good topology for that job. But we need to worry since some light is already going to angle tracking. If we take some more for co-phasing, that leaves very little for science.

14.5 Comparison of Beam Combination Techniques

Many of the tradeoffs between different approaches to beam combination are difficult to quantify. For example how difficult it is to manufacture or maintain can depend on the person doing the work. The susceptibility of the data products to biases can depend on the types of observations planned for the instrument. As a result, this section ignores the truly important differences between beam combiners and focus on the one aspect easiest to quantify, sensitivity. We also ignore important details such as read noise and note that for photon-noise-limited performance, the signal-to-noise of an observation is a monotonically increasing a function of NV^2 , where N is the number of photons in a single observation and V is the observed visibility amplitude. This is reasonable in the high signal-to-noise case.

$$\frac{\text{SIGNAL}}{\text{NOISE}} = \frac{NV}{\sqrt{N}} = \sqrt{NV^2}. \quad (14.4)$$

With this background, comparing the sensitivity of two beam combiners consists of calculating NV^2 for each configuration. The highest value wins. In everything that follows, N is the number of photons that reach a detector, N_0 is the number of photons from a single aperture, V is the visibility of a fringe at the detector, V_0 is the visibility that we would see if there were only two telescopes, and E is the number of telescopes contributing to the beam combiner.

Pairwise Combination

For pairwise combination, the photons from E telescopes are distributed between $E(E-1)/2$ detectors.

$$\begin{aligned} N &= \frac{EN_0}{E(E-1)/2} \\ &= 2\frac{N_0}{E-1}. \end{aligned} \quad (14.5)$$

The detector only sees light from the two telescopes contributing to the baseline so the visibility is not reduced. Therefore

$$NV^2 = (2N_0)V_0^2 \left(\frac{1}{E-1} \right). \quad (14.6)$$

All-on-One Combination

For all-on-one combination, all of the detectors see the same signal. Since there is only photon noise in this example, the signals from all the detectors can be combined before being processed. The E telescopes contribute to a single detection, so $N = EN_0$.

The visibility is reduced because the fringe is formed between the light from two telescopes. As far as this baseline is concerned, the light from the other telescopes does not contribute to the fringe, it just forms a background against which the fringe detection has to be made. The visibility amplitude is reduced to $V = 2V_0/E$ and

$$NV^2 = (2N_0)V_0^2 \left(\frac{2}{E} \right). \quad (14.7)$$

since $2/E$ is larger than $1/(E-1)$, all-on-one combination is more sensitive than pairwise combination. Though this may seem like an odd result, it is telling us that the decrease in visibility amplitude due to adding in all those extra photons from the unused telescopes is more than offset by being able to use *all* the light from each telescope for each baseline. The gain is modest. The sensitivity advantage of an all-on-one combiner is 0.3 magnitudes for three telescopes, increasing with the number of telescopes to an asymptotic limit of 0.7 magnitudes.

CHARA: Partial-Pairwise Combination

The CHARA Array advocates the use a partial pairwise scheme (ten Brummelaar and Bagnuolo, 1994). Since, in general, this combiner will be used to co-phase an array, assume that only a fraction f of the light from each telescopes is used for beam combination. There are E telescopes and E detectors giving $N = fN_0$. Since this is a pairwise combiner, the visibility amplitude is not reduced and

$$NV^2 = (2N_0)V_0^2 \left(\frac{f}{2} \right). \quad (14.8)$$

The major advantage of this approach is now apparent; the sensitivity does not drop as more telescopes are added. The disadvantage is that it should be used in combination with a science fringe detector that detects all the baselines. As more photons are taken for the co-phasing (partial-pairwise) beam combiner, fewer photons reach the science beam combiner and the sensitivity is increased at the expense of pushing the science detectors into the photon-starved regime. But even with very few photons per integration for science, the phase of the fringe can be determined from the co-phasing beam combiner and observations can be combined to force the science data back into the photon-rich regime. This is clearly an interesting way to build an instrument.

Acknowledgments

The preparation of this talk made use of NASA's Astrophysics Data System Bibliographic Services.

References

- J.T. Armstrong, D. Mozurkewich, L.J. Rickard, D.J. Hutter, J.A. Benson, P.F. Bowers, N.M. Elias II, C.A. Hummel, K.J. Johnston, D.F. Buscher, J.H. Clark III, L. Ha, L.-C. Ling, N.M. White, and R.S. Simon, "The Navy Prototype Optical Interferometer (NPOI)," *Astrophys. J.* **496**, 550–571 (1998).
- T.A. ten Brummelaar and W.G. Bagnuolo, Jr., "The CHARA beam combiner design", in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., Proc. SPIE **2200**, 140–151 (1994).
- Y.L. Chow, "On designing a supersynthesis antenna array," *IEEE Trans. Antennas Propagat.* **20**, 30–35 (1972).
- T.J. Cornwell, "A novel principle for optimization of the instantaneous Fourier plane coverage of correlation arrays," *IEEE Trans. Antennas Propagat.* **36**, 1165–1167 (1988).
- M.J.E. Golay, "Point arrays having compact, nonredundant autocorrelations," *J. Opt. Soc. Am.* **61**, 272–273 (1970).
- E. Keto, "The shapes of cross-correlation interferometers," *Astrophys. J.* **475**, 843–852 (1997).
- E.M. Kong, D.W. Miller, R.J. Sedwick, "Optimal Trajectories and Orbit Design for Separated Spacecraft Interferometry," SERC No. 13-98 (November 1998).
- N.C. Mathur, "A pseudodynamic programming technique for the design of correlator supersynthesis arrays," *Radio Sci.* **4**, 235–244 (1969).
- D. Mozurkewich, "Hybrid design for a six way beam combiner," in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., Proc. SPIE **2200**, 76–80 (1994).
- A. Quirrenbach, D. Mozurkewich, D.F. Buscher, C.A. Hummel, and J.T. Armstrong, "Angular diameter and limb darkening of Arcturus," *Astron. Astrophys.* **312**, 160–166 (1996).
- W.J. Tango and R.Q. Twiss, "Michelson stellar interferometry", *Prog. Opt.* **17**, 239–277 (1980).

Chapter 15

Data Reduction for Synthesis Imaging

J. THOMAS ARMSTRONG

NAVAL RESEARCH LABORATORY
WASHINGTON, D.C.

15.1 Preliminaries

I will start with a collection of ideas that underlie the rest of this Chapter. The first idea is that in optical interferometry the phase reference is the star itself, because the fringe motions are so rapid (roughly a radian of fringe phase in ~ 10 msec at visual wavelengths) and so large (tens of microns over a few seconds) that using another source as the phase reference is impossible.* This leads to a second basic notion, that the targets of current optical interferometers are stars. No other objects have enough surface brightness to act as their own phase calibrators.

Now, fringe visibilities from a stellar image are high only in the inner part of the (u, v) plane. For a star whose image is a uniform disk, $V^2(\lambda) \propto [J_1(\mathbf{x})/\mathbf{x}]^2$ i.e., $V^2(\mathbf{x})$ has the radial profile of an Airy disk, with its first zero at $\mathbf{x} = 3.83$ (refer to Table 15.1 for notation). Consequently, we arrive at a third basic notion, that there is a longest baseline length on which the fringe visibility is high enough to allow us to track the fringe. The smallest V^2 that can be tracked depends on the signal-to-noise ratio, which is a function of NV^2 (cf. Equation 14.1) and thus depends on both the brightness and angular size of the star.

If, for instance, we can track fringes when V^2 is $\simeq 0.2$, the constraint on array-element separations means that the elements must form a chain in which no link is longer than

*The Palomar Testbed Interferometer (PTI) circumvents this problem by looking at two stars simultaneously, but no other interferometers currently under development will use this technique.

Table 15.1: Notation

Symbol	Meaning
B, \mathbf{B}	Baseline length; baseline vector
FASN	Fringe amplitude SNR
$J_i(\cdot)$	Bessel function of first kind of order i
LD	Limb-darkened disk
N	Photons per integration
SNR	Signal to noise ratio
UD	Uniform disk
(u, v)	Spatial frequency coordinates; $\sqrt{u^2 + v^2} = B/\lambda$
V^2	Squared fringe visibility; loosely, “visibility”
V_*^2	True visibility of a star
V_{raw}^2	Observed visibility
$V^{(3)}$	Amplitude of triple product
\mathbf{x}	$= \pi\theta\mathbf{B}/\lambda$, so $V_{*,\text{UD}}^2 = [J_1(\mathbf{x})/\mathbf{x}]^2$ has its first zero at $\mathbf{x} = 3.83$
θ	Stellar angular diameter
λ	Wavelength
ϕ_{cl}	Closure phase (= phase of triple product)
$\vec{\rho}$	Binary separation vector

$\sim 125\lambda_{\mu\text{m}}/\theta_{\text{mas}}$. For future convenience, I will designate this greatest minimum separation as B_{track} , the longest baseline on which fringes from a given star can be detected within the coherence time of the atmosphere so they can be tracked.

The final notion is that an interferometer ought to be designed with both wavelength and baseline bootstrapping in mind, for the reasons Dave Mozurkewich went into in Chapter 14. The following discussion assumes that both are available. My discussion will also draw on several of the V^2 models that Dave generated, so you should be prepared to refer frequently to Chapter 14.

15.2 Planning the Observations

In planning the observations, you want to choose the array configuration that fits the source you’re looking at, with baseline lengths appropriate to the angular scale you’re interested in. Keeping the array phased up is a major constraint in some circumstances. You also want to observe the star at enough hour angles to get sufficient (u, v) coverage, and get enough scans on calibrator stars to monitor seeing changes.

The first thing to determine when choosing a configuration is whether the angular scale of the structure that you are interested in is larger or smaller than the stellar disk.

When the structure of interest is small compared to the stellar disk, the requirements of phasing up the array are paramount. The most obvious example of small structure is stellar surface structure, but a second is limb darkening. It is when we are observing small structure that wavelength and baseline bootstrapping are vital. In order to bootstrap, we need a configuration for which every link in the chain of array elements is shorter than, but as close as possible to, B_{track} in order to maximize the longest baseline of the array.

If the structure of interest is larger than the star, we have a “source with high visibility” in Dave Mozurkewich’s terminology (Section 14.1.2). Examples of observing large structure include measuring a binary orbit and imaging circumstellar dust or H α emission. For these sources, we choose the configuration that best constrains the image or the model, rather than the one that maximizes the longest baseline. In general, that means choosing the most uniform (u, v) coverage, which in turn means choosing a variety of telescope separations.

Dave touched on how to deal with the various cases of observing small and large structure. In this section, I will go into more detail with four examples of the effects of differing observing goals on the choice of configuration: determining a stellar uniform-disk diameter; observing stellar surface structure, including limb darkening; determining a binary-star separation; and observing circumstellar emission.

15.2.1 Stellar Diameters

If all that we want is an equivalent uniform disk diameter, θ_{UD} , we do not need a multi-element interferometer; we need only a single baseline of length $\sim B_{\text{track}}$. Choosing a configuration reduces to choosing a baseline long enough to partially resolve the star, but still short enough to track the fringes. If enough wavelength bootstrapping is available, the bluest channels can reach the first null of V^2 and give us a very precise measurement of θ_{UD} .

To put it in more concrete terms, I will show an example with the NPOI. Consider some observations of ϵ Gem (G8 Ib; $m_V = 4.38$, $(V - R) = 0.96$, $(R - I_C) = 0.60$; $\theta_{LD} = 4.89$ mas [Pauls 1998]) made with the NPOI on 26 March 1997 with three elements of the astrometric array. Figure 15.1 shows uncalibrated visibilities from the longest baseline and one of the short baselines, as well as uncalibrated closure phases. The baseline lengths were 18.9, 22.2, and 37.5 m; about 500 photons were received per 2-ms data frame; and uncalibrated V^2 values ranged from 0.4 at the smallest (u, v) spacings (the 18.9 m baseline at $\lambda = 850$ nm), to zero at the longest, varying with time because the projected baseline length changes as the Earth rotates. The resulting fringe amplitude SNR (FASN), derived from Equation 14.1 in Chapter 14, was ~ 8 on the shorter baselines, but was as low as ~ 1.5 on the long baseline.

If we pretend that the 18.9 m baseline data shown in Figure 15.1(a) is from our one-baseline interferometer, we have plenty of FASN to track the fringes, yet the decline in visibility with wavelength (i.e., with increasing (u, v) distance) is enough to fit a diameter to the

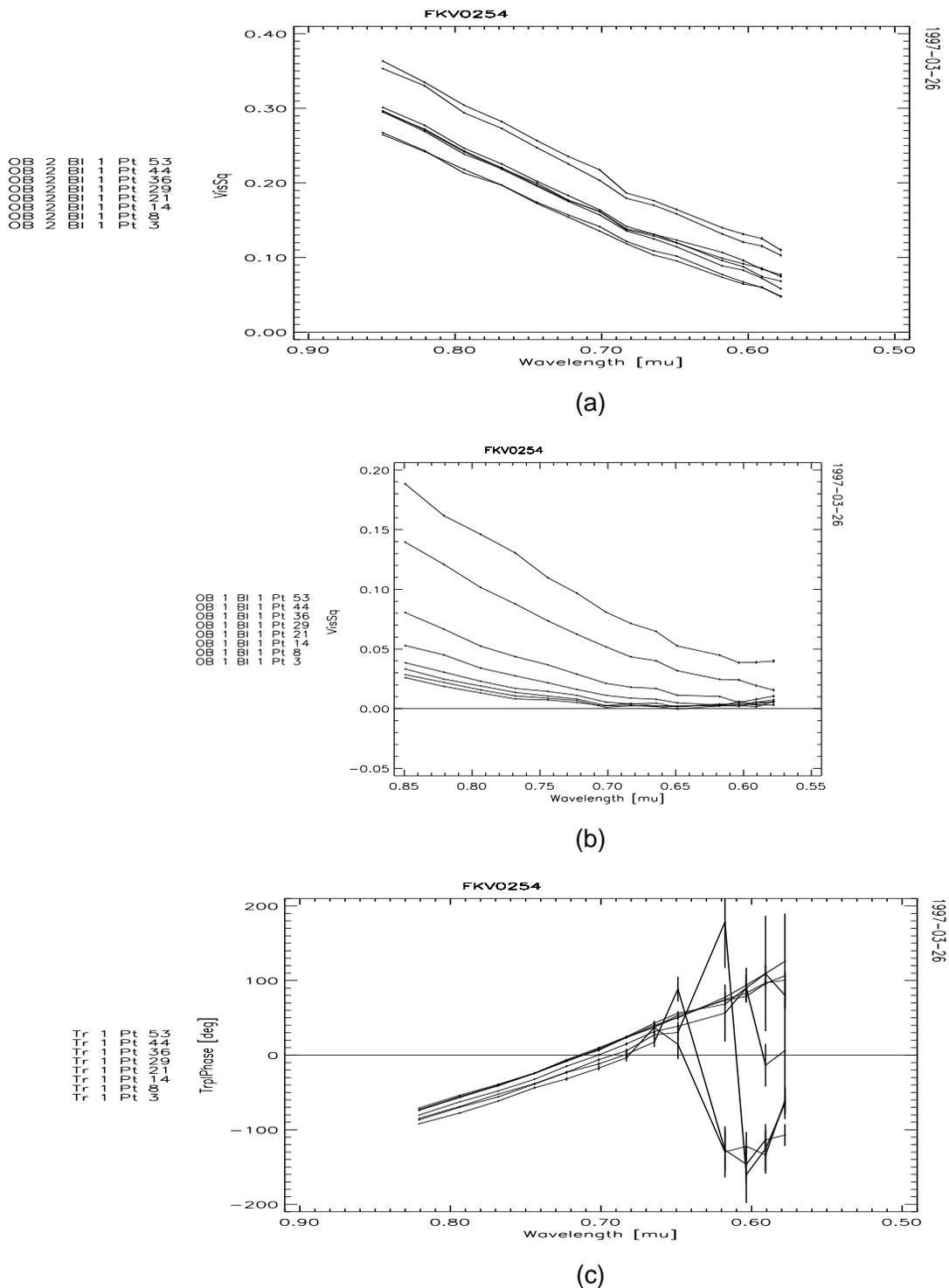


Figure 15.1: Raw visibilities and closure phases of ϵ Gem from the NPOI. (a) Data from the 19 m baseline. (b) Data from the 38 m baseline. (c) Closure phase. Note that the point at $\lambda = 0.62 \mu\text{m}$, $\phi_{\text{cl}} = 170^\circ$ should be shifted down 360° —which is permissible because the phase is determined modulo 360° —to be consistent with the other points.

star. Whether the diameter we fit is a uniform disk or a limb-darkened disk depends only on the model we choose, not on the data, since the shapes of the two models are virtually identical within the first zero of V^2 .

On the other hand, if we pretend that the 37.5 m baseline data shown in Figure 15.1(b) is our interferometer, we can do the observations only at some hour angles. The best of the scans has FASN $\simeq 5$, and still just reaches $V^2 = 0$, demonstrating the utility of wavelength bootstrapping. Andreas Quirrenbach used just this approach on the Mark III interferometer to measure the diameter—and the limb darkening—of Arcturus (Quirrenbach *et al.*, 1996). However, the scan that best detects the zero of V^2 has FASN $\simeq 1.5$, probably too low to fringe track. The fact that we were able to get these data anyway demonstrates the utility of combining wavelength and baseline bootstrapping.

What was B_{track} for these observations? Assuming that we need FASN ≥ 2.5 and using the observed number of photons per frame, we need $V^2 \geq 0.042$ averaged across the frame. The system visibility (i.e., the observed V^2 of a point source, given the imperfections in the instrument and the effects of the atmosphere) was ~ 0.6 , so we need to know the spatial frequency at which $V_*^2 = 0.07$. For a star with this amount of limb darkening, that visibility occurs at $\mathbf{x} = 3.15$. Taking $\lambda = 750$ nm as the representative wavelength, we find $B_{\text{track}} = 32$ m. That is satisfyingly close to the actual baseline length of 37.5 m, since it implies that we could fringe track on that baseline only at hour angles great enough to make the projected baseline less than B_{track} .

Since we actually have a three-element interferometer, we have one more piece of information to help us fit the angular diameter: the closure phase, ϕ_{cl} . Figure 15.1(c) shows the 180° phase jumps at the wavelengths at which $V^2(\lambda)$ reaches zero on the longest baseline. The spatial frequency of that V^2 null is constant; the wavelength at which it occurs changes as the projection of the baseline on the sky changes. In some circumstances, the signal-to-noise ratio for ϕ_{cl} is better than that for V^2 or $V^{(3)}$ (see Section 15.4), so the phase is the best indication of the angular diameter.

15.2.2 Stellar Surface Structure

The hardest problem is imaging a stellar surface, because all the information about surface structure is far out in the (u, v) plane where the fringes are too weak to track. By definition, any surface structure shows up at spatial frequencies that are higher than that of the first zero of V^2 , i.e., for $\mathbf{x} > 3.83$, where \mathbf{x} is calculated using θ_{UD} , the best fit uniform-disk diameter.

So do the data shown in Figure 15.1 tell us anything about surface structure? If we include limb darkening in this category, the answer is that they do tell us something, but not very much. For some scans, the data extend out to somewhere in the neighborhood of the first maximum after the null, but the signal-to-noise ratio of the visibilities is not good enough

there to determine the limb darkening. And certainly the data do not extend to large enough values of \mathbf{x} to detect a spot.

How many array elements do we need to detect a spot? It depends on the size of the spot, of course, so let's say we're trying to detect a spot whose diameter is 20% of the stellar diameter. You can make a rough estimate by imagining that you're trying to measure the diameter of a "star" that is the size of the spot, but with negative flux: you should need about five times the baseline needed for the measurement of the real star, which calls for six array elements.

It is clear that to do these observations, bootstrapping is indispensable. You want to make maximum use of baseline bootstrapping by choosing an array that has the longest possible chain of fringe-tracking baselines. You can reduce the necessary baseline length somewhat by using wavelength bootstrapping. If you're willing to settle for a detection of the spot rather than a measurement of its size, you may be able to reduce the baseline a bit more.

15.2.3 Binary-Star Separation

The choice of array configuration to observe binary stars requires a little more thought. The binary separation $|\rho|$ is larger than either star, but $B_{\text{track}} \simeq \lambda/\theta$; therefore, B_{track} is always more than long enough to resolve the binary, and keeping the array phased up is not the limiting factor. The choice of configuration and/or observing schedule is determined by the pattern of $V^2(u, v)$ values due to the binary. This pattern is formed by the fringe patterns of the two stars beating against one another, causing an oscillation of V^2 whose maxima in the (u, v) plane lie along lines perpendicular to the vector $\vec{\rho}$ separating the stars. The distance between crests in the (u, v) plane is $1/|\rho|$, and one of the crests passes through the origin.

Our goal is to sample this pattern of maxima well enough to determine its spacing and orientation, and thus to determine $\vec{\rho}$. We can sample the pattern radially, by observing in multiple spectral channels on each baseline, or circumferentially, by using Earth rotation to measure V^2 along an elliptical arc in the (u, v) plane at each wavelength, or both. With a three-element array, each observation gives us V^2 along three rays in different directions in the (u, v) plane, which will usually give us a unique solution for $\vec{\rho}$ even without using Earth rotation. Obviously, at least one baseline should be longer than $\lambda/2|\rho|$ or so; otherwise, the binary is unresolved. Not so obviously, at least one baseline should be shorter than $(\lambda^2/\delta\lambda)(|\rho|/n)$, where n is maybe 4 or so; otherwise, the spectral channels (width $\delta\lambda$) will sample too much of the oscillation, washing it out and failing to detect the binary. A more familiar way of looking at this constraint is as a condition on the fractional bandwidth of the channels. You want $\lambda/\delta\lambda \lesssim (\lambda/B)(|\rho|/n)$; otherwise, you get too much bandwidth smearing to see the binary.

The visibility data we end up with are shown well in Figure 14.3: we see an oscillation of V^2 as a function of (u, v) distance. As Dave said, there is a lot of information in these

data. The envelope of V^2 tells us the diameter. The depth of the troughs tells us the magnitude difference between the components, Δm : when $\Delta m = 0$, the troughs go down to zero visibility, while nonzero Δm produces shallower troughs.

We also get ϕ_{cl} data if there are three baselines or more. The detailed shape of $\phi_{\text{cl}}(u, v)$ also contains information about Δm . When Δm is close to zero, ϕ_{cl} in fact is the better indicator of Δm .

15.2.4 Circumstellar Material

Yet another type of observation is imaging extended circumstellar structure. John Monnier showed several examples (Figure 13.13), and Dave Mozurkewich also showed a model of the shape of the visibility with (u, v) distance (Figure 14.4). The information about the extended emission is in the bumps and wiggles in both the visibilities and the closure phases at low spatial frequencies. As with a binary, extended structure is relatively easy to deal with, because the star, which is the phase reference, is relatively compact compared to the structure you want to image, so once again, we may choose a configuration in which all the baselines can be shorter than B_{track} , and in which we have a variety of spacings to optimize the coverage.

15.3 Closure Phase and Calibration

So much for selecting the array. Now I will show what can happen to closure phases with some more data from the NPOI. There are two features that distinguish the NPOI data from the Keck aperture-masking results that John Monnier showed. First, we have only one triangle, since we are currently working with only three array elements, while John was working with 15 or 21 holes (sub-apertures). Second, we take data in 32 spectral channels, so we get 32 closure phases from our three-element array. Typically only the first 16 of those have sufficient signal-to-noise to be useful.

Figure 15.2 shows some uncalibrated NPOI ϕ_{cl} data from unresolved sources. These are stars that are small and single, so we're not getting any data from beyond the first null. Now, closure phase is supposed to be a good observable, that is, it's unaffected by telescope errors (including the atmosphere above the telescopes), and it ought to be zero for an unresolved single source. But ϕ_{cl} runs from -180° to -100° . So what's going on here?

The answer of course is that ϕ_{cl} is immune to telescope-based errors, but not to baseline-based errors. With the type of beam combiner we have at the NPOI [see Figure 14.12(a)], there are opportunities for baseline-based errors to arise. For instance, the light from input beam C is divided into reflected and transmitted beams at the first beam splitter. The reflected beam transits the first beam splitter twice before it combines with beam A at the second beam splitter, while the transmitted beam transits each beam splitter once before it combines with beam B . If the two beam splitters are not exactly the same thickness, the

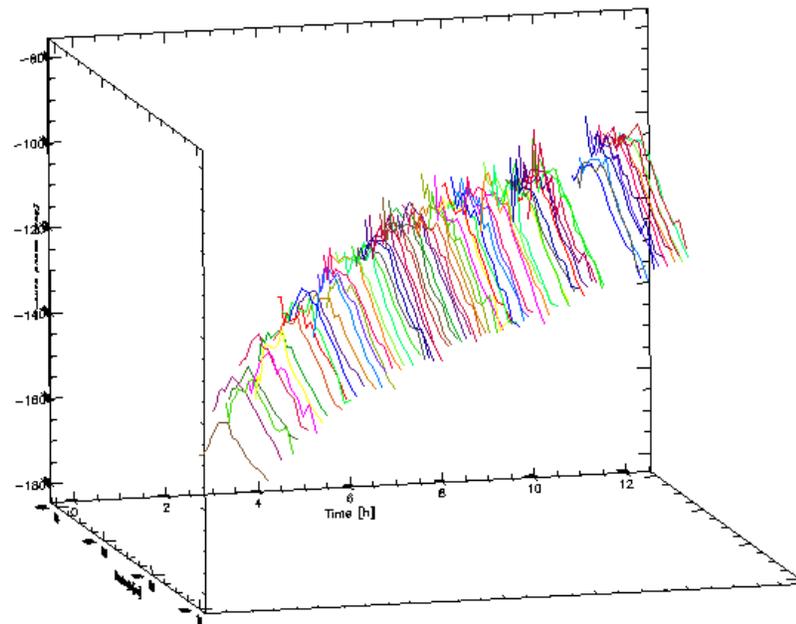


Figure 15.2: Raw closure phases of unresolved stars observed with the NPOI as a function of time and wavelength. The wavelength scale increases along the axis coming out of the page, running from 550 to 850 nm. The closure phase axis runs from -180° to -80° .

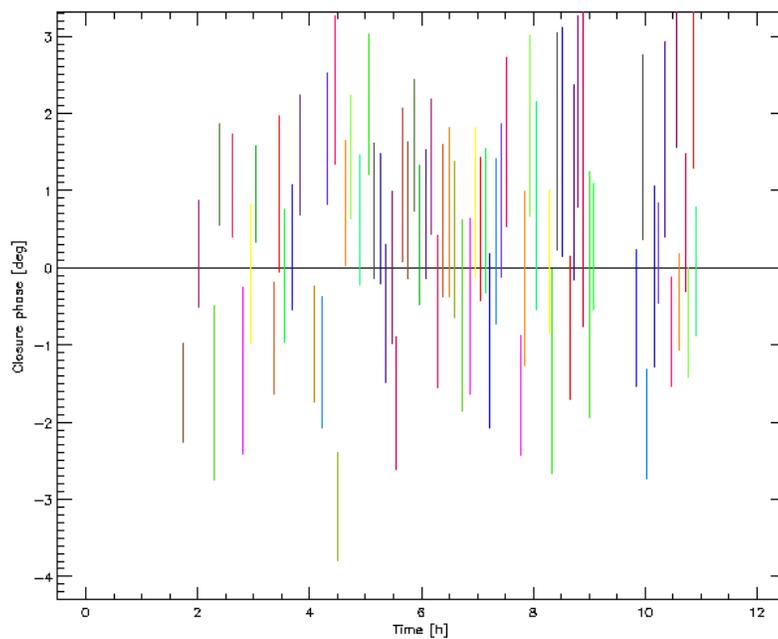


Figure 15.3: Calibrated closure phases as a function of time for the reddest channel of the NPOI, $\lambda = 850$ nm.

two parts of beam C experience different delays. The two baselines for which C is one of the elements will have different phase contributions, creating a baseline-based error.

Another possible source of closure-phase error is nonclosure of apparently closed triangles of array elements. In general, any three apertures A , B , and C will have different wavefront distortions across them. Some of those distortions are atmospheric and eventually average to zero, and some are due to optical imperfections and are constant. The parts of apertures A and B that correlate well—call them subapertures A_1 and B_1 —and that therefore have the dominant contributions to ϕ_{cl} will be different from the parts of B and C that correlate well (B_2 and C_2) and from C and A that correlate well (C_3 and A_3). Now, $A_1\vec{B}_1 + B_2\vec{C}_2 + C_3\vec{A}_3 \neq 0$, i.e., they don't form a triangle, so the phase that we measure on this trio of array elements is not really a closure phase.

Whatever the causes, Figure 15.2 shows that ϕ_{cl} changes smoothly with time and wavelength. In Figure 15.3, we see calibrated ϕ_{cl} from the $\lambda = 850$ nm (i.e. the reddest, and therefore the best behaved) channel. The rms deviation from zero is only about 1.5° , demonstrating how accurate the calibration can be.

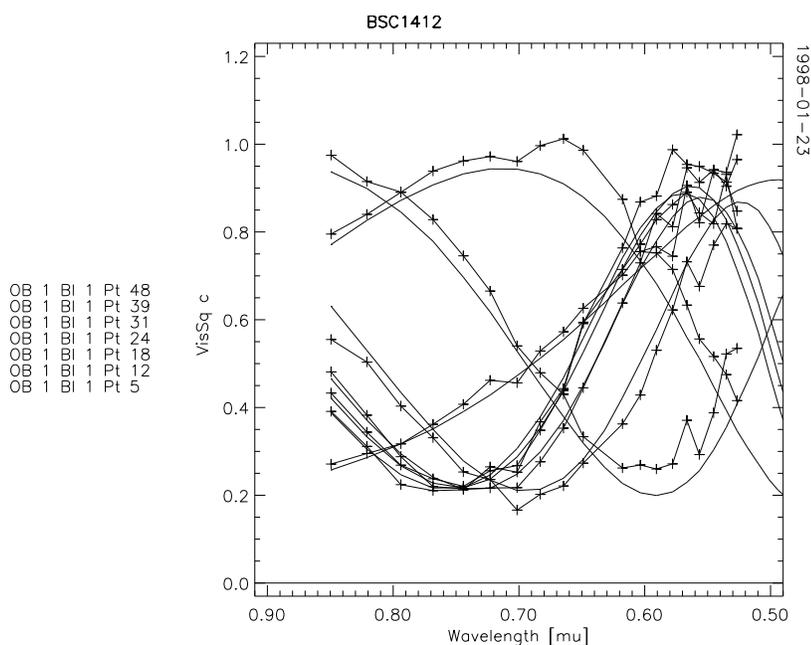
In this case, we generated phase corrections by fitting a paraboloid in the wavelength dimension and by smoothing in the time dimension. One of the improvements that we are investigating is generating a correction as a function of wavelength by determining the amount of air-path and glass-path mismatches within the beam combiner needed to produce the observed curvature of phase versus wavelength. Since the beam combiner is in a temperature-stabilized room, that correction should be quite stable during a night, and could significantly improve the phase calibration.

15.4 Model Fitting and Imaging

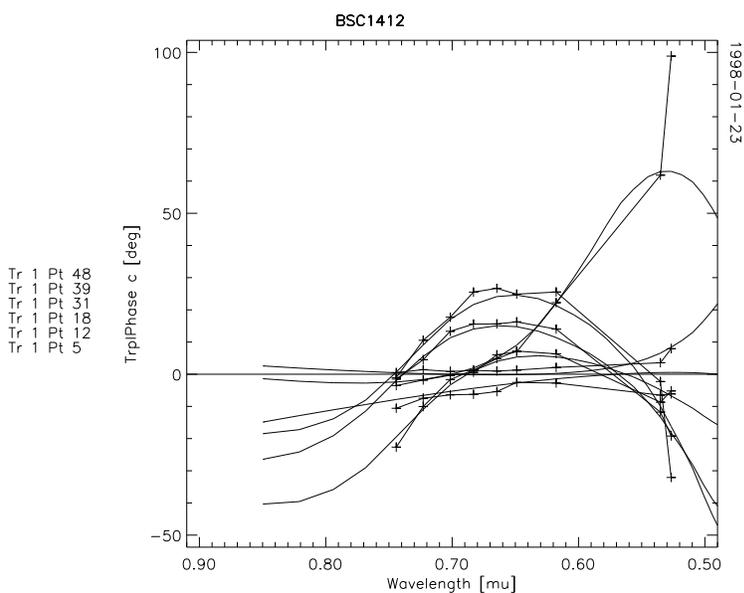
Once we have calibrated data in hand, the next thing is to interpret it. There are essentially two ways of doing this: model fitting or imaging. In fitting a model, of course, we minimize the difference between the data and the model predictions.

Figure 15.4 shows an example of fitting a model of θ^2 Tau (a spectroscopic binary with $\Delta m = 1.1$ mag) to the calibrated V^2 and ϕ_{cl} data of 23 January 1998, when $|\rho| = 9.0$ mas. A model including the orbital elements and Δm was fitted to 30 nights of Mark III data and six nights of NPOI data. The superimposed model curves in Figure 15.4 are derived from this 30-night data set rather than from this night alone. Conveniently, a single Δm for all wavelengths fits this binary well: it comprises two A stars, one of which has evolved off the main sequence just enough to lie just above the other one on the H-R diagram.

The V^2 data clearly show an oscillation as a function of spatial frequency, with maxima occurring at different spatial frequencies for different scans as the Earth rotates the projected baseline against the pattern of V^2 maxima.



(a)



(b)

Figure 15.4: NPOI data on θ^2 Tau on 23 January 1998, with model curves superimposed. Error bars have been suppressed for clarity. The data are shown as a function of decreasing wavelength. For V^2 , this corresponds to increasing spatial frequency. (a) $V^2(\lambda)$ data for seven scans on the 37.5 m baseline. (b) $\phi_{cl}(\lambda)$ data for six scans. The four reddest channels are missing due to bad detectors on one of the baselines.

The reason for the shapes of the ϕ_{cl} curves is less intuitively obvious. As we move from short wavelengths to long, the visibilities are sampled along rays in the (u, v) plane. Whenever we cross a minimum of V^2 along one of those rays, the phase on that baseline—and thus ϕ_{cl} —changes by an amount that depends on Δm : for zero magnitude difference, the phase change is 180° . As Δm increases, the size of the phase jump decreases, and in addition, the transition becomes less abrupt.

For an example of modeling the triple product, a different form of the data, I return to the example of ϵ Gem. Figure 15.5 shows the calibrated amplitudes and phases of the triple product as a function of wavelength, with model curves superimposed.

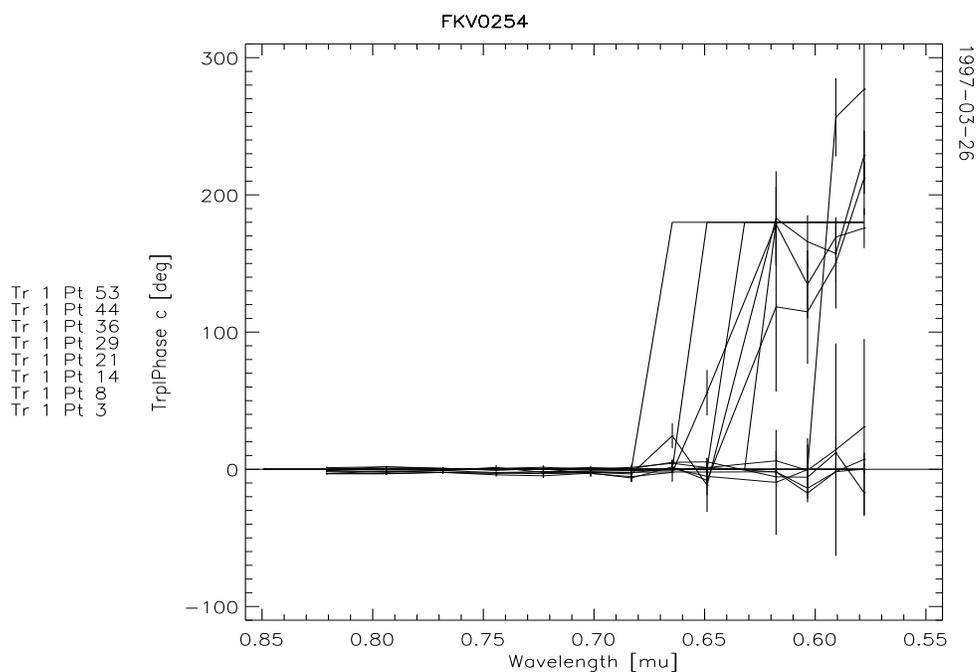
You will recall that the closure phase is the phase of the product of the complex visibilities on three baselines forming a triangle (Section 13.2). Since V data that are measured beyond a null are 180° out of phase with those inside the null, the values of ϕ_{cl} show a step function at the wavelength where one of the contributing baselines crosses the null. The amplitude of the complex product is the triple amplitude $V^{(3)}$. When one of the baselines is sampling near the first null, $V^{(3)}$ may offer a better signal-to-noise ratio than does that V^2 , as appears to be the case in Figure 15.5. In this instance a model diameter for the star should be compared to $V^{(3)}$ rather than to V^2 data.

The other route to interpreting the data is making an image from it. The advantage of imaging is that you're not restricted to your preconceptions about what the source might look like. Rather than impose any constraints on the data, we use the data themselves to tell us what the star looks like.

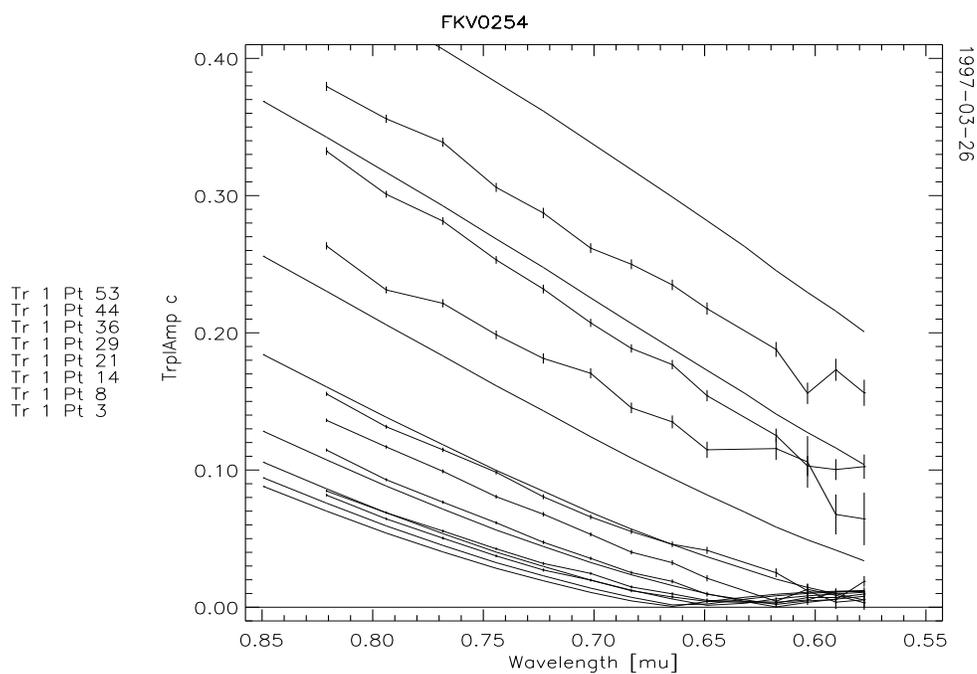
There are two difficulties with this approach. The first is that we need complex visibilities to Fourier transform into an image, but instead we have V^2 and ϕ_{cl} . We would like to recreate the $N(N-1)/2$ baseline phases, but there are only $(N-1)(N-2)/2$ independent closure phases. John Monnier described in Section 13.3.4 how one starts with an initial guess for the baseline phases and then uses the radio-astronomy based self-calibration and CLEAN routines to iterate to a best fit between the V^2 and ϕ_{cl} data and the model. This technique is still being worked out, and mostly with tools built for radio astronomy that assume that we're working with complex visibilities. There is a good deal of development to be done in this direction.

The second difficulty is that we must deconvolve the point-spread function of the array from the image. This is an art that has been extensively developed in radio astronomy, as well as in a number of other fields. It is here that the distinction between modeling and imaging starts to break down. For instance, the CLEAN algorithm treats an image as an ensemble of delta functions. The model is convolved with the point-spread function to produce the final product. People think of the process of producing a CLEANed, self-calibrated result as imaging, but it is imaging with a strong modeling component.

At NPOI we have produced images from data of V^2 and ϕ_{cl} by first assigning ϕ_{cl} to one of the baselines, then self-calibrating and CLEANing. Because we have been working with



(a)



(b)

Figure 15.5: Calibrated closure phases and triple amplitudes of ϵ Gem compared to a limb-darkened model with $T_{\text{eff}} = 4730$ K and $\log g = 1.5$ (Evans & Teays, 1996), and $\theta_{\text{LD}} = 4.89$ mas (Pauls *et al.*, 1998). (a) Closure phases and model curves. (b) Triple amplitudes and model curves.

only three array elements, we don't have the problem that John Monnier had in his Keck aperture masking data, of choosing which set of independent closure phases to use. But with only three array elements commissioned so far, we are also limited to much simpler images, those of binary stars. It is encouraging, however, that even in these simple images, we have attained a dynamic range of between 100:1 and 200:1, which is about the same as in the aperture masking images.

15.5 Current Capabilities and Limitations

The current capabilities and limitations of optical interferometry include high angular resolution, low sensitivity, moderate complexity, and moderate dynamic range. The highest resolutions that have been attained by long-baseline interferometers are between 1 and 2 mas for measurements of stellar angular diameters, and 2 mas or a little more for separations of binary stars, far outstripping any other techniques available at visual and infrared wavelengths. But interferometers demand a lot of light. The magnitude limit at visual wavelengths is currently about 6^m for long-baseline interferometers, and about 3^m for aperture masks.

The current long-baseline optical interferometers can produce images of relatively low complexity because they sample a small number of (u, v) points; even with the NPOI's 32 spectral channels on each of three baselines, we get only 33% of the phase information. For comparison, the VLA has 351 baselines. The Keck aperture masks in the observations John Monnier described used 210 baselines and yield 190 independent closure phases, or 90% of the phase information available before the light hit the atmosphere. The increased information content is clear in the complexity of the resulting images.

The dynamic range in images from both long-baseline and aperture-masking interferometry is in the range of 100:1 to 200:1, largely determined by the precision of the calibration. Improving the calibration is difficult, but a promising approach is to use single-mode fibers as spatial filters, as is being tried by the FLUOR group using IOTA and by the group at PTI (cf. Traub, Section 3.4.2).

15.6 Future Developments

There are two directions in which the field of optical interferometry is growing. One is toward aperture masking of the kind that the Cambridge group and the Keck group are doing, using masks with a large number of holes and thus taking data from a large number of baselines and retrieving upwards of 80% of the phase data. There is a lot of interesting work to be done with this technique, even though it has two distinct limitations. One is that the maximum baseline is 10 m; the other is that the images are monochromatic.

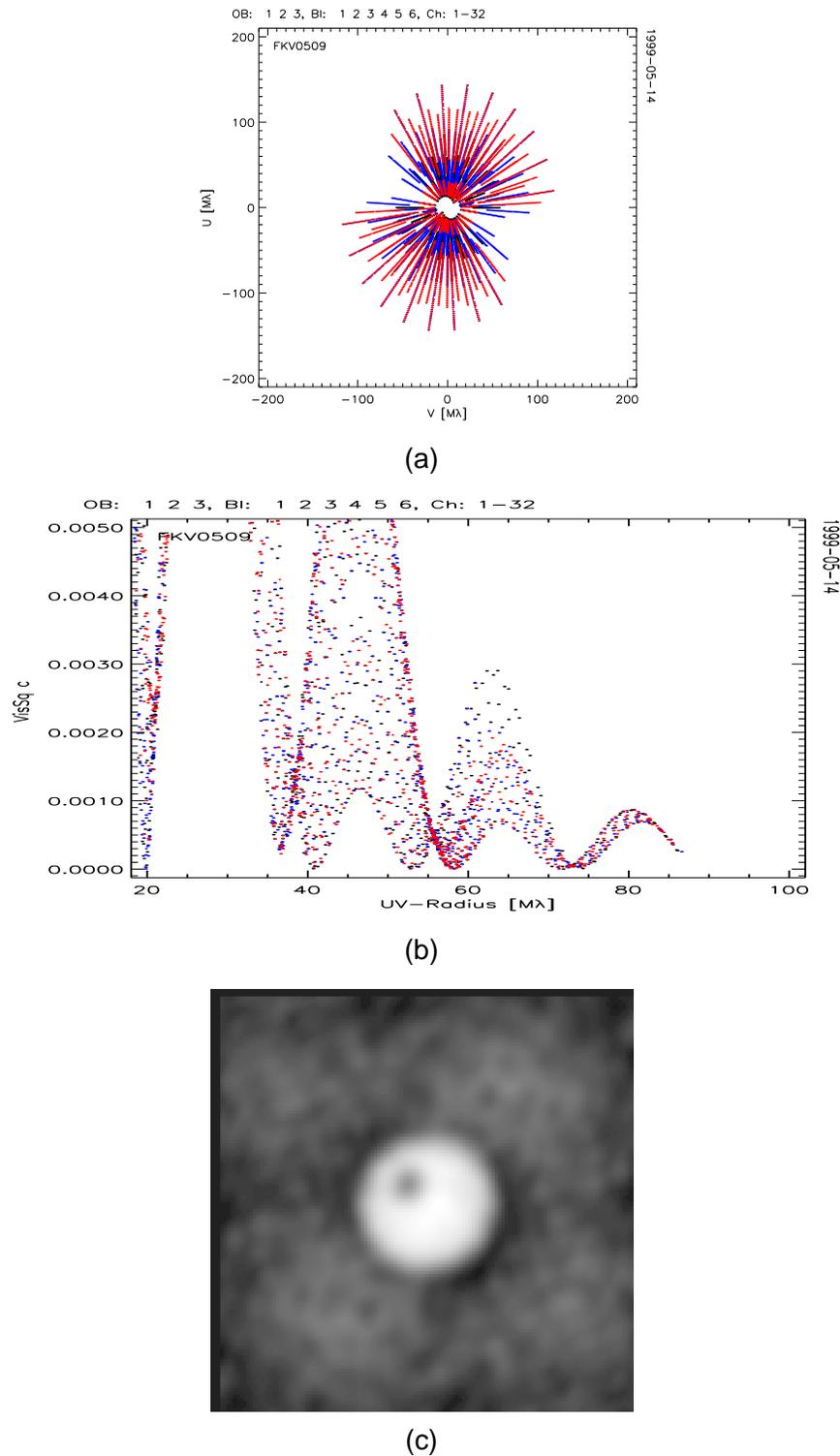


Figure 15.6: Simulation of a 12-mas limb-darkened star with a spot imaged with a six-element array in 32 spectral channels. (a) (u, v) coverage. (b) V^2 vs. $\sqrt{u^2 + v^2}$ beyond the first null, before noise addition. (c) Image restored with a maximum-entropy routine. [Simulation courtesy of Christian Hummel.]

The other direction is toward increasing the number of elements and the maximum baseline in such long-baseline interferometers as COAST, the NPOI, the CHARA array, or the GI2T. Currently, COAST has five elements working and can use four at any given time; the NPOI has three, but should have six by the end of 2000 with a maximum baseline of 65 m (and ultimately 437 m); CHARA has had first light on one baseline, and will ultimately have as many as eight elements and a 400 m maximum baseline; and IOTA and GI2T will have three elements within about the same period, with somewhat shorter baselines. The longest baseline may eventually be that of SUSI, at 640 m, although for the next few years it is likely to remain a two-element interferometer.

These long-baseline interferometers overcome both of the limitations of the aperture masking technique. First, baseline length is no longer limited by the size of a single telescope. Second, separating the light-gathering from the imaging functions of the instrument makes it possible to take data in a large number of spectral channels. Increasing the number of data channels can offset the advantage the aperture masking technique has in number of baselines.

Figure 15.6 shows an indication of the image complexity that a six-element multi-channel array will be capable of. We simulated observations of a 12-mas limb-darkened star with 5% noise added to the V^2 data. We used six elements in one of the baseline-bootstrapped arrays that Dave Mozurkewich showed (Figure 14.8), with three telescopes evenly spaced along each of two arms of the NPOI Y-shaped imaging array, and simulated 12 scans, using Earth rotation to improve the (u, v) coverage. We then deconvolved the image with a maximum-entropy routine. The result is comparable to the images from aperture masking observations.

Two and three element optical interferometers have already shown their ability to measure stellar diameters and binary orbits. The multiple-baseline interferometers that are coming into operation over the next few years should allow us to see the details of what is happening on and around the surfaces of stars.

References

- N.R. Evans and T.J. Teays, “Temperatures of δ Cep and nonvariable supergiants,” *Astrophys. J.* **112**, 761–771 (1996).
- T.A. Pauls, D. Mozurkewich, J.T. Armstrong, C.A. Hummel, J.A. Benson, and A.R. Hajian, “Observations of stellar limb darkening with the Navy Prototype Optical Interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg, ed. Proc. SPIE, **3350**, 467–470 (1998).
- A. Quirrenbach, D. Mozurkewich, D.F. Buscher, C.A. Hummel, and J.T. Armstrong, “Angular diameter and limb darkening of Arcturus,” *Astron. Astrophys.* **312**, 160–166 (1996).

Part V

Nulling Interferometry

Chapter 16

Nulling Interferometry and Planet Detection

EUGENE SERABYN

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

16.1 Introduction

The small angular separation and high brightness contrast characterizing star-planet systems has thus far prevented the direct detection of planets beyond our solar system. If our solar system were viewed from a distance of 10 pc, a mere 0.1'' would separate the Earth from the Sun (Figure 16.1), while Saturn would lie 1'' out. Moreover, the Sun/Earth contrast ratio at visual wavelengths would exceed 10^9 . However, while the angular separations are fixed constraints, the contrast ratio is amenable to modification in at least two ways. First, the observational wavelength can be shifted into the infrared, where planetary thermal emission spectra peak (Bracewell, 1978; Angel *et al.*, 1986). Second, it is possible to alter the intrinsic contrast ratio with an optical approach capable of selectively dimming the star relative to its surroundings.

In the latter regard, two techniques are quite promising: coronagraphy and nulling interferometry. Because coronagraphy relies on physically blocking the incident starlight with a small obscuring spot located at a focal plane stellar image, useful coronagraphic imaging is possible only beyond several Airy radii from the on-axis stellar source. This approach is thus likely optimal primarily for the nearest stars. On the other hand, as is elaborated in the succeeding sections, nulling interferometry is expected to be effective only within the core of a telescope's point-spread function, and so can be employed for stars at greater distances, where a larger sample is available.

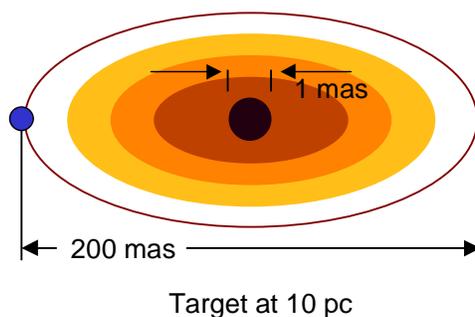


Figure 16.1: Schematic diagram of the inner part of our own solar system as seen from a distance of 10 pc. Included are the Sun, Earth, and zodiacal cloud. The centrally peaked zodiacal emission actually extends to radii well beyond 1 AU.

In this paper, an overview of the techniques of “nulling” interferometry is presented. As a concrete example, the operation and performance of fiber-coupled rotational shearing interferometers is described in some detail. However, it should be borne in mind that at this stage few of the suggested nulling approaches have been demonstrated experimentally, and so the optimal approach may not yet be identified.

16.2 Exozodiacal Light

Beside proximity and contrast, a third obstacle to direct exoplanet detection is the possible emission from “exozodiacal” dust grains congregating in and near the orbital plane of the target solar system (Figure 16.1). Such dust disks can be expected both by analogy with our own solar system’s zodiacal cloud (Reach *et al.*, 1995), and by extrapolation of the few known cases of very bright circumstellar disk emission around young main sequence stars (Backman and Paresce, 1993). In fact, due to the extended nature of such exozodiacal disk emission, a disk’s integrated thermal emission is likely to exceed even that of gas giant planets by orders of magnitude. For example, the integrated thermal emission from our own solar system’s zodiacal dust is about 10^{-4} that of the sun’s luminosity at a wavelength of $10 \mu\text{m}$, while the emission from Jupiter is two orders of magnitude lower, and the Earth’s emission yet another order of magnitude lower (Traub *et al.*, 1996; Angel, 1998; Beichman *et al.*, 1999; Serabyn *et al.*, 2000).

While observations of nearby stars have established the presence of dust in a few systems, the high contrast ratio involved means that the dust detected to date tends to be located at large distances from the central stars. It is thus largely colder dust at Kuiper-belt scales ($> 30 \text{ AU}$ radius) which has been detected in these systems. Information on warmer dust at the smaller radial offsets commensurate with habitable zones is therefore almost nonexistent. Prior to conducting extensive searches for very weak emission from terrestrial planets around nearby stars, it would therefore be advantageous to establish the emission levels

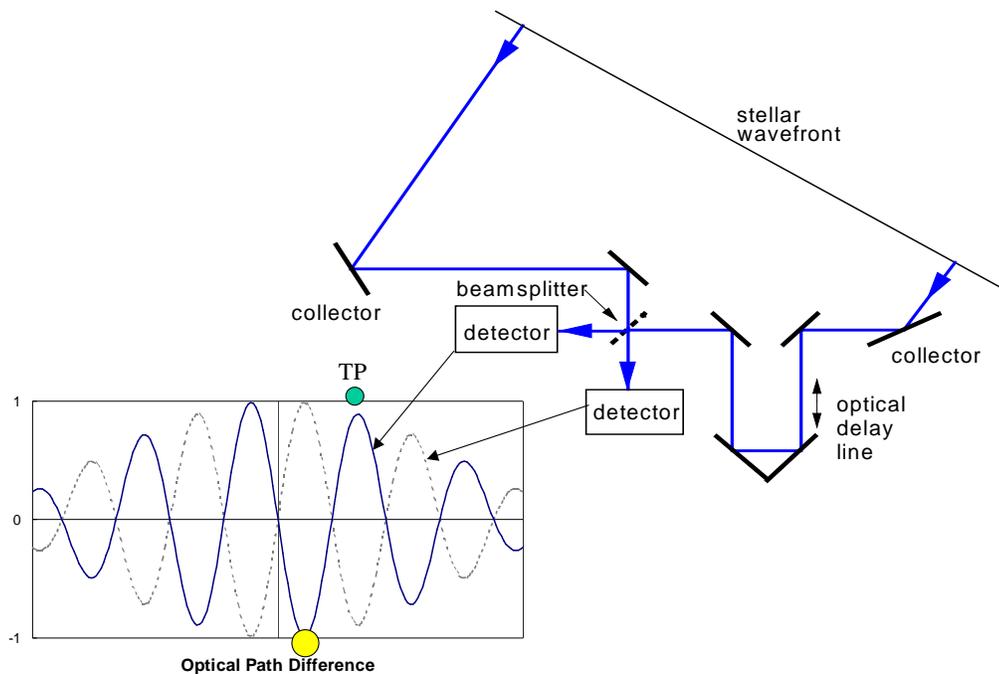


Figure 16.2: Schematic diagram of a standard stellar interferometer and the resultant fringe pattern at the two beam-splitter outputs. Conceptually the goal is to place the star at the bottom of a deep destructive fringe, and an accompanying terrestrial planet (TP) near the top of a constructive fringe.

from exozodiacal dust around these stars, and indeed, first-generation nulling experiments are being designed primarily to determine these exozodiacal emission levels. In particular, both the Keck Interferometer and Large Binocular Telescope nulling experiments are being designed to search for the warm dust at AU scales (Angel, 1998; Colavita *et al.*, 1998; Booth *et al.*, 1999; Serabyn *et al.*, 2000).

16.3 Nulling Interferometry Basics

The basic premise of nulling interferometry is conceptually quite simple: combine the light incident on a pair of telescopes so that at zero optical path difference (OPD) between the incident beams, the two electric-field vectors are exactly 180° out of phase, thus allowing near-perfect starlight subtraction. In the language of interferometry (Bracewell, 1978), a deep destructive fringe is to be placed across the star (Figure 16.2). However at the same time, off-axis emission from sources located near constructive fringe maxima can be transmitted through the system (Figure 16.2), and so even though the star is nulled to a deep level, appropriately situated planets are not attenuated greatly. The trick then is to devise an optical scheme which can null starlight of both polarizations and at all wavelengths across the passband of interest simultaneously.

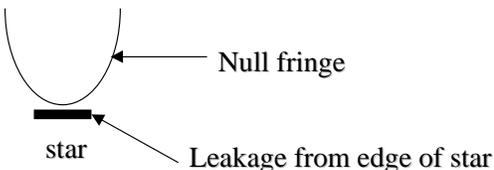
$$N = \frac{\pi^2}{16} \left(\frac{\theta_{dia}}{\lambda / b} \right)^2$$


Figure 16.3: Cross-sectional view through the null fringe and star. The finite stellar diameter allows for light from the edges of the star to “leak” through the fringe pattern. The null depth, N , is the ratio of transmitted powers in the destructive and constructive states.

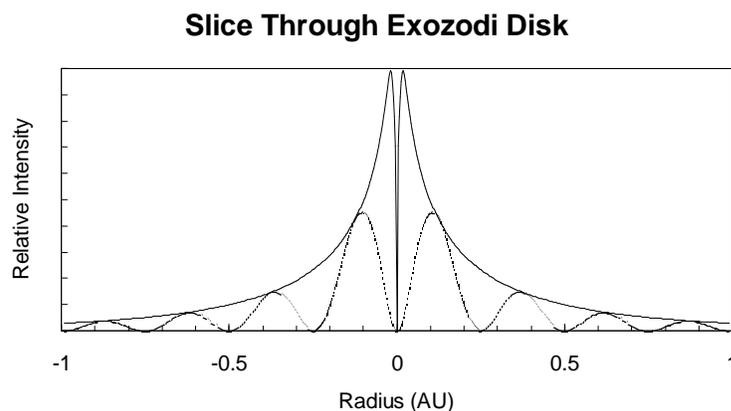


Figure 16.4: Cross-sectional view through an edge-on exozodiacal cloud at a distance of 10 pc, showing the emergent flux (solid curve), and the flux transmitted by the $10\ \mu\text{m}$ fringe pattern of the Keck Interferometer in the nulling mode (dotted curve).

The next section addresses the technical challenges which must be met in order for the potential of deep stellar nulling to be realized, but there is one fundamental null depth limitation which cannot be overcome, because it arises in a property of the source itself: the finite stellar diameter. For a uniform circular disk source, and a single pair of nulling telescopes, stellar radiation from points off the central axis of symmetry can leak around the central destructive fringe, as indicated in Figure 16.3. Thus, for a single baseline, the stellar cancellation cannot be very deep unless the baseline is short enough to spread the central destructive fringe across the entire stellar disk. For example, a sun-like G2 star at a distance of 10 pc has a diameter of 0.93 mas, implying that on the 85-m baseline between the two Keck telescopes, $10\text{-}\mu\text{m}$ null depths (where the null depth is defined as the ratio of transmitted powers in the destructive and constructive states) for solar-equivalent stars are limited to about 10^{-3} . At the same time, the off-axis fringes can transmit a substantial fraction of the exozodiacal and exoplanetary radiation. Figure 16.4 shows the effect of the Keck-Keck fringe pattern (of period 24 mas) on an exozodiacal cloud like that of our own solar system’s, at a distance of 10 pc: approximately half the arriving exozodiacal flux is transmitted by the nulling interferometer’s off-axis fringes.

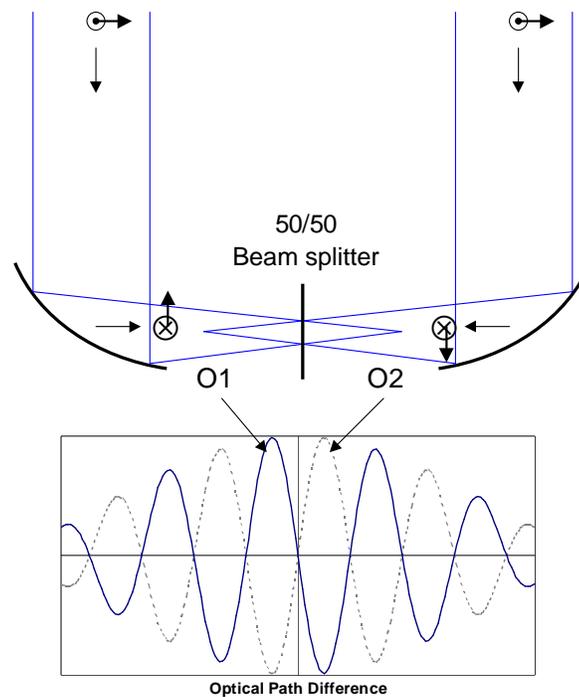


Figure 16.5: Original nulling scheme proposed by Bracewell & MacPhie (1979).

For the nulling approach to be effective, the stellar cancellation must be very deep, on the order of a part in 10^6 for planet detection, and a part in 10^{3-4} for exozodi detection (Beichman *et al.*, 1999). Of course the null depths given refer to the net nulls integrated across the passband. Such a deep achromatic null fringe cannot possibly be provided by a standard stellar interferometer, because as Figures 16.2 and 16.5 show, the inherent symmetry of such systems implies that at zero OPD, the two beam-splitter outputs are equivalent, with each receiving half the incident power. To effect cancellation, a finite OPD must therefore be introduced. An offset of a quarter of the average wavelength in the passband will optimize the cancellation (Figures 16.2 and 16.5), but since a quarter wavelength is clearly a chromatic quantity, it is evident that all wavelengths cannot cancel simultaneously at a non-zero OPD setting. Indeed, this is the reason the original nulling scheme (Figure 16.5) of Bracewell and MacPhie (1979) fails to provide a very deep null. (Of course, there is an even more serious problem with this scheme, in that the fringes in the two polarization states actually cancel each other completely for such a perfectly symmetric system).

A more achromatic approach is therefore needed, the ideal case being completely wavelength-independent cancellation at zero OPD. What is thus desired is a set of fringes which are complementary to the case of a standard laboratory Michelson interferometer (Figure 16.6). Such an ideal achromatic destructive fringe at zero OPD (solid curve in Figure 16.6) can be generated by subtracting, instead of adding, the two incident electric fields. In fact, by con-

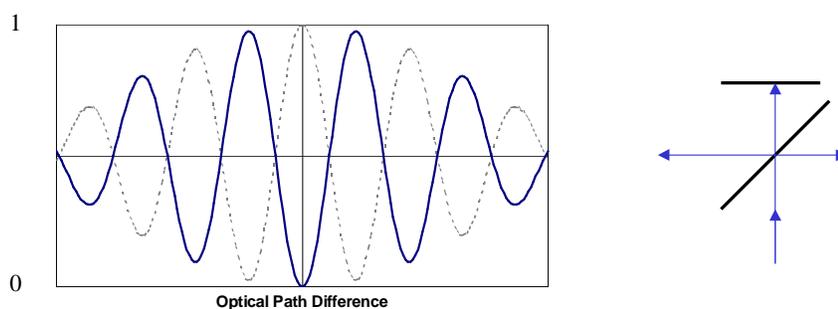


Figure 16.6: Standard laboratory Michelson interferometer, and the associated fringe pattern at the output (dotted line). For deep achromatic nulling, what is needed is an inverted fringe pattern (solid line), with achromatic cancellation zero OPD.

ervation of energy such subtraction applies at a Michelson interferometer's complementary output. However, as this output sends the light back to the input port, it is not readily accessible. To enable the necessary subtraction at a more accessible output port, two approaches have been considered: a relative flip of the electric-field vectors (Diner, 1990; Shao, 1990), which is intrinsically achromatic, and a phase retardation approach (chromatic by design) in which light at each wavelength is delayed by exactly the distance necessary for all waves to arrive exactly 180° out of phase (Woolf and Angel, 1998).

Two approaches have been proposed for implementing a field-flip, both based on variants of rotational shearing interferometers. In the first approach (Diner, 1990; Shao, 1990), the flat end mirrors in the two arms of a Michelson interferometer are replaced by a pair of rooftop mirrors, oriented so that they appear orthogonal when viewed in projection in the common output beam (Figure 16.7). This rotational shearing interferometer concept will be described more fully below, but in brief, each rooftop flips one component of the incident field (Figure 16.8), so that a relative flip of the total electric-field vector, \mathbf{E} , is induced by the pair of rooftops. This approach clearly has the advantage of relying solely on flat mirrors.

In a second configuration (Baudoz *et al.*, 1998a,b), the basic Michelson interferometer is again modified by replacing the two flat end mirrors, but this time with focusing cat's eye assemblies (Figure 16.9). In one of the interferometer's two arms, the cat's eye secondary is a flat located at the focus of the primary mirror, while in the second arm, the flat secondary is replaced by a curved mirror which reflects the light back on itself before focus is reached. Thus, a focus is present in only one of the interferometer's two arms. Since passage through focus introduces an achromatic phase shift of π radians into a beam of light (Born and Wolf, 1980), upon recombination at the second beam splitter pass the two beams emerge from the interferometer with their electric-field vectors flipped relative to each other. Because of the different secondary mirrors in the two cat's eyes, the optical system is not quite symmetric, but for focal ratios greater than about 10, the limitations incurred thereby are insignificant.

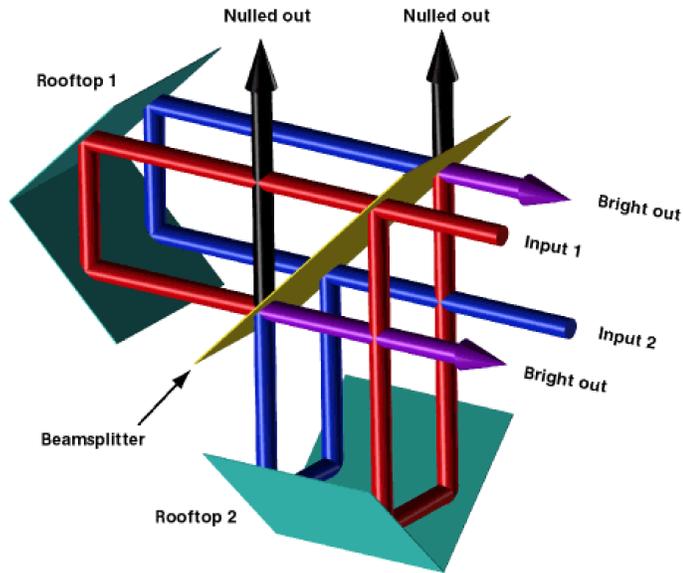


Figure 16.7: Orthogonal-rooftop-based nulling interferometer. Each rooftop flips only that field component which is perpendicular to the rooftop apex line, resulting in a net field flip between the two arms. In this configuration, the two inputs lead to two nulled and two “bright” outputs.

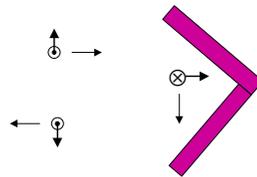


Figure 16.8: Side view of a rooftop (dihedral) mirror, illustrating the fact that only the E -field component normal to the rooftop apex emerges flipped in direction.

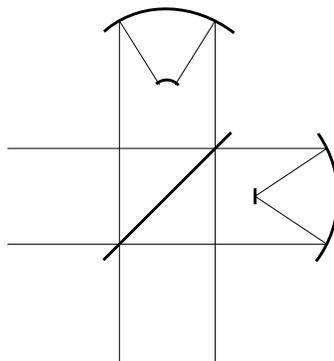


Figure 16.9: Cat's eye-based nulling interferometer. A focus is present on the flat cat's eye secondary in the right arm, but the upper arm has no focus. By virtue of passage through focus, an extra phase shift of π radians is present in the right arm.

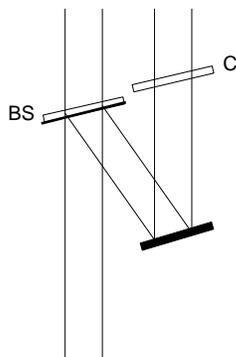


Figure 16.10: Phase retardation scheme. The dielectrics comprising the compensator are designed to add a quarter wavelength of phase across the band. The ideal, lossless beam splitter introduces a second $\pi/4$ of phase.

Finally, instead of simply flipping the \mathbf{E} -field vectors, there is the alternative of introducing an equivalent wavelength-dependent phase retardation (Hinz *et al.*, 1999; Morgan and Burge, 1999). In this approach, the dielectric indices of a beam splitter/compensator pair are carefully selected to delay each wavelength in the passband by the requisite half wavelength. Since in this approach the beam splitter is used in single-pass, the goal is to introduce an achromatic $\pi/4$ phase shift at the beam splitter, with another $\pi/4$ introduced by the compensator (Figure 16.10).

The phase retardation approach is thus quite different from the field-flip approach, and one essential difference in particular is worth noting: in both of the field-flip approaches the beam splitter is used in double pass inside a small beam-combining interferometer, while in the phase retardation approach, the beam splitter is used in single pass (Figure 16.11). Thus, in the first case, the two input beams are both modulated by similar beam-splitter reflection-transmission products, and so they retain their original intensity ratio, while in the latter case, one beam is modulated by the beam splitter's transmission coefficient and the other by its reflection coefficient. In the first case, power balance between the two beams is automatically maintained (for equal input powers), independent of the detailed beam-splitter properties, thus allowing for high-accuracy subtraction. In the latter case, power balance requires equality of the beam-splitter reflection and transmission coefficients.

16.4 Symmetry and Stability Requirements

The on-axis cancellation of stellar radiation requires a very symmetric and stable optical system. In particular, for a given polarization state, the two beam intensities, electric-field rotation angles (Figure 16.12), and phase delays must all be matched to $2\sqrt{N}$ (Serabyn, 2000), where N is the null depth as defined earlier. Thus for a net null depth of 10^{-6} , the allowable intensity mismatch (the deviation of either beam intensity from the mean), the relative rotation angle, and the phase difference must all be individually less than about

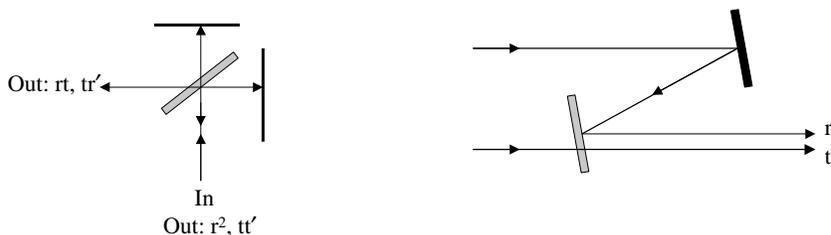


Figure 16.11: Comparison of beam splitter roles in the two nulling approaches. In the field-flip approach (left) the beam splitter is used in double pass, and the powers are automatically balanced; in the phase-retardation approach (right), one input field is multiplied by the beam-splitter reflection coefficient, and the other input by the transmission coefficient.



Figure 16.12: In the plane normal to the propagation direction, the angle between the two \mathbf{E} -field vectors must be within $2\sqrt{N}$ radians of π .

10^{-3} (the latter two in radians). This fine degree of balancing can only be achieved if a suitable means of adjusting the relevant parameters exists.

The rotation angle can be adjusted fairly accurately by adjusting the accompanying image rotation. As this is a static quantity, it is the easiest of the three to deal with. The relative beam intensities can be modified by introducing slight relative pointing offsets, or by introducing adjustable small-area obscurations across the centers of the beams. Possibilities in the latter category include a small scissor-like device (Figure 16.13, left), or a small Venetian-blind-like rotating vane (Figure 16.13, right). Finally, the relative phases (path delays) can be adjusted by means of precision optical delay lines, as are typically used in stellar interferometers, but with accuracies on the order of 1 nm to allow 10^{-6} nulling at $\lambda = 10 \mu\text{m}$. Of course, also needed for the latter is a control algorithm, i.e., a means of sensing non-zero phase errors. The simplest control approach, and the only one which has been experimentally demonstrated to date, is cavity-length dithering (Serabyn *et al.*, 1999a), but this approach loses efficacy as the photon flux is diminished. Possible alternatives include control of one nuller output by means of the second (Serabyn, 1999), control by means of light in a non-nulled waveband (Hinz *et al.*, 1999), and control of one polarization component by means of the second. All three of these approaches require experimental demonstration. A final possibility is of course laser metrology.

Moreover, these conditions must be met simultaneously for both polarization components, at every point across the aperture, and for all wavelengths in the band. Thus it is also critical to avoid effects such as “differential birefringence”: differing s-p phase delays between the two interferometer arms. The problem is illustrated graphically in Figure 16.14, where it can

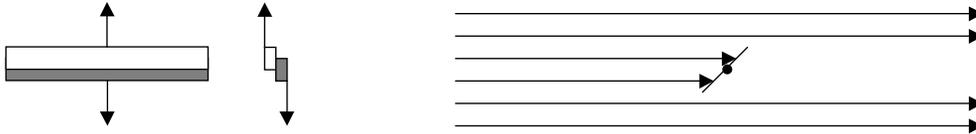


Figure 16.13: Two possible intensity modulation schemes, both based upon a linear obstruction of variable width crossing the center of the aperture of the beam. On the left is a scissor-like mechanism, and on the right is a rotating vane which resembles a Venetian-blind.

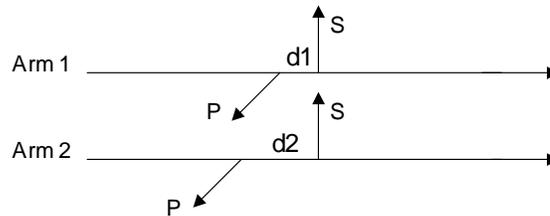


Figure 16.14: Diagram illustrating “differential birefringence”, $d_2 - d_1$. In this example, the s-waves have been phased up, but the p-waves are out of phase by $d_2 - d_1$. Alternatively, by adding a delay of this magnitude to arm 1, the p-waves could be phased up, but at the cost of the s-waves going out of phase by the same amount.

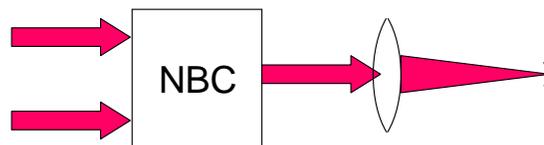


Figure 16.15: Schematic showing the use of a single-mode spatial filter to clean up the wavefront and deepen the null.

be seen that in the presence of differing s-p phase delays, either one of the two polarization states can be phased up, but not both simultaneously. The limit on the allowable differential birefringence is $4\sqrt{N}$ (Serabyn, 2000).

Satisfaction of the condition of simultaneous nulling across the full beam aperture brings a further experimental requirement, as this condition effectively means translating the aforementioned phase error requirement into a surface accuracy requirement. As a phase accuracy of 10^{-3} radians at a wavelength of $10\ \mu\text{m}$ implies a surface accuracy of order 1 nm, it is clear that this goal cannot be met even with the finest optical surfaces available. Indeed, if S is the beam Strehl ratio, attainable null depths would be limited to $\sim 1 - S$, or 10^{-2} . Another approach is therefore necessary. However, this limitation can be overcome (Shao, 1991; Shao and Colavita, 1992; Ollivier and Mariotti, 1997) by means of spatial filtering in the output focal plane (Figure 16.15), as the center of the focal-plane point spread function is the Fourier transform of the *average* aperture-plane field. Thus the focused nuller output beam needs to be passed through a single-mode spatial filter, i.e., a single mode fiber at optical wavelengths, or a pinhole in the mid-infrared. This effectively limits operation to a single diffraction-limited mode of the telescope aperture, and so an analogy to single-dish, single-detector radio astronomy is not inappropriate.

Finally, the need to satisfy all of the above conditions across the passband of interest calls for a minimization of dispersion. In particular (Serabyn, 2000), the mean square phase dispersion across the passband also cannot exceed $2\sqrt{N}$. This implies mainly careful matching of the beam splitter/compensator pair, although in laboratory experiments, the injection of the light must also be carried out with minimal dispersion. The effects of atmospheric dispersion must also be considered in this regard.

16.5 Rooftop-Based Rotational Shearing Interferometers

The use of a fiber-coupled rotational shearing interferometer (RSI) as a nulling beam combiner bears further examination. As pointed out in Figure 16.8, a single rooftop mirror acts to reverse only one component of the incident field—that component perpendicular to the rooftop apex. With two orthogonal rooftop mirrors located behind a beam splitter, as in Figure 16.7, both polarization components are then flipped relative to each other (Figure 16.16). In addition to flipping the \mathbf{E} -fields, the rooftop mirrors also shear the return beams symmetrically across the rooftop centerlines (Figure 16.17), so that two input beams yield two nulled and two bright (or constructive) output beams (Figure 16.7), all of which are separated from the input beams. Thus, in comparison to the phase-retardation approach of Figure 16.10, the number of output beams is doubled. The RSI approach thus appears somewhat more complicated, but on the other hand, it cleanly separates the field-flip and phase-delay issues, and also brings with it the ability to implement a pathlength control scheme in which one nulling output can be used to control the second (Serabyn, 1999).

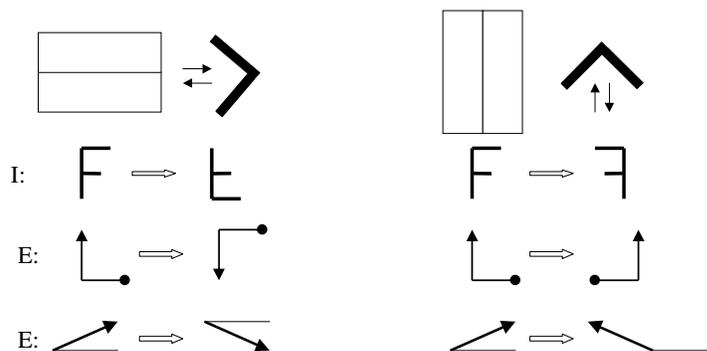


Figure 16.16: Illustration of the action of two orthogonal rooftop mirrors. I refers to the image, and E to the E-field.

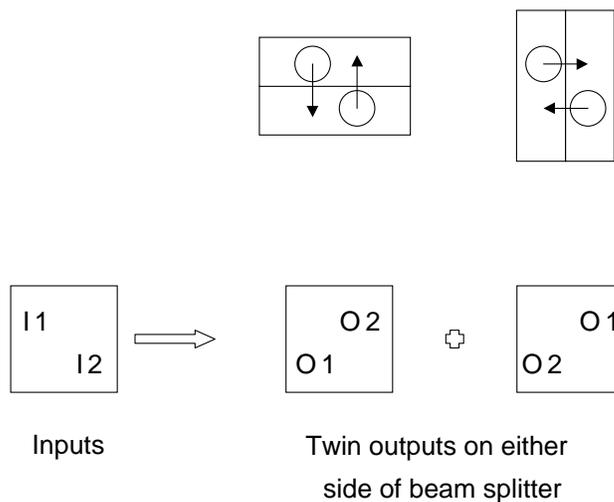


Figure 16.17: Illustration of the lateral beam shear obtained in a beam combiner based on an orthogonal-rooftop RSI.

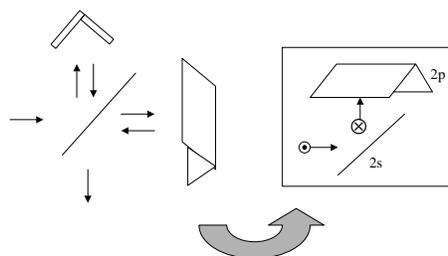


Figure 16.18: Diagram illustrating the fold mirrors needed to symmetrize the two arms of an orthogonal rooftop-based RSI.

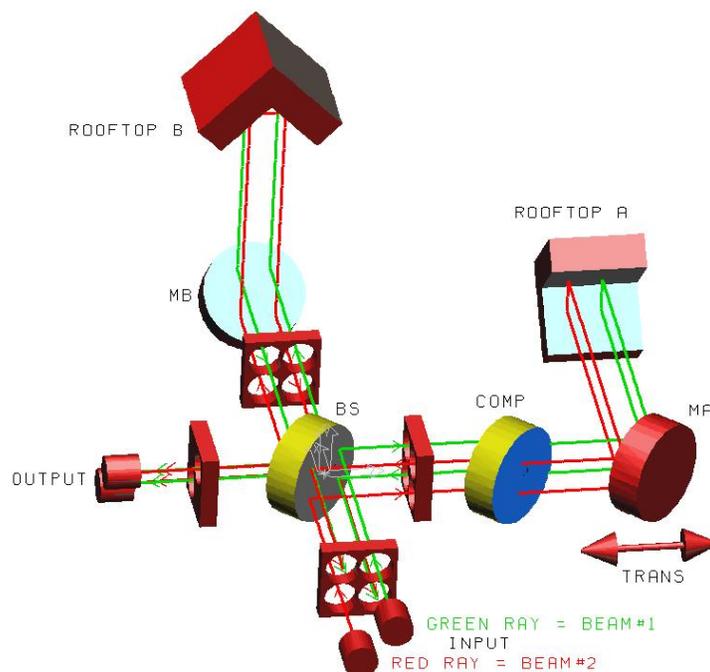


Figure 16.19: Optical layout of a polarization-compensated rooftop-based RSI. The two 45° fold mirrors, MA & MB, are needed to obtain complete symmetry between s- and p-reflections. The aperture plates with the four holes are guides for the eye.

However, one problem with the basic orthogonal-rooftop RSI shown in Figure 16.7 is that a significant asymmetry exists: light of a given polarization state (s- or p-plane polarized at the beam splitter) undergoes two s-plane reflections at one of the rooftop mirrors, but two p-plane reflections at the other. Thus incident light in the two polarization states would emerge with nonzero and opposite s-p phase delays. Luckily, it is possible to symmetrize the situation by inserting properly oriented 45° fold mirrors prior to the rooftops, as in Figure 16.18, so that light in each arm of the modified RSI sees two s-plane and two p-plane reflections, all at the same 45° angle of incidence (Shao, 1991; Diner *et al.*, 1991). This results in the overall optical layout shown in Figure 16.19.

16.6 Experimental Results

A fiber-coupled polarization-compensated rotational-shearing interferometer of the type described in the last section has been built at JPL (Figure 16.20) in order to test the potential of the approach. To date, the experiments at visible wavelengths have proven quite successful, verifying essentially all aspects of the fiber-coupled RSI approach except for dual-polarization operation. Indeed, after many cycles of component and environmental improvements, null depths of a part in 10^5 are now achieved regularly with a narrowband (0.5%) visible-wavelength laser diode source (Figure 16.21), and white light nulls of order a part in 10^4 have been achieved for 10% bandwidth, single-polarization thermal (red) light

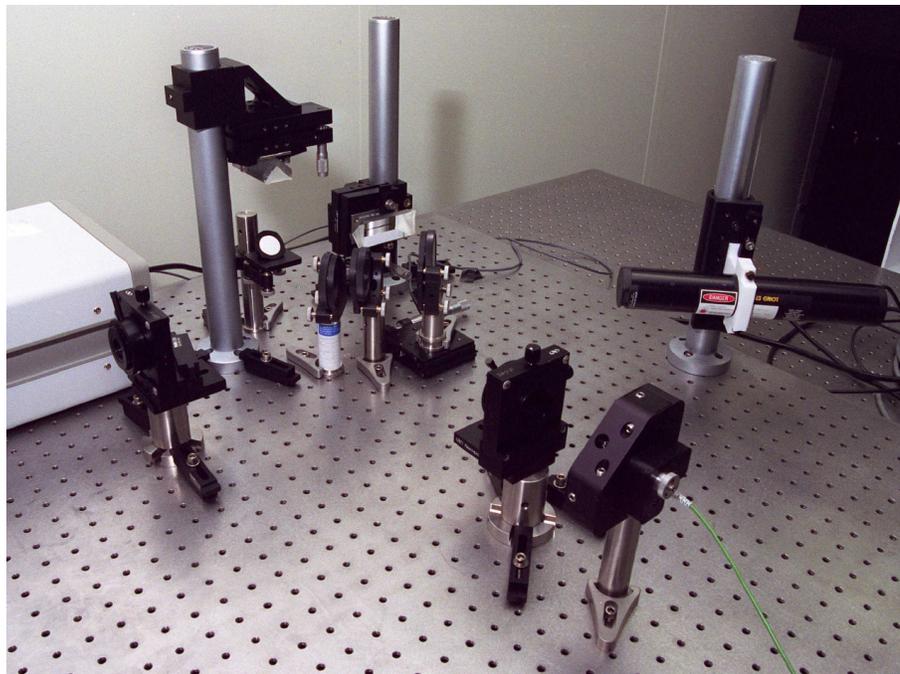


Figure 16.20: Experimental table-top fiber-coupled nuller developed at JPL. The orientation is similar to the previous figure. The input fiber and collimating lens are at the lower right, the two rooftops toward the center and upper left, and the output lens toward the left.

(Serabyn *et al.*, 1999a,b). Stabilization of the null to a part in 10^4 has also been achieved with simple path-length dithering (Serabyn *et al.*, 1999a), implying a path length stability of about 1 nm, or $\lambda/600$.

The white light nulls achieved to date in these ongoing experiments are already significantly deeper than a standard Michelson interferometer could provide, thus verifying the achromatic nature of the null provided by the field-flip approach. The experiments have also verified the need for spatially filtering the nulled output, as coupling to an output fiber indeed improves the null by several orders of magnitude. Finally, the requisite nanometer-level path-length stability (for 10^{-6} nulls in the mid-infrared) has also been demonstrated. Thus, these experiments have largely verified the viability of the fiber-coupled-RSI approach, with only simultaneous dual-polarization nulling remaining in need of demonstration. Of course, much work remains to be done, including broadening the band over which deep nulling occurs, deepening the nulls, stabilizing the nulls for much weaker input signal levels, and demonstrating all of the above in the mid-infrared.

The first nulling experiments to be carried out on a telescope were in fact in the infrared (Baudoz *et al.*, 1998b; Hinz *et al.*, 1998), because of the higher wavefront qualities available. Transient destructive interference of stellar light to residuals of about 5% was achieved

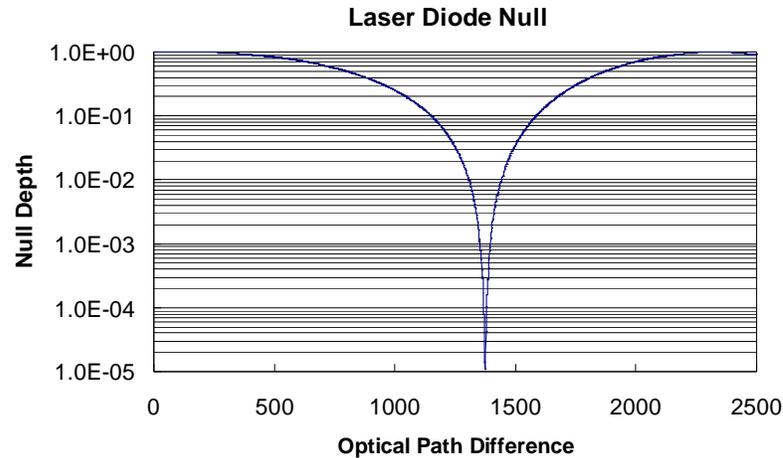


Figure 16.21: Experimental results from JPL's fiber-coupled RSI. An OPD scan through the expected central null fringe shows a null of 1×10^{-5} . The laser diode bandwidth is 0.5%.

with both the cat's eye and phase retardation approaches. The level of stellar rejection achieved in both cases is consistent with limitations imposed by optical surface quality and atmospheric issues.

16.7 Multi-Baseline Nulling

The preceding sections have focused largely on the optical physics of the nulling process. However, there is one further aspect to consider, that being whether a two-telescope, i.e., single-baseline, interference pattern is sufficient to successfully detect planetary companions to nearby stars. In particular, two goals are in opposition: to null a star out to its rim, a deep and wide central fringe is desirable, suggesting short baselines. On the other hand, short baselines do not provide sufficient angular resolution to resolve near-in planets from possibly bright exozodiacal emission. Thus, to generate both a deep and wide central null fringe for stellar nulling, and to retain high angular resolution off the central null, multi-baseline nulling configurations have been considered (e.g. Angel, 1990; Angel and Woolf, 1997a,b; Beichman *et al.*, 1999; Leger *et al.*, 1996; Mennesson and Mariotti, 1997; Velusamy *et al.*, 1999; Woolf and Angel, 1997). The basic idea is to combine the outputs of single baseline nulls to generate cancellation near the optical axis to higher powers of the off-axis field angle, θ , than the θ^2 nulls that a single baseline can supply.

Several specific cases have been developed which are capable of modifying the basic single baseline null to broader and more effective θ^4 , θ^6 , etc. nulls (Angel, 1990; Woolf and Angel, 1997; Mennesson and Mariotti, 1997). One particularly promising configuration is the combination of short nulling baselines with longer post-nulling imaging baselines (Velusamy *et al.*, 1999). Another promising variant is the reflection-asymmetric fringe patterns provided by the modulation scheme of Angel and Woolf (1997a). A recent tally of

configuration concepts is provided by Lawson *et al.* (1999). However, the general question of the optimal nulling configuration remains without a clear answer, as the field is still quite young, with novel configurations no doubt awaiting discovery. Indeed, envisaged space missions based on nulling interferometry, such as NASA's proposed Terrestrial Planet Finder (Beichman *et al.*, 1999), and ESA's proposed Infrared Space Interferometer (Darwin) mission (Leger *et al.*, 1996) are still in very early developmental phases, with no clear favorite for the configuration as yet. Thus, given recent progress in the experimental demonstration of deep nulling, such configuration level questions can now be considered to be one of the most pressing unresolved issues in the development of nulling into a tool for exozodiacal light and extrasolar planet detection.

Acknowledgments

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

- J.R.P. Angel, A.Y.S. Cheng, and N.J. Woolf, "A space telescope for infrared spectroscopy of earth-like planets," *Nature* **322**, 341–343 (1986).
- J.R.P. Angel, "Use of a 16 m telescope to detect earthlike planets," in *The Next Generation Space Telescope*, P.Y. Bely, C.J. Burrows, and G.D. Illingworth, eds., Proc. NASA/STSI Workshop, 81–95 (Baltimore: Space Telescope Institute, 1990).
- J.R.P. Angel and N.J. Woolf, "An imaging nulling interferometer to study extrasolar planets," *Astrophys. J.* **475**, 373–379 (1997a).
- J.R.P. Angel and N.J. Woolf, "The Large Binocular Telescope: a unique scientific and technology precursor to Planet Finder," in *Planets Beyond the Solar System and the Next Generation of Space Missions*, D.R. Soderblom, ed., ASP Conf. Ser. **119**, 207–222 (Provo, Utah: Brigham Young University, 1997b).
- R. Angel, "Sensitivity of nulling interferometers to extra-solar zodiacal emission (EZE)," in *Exozodiacal Dust Workshop*, D.E. Backman, L.J. Caroff, S.A. Sanford, and D.H. Wooden, eds. NASA/CP, **1998-10155**, 209–218 (1998).
- D.E. Backman and F. Paresce, "Main-sequence stars with circumstellar solid material: the Vega phenomenon," in *Protostars and Planets III*, F.H. Levy and J.I. Lunine, eds., 1253–1304 (Tucson, AZ: Univ. of Arizona Press, 1993).
- P. Baudoz, J. Gay, and Y. Rabbia, "Interfero-coronagraphy: a tool for detection of faint companions," in *Brown Dwarfs and Extrasolar Planets*, R. Rebolo, E.L. Martin, and M. Osorio, eds., ASP Conf. Ser. **134**, 254–261 (Provo, Utah: Brigham Young University, 1998a).
- P. Baudoz, Y. Rabbia, J. Gay, E. Rossi, L. Petro, S. Casey, P. Bely, R. Burg, J. MacKenty, B. Fleury, and P.-Y. Madec, "First results with the achromatic interfero coronagraph," in *Adaptive Optical System Technologies*, D. Bonaccini and R.K. Tyson, eds., Proc. SPIE **3353**, 455–462 (1998b).

- C.A. Beichman, N.J. Woolf, and C.A. Lindensmith, eds., *Terrestrial Planet Finder* (Pasadena, California: Jet Propulsion Laboratory, 1999).
- A.J. Booth, M.M. Colavita, M. Shao, P.N. Swanson, G.T. van Belle, S.L. Crawford, D.L. Palmer, L.J. Reder, E. Serabyn, M.R. Swain, G. Vasisht, and J.K. Wallace, “The Keck Interferometer: Instrument overview and proposed science,” in *Working on the Fringe: Optical and IR Interferometry from Ground and Space*, S. Unwin and R. Stachnik, eds., ASP Conf. Ser. **194**, 256–263 (Provo, Utah: Brigham Young University, 1999).
- M. Born and E. Wolf, *Principles of Optics*, Section 8.8.4, 6th edn. (Oxford, UK: Pergamon Press, 1980).
- R.N. Bracewell, “Detecting nonsolar planets by spinning infrared interferometer,” *Nature* **274**, 780–781 (1978).
- R.N. Bracewell and R.H. MacPhie, “Searching for nonsolar planets,” *Icarus* **38**, 136–147 (1979).
- M.M. Colavita, A.F. Boden, S.L. Crawford, A.B. Meinel, M. Shao, P.N. Swanson, G.T. van Belle, G. Vasisht, J.M. Walker, J.K. Wallace, and P.L. Wizinowich, “The Keck Interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg, ed., Proc. SPIE **3350**, 776–784 (1998).
- D.J. Diner, “IBIS: an interferometer-based imaging system for detecting extrasolar planets with a next generation space telescope,” in *The Next Generation Space Telescope*, P.Y. Bely, C.J. Burrows, and G.D. Illingworth, eds., Proc. NASA/STSI Workshop, 133–141 (Baltimore: Space Telescope Science Institute, 1990).
- D.J. Diner, E.F. Tubbs, S.L. Gaiser, and R.P. Korechoff, “Infrared imaging of extrasolar planets,” *J. British Interplanetary Soc.* **44**, 505–512 (1991).
- P.M. Hinz, J.R.P. Angel, W.F. Hoffmann, D.W. McCarthy Jr., P.C. MacGuire, M. Cheselka, J.L. Hora, and N.J. Woolf, “Imaging circumstellar environments with a nulling interferometer,” *Nature* **395**, 251–253 (1998).
- P.M. Hinz, J.R.P. Angel, N.J. Woolf, W.F. Hoffman, and D.W. McCarthy Jr., “Imaging extrasolar systems from the ground: the MMT and LBT nulling interferometers,” in *Working on the Fringe: Optical and IR Interferometry from Ground and Space*, S. Unwin and R. Stachnik, eds., ASP Conf. Ser. **194**, 401–407 (Provo, Utah: Brigham Young University, 1999).
- P.R. Lawson, P.J. Dumont, and M.M. Colavita, “Interferometer designs for the Terrestrial Planet Finder,” in *Working on the Fringe: Optical and IR Interferometry from Ground and Space*, S. Unwin and R. Stachnik, eds., ASP Conf. Ser. **194**, 423–429 (Provo, Utah: Brigham Young University, 1999).
- A. Leger, J.-M. Mariotti, B. Mennesson, M. Ollivier, J.L. Puget, D. Rouan, and J. Schneider, “Could we search for primitive life on extrasolar planets in the near future?: The DARWIN project,” *Icarus* **123**, 249–255 (1996).
- B. Mennesson and J.-M. Mariotti, “Array configurations for a space infrared nulling interferometer dedicated to the search for earthlike extrasolar planets,” *Icarus* **128**, 202–212 (1997).
- R.M. Morgan and J.H. Burge, “Initial results of a white light nulled fringe,” in *Working on the Fringe: Optical and IR Interferometry from Ground and Space*, S. Unwin and R. Stachnik, eds., ASP Conf. Ser. **194**, 396–400 (Provo, Utah: Brigham Young University, 1999).

- M. Ollivier and J.-M. Mariotti, “Improvement in the rejection rate of a nulling interferometer by spatial filtering,” *Appl. Opt.* **36**, 5340–5346 (1997).
- W.T. Reach, B.A. Franz, J.L. Weiland, M.G. Hauser, T.N. Kelsall, E.L. Wright, G. Rawley, S.W. Stemwedel, and W.J. Spiesman, “Observational confirmation of a circumsolar dust ring by the COBE satellite,” *Nature* **374**, 521–523 (1995).
- E. Serabyn, “Nanometer-level path-length control scheme for nulling interferometry,” *Appl. Opt.* **38**, 4213–4216 (1999).
- E. Serabyn, J.K. Wallace, G.J. Hardy, E.G.H. Schmidtlin, and H.T. Nguyen, “Deep nulling of visible laser light,” *Appl. Opt.* **38**, 7128–7132 (1999a).
- E. Serabyn, J.K. Wallace, H.T. Nguyen, E.G.H. Schmidtlin, and G.J. Hardy, “Nulling interferometry: working on the dark fringe,” in *Working on the Fringe: Optical and IR Interferometry from Ground and Space*, S. Unwin and R. Stachnik, eds., ASP Conf. Ser. **194**, 437–442 (Provo, Utah: Brigham Young University, 1999b).
- E. Serabyn, “Nulling interferometry: symmetry requirements and experimental results,” in *Interferometry in Optical Astronomy*, P.J. Léna and A. Quirrenbach, eds., Proc. SPIE **4006**, 328–339 (2000).
- E. Serabyn, M.M. Colavita, and C.A. Beichman, “Exozodiacal disk detection potential with the Keck Interferometer,” in *Thermal Emission Spectroscopy of Dust, Disks, and Regoliths*, M.L. Sitko, A.L. Sprague, and D.K. Lynch, eds., ASP Conf. Ser. **196**, in press (Provo, Utah: Brigham Young University, 2000).
- M. Shao, “Direct IR interferometric detection of extra solar planets,” in *The Next Generation Space Telescope*, P.Y. Bely, C.J. Burrows, and G.D. Illingworth, eds., Proc. NASA/STSI Workshop, 160–168 (Baltimore: Space Telescope Institute, 1990).
- M. Shao, “Hubble extra solar planet interferometer,” in *Space Astronomical Telescopes and Instrumentation*, P.Y. Bely and J.B. Breckinridge, eds., Proc. SPIE **1494**, 347–356 (1991).
- M. Shao and M.M. Colavita, “Long-baseline optical and infrared stellar interferometry,” *Ann. Rev. Astron. Astrophys.* **30**, 457–498 (1992).
- W.A. Traub, N.P. Carleton, and J.R.P. Angel, “On the detection of exo-zodiacal light by nulling interferometry with the Magellan telescopes,” in *Science with the VLT Interferometer*, F. Paresce, ed., 80–85 (Berlin: Springer-Verlag, 1996).
- T. Velusamy, C.A. Beichman, and M. Shao, “A dual 3-element nulling interferometer for TPF,” in *Working on the Fringe: Optical and IR Interferometry from Ground and Space*, S. Unwin and R. Stachnik, eds., ASP Conf. Ser. **194**, 430–436 (Provo, Utah: Brigham Young University, 1999).
- N. Woolf and J.R. Angel, “Astronomical searches for earth-like planets and signs of life,” *Ann. Rev. Astron. Astrophys.* **36**, 507–537 (1998).
- N.J. Woolf and J.R.P. Angel, “Planet Finder options I: new linear nulling array configurations,” in *Planets Beyond the Solar System and the Next Generation of Space Missions*, D.R. Soderblom, ed., ASP Conf. Ser. **119**, 285–292 (Provo, Utah: Brigham Young University, 1997).

Part VI

Stellar Astronomy and Astrophysics

Chapter 17

Binary Stars

Five (5) Reasons Why the Most Interesting, Most Exciting, & Most Important Objects to Observe (Interferometrically or Otherwise) are Binary Stars

WILLIAM I. HARTKOPF

U.S. NAVAL OBSERVATORY
WASHINGTON, DC

These course notes are primarily concerned with answering five of the six standard “reporter’s questions”—the Who, What, Where, When, and How of the field. In this Chapter we’re attempting to answer the last of the questions—the “Why.” Hal McAlister, in Chapter 1, discussed some of the whys, but it’s worthwhile to repeat and expand upon some of the topics he discussed. After all, the most important aspect of this field isn’t the great hardware or clever software being developed; in the end it’s the science which will come out of it.

My title is rather tongue in cheek, but I hope to convince you that the study of binary stars is indeed an important and worthwhile topic. I might note that binary stars may be considered the reason behind the existence of CHARA and the CHARA Array. Hal McAlister became involved in the study of binary stars using the technique of speckle interferometry in the mid-1970s. The success of this speckle effort led to the creation of the CHARA research group and later to dreams of higher resolution; it also gave CHARA the credentials to raise the money for its array project. The goals of the CHARA Array go far beyond just binary stars, of course, but they are an important part of its history (and future).

So, what are the five reasons? (Let me note that at no time do I say these are the *only* five reasons!) Much of what I discuss will be familiar, but I want to remind you of these topics and give examples of the contribution interferometry can make to each.

17.1 Reason One: Binaries as Scales

This is arguably the most important reason, since mass is **the** fundamental quantity which determines a star’s luminosity, size, lifetime, heavy-element generation, and ultimate fate. However, you cannot determine the mass of a single star any more than you can directly measure your weight without the use of bathroom scales or some other means of measuring the gravitational force you exert. To determine stellar masses, then, we need to derive the orbital motions of binary stars.

But why is interferometry important in binary star work? The answer is a two-parter:

17.1.1 Part 1

No single technique for studying binary star orbits gives us all the information we need. For example, an astrometric or “visual” orbit (I will use these terms interchangeably) yields the elements P , a'' , T , and e which define the size and shape of the orbit and the rate of motion of the stars, plus the three angles i , Ω , and ω which define the orientation of the orbit in space. However, Kepler’s Third Law requires the linear separation a between the stars, rather than the angular separation.

On the other hand, a spectroscopic orbit yields P and $a \sin i$ ($a_1 \sin i$ and $a_2 \sin i$ if it’s a double-lined spectroscopic binary, or SB2). We now get the linear separation, but convolved with the inclination.

We therefore need to gather data by at least two complementary techniques. For example, distance + astrometric orbit yields a , thus the mass sum. More on this combination later. A particularly useful combination is spectroscopic + astrometric data, which (if SB2) will give us individual masses.

17.1.2 Part 2

Different observing techniques are applicable in different separation or period regimes. Consider the two techniques just discussed:

- Astrometric observations—measuring separations and position angles—require binaries which are wide enough to be seen as separate images. This means wide, mostly long-period systems.
- Spectroscopic observations—measuring Doppler shifts of spectral lines—requires radial velocities high enough to noticeably move those spectral lines. This means close, short-period systems.

The histogram in Figure 17.1 illustrates the problem. Here I have tallied orbits from the spectroscopic orbit catalogue of Batten *et al.* (1978; data from Griffin 1992), as well as orbits

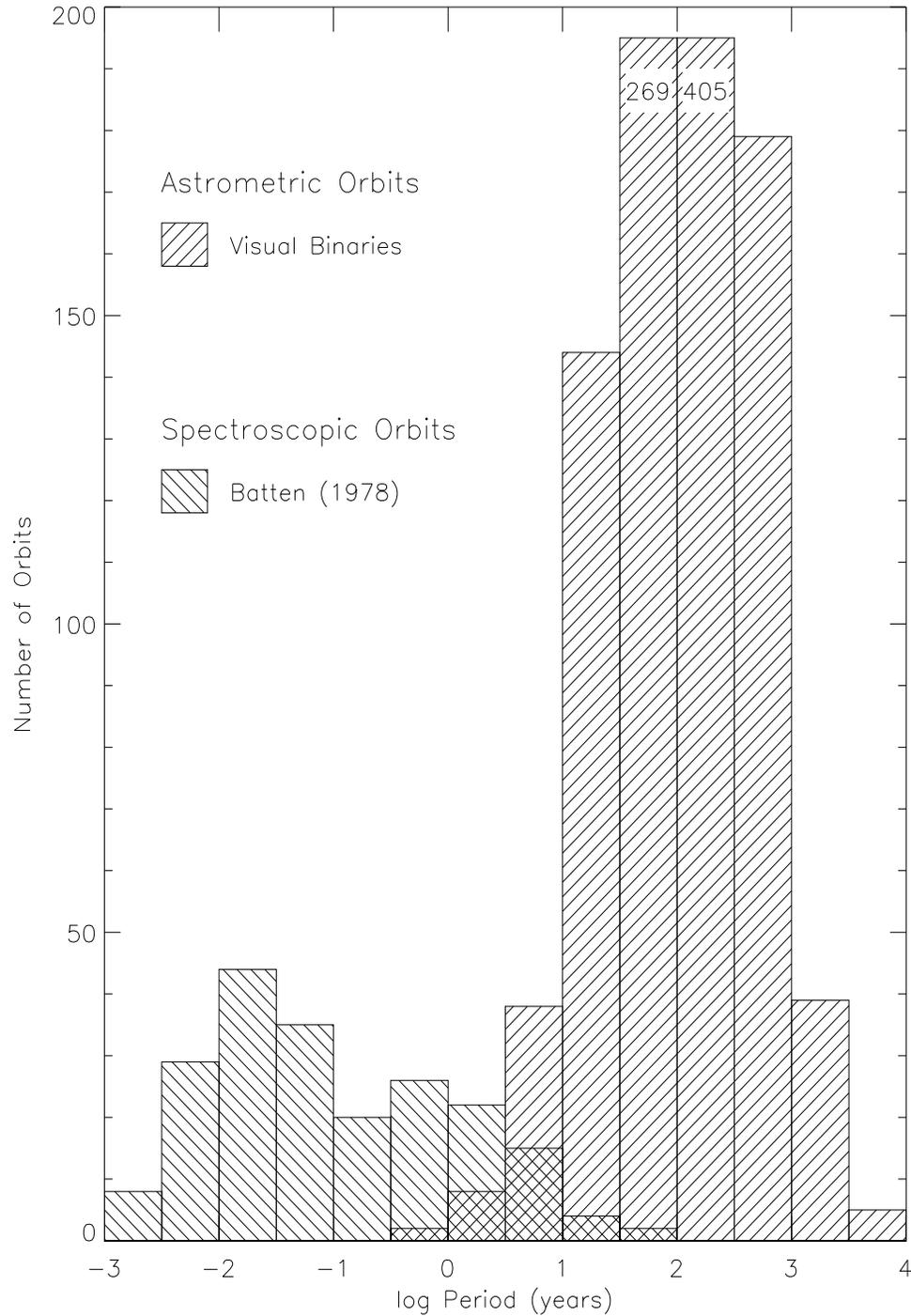


Figure 17.1: Published binary star orbits, binned by period. Shown here are spectroscopic orbits from Batten *et al.* (1978, poorer quality orbits removed), as well as astrometric orbits of all visual binaries in the USNO orbit database. Two columns of visual orbits were truncated for display purposes; the numbers near the top of these columns indicate the true number of published orbits in these period ranges.

of visual binaries from the current catalogue maintained at the USNO. (Note: the poorer quality Batten orbits—grades 4 and 5—have been removed from this histogram. Poorer quality visual orbits have not been removed, however, as most of these orbits have not yet been graded. The poor orbits are generally for systems with extremely long periods—centuries or longer—so their coverages are incomplete.)

We see little overlap. Of course, even this overlap is misleading, since these are not all stars in common to both lists. Dimitri Pourbaix (1998) searched the literature for visual binaries which were also double-lined spectroscopic systems. His results: there are some 500 SB2s with orbits (many very poor), but only 174 of these with astrometry as well and only 38 with sufficient astrometry for a combined solution. Think of the overlap region, then, as illustrating the upper limit to the number of possible combined-solution systems.

Improvements in spectroscopic equipment and reduction techniques (coravel, other cross-correlation methods) have enabled spectroscopists to measure much smaller velocity shifts, thus longer-period systems. In Figure 17.2, I include orbits derived by Roger Griffin using his cross-correlation technique. This collection is as of 1992, so matters have improved a bit over what is shown here. There is a limit, however! As Griffin (1992) has pointed out, in order for spectroscopists to increase their overlap with visual orbits significantly they must observe objects for a century or more. However, Roger says he plans to be doing other things by then! Most of the improvement must come from the “visual” side—this is where interferometry can make its mark.

Speckle interferometry has been in routine use for about 25 years now, discovering new systems with separations down to a few tens of milliarcseconds (corresponding to periods in the years to decades range). The Mark III did a bit better (periods of weeks to years), although for a small number of stars and over a shorter period of time, and NPOI is beginning to obtain data for much closer systems (periods of days). Orbits of binaries discovered by interferometric techniques have also been added to Figure 17.2. The overlap has improved, but there is still a long way to go. This is where some of the array projects you have seen or heard discussed this week come in. Arrays such as NPOI and CHARA have the capability of resolving essentially all of these SBs (at least of those visible to northern observers). I must stress that the reason for making these observations is not to just create a fat catalog of stellar masses, however! Figure 17.3 shows a mass-luminosity plot (Mason, private communication) for binaries whose masses were derived from speckle orbits. Note the lack of coverage, especially at the low- and high-mass ends. The number of masses known to even 5% accuracy is too small to accurately define the M-L relation for “typical” solar-neighborhood main-sequence stars, let alone see how that relation is effected by, say, metallicity, age, etc. Such knowledge would be of great importance to stellar evolution theorists.

So masses are an obvious reason to study binaries. This brings us to ...

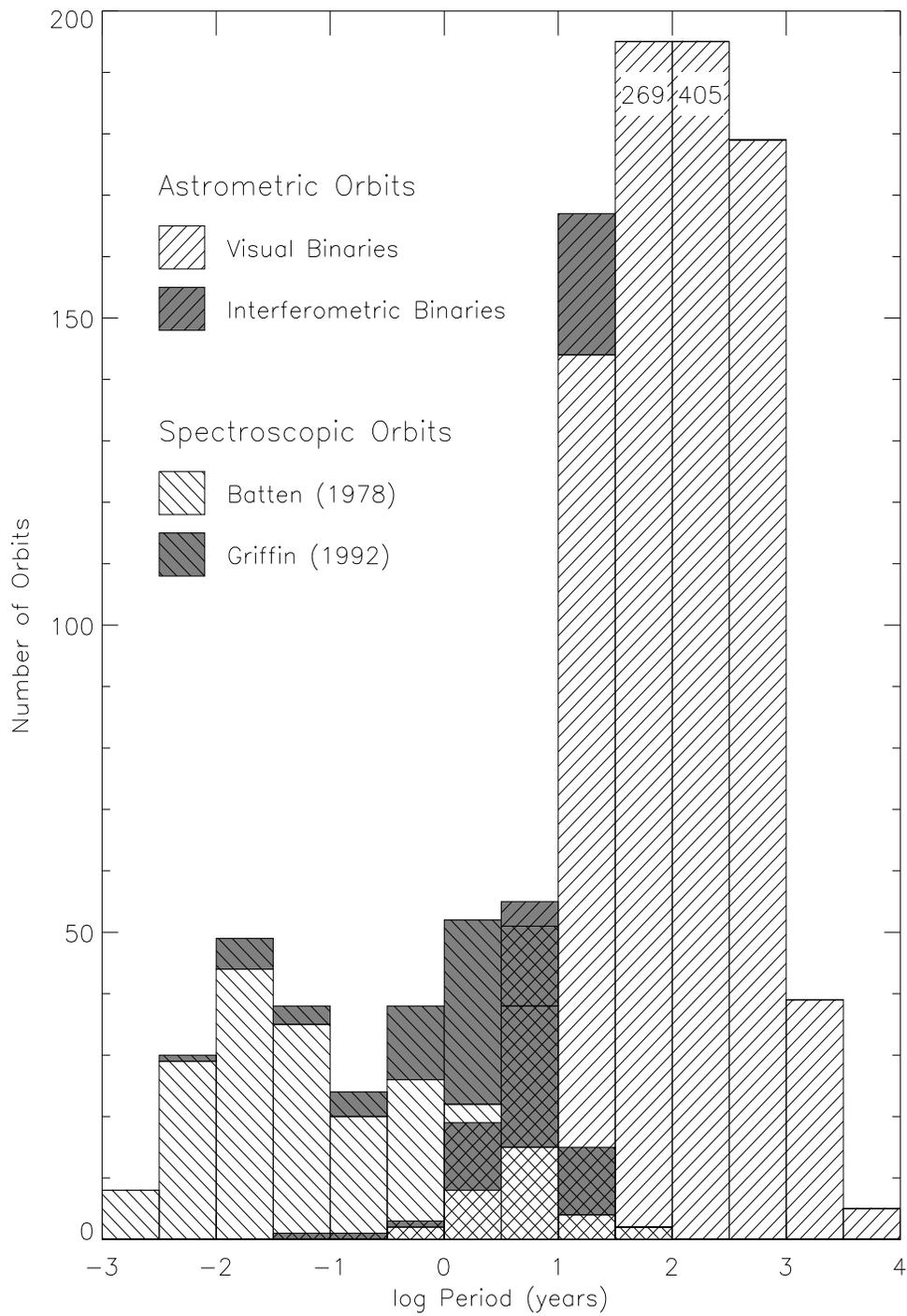


Figure 17.2: Same as Figure 17.1, with the addition of spectroscopic orbits published by Griffin as of 1992, as well as astrometric orbits of binaries discovered by interferometric techniques and tabulated at the USNO. These additional data are shown as shaded hatched regions in the histogram.

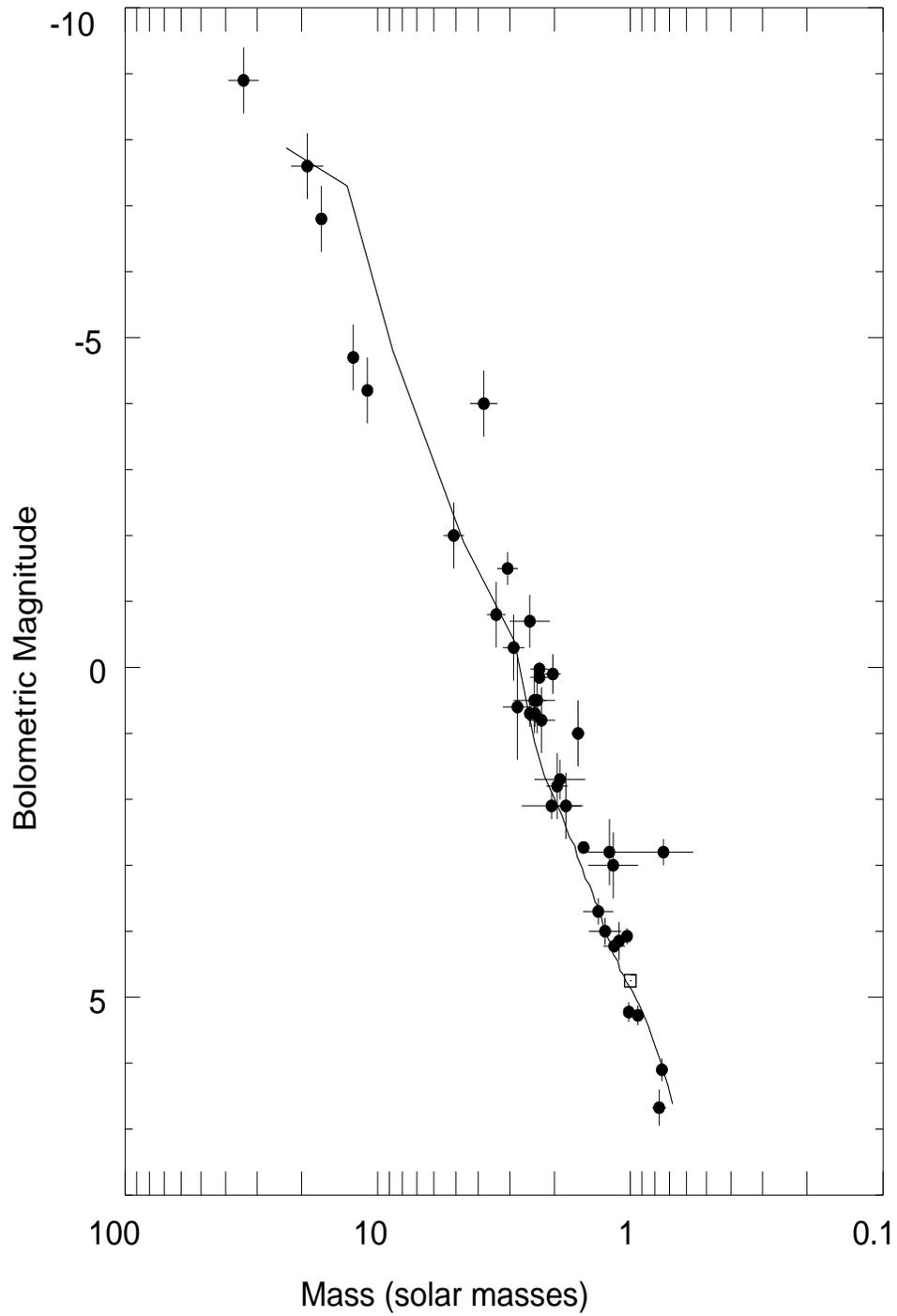


Figure 17.3: Mass–Luminosity diagram, from astrometric orbits defined primarily by interferometric observations (B. Mason, private communication).

17.2 Reason Two: Binaries as Yardsticks

Note that the above-mentioned combination of spectroscopic and astrometric orbital elements yields both a'' and a . From these two values we can immediately derive the binary's distance (this is often expressed as the “orbital parallax”). There are several advantages to this technique: First, no assumptions are needed regarding spectral type, interstellar extinction, etc. (as are needed with methods using stars' apparent magnitudes and colors). This technique works for stars even when standard parallax measurements fail (due to blended images of components). Finally, accuracy is not directly a function of distance.

17.3 Reason Three: Binaries and Stellar Evolution

The term “stellar evolution” is a broad one, and I will only touch on a couple aspects, by asking a few questions:

1. What role does duplicity play in stellar evolution?
2. Are ALL stars created in sets of two or more?
3. Do all stars have a choice—either companions or planetary systems? Can they have both?
4. Do stars of all spectral classifications show similar duplicity rate?
5. How does duplicity change with time? —i.e., once formed, how often are binaries dissociated?

The standard number we all hear is that about half of all stars are actually binaries; in other words, about two-thirds of the stars are members of binary or multiple systems. If a person glanced at the WDS (the Washington Double Star database, repository for essentially all binary star astrometric measurements ever made) they would think we have a pretty good handle on the binary star fraction. The WDS contains 450,000+ observations of ~80,000 binaries, with 200+ years worth of data.

Duplicity surveys are incomplete, however, so the true numbers are not very well known! A quick example: through much of the 1980s the CHARA group attempted to make a speckle survey of the 9,000+ stars in the Bright Star Catalogue (see McAlister *et al.* 1987, 1993). In the process they discovered numerous binaries wide enough to have been seen by visual observers—in other words, naked eye stars never noticed before! Only about one third of the BSC has been surveyed, however, so there are probably dozens of as-yet undiscovered bright companions. Earlier in this decade Hipparcos (ESA 1997) found nearly 3,500 new binaries, many of them also observable visually. Samples of some specific stellar types—bright O stars, Be stars, B giants, a few white dwarfs—have also been surveyed by speckle, but much more is needed. I will return to this point at the end of this talk.

The problem is even worse than that, however. Just as one technique can't provide all the orbital information needed for mass determination, one technique can't make a complete survey for duplicity at all separations. For example, due to both theoretical and equipment limitations, speckle at a 4-m telescope can only detect binaries within the separation range ~ 30 mas to a few arcseconds. A survey using multiple techniques, however, is time- and telescope- and money-consuming. As a result, there have been very few thorough duplicity surveys. I will discuss this further in a moment. Let me return to the questions presented a moment ago and give you one tantalizing result.

Surveys of pre-main sequence (PMS) stars in young star-formation regions [for example, the Taurus–Auriga complex, age 0.002 Gyr, by Mathieu (1994); also other surveys by Ghez *et al.* (1992), Leinert *et al.* (1992), etc.] have found multiplicity rates roughly twice those of their older (~ 5 Gyr) solar-neighborhood counterparts. Patience *et al.* (1998) found that the multiplicity fraction in the Hyades (age 0.7 Gyr) was somewhere in between those of the PMS and older stars. A possible reason—Leonard (1995) has suggested that binary-binary collisions within clusters and associations eject stars, resulting in lower binary frequencies in clusters than in the initial stellar distribution.

A speckle survey of all O stars brighter than $V=8.5$ (Mason *et al.*, 1998a) found a much lower binary frequency for stars in clusters and associations compared to field stars, supporting Leonard's assertion. Just when do these ejections or other binary disruptions occur, however? Little information is known for that age gap from 0.7 and 5 Gyr. Mason *et al.* (1998b) surveyed ~ 200 solar-type stars, combining archival visual micrometry with speckle interferometry to survey the region from about 2 to 120 AU around each star. Chromospheric activity was used as an indication of age to segregate the stars into a couple different groups. Their findings: for a sample of 84 chromospherically active stars (age ~ 1 Gyr) they find a duplicity fraction of 18%. A sample of 118 less-active stars (average age ~ 4 Gyr) yielded a multiplicity fraction of 9%. Although intriguing, the samples are too small to draw many conclusions yet. More complementary data at both closer and wider separations are also needed.

17.4 Reason Four: Binaries in Other Guises

The effects of duplicity are not always obvious! Let me give you an example.

There are a class of variable stars called λ Boo variables, first recognized by Morgan in the early 1940s. As a class they are rather poorly defined—definitions vary from author to author—but are usually denoted as having weak metal lines (especially one Mg II line), while C, N, O, and S are nearly solar in abundance. Most but not all have moderate to high projected rotational velocities. The stellar types are rather uncertain, as well. Farraggiana and Bonnifacio (1999) find several published hypotheses:

1. Very young stars which have not reached the main sequence
2. Main-sequence dwarfs, with ages of 10^7 – 10^9 years, or
3. Quite old objects resulting from mergers of W UMa type binaries.

In other words, the stars have been narrowed down to either pre-MS, MS, or post-MS stars!

In collecting the available literature on λ Boo candidate stars, Farraggiana and Bonnifacio find evidence of duplicity for 1/4 to 1/3 of these stars, largely from speckle or Hipparcos observations. They hypothesize that most λ Boo stars may in fact be normal binary stars, and that the abundance anomalies are simply due to “veiling”—filling in of spectral lines by the other component’s continuum.

How many other types of variable stars are thought to be binaries? I checked Sterkin and Jaschek’s book *Light Curves of Variable Stars*. The results were surprising—some 40 variable classifications! Here they are:

- Eruptive variables:
 - ★ RS CVn — close binaries with H and K Ca II in emission
 - ★ IN(YY) — matter-accreting Orion variables
- Eruptive supernovae and cataclysmic variables:
 - ★ Novae — massive white dwarf/cool dwarf binaries — include fast, slow, very slow, recurrent types
 - ★ Nova-like systems — WD+WD, WD+MS, etc. — include AM CVn, AM Her, DQ Her, UX UMa, VY Scl systems
 - ★ Type I supernovae
 - ★ Dwarf novae or U Gem variables — include SS Cyg, Z Cam, SU UMa, and Z And or symbiotic stars
- Eclipsing variables:
 - ★ EA — Algol types
 - ★ W Ser systems — long-period Algol-like mass-transferring binaries
 - ★ EB — β Lyr types
 - ★ EW — W UMa types
 - ★ GS — have one or more giant components
 - ★ PN — one component is nucleus of a planetary nebula
 - ★ WD — have white dwarf component
 - ★ WR — have Wolf-Rayet component
 - ★ AR — AR Lac type detached systems
 - ★ DM — detached MS systems
 - ★ DS — detached systems with subgiant

- ★ DW — detached systems like W UMa system
- ★ KE — contact systems of early spectral type
- ★ KW — contact systems of late spectral type
- ★ SD — semi-detached systems
- X-ray sources — 9 categories of bursters, novae, pulsars

What can interferometry contribute to the study of these objects? Other speakers this morning will discuss many of these contributions—sizes and shapes of component stars, hot spots and dark spots, limb-darkening and other effects, eventually imaging of accretion disks or matter streams. “Simple” interferometry can yield other useful information as well. Consider a few examples which come to mind (there are MANY more!):

- Masses and distances—basic but essential information, obtained as mentioned in Reason 1. Note that these data can be obtained for other variables in “normal” binaries, as well.
- Knowledge of the orbital inclination gives the exact trajectory of one component during an eclipse. Coupled with photometric and spectroscopic data this may allow one to determine sizes and compositions of extended atmospheres, accretion disks, etc.
- Some longer-term variability is believed to be caused by orbital precession presenting us with different viewing angles of a close binary. Precise astrometry will allow us to measure these precessional changes.

17.5 Reason Five: Binaries as “Vermin”

Some people despise binary stars (poor misguided fools)! These are often people who need point sources for calibration or as guide stars for pointing satellites, etc. As I have been discovering since joining the USNO in the summer of 1999, the celestial reference frame is also of vital importance for guidance of aircraft and missiles. A jet airplane that uses a binary star as a reference point, may have huge navigational errors that arise solely from the changing position of binary’s photocenter. As an example closer to most of our interests, guidance sensors on a satellite such as HST may be unable to lock onto a binary guide star, causing the satellite to waste time searching for another guide star or to just lose an observing cycle altogether.

Surveys for duplicity, then, are important from a technical, non-astronomical standpoint. Consider one example—SIM (the Space Interferometry Mission). This mission involves narrow-angle astrometry of target stars relative to a stable framework of grid stars. For SIM to succeed, it needs some 4,000 grid stars (plus an additional 2,000 backup stars)

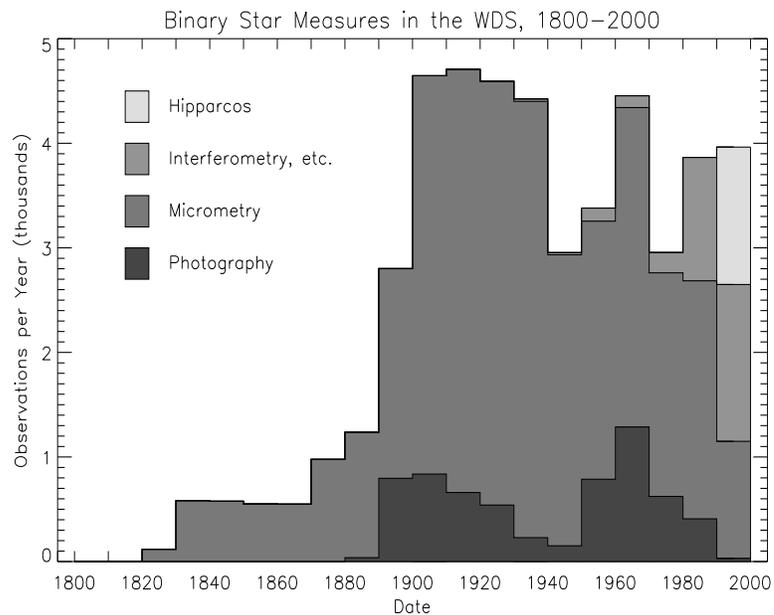


Figure 17.4: Number of binary star measurements per year from 1800 to 2000, as tabulated in the Washington Double Star database. Observations are categorized by observing technique and averaged by decade.

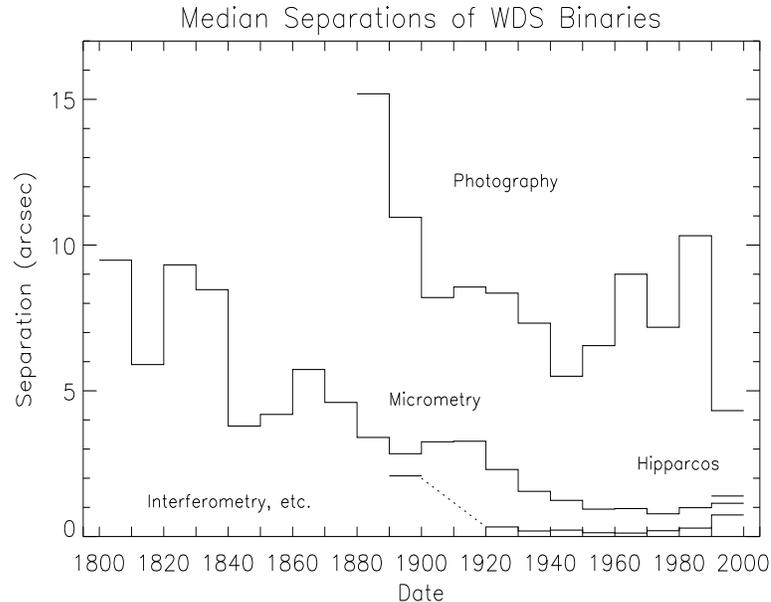


Figure 17.5: Median separation of binaries in the Washington Double Star database, 1800–2000. Observations are again categorized by observing technique and binned by decade. The first interferometric measurements were made by Schwarzschild in the mid-1890s; no further binary star interferometry was done until Anderson and Merrill’s work in the early 1920s—hence the gap in the data.

distributed over the entire sky which are astrometrically stable at the $4 \mu\text{as}$ level over the 5-year lifetime of the project. Obviously, double and multiple stars must be avoided! The USNO has undertaken a project to observe 7,200 stars for duplicity, using a combination of techniques. After checks against the WDS for known doubles, CCD imaging by the USNO astrograph will search for wide pairs, speckle interferometry will look for pairs down to $\sim 30 \text{ mas}$, and an Fourier transform spectrometer currently under construction will look for the closest pairs. Some samples of stars will also be surveyed using NPOI, although this instrument will be limited to the brightest few hundred northern targets.

I should note that interferometry has a distinct advantage in duplicity survey work. Unless a system is double-lined, a single spectroscopic observation won't tell you if a star is single or double—you must make two and probably more observations and look for line shifts. However, unless that system contains a variable star (so Δm becomes too large), a single astrometric observation will usually tell you if the object has a companion within the separation regime accessible to the instrument.

Okay, there we have five reasons why binary stars are great! Let me end with a couple figures to illustrate the current state of affairs regarding binary star observations. The histogram in Figure 17.4 shows the number of binary star measurements obtained per year over the past two centuries. We can see, for example, that the observing rate has remained about constant throughout the 20th century. At first glance this looks fine, until one considers how the number of astronomers has changed during this period! The major thing which I want you to notice, however, is the change in technique. Measurements made photographically or using filar micrometry have plummeted during the past couple decades, as older visual observers retired or died. High-resolution techniques have virtually taken over the field, with 70% of data obtained thus far in the 1990s coming from either Hipparcos or interferometry (almost exclusively speckle interferometry). I expect interferometry to become even more dominant in the next few years.

Figure 17.5 shows both good and bad news. Median separations have for the most part continually decreased for most of this century. The bad news, however, is that median separations for interferometric observations have increased significantly. The reason—little interferometry is now being done at large telescopes. The CHARA speckle effort, which used 4-m and later 2.5-m telescopes, has essentially ended. Virtually all speckle measurements being published now are those obtained at the USNO on its 26-inch refractor. The earlier histogram is rather misleading, then. Yes, interferometry is taking over binary star astrometry, but unless that astrometry can be obtained in significant numbers at larger telescopes or using multi-aperture arrays, many of the benefits of interferometry will not be realized.

References

- European Space Agency, *The Hipparcos and Tycho Catalogues (ESA SP-1200)* (Noordwijk: European Space Agency, 1997).
- A.H. Batten, J.M. Fletcher, and P.J. Mann, "Seventh catalogue of the orbital elements of spectroscopic binary systems," *Pub. Dom. Astrophys. Obs.* **15**, 121 (1978).
- R. Faraggiana and P. Bonnifacio, "How many λ Bootis stars are binaries?" *Astron. Astrophys.* **349**, 521–531 (1999).
- A. Ghez, G. Neugebauer, and K. Matthews, "A two-dimensional near-infrared speckle imaging survey of T Tauri stars in Taurus and Ophiuchus," in *Complementary Approaches to Double and Multiple Star Research*, H.A. McAlister and W.I. Hartkopf, eds., ASP Conf. Ser. **32**, 1–9 (San Francisco: Astronomical Society of the Pacific, 1992), IAU Colloquium 135.
- R.F. Griffin, "Spectroscopic studies of binary and multiple star systems," in *Complementary Approaches to Double and Multiple Star Research*, H.A. McAlister and W.I. Hartkopf, eds., ASP Conf. Ser. **32**, 98–109 (San Francisco: Astronomical Society of the Pacific, 1992), IAU Colloquium 135.
- Ch. Leinert, N. Weitzel, M. Haas, R. Lenzen, H. Zinnecker, J. Christou, S. Ridgway, R. Jameson, and A. Richichi, "A survey of young stars in Taurus for multiplicity," in *Complementary Approaches to Double and Multiple Star Research*, H.A. McAlister and W.I. Hartkopf, eds., ASP Conf. Ser. **32**, 21–29 (San Francisco: Astronomical Society of the Pacific, 1992), IAU Colloquium 135.
- P.J.T. Leonard, "Merged dynamically ejected OB runaway stars," *Mon. Not. R. Astron. Soc.* **277**, 1080–1086 (1995).
- B.D. Mason, D.R. Gies, W.I. Hartkopf, W.G. Bagnuolo Jr., T. ten Brummelaar, and H.A. McAlister, "ICCD speckle observations of binary stars. XIX. An astrometric/spectroscopic survey of O stars," *Astron. J.* **115**, 821–847 (1998a).
- B.D. Mason, T.J. Henry, W.I. Hartkopf, T. ten Brummelaar, and D.R. Soderblom, "A multiplicity survey of chromospherically active and inactive stars," *Astron. J.* **116**, 2975–2983 (1998b).
- R.D. Mathieu, "Pre-main-sequence binary stars," *Ann. Rev. Astron. Astrophys.* **32**, 465–530 (1994).
- H.A. McAlister, W.I. Hartkopf, D.J. Hutter, M.M. Shara, and O.G. Franz, "ICCD speckle observations of binary stars. I. A survey for duplicity among the bright stars," *Astron. J.* **93**, 183–194 (1987).
- H.A. McAlister, B.D. Mason, W.I. Hartkopf, and M.M. Shara, "ICCD speckle observations of binary stars. X. A further survey for duplicity among the bright stars," *Astron. J.* **106**, 1639–1655 (1993).
- J. Patience, A.M. Ghez, I.N. Reid, A.J. Weinberger, and K. Matthews, "The multiplicity of the Hyades and its implications for binary star formation and evolution," *Astron. J.* **115**, 1972–1988 (1972).
- P. Pourbaix, *Simultaneous adjustment of visual and spectroscopic observations of binary stars*, Ph.D. thesis, Université de Liege (1998).
- C. Sterkin and C. Jaschek, *Light Curves of Variable Stars* (Cambridge University Press, 1996).

Part VII

Future Ground-Based Interferometry

Chapter 18

Future Ground-Based Interferometry

STEPHEN T. RIDGWAY

CHARA ARRAY, GEORGIA STATE UNIVERSITY
MOUNT WILSON, CALIFORNIA

As I write, the National Academy of Science Decade Review for Astronomy is in progress. The work has been mostly done and the reports are being written. As input to this once-in-a-decade process, a lot of us working in interferometry thought there should be something on the table for discussion by this group in the area of ground-based interferometry. After asking around, it was clear that nothing was being proposed so François Roddier and I made up some notes and circulated them to colleagues. We received a few responses, and prepared a report which we submitted to the review committee. What I'd like to do in this Chapter is to describe the thinking process we went through and the conclusions we drew, and to lead you through our thoughts concerning what might be done in interferometry in the future. On the scale of the decade review, we won't concern ourselves with minor details but consider the big scale of things.

18.1 Existing Facilities and Scientific Output

The first thing we did was review what has been done in interferometry to date. I have combined in Table 18.1 a couple of tables listing the interferometric facilities that you have been reading about which are in operation, a few you probably haven't heard about, and ones which are coming into operation shortly. It's quite a list and doesn't even include prototype facilities which have closed down, or have done their thing and passed away.

The interferometers listed in Table 18.1 are on the whole, rather special purpose facilities. There isn't anything listed that one would call a general purpose interferometric facility.

Table 18.1: Interferometric Facilities of Today and Tomorrow

Facility	Name	Number of Telescopes	Aperture (meters)	Baseline (meters)
CHARA	Center for High Angular Resolution Astronomy	6	1.00	350
COAST	Cambridge Optical Aperture Synthesis Telescope	5	0.40	20
GI2T	Grand Interféromètre à 2 Télescopes	2*	1.52	65
IOTA	Infrared Optical Telescope Array	2*	0.40	38
ISI	Infrared Spatial Interferometer	2*	1.65	85*
MIRA-I	Mitaka Infrared Array	2*	0.25*	4*
NPOI	Navy Prototype Optical Interferometer	3*	0.12*	35*
PTI	Palomar Testbed Interferometer	3	0.40	110
SUSI	Sydney University Stellar Interferometer	2	0.14	640
Keck	K1–K2	2	10.0	60
Keck	Auxiliary array upgrade (NASA Origins)	4	1.8	140
LBT	Large Binocular Telescope	2	8.4	23
VIMA	VLT Interferometer Main Array	4	8.0	130
VISA	VLT Interferometer Sub-Array	4	1.8	202

Some examples:

1. The Sydney University Stellar Interferometer was built to measure stellar diameters. In fact, it has long baselines—but only on a single line. It was built under the assumption that stars are circular.
2. The CHARA Array is very well laid out for binary star work. It has large baselines, relatively large apertures for sensitivity, but not very good (u, v) coverage because the number of telescopes versus resolution is not very large.
3. The Palomar Testbed Interferometer is well suited for astrometric tests.
4. The Navy Prototype Optical Interferometer, at least the imaging part, is very well suited for imaging stars, but not for particularly faint sources.
5. The Keck and VLTI Interferometers each consist of one or more large telescopes and a few small ones but with very specific capabilities and limitations. Limitations being small baselines, few telescopes, and limited potential for expansion.

Now, with all those facilities, what has been done? Table 18.2 and Figure 18.1 show an overview of the scientific work that has come out of these interferometers. Much of this material has been described in other Chapters, so there aren't any surprises here.

Most of the work has been stellar diameters. In fact, virtually all the work has been diameter-type measurements; that is, characterization by 1 or 1.5 parameters of a distribution of generally circular, stellar, or shell-type objects. This is of course the stellar bright-source astronomy that best suits these facilities.

Table 18.2: Refereed Papers in Optical Interferometry Science

Papers by facility:		Papers by Topic:	
I2T	14	Stellar Angular Diameter	23
ISI	11	Shells of Late Type Stars	11
MkIII	16	Be/P Cyg Shells	8
GI2T	9	Binary Star Orbits	10
IRMA	3	Stellar Atmospheres	6
IOTA	5	Wide Angle Astrometry	2
COAST	3	Novae	1
NPOI	3	Cepheids	1

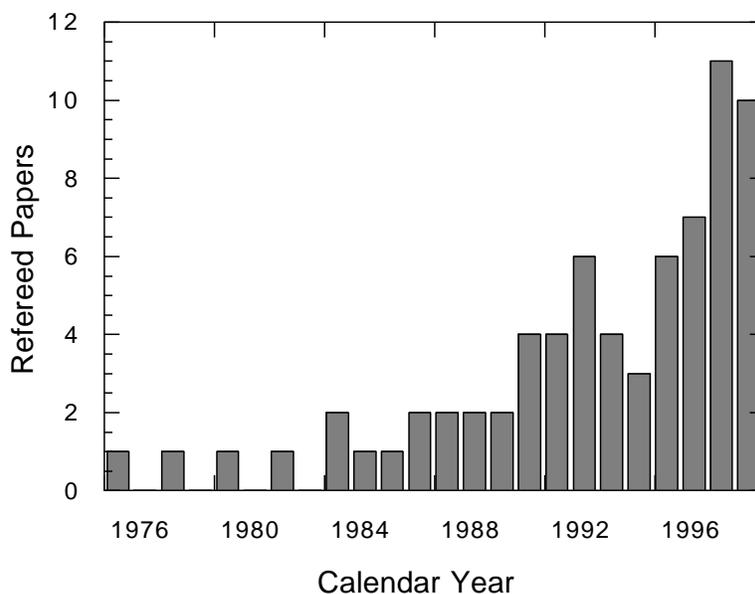


Figure 18.1: Publications of astrophysical results as a function of time.

18.2 Prospects for Astrophysics with Interferometers

We also looked over the literature of the dreams that people have described for interferometry in the future and looked at some of the scientific areas, and tried to lay out, in a visual way, some of the types of sources that one talks about observing.

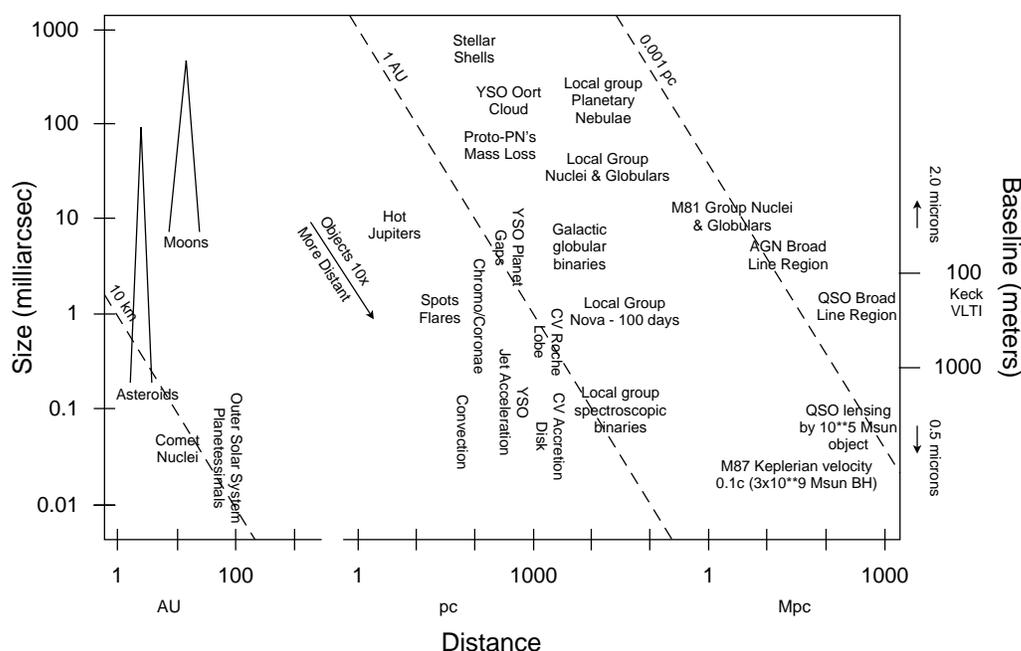


Figure 18.2: Potential astrophysical targets as a function of angular size and distance. [Reprinted from *Allen's Astrophysical Quantities*, Arthur Cox ed. (New York; Springer-Verlag, 2000).]

Figure 18.2 illustrates these thoughts. It is partly astrophysics, partly observational. For various objects the distance, from solar system out to cosmological distances, is shown along with the apparent sizes from microarcseconds to milliarcseconds. On the right-hand side there are tickmarks indicating some typical facilities, with a baseline of 1000 m indicated. Figure 18.2 is scaled for a wavelength of $1 \mu\text{m}$, and there are changes in the scale for corresponding changes in wavelength. It is interesting to note and perhaps obvious, that the capability of high resolution beyond the 10-m aperture range (but less than something like a kilometer) does not have an infinitely broad range, and so not all science falls into an area where that particular resolution range is critical.

These are perhaps some of the typical areas we might consider. Interestingly enough, the targets that tend to push the angular resolution are small sources in the solar system and phenomena on the surface of stars. As an extreme example we might consider observations of convection on the surface of lower-luminosity stars. Phenomena that might require driving the resolution up to high or even implausibly high values for ground-based interferometry include the acceleration of jets around stars, high-energy phenomena in close binaries, and interacting binaries. What I've noted here is the resolution to start to resolve a source, but not necessarily to see lots of details on it. It is also interesting to note that a large number of potential targets fall at baselines which are not much greater than baselines provided by a large telescope.

18.3 Next Generation Ground-Based Array

Let me go through the thinking process of developing a concept for a next generation interferometric array. There are all kinds of directions one could go, for example picking an interferometer from the list of existing facilities and expanding it with more telescopes or longer baselines or putting adaptive optics on it. We thought however that we should focus on something that was more of a quantum step in capability and resolution, adequate to encompass a substantial increment in science—which is to say a capability for substantially better imaging than arrays that are now in development.

18.3.1 Near Infrared Capability

We are discussing a ground-based facility, and so it would have to be a near-infrared facility. We would, of course, need high sensitivity, because so much of astronomy deals with faint sources. If it is to be sensitive, it will have to have adaptive optics, which will work well at infrared wavelengths, but only by a stretch of the imagination perhaps at visible wavelengths. On the other hand, at the longer wavelengths as you get into thermal infrared, interferometry on the ground starts to get really tough because of the large number of reflections, and the thermal background is changing—all kinds of bizarre things going on in the facility. So the interferometer would operate at near-infrared wavelengths.

18.3.2 Baseline of ~ 1000 m

What maximum baseline is reasonable for a next generation ground-based interferometer? We suggest 1000 m, maybe a little more, perhaps a little less. You don't want to be in a site where the atmospheric conditions are actually hostile to astronomy. You would like to be on the best possible site, so 1000 m is a compromise, but which provides a tremendous resolution—straight brute force resolution of $200 \mu\text{as}$ at $1 \mu\text{m}$. As you have probably seen already, if you want to simply parameterize a source, let's say a characteristic diameter, with a baseline of 1000 m you can easily expect to go a factor of 10 beyond simply resolving it. We're talking about measuring sources of $20 \mu\text{as}$ size; that is a substantial change.

18.3.3 General Array Capabilities

Now in order to attack a wide range of science, we have to imagine a dual-beam facility so that it can be phased on offset reference stars and integrate on faint sources.

It would have to be configurable. Although, with a large number of telescopes, you might be able to arrange them such that they are well disposed for one science problem, you would have to reconfigure them for a target with an angular size 10 times smaller or larger. Examples that come to mind include the GI2T at the Observatoire de la Côte d'Azur, or the VLA where the telescopes move happily on rails.

The array would have to be a facility dedicated to interferometric observations, and not an array attached to a bigger facility which has a bigger use, such as the Keck or VLT observatories.

The array would have to be affordable, and would probably be a multi-national collaboration.

18.3.4 Number of Telescopes and Array Geometry

An imaging capability requires quite a few telescopes. There are a number of interferometric arrays in progress now with ~ 6 telescopes, which is far fewer telescopes than we would consider.

It would also have to have a snapshot capability with good (u, v) coverage in an instantaneous mode. That implies a large number of telescopes, baselines, and closure phases from one moment to the next. We had a discussion of how rapidly things can be expected to change with these high resolutions, and so you're now in the realm where sources change structure overnight. Furthermore, whereas with the VLA it's customary to build up (u, v) coverage by tracking down to the horizon, that's not as useful for optical arrays where the atmosphere is such a strong factor. So you can't rely on supersynthesis to the extent you can with radio.

We can say a few more things about what an array like this probably would look like. By now you probably understand the problems, probably have read about it repeatedly, involved in phasing an array on a resolved source. If you have a source which is approximately a disk resolved at a baseline Δb , and you'd like to further resolve 10 pixels across the disk, you need to observe the fringe pattern out to baseline b , where $b/\Delta b = 10$. However, if this is a faint source, you're going to have a hard time tracking fringes at baselines comparable to b , so you now require that no two telescopes be separated by more than a distance Δb . Therefore when you start laying out the array of telescopes, you must keep in mind the ratio $b/\Delta b$, and keep to a relation something like

$$\frac{\pi b}{N} \simeq \Delta b, \quad (18.1)$$

where N is the number of telescopes. Now if you want $b/\Delta b = 10$, you need ~ 30 telescopes to get 10 pixels across your source. It follows that the number 27, even though it is the same as the VLA uses, isn't arbitrary. The configuration of the telescope layout then quite logically drops out to be a circle, because that gives you the short baselines all the way around including an extra closure as you come back to where you started, and the advantages of that have been recognized for a long time—although not before the VLA was built. Tim Cornwell, who studied this analytically, described a family of array geometries such as this that are now called Cornwell circles (Cornwell, 1988).

Let me comment on the large baseline and the number of telescopes, and now compare a facility of that description to the other facilities that are existing or impending. Figure

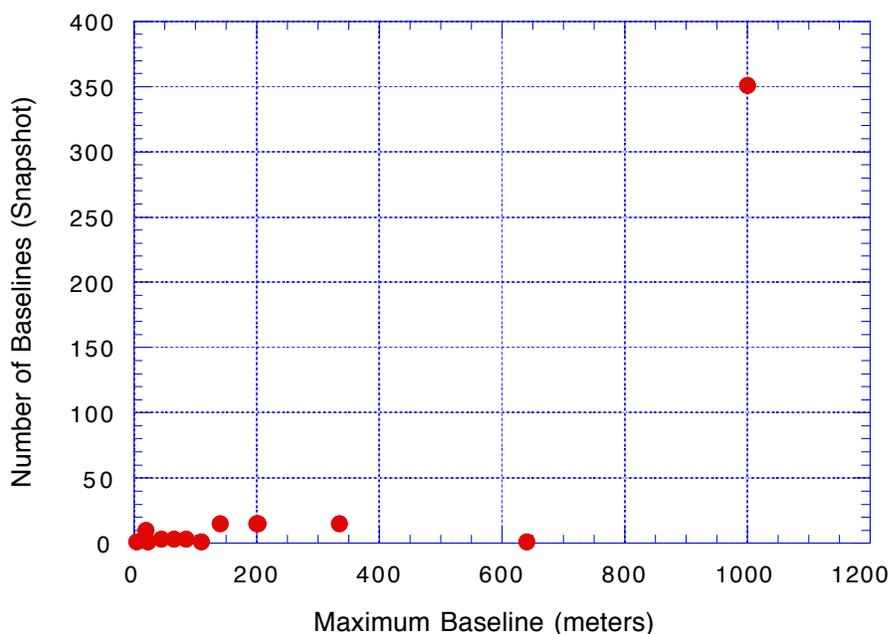


Figure 18.3: A comparison of array characteristic of existing and planned facilities with those of an Infrared Very Large Array.

18.3 illustrates the most characteristic interferometric parameters: the maximum baseline and the number of baselines. The number of baselines goes quadratically with the number of telescopes, and I could have shown the number of phase closures but they would have disappeared along the bottom line. It is obvious that in going from ~ 6 telescopes, which is the number in existing or planned facilities, to 27, makes a huge difference. This is an enormous jump. You don't have to look for subtleties to see advantages in a facility with that capability.

18.3.5 Adaptive Optics and Aperture Size

In order to reach faint sources, each telescope will have to be reasonably large since there have to be enough photons to phase the telescopes into a coherent aperture, assuming the telescopes use adaptive optics. And again, a fair number of telescopes will be required in order to reach faint integrated fluxes.

An aperture size of the order of 3 to 4 m is appropriate for reasons related to the adaptive optics which is now essentially off-the-shelf in that telescope size range. Table 18.3 illustrates the actual demonstrated performance of adaptive optics on the CFHT for example, and the relevant parameters versus wavelength. The performance is essentially the Strehl loss as a function of the reference star magnitude. This is telling you that the current systems can be used to phase-up a 3- or 4-m telescope to a K magnitude of 15 or 16 and gives you an

	ADAPT				PUEO		
	B	V	R	I	J	H	K
Median Strehl =	0.02	0.07	0.15	0.30	0.27	0.41	0.56
Strehl attenuation 50% at magnitude R =		10.0			14.3	15.0	15.7
Strehl attenuation 50% at field angle =					20"	30"	40"

Table 18.3: Demonstrated Performance of Adaptive Optics. Performance of operational natural guide star adaptive optics systems ADAPT at Mount Wilson (Mount Wilson Institute, private communication from C. Shelton), and PUEO (Canada-France-Hawaii Telescope on Mauna Kea, from the CFHT Web pages.)

idea, in what would be a simple-minded sense, of the limiting performance of a facility put together with off-the-shelf parts.

And this can be expected to go a little fainter with further developments in natural guide star adaptive optics. I'm not going to talk about laser-beacon adaptive optics which of course would improve the performance considerably but has not yet been demonstrated.

These are the ideas that we discussed with people in the community. Table 18.4 summarizes the concept that came out as a strawman—no more than that. Something we might call the Infrared Very Large Array, just to give it a designation, with a large total baseline of ~ 1000 m, 27 telescopes with aperture diameters of 3 to 4 m.

18.3.6 Budget Estimate

Let's do a quick budget. Suppose we have, just to round it off, 30 telescopes. Four-meter telescopes sell for about \$20 million now, and we can probably get a 50% discount for building 30 of them—in this calculation the exact discount doesn't matter. We add to that \$100 million for delay lines and beam-combining laboratory for a subtotal now of \$400 million. The facility is probably going to be at a somewhat exotic site so we'll add a factor of 1.5, and we now arrive at a \$600 million project. That's probably right within 50%. So does this budget make sense? Or does it immediately say that this is just a stupid idea? Well, the answer is not necessarily obvious.

18.4 Astrophysical Potential of an Infrared Very Large Array

The answer depends on what science can be done with the facility, and so it becomes at some level a subjective matter depending on one's ranking of topics in astrophysics. Current astronomical research is strongly leaning towards dark-sky, extragalactic, cosmological areas

Table 18.4: Concept Summary

Infrared Very Large Array (IRVLA)	
Large baseline	1000 m
Many telescopes	27
Large unit telescope apertures	3–4 m
Adaptive optics equipped	Curvature sensing based
Dual beam operation	Similar to PTI/Keck
Reconfigurable	Telescopes on rails (OCA, VLA)
Dedicated facility	
Multi-national collaboration	

and so many astronomers would consider that anything closer than 20 Mpc is irrelevant.

There's been a lot of stellar science during this course, but I don't think that the above budget is going to be applied to stellar astronomy or stellar physics in the real near future. I could imagine a further, more distant future after a couple of time constants and die-off of the cosmologists. There is a seriousness to that; interest and emphasis in sciences do rise, fall, and change. The qualitative era of astronomy gave way to the quantitative era, and stellar physics faded out as people did what they could with high-resolution spectroscopy—what hasn't really started up yet, is people taking advantage of what can be done with high spatial resolution.

Table 18.5 lists a number of characteristic source types for which interferometry would clearly have an interesting application, as well as the typical distances to the sources, and the size of partially and fully-resolved features that a kilometer baseline would give you at a wavelength $\lambda = 1 \mu\text{m}$ —partially-resolved in the sense of characterizable.

Take for example Seyfert galaxies, listed in Table 18.5. They are a very interesting area of application for interferometry, which would be of considerable interest to many astronomers. The picture of a Seyfert galaxy as a high-energy galactic nucleus consists of a central source sometimes called a continuum region, or (u, v) continuum, with a size of less than one-tenth of a parsec, probably quite a bit less, but (with more confidence) surrounded by a clumpy region called the broadline region with a size on the order of one parsec. Interferometry is in fact a very good candidate for finding out what's really going on in a source like that. Because not only does it have the potential to resolve structure well inside the broadline region but to even distinguish specific models involving processes that are going on in here, such as orbital motion of clumps of material versus ejection versus other mechanisms and the potential for actually looking at internals and following them with time. And a similar argument applies to the internal workings of quasars.

But of course this type of astronomy involves relatively faint sources and starts to run into the co-phasing and phasing limits for an array based on natural guide-star adaptive optics.

As I indicated, an adaptive-optic system is pretty well guaranteed to work to a visual magnitude of about 16 but then it starts to get tough after that. This limiting magnitude is a little bright for extragalactic exotic-source astronomy. I would like to see that pushing up to 18 or so, but a magnitude limit of 16 is in the ballpark.

Table 18.5: Potential targets for high-resolution infrared interferometry.

Source Type	Typical Distance	Resolved	Partially Resolved
Asteroid	2 AU	200 m	20 m
M dwarf	5 pc	0.5 R	0.05 R
Evolved star	100 pc	0.02 R	0.002 R
YSO	400 pc	0.04 AU	0.004 R
Recurrent Nova	2 kpc	0.4 AU	0.04 AU
Relativistic disk	5 kpc	1 AU	0.1 AU
Local group galaxy	16 kpc	3.2 AU	0.3 AU
M81 group	3.4 Mpc	0.003 pc	6 AU
Seyfert, BL Lac	10 Mpc	0.01 pc	18 AU
Virgo cluster	19 Mpc	0.02 pc	0.002 pc
Quasar	1500 Mpc	1.4 pc	0.14 pc

18.5 Comments and Conclusion

Just to finish off the story, we presented our proposal to the Decade Review Committee, a group of primarily dark sky astronomers, who were predominantly, I would say, somewhat neutral to this general plan. But quite cordial. Their recommendations go into the report which then gets reviewed by the National Academy of Sciences, and the review process takes about 6 months and it's all held in great secrecy.

I wound up my presentation to them by reviewing many of the concepts that have been described in these course notes and showing that there are no technical blocks to the construction of an Infrared Very Large Array. It will work and we can now go ahead. Interferometry has been developing very rapidly for the last decade, a lot of money has been invested in it. The science productivity from interferometers has been strong and is accelerating. The intermediate facilities are coming on-line now, and it is logical to look at a further more ambitious step in ground-based interferometry. The scientific rationale is, of course, the level of detail that a new array would bring to studies of compact sources. Going from a star that appears as a point-source observationally, to a resolvable source is an enormous advance, and it's also an enormous complication because the physics you need to understand it becomes a lot more complicated—the spherical symmetry, homogeneity, all those assumptions suddenly go away. There's a tremendous jump to actually processing the advances that are possible.

I made the argument that the development of such a facility could begin any time—tomorrow even. My own feeling is that the huge step in capability is so large, and the power of it is so obvious even to a non-scientist, that I'm sure this array will happen.

References

- T.J. Cornwell, "A novel principle for optimization of the instantaneous Fourier plane coverage of correlation arrays," *IEEE Trans. Antennas Propagat.* **36**, 1165–1167 (1988).
- H.M. Dyck and E.J. Kibblewhite, "Giant infrared telescopes for astronomy: a scientific rationale," *Pub. Astron. Soc. Pac.* **98**, 260–267 (1986).
- M. Mountain, "What is beyond the current generation of groundbased 8m – 10m class telescopes and the VLT-I?," in *Optical Telescopes of Today and Tomorrow*, A. Ardeberg, ed. *Proc. SPIE* **2871**, 597–606 (1997).
- S.T. Ridgway, "Astrophysical programs for high angular resolution optical interferometry," in *Diffraction-Limited Imaging with Very Large Telescopes*, D.M. Alloin and J.-M. Mariotti, eds. *NATO ASI Series* **274**, 307–326 (Dordrecht: Kluwer Academic, 1989).
- S.T. Ridgway and F.J. Roddier, "Infrared Very Large Array for the 21st century," in *Interferometry in Optical Astronomy*, P.J. Léna and A. Quirrenbach, eds., *Proc. SPIE* **4006**, 940–950 (2000).
- F. Roddier, "Prospects for imaging interferometry," *Astrophys. Space Sci.* **223**, 109–116 (1995).
- F. Roddier and S.T. Ridgway, "Filling factor and signal-to-noise ratios in optical interferometry," *Pub. Astron. Soc. Pac.* **111**, 990–996 (1999).

Appendices

Appendix A

Notes on the History of Stellar Interferometry

PETER R. LAWSON

JET PROPULSION LABORATORY
PASADENA, CALIFORNIA

In the following I discuss what influences Michelson may have had relating to his early work on stellar interferometry, and I also include a timeline of interferometry from 1868 to the present day.

A.1 Fizeau and Michelson

Hippolyte Fizeau (1819–1896) was the first to suggest that it might be possible to measure the angular diameters of stars through observations of interference fringes produced from starlight (Fizeau, 1868). Based on this suggestion, and through subsequent correspondence, Edouard Stéphan, the director of the Marseilles Observatory, conducted the first interferometric measurements of stars between 1872 and 1873. Stéphan published his initial thoughts on this technique and his subsequent results in the Proceedings of the French *Academie des Sciences* (Stéphan, 1873, 1874).*

Albert Michelson (1852–1931) was the first to fully describe the mathematical foundations of stellar interferometry and the first to measure a stellar diameter (Michelson, 1890; Michelson and Pease, 1921). His earliest work in this field took place around 1890, 17 years after

*The telescope that Stéphan used was the largest reflecting telescope then in existence, built by Foucault. The history of Foucault's 80 cm reflector, described by Tobin (1987), is interesting in its own right and is well worth reading. Biographical material concerning Stéphan's life and work may be found in his own history of the Marseille Observatory (Stéphan, 1914).

Stéphan's observations in Marseille. Although Michelson appears to have invented stellar interferometry independently, because he never once refers to the prior work of Fizeau, Stéphan, or others, the obvious question arises: to what extent was Michelson influenced by Fizeau? In the following section I will attempt to address this question. For more on the history of Michelson's interest in stellar interferometry, the reader is encouraged to read the paper by de Vorkin (1975) and the biography by Livingston (1973).

A.1.1 American References to European Work in Stellar Interferometry

Michelson's early work in stellar interferometry, and in particular his measurement of the satellites of Jupiter (Michelson, 1891), soon became known to European astronomers, as can be seen in the preamble of papers by Hamy (1893) and Schwarzschild (1896). However, European work in stellar interferometry, including the pioneering observations by Stéphan, only begins to appear in scientific papers in English around 1928.

The French work on stellar interferometry seems to have been unknown to Michelson and his collaborators until sometime after the measurement of the diameter of Betelgeuse (Michelson and Pease, 1921). Anderson seems to be ignorant of anything but Michelson's work when he describes his own binary star interferometer (Anderson, 1920). He states that

In view of the great beauty and simplicity of the method, it is surprising to find that the only application it has had up to the present time is to the determination of the diameters of Jupiter's satellites, and this was done by Professor Michelson himself.

Neither Michelson nor Pease (Michelson, 1920; Michelson and Pease, 1921) make any reference to prior work by others, apart from the work of Anderson.

Only in 1928, when Henroteau reviews measurements of double and multiple stars, do we find what I believe to be the first references in English to Fizeau, Stéphan, and Hamy (Henroteau, 1928). The references that Henroteau cites are exhaustive, and to my knowledge cover all sources of work in this field up to 1928, and further include some obscure theoretical work and early work on double star interferometry.^{†‡}

It is perhaps not surprising then that in 1931 when Pease writes about the 50-ft interferometer (Pease, 1931) he now reviews previous work citing Fizeau and Stéphan:

Stephan had already shown that fringes were conspicuous in the 80-cm Marseilles refractor, but his results were hidden in the volumes of the *Comptes Rendus*...

[†]The history of work on double star interferometry prior to the invention of speckle interferometry is covered in detail by Finsen (1971).

[‡]Henroteau's description of Michelson is as follows: "In the *Philosophical Magazine* for July 1890, Michelson, carrying both theory and application further, describes his method for measuring the angular magnitude of celestial objects when these are beyond the powers of the largest telescopes." This seems to be worded to avoid the issue of whether Michelson was aware of Fizeau's work.

It remained for Michelson in 1890, then at Clark University, to work out this idea independently and to add to it an essential element which led to the construction of the modern stellar interferometer.

The simplest conclusion is therefore that Michelson independently invented stellar interferometry, and that (taken in context with the rest of his career) it was simply one of the many applications he invented for his interferometer.

Notwithstanding the above, in my own opinion it is entirely possible that Fizeau's scientific career was an inspiration to Michelson, as they both shared an interest in the measurement of the speed of light and the detection of relative motions of the ether. Moreover, it is likely that the two met in Paris long before Michelson elaborated his thoughts "On the application of interference methods to astronomical measurement" (Michelson, 1890).[§]

A.1.2 Could Michelson Have Been Influenced by Fizeau Prior to 1890?

It is certainly possible that Michelson met Fizeau at a period where Michelson was still establishing his career.* The earliest opportunity would have been during Michelson's stay in Europe, beginning in the fall of 1880. Michelson had previously conducted experiments on the speed of light at the Naval Academy, and was given a year's leave of absence to take courses at the University of Berlin and study under Hermann von Helmholtz.† During this period he also visited France, a visit that is described by Michelson's daughter in *The Master of Light* (Livingston, 1973):‡

In the interim, Michelson turned his mind to opportunities at hand. He was well aware that the French were making important strides in the field of optics, and Paris was becoming the center of important work in several scientific fields. He had long been anxious to meet with the eminent Frenchmen whose work had originally brought him to a career in physics, and he made arrangements to leave for a winter's study in the French capital.

Foucault was dead, but there were others who had carried on the study of optics. Michelson felt that he could learn much from a winter of study at the College de France and the Ecole Polytechnique. There were three men in particular with whom he wished to study: Cornu, Mascart, and Lippmann.

...When Michelson arrived in Paris in the fall of 1881 he found that his reputation had preceded him. His accurate determination of the speed of light, followed by his daring challenge to the received theory of astronomical aberration as Fresnel had established it, made his French colleagues eager to meet him. When he presented himself at the Ecole Polytechnique, he was asked if he were by chance the son of the famous Michelson.

[§]In 1881 Michelson would have been 29 years old and Fizeau 62.

*Fizeau became a member of the *Academie des Sciences* in 1860. In 1863 he was also appointed Inspector of Physics at the *Ecole Polytechnique*. He died in Nanteuil, France, in 1896.

†In Berlin he put together the earliest version of what would later be known as the Michelson-Morley experiment. In 1881 the Michelsons moved to Heidelberg where he studied as well.

‡Dorothy Livingston was Michelson's youngest daughter from his second marriage. She had not yet been born at the time of these events.

Table A.1: Events in Early Stellar Interferometry

Year	Event	Authors and Reference
1868	Stellar interferometry suggested	H. Fizeau, C. R. Acad. Sci. 66 , 932 (1868)
1872–73	Stellar diameters need be $\ll 0.158$ arcsec	E. Stéphan, C. R. Acad. Sci. 78 , 1008 (1874)
1890	Mathematical theory of stellar interferometry	A.A. Michelson, Phil. Mag. 30 , 1 (1890)
1891	Satellites of Jupiter measured	A.A. Michelson, Nature 45 , 160 (1891)
1896	Binary star measurements	K. Schwarzschild, Astron. Nachr. 139 3335 (1896)
1920	Orbit of Capella measured	J.A. Anderson, Astrophys. J. 51 , 263 (1920)
1921–31	First stellar diameter measured	A.A. Michelson, F.G. Pease, Astrophys. J. 53 , 249 (1921)
1931–38	50-ft interferometer	F.G. Pease, Erg. Exakt. Natur. 10 , 84 (1931)

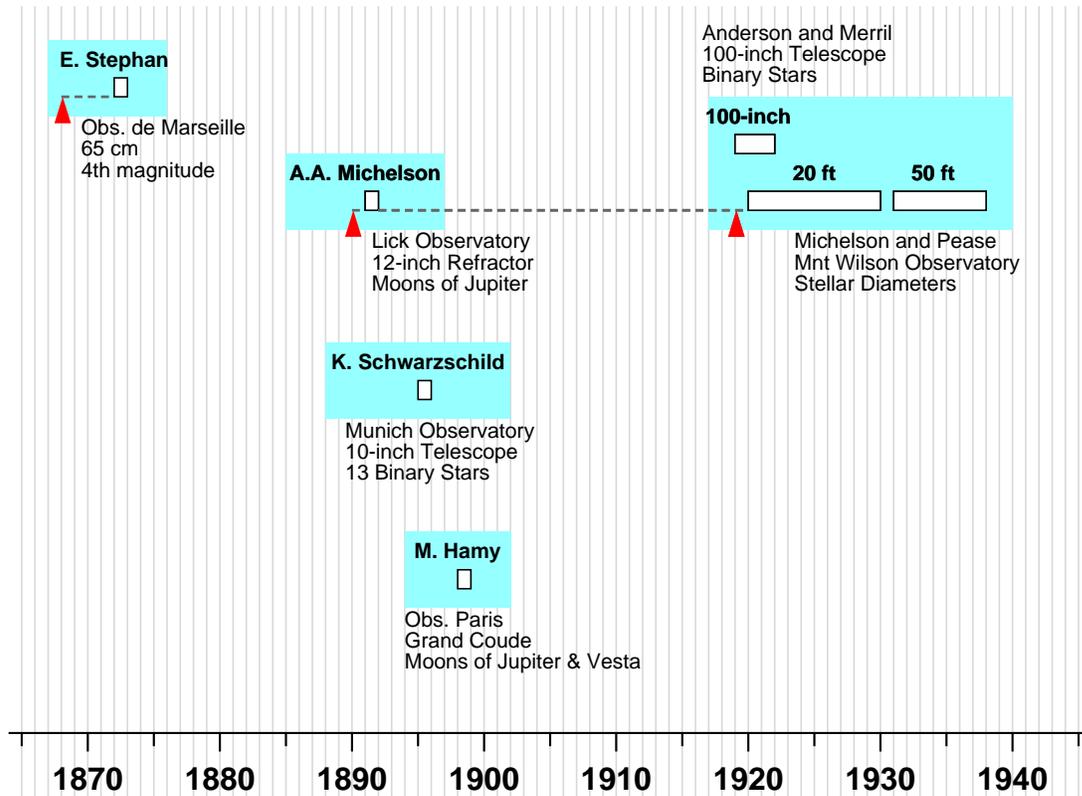


Figure A.1: Stellar Interferometry from 1868 to 1940.

When he laughingly replied that he was himself that Michelson, invitations were pressed upon him to meet the other professors of the science faculty. Discussions soon turned to the instrument he had devised.

...M. Andre Potier, a former pupil of Cornu, called Michelson’s attention to an error in his calculations in neglecting the effect of the earth’s motion on light travelling in the arm of the interferometer at right angles to that motion, which Potier said would reduce the fringe shift to zero. Michelson acknowledged his error at the February 20, 1882 meeting of the Paris *Academie des Sciences* but pointed out that the fringe displacement would be reduced only by half, not to zero.

Fizeau was actively involved in the *Academie des Sciences*, and could possibly have been present at the meeting Michelson attended. Seeing as Michelson had improved upon Fizeau’s measurement of the speed of light, it would be surprising if Michelson did not seek him out to introduce himself, although no meeting is specifically mentioned in the biography.^{†*}

It is clear that Michelson admired Fizeau and was indeed influenced by him. Michelson wrote that Fizeau’s measurement of the speed of light in moving water was “in my opinion one of the most ingenious experiments that have ever been attempted in the whole domain of physics” and “on this account, and also for the reason that the experiment was regarded as one of the most important in the entire subject of optics, it seemed to me that it was desirable to repeat it...” (Michelson 1902, pp. 152 and 155). Although it is tempting to believe that a connection existed between these two, and despite the similarity of their work and their many shared professional interests, there does not seem to be any clear evidence that Fizeau guided Michelson in his work in stellar interferometry.

A.2 A Timeline of Stellar Interferometry: 1868–2000

I now include two timelines of stellar interferometry, shown in Figures A.1 and A.2, to illustrate the development of the field through its early years, up until about 1940, and from the period beginning with the invention of intensity interferometry up until the present day. Tables A.1 and A.2 provide comments and references to accompany the figures. A more detailed account of the history of stellar interferometry can be found in the introduction to *Selected Papers on Long Baseline Stellar Interferometry* (Lawson, 1997).

In the figures I have attempted to group together projects where there has been a unified effort, or where several interferometers have been strongly associated. I have started each

[†]Labeyrie (1982) has written that “Years after H. Fizeau’s 1868 suggestion, the American physicist A.A. Michelson went to work with him in Paris and also became interested in stellar interferometry.” When I asked Prof. Labeyrie for further information, he could not remember where he had read about the meeting, and so I believe he was instead referring to Michelson’s time in Paris, as described above.

*After Fizeau’s death, the *secrétaire perpétuelle* of the *Academie des Sciences*, wrote a biography of Fizeau (Picard, 1924) citing Fizeau’s original role in stellar interferometry, but without any mention made of a meeting with Michelson.

Table A.2: Milestones in Long Baseline Stellar Interferometry

Year	Milestone	Authors and Reference
1956	Fringes with the prototype intensity interferometer	R. Hanbury Brown and R.Q. Twiss, <i>Nature</i> 177 , 27 (1956)
1970	Invention of speckle interferometry	A. Labeyrie, <i>Astron. Astrophys.</i> 6 , 85 (1970)
1972	10-micron heterodyne fringes	J. Gay and A. Journet, <i>Nature Phys. Sci.</i> 241 , 32 (1973)
1974	10-micron heterodyne fringes with separated telescopes	M.A. Johnson et al., <i>Phys. Rev. Lett.</i> 33 , 1617 (1974)
1974	Direct detection visible fringes with separated telescopes	A. Labeyrie, <i>Astrophys. J.</i> 196 , L71 (1975)
1979	Phase tracking stellar interferometer	M. Shao and D.H. Staelin, <i>Appl. Opt.</i> 19 , 1519 (1980)
1982	Fringe measurements at 2.2 microns	G.P. Di Benedetto and G. Conti, <i>Astrophys. J.</i> 268 , 309 (1983)
1985	Measurement of closure phase at optical wavelengths	J.E. Baldwin et al., <i>Nature</i> 320 , 595 (1986)
1986	Fully automated interferometer for astrometry	M. Shao, M.M. Colavita et al., <i>Astron. Astrophys.</i> 193 , 357 (1988)
1991	Use of single-mode fibers with separated telescopes	V. Coudé du Foresto and S.T. Ridgway, <i>ESO Proc.</i> 39 , 731 (1992)
1995	Optical synthesis imaging with separated telescopes	J.E. Baldwin et al., <i>Astron. Astrophys.</i> 306 , L13 (1996)

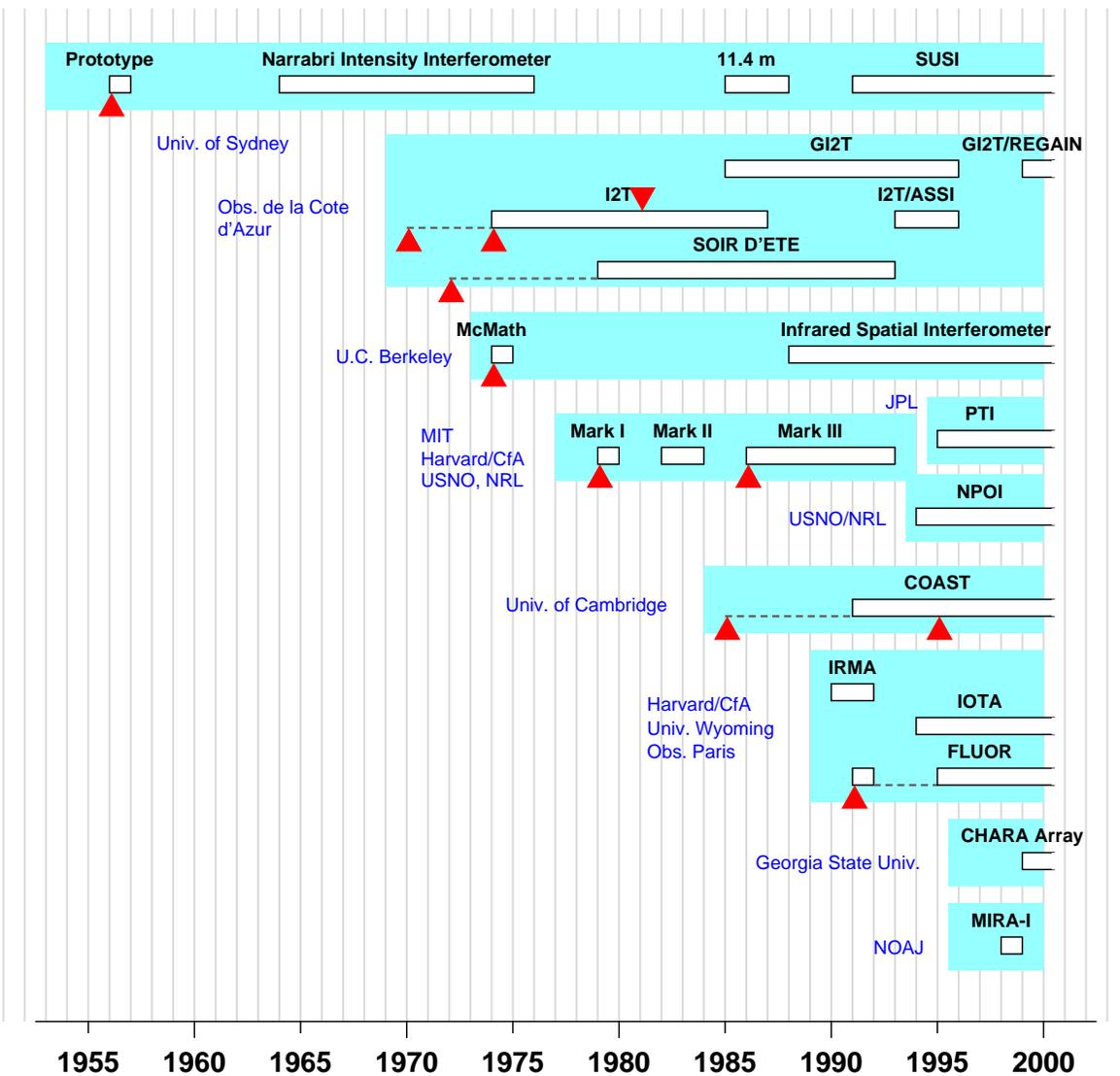


Figure A.2: Stellar Interferometry from 1950 to 2000.

interferometer in the year it first acquired fringes and ended it (if applicable) in the year it ceased observations. Some of these groupings are subjective: for example, three quite separate interferometer projects have existed at the Observatoire de la Côte d'Azur—although here I group them together. What I consider to be milestones in the timeline are indicated by the triangles and described in the Tables.

Acknowledgments

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

- J.A. Anderson, “Application of Michelson’s interferometer method to the measurement of close double stars,” *Astrophys. J.* **51**, 263–275 (1920).
- W.S. Finsen, “Twenty years of double-star interferometry and its lessons,” *Astrophys. and Space Sci.* **11**, 13–19 (1971).
- H. Fizeau, “Prix Bordin: Rapport sur le concours de l’année 1867,” *C. R. Acad. Sci.* **66**, 932–934 (1868).
- M. Hamy, “Sur la mesure des faibles diamètres,” *Bulletin Astronomique* **10**, 489–504 (1893).
- F.C. Henroteau, “Double and multiple stars,” in *Handbuch der Astrophysik 6*, 299–468 (Berlin, Germany: Julius Springer, 1928).
- A. Labeyrie, “Stellar interferometry: A widening frontier,” *Sky and Telescope* **63**, 334–338 (1982).
- P.R. Lawson, *Selected Papers on Long-Baseline Stellar Interferometry* MS **139** (Bellingham WA: SPIE Press, 1997).
- D.M. Livingston, *The Master of Light* (New York: Scribners, 1973).
- A.A. Michelson, “On the application of interference methods to astronomical measurements,” *Phil. Mag.* **30**, 1–20 (1890).
- A.A. Michelson, “Measurement of Jupiter’s satellites by interference,” *Nature* **45**, 160–161 (1891).
- A.A. Michelson, *Light Waves and Their Uses* (Chicago: University of Chicago Press, 1902).
- A.M. Michelson and F.G. Pease, “Measurement of the diameter of α Orionis,” *Astrophys. J.* **53**, 249–259 (1921).
- F.G. Pease, “Interferometer methods in astronomy,” *Ergebnisse der Exakten Naturwissenschaften* **10**, 84–96 (1931).
- E. Picard, *Les Theories de l’Optique et l’Oeuvre d’Hippolyte Fizeau* (Paris, France: Academie des Sciences, 1924).

- K. Schwarzschild, “Ueber messung von doppelsternen durch interferenzen,” *Astron. Nachr.* **139**, No. 3335 (1896).
- E. Stéphan, “Extrait d’une lettre de M. Stéphan à M. Fizeau,” *C. R. Acad. Sci.* **76**, 1008–1010 (1873).
- E. Stéphan, “Sur l’extrême petitesse du diamètre apparent des étoiles fixes,” *C. R. Acad. Sci.* **78**, 1008–1012 (1874).
- E. Stéphan, “L’Observatoire de Marseille: histoire depuis la révolution,” in *Encyclopédie des Bouches du Rhône* **6** (1914). <http://www-obs.cnrs-mrs.fr/patrimoine/histoire-f.html>.
- W. Tobin, “Foucault’s invention of the silvered-glass reflecting telescope and the history of his 80-cm reflector at the Observatoire de Marseille,” *Vistas in Astronomy* **30**, 153–184 (1987).
- D.H. De Vorkin, “Michelson and the problem of stellar diameters,” *Journal for the History of Astronomy* **6**, 1–18 (1975).

Appendix B

Suggested Reading

B.1 General References

Engineering and Technology for Stellar Interferometry

P.R. Lawson, *Selected Papers on Long Baseline Stellar Interferometry*, MS 139 (SPIE Press: Bellingham, WA, 1997).

J. Davis, “Observing with optical/infrared long baseline interferometers,” in *High Angular Resolution in Astrophysics*, A.-M. Lagrange, D. Mourard and P. Léna, eds. (Kluwer Academic: Dordrecht, 1997), pp. 49–79.

M. Shao and M.M. Colavita, “Long-baseline optical and infrared stellar interferometry,” *Ann. Rev. Astron. Astrop.* **30**, 457–498 (1992).

V. Coudé du Foresto, “Integrated optics in astronomical interferometry,” in *Very High Angular Resolution Imaging*, J.G. Robertson and W.J. Tango, eds., IAU Symp. **158** (Kluwer Academic: Dordrecht, Netherlands, 1994), pp. 261–271.

Recent Conferences and Workshops on Stellar Interferometry

Interferometry in Optical Astronomy, P.J. Léna and A. Quirrenbach, eds., Proc. SPIE **4006** (SPIE Press: Bellingham, WA, 2000). Munich, Germany, 27–29 March 2000.

Working on the Fringe: Optical and IR Interferometry from the Ground and Space, S. Unwin and R. Stachnik, eds., ASP Conf. Ser. **194** (Brigham Young University Press: Provo, UT, 1999). Dana Point, California, USA, 24–28 May 1999.

Catching the Perfect Wave: Adaptive Optics and Interferometry for the 21st Century, S. Restaino, W. Junor, and N. Duric, eds., ASP Conf. Ser. **174** (Brigham Young Univ. Press: Provo Utah, 1999). Albuquerque, New Mexico, USA, 28 June – 1 July 1998.

Astronomical Interferometry, R.D. Reasenberg, ed., Proc. SPIE **3350** (SPIE Press: Bellingham, WA, 1998). Kona, Hawaii, USA, 20–24 March 1998.

Astrophysics with Stellar Interferometers

Fundamental Stellar Properties: The Interaction Between Observation and Theory, T.R. Bedding, A.J. Booth, and J. Davis, eds., IAU Symposium No. 198 (Kluwer Academic : Dordrecht, 1997).

Complementary Approaches to Double and Multiple Star Research, H.A. McAlister and W.I. Hartkopf, eds. IAU Colloquium 135, ASP Conf. Ser. **32** (San Francisco: Astronomical Society of the Pacific, 1992).

Stellar Surface Structure, K.G. Strassmeier and J.L. Linsky eds., IAU Symposium No. 179 (Kluwer Academic: Dordrecht, 1996).

K.J. Johnston, “Reference frames in astronomy,” *Ann. Rev. Astron. Astrophys.* **37**, 97–125 (1999).

Radio Astronomy

A.R. Thompson, J.M. Moran, and G.W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy* (John Wiley and Sons: New York, 1986).

P.F. Goldsmith ed., *Instrumentation and Techniques for Radio Astronomy*, an IEEE volume of collected reprints (IEEE Press: New York, 1988).

G. B. Taylor, C. L. Carilli, and R. A. Perley, eds., *Synthesis Imaging in Radio Astronomy II*, ASP Conf. Ser. **180** (Brigham Young Univ. Press: Provo Utah, 1999)

E.B. Fomalont and M.C.H. Wright, “Interferometry and aperture synthesis,” in *Galactic and Extra-Galactic Radio Astronomy* (1st edition), G.L. Verschuur and K.I. Kellermann, eds., (Springer-Verlag: New York, 1974), pp. 256–290.

J.L. Yen, “Image reconstruction in synthesis radio telescope arrays” in *Array Signal Processing*, S. Haykin ed., (Prentice-Hall: Englewood Cliffs, NJ, 1985), pp. 293–350.

Resources on the Web

<http://huey.jpl.nasa.gov/olbin/>

Optical Long-Baseline Interferometry Newsletter, P.R. Lawson ed., (Jet Propulsion Laboratory, Pasadena, California).

B.2 Long Baseline Stellar Interferometers

In this section are listed references for the interferometers that have been mentioned in these course notes. In each case the first reference is a paper that describes the engineering of the interferometer and the other references are papers that describe astrophysical results.

Mark III Interferometer

<http://www.mtwilson.edu/Tour/NRL/>

M. Shao, M.M. Colavita, B.E. Hines, D.H. Staelin, D.J. Hutter, K.J. Johnston, D. Mozurkewich, R.S. Simon, J.L. Hershey, J.A. Hughes, and G.H. Kaplan, “The Mark III stellar interferometer,” *Astron. Astrophys.* **193**, 357–371 (1988).

C.A. Hummel, D. Mozurkewich, N.M. Elias, A. Quirrenbach, D.F. Buscher, J.T. Armstrong, K.J. Johnston, R.S. Simon, and D.J. Hutter, “Four years of astrometric measurements with the Mark III optical interferometer,” *Astron. J.* **108**, 326–336 (1994).

A. Quirrenbach, D. Mozurkewich, D.F. Buscher, C.A. Hummel, and J.T. Armstrong, “Angular diameter and limb darkening of Arcturus,” *Astron. Astrophys.* **312**, 160–166 (1996).

Cambridge Optical Aperture Synthesis Telescope (COAST)

<http://www.mrao.cam.ac.uk/telescopes/coast/>

J.E. Baldwin, R.C. Boysen, G.C. Cox, C.A. Haniff, J. Rogers, P.J. Warner, D.M.A. Wilson, and C.D. Mackay, “Design and performance of COAST,” in *Amplitude and Intensity Spatial Interferometry II*, J.B. Breckinridge, ed., *Proc. SPIE* **2200**, 118–128 (1994).

J.E. Baldwin, M.G. Beckett, R.C. Boysen, D. Burns, D.F. Buscher, G.C. Cox, C.A. Haniff, C.D. Mackay, N.S. Nightingale, J. Rogers, P.A.G. Scheuer, T.R. Scott, P.G. Tuthill, P.J. Warner, D.M.A. Wilson, and R.W. Wilson. “The first images from an optical aperture synthesis array: mapping of Capella with COAST at two epochs,” *Astron. Astrophys.* **306**, L13–L16 (1996).

D. Burns, J.E. Baldwin, R.C. Boysen, C.A. Haniff, P.R. Lawson, C.D. Mackay, J. Rogers, T.R. Scott, D. Saint Jacques, P.J. Warner, D.M.A. Wilson, J.S. Young, “Large amplitude periodic variations in the angular diameter of R Leonis,” *Mon. Not. R. Astron. Soc.* **297**, 462–466 (1998).

Grand Interféromètre à 2 Télescopes (GI2T)

<http://www.obs-nice.fr/fresnel/gi2t/en/>

D. Mourard, I. Tallon-Bosc, A. Blazit, D. Bonneau, G. Merlin, F. Morand, F. Vakili, and A. Labeyrie, “The GI2T interferometer on Plateau de Calern,” *Astron. Astrophys.* **283**, 705–713 (1994).

Ph. Stee, F. Vakili, D. Bonneau, and D. Mourard, “On the inner envelope of the Be star γ Cas,” *Astron. Astrophys.* **332**, 268 (1998).

F. Vakili, D. Mourard, Ph. Stee, D. Bonneau, P. Berio, O. Chesneau, N. Thureau, F. Morand, A. Labeyrie and I. Tallon-Bosc, “Evidence for one-armed oscillations in the equatorial disk of ζ Tau,” *Astron. Astrophys.* **335**, 261–265 (1998).

Infrared/Optical Telescope Array (IOTA)

<http://cfa-www.harvard.edu/cfa/oir/IOTA/>

W.A. Traub, “Recent results from the IOTA Interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg ed., Proc. SPIE **3350**, 848–855 (1998).

H.M. Dyck, J.A. Benson, G.T. van Belle, and S.T. Ridgway. “Radii and effective temperatures for K and M giants and supergiants,” *Astron. J.* **111**, 1705–1712 (1996).

R. Millan-Gabet, F. P. Schloerb, W. A. Traub, F. Malbet, J. P. Berger and J. D. Bregman. “Sub-AU Structure of the Near Infrared Emission from AB Aurigae,” *Astrophys. J.* **513**, L131–L134 (1999).

FLUOR

<http://g4-fluor.obspm.fr/fluor/>

V. Coudé du Foresto, G. Perrin, C. Ruilier, B. Mennesson, W. Traub, and M. Lacasse, “FLUOR fibered instrument at the IOTA interferometer,” in *Astronomical Interferometry*, R.D. Reasenberg ed., Proc. SPIE **3350**, 856–863 (1998).

G. Perrin, V. Coudé du Foresto, S.T. Ridgway, J.-M. Mariotti, W.A. Traub, N.P. Carlton, and M.G. Lacasse, “Extension of the effective temperature scale of giants to types later than M6,” *Astron. Astrophys.* **331**, 619–626 (1998).

G. Perrin, V. Coudé du Foresto, S.T. Ridgway, B. Mennesson, C. Ruilier, J.-M. Mariotti, W.A. Traub, and M.G. Lacasse, “Interferometric observations of R Leonis in the K-band: First direct detection of the photospheric pulsation and study of the atmospheric intensity distribution,” *Astron. Astrophys.* **345**, 221–232 (1999).

Infrared Spatial Interferometer (ISI)

<http://isi.ssl.berkeley.edu/>

D.D.S. Hale, M. Bester, W.C. Danchi, W. Fitelson, S. Hoss, E.A. Lipman, J.D. Monnier, P.G. Tuthill, and C.H. Townes, “The Berkeley Infrared Spatial Interferometer: a heterodyne stellar interferometer for the mid-infrared,” *Astrophys. J.* **537**, 998–1012 (2000).

W.C. Danchi, M. Bester, C.G. Degiacomi, L.J. Greenhill, and C.H. Townes. “Characteristics of dust shells around 13 late-type stars,” *Astron. J.* **107**, 1469–1513 (1994).

M. Bester, W.C. Danchi, D. Hale, C.H. Townes, C.G. Degiacomi, D. Mekarnia, and T.R. Geballe. "Measurement at 11 micron wavelengths of the diameters of α Orionis and α Scorpii, and changes in effective temperature of α Orionis and very recent dust emission," *Astrophys. J.* **463** 336–343 (1996).

Navy Prototype Optical Interferometer (NPOI)

<http://ad.usno.navy.mil/npoi/>

J.T. Armstrong, D. Mozurkewich, L.J. Rickard, D.J. Hutter, J.A. Benson, P.F. Bowers, N.M. Elias II, C.A. Hummel, K.J. Johnston, D.F. Buscher, J.H. Clark III, L. Ha, L.-C. Ling, N.M. White, and R.S. Simon, "The Navy Prototype Optical Interferometer," *Astrophys. J.* **496**, 550–571 (1998).

A.R. Hajian, J.T. Armstrong, C.A. Hummel, J.A. Benson, D. Mozurkewich, T.A. Pauls, D.J. Hutter, N.M. Elias II, K.J. Johnston, L.J. Rickard, and N.M. White, "Direct confirmation of stellar limb darkening with the Navy Prototype Optical Interferometer," *Astrophys. J.* **496**, 484–489 (1998).

C.A. Hummel, D. Mozurkewich, J.T. Armstrong, A.R. Hajian, N.M. Elias II, D.J. Hutter, "NPOI observations of the double stars Mizar A and Matar," *Astron. J.* **116**, 2536–2548 (1998).

Palomar Testbed Interferometer (PTI)

<http://huey.jpl.nasa.gov/palomar/>

M.M. Colavita, J.K. Wallace, B.E. Hines, Y. Gursel, F. Malbet, D.L. Palmer, X.P. Pan, M. Shao, J.W. Yu, A.F. Boden, P.J. Dumont, J. Gubler, C.D. Koresko, S.R. Kulkarni, B.F. Lane, D.W. Mobley, G.T. van Belle, "The Palomar Testbed Interferometer," *Astrophys. J.* **510**, 505–521 (1999).

A.F. Boden, G.T. van Belle, M.M. Colavita, P.J. Dumont, J. Gubler, C.D. Koresko, S.R. Kulkarni, B.F. Lane, D.W. Mobley, M. Shao, J.K. Wallace, "An interferometric search for bright companions to 51 Pegasi," *Astrophys. J. Lett.* **504**, L39–L42 (1998).

G.T. van Belle, B.F. Lane, R.R. Thompson, A.F. Boden, M.M. Colavita, P.J. Dumont, D.W. Mobley, D. Palmer, M. Shao, G.X. Vasisht, J.K. Wallace, M.J. Creech-Eakman, C.D. Koresko, S.R. Kulkarni, X.P. Pan, and J. Gubler. "Radii and Effective Temperatures for G, K and M Giants and Supergiants," *Astron. J.* **117**, 521–533 (1999)

Sydney University Stellar Interferometer (SUSI)

<http://www.physics.usyd.edu.au/astron/astron.html>

J. Davis, W.J. Tango, A.J. Booth, T.A. ten Brummelaar, R.A. Minard, S.M. Owens, "The Sydney University Stellar Interferometer I: The Instrument," *Mon. Not. R. Astron. Soc.* **303**, 773–782 (1999).

J. Davis, W.J. Tango, A.J. Booth, E. Thorvaldson, and J. Giovannis, "The Sydney University Stellar Interferometer II: Commissioning observations and results," *Mon. Not. R. Astron. Soc.* **303**, 783–791 (1999).

J. Davis, A.J. Booth, W.J. Tango, “Observations of β Centauri with the Sydney University Stellar Interferometer,” *Mon. Not. R. Astron. Soc.*, in press (2000).

CHARA Array

<http://www.chara.gsu.edu/CHARA/array.html>

H.A. McAlister, W.G. Bagnuolo, T. ten Brummelaar, W.I. Hartkopf, M.A. Shure, L. Sturmann, and N.H. Turner. “Progress on the CHARA Array,” in *Astronomical Interferometry*, R.D. Reasenberg ed., Proc. SPIE **3350**, 947–950 (1998).

Keck Interferometer

<http://huey.jpl.nasa.gov/keck/>

M.M. Colavita and P.L. Wizinowich, “The Keck Interferometer: progress report,” in *Interferometry in Optical Astronomy*, P.J. Léna and A. Quirrenbach, eds., Proc. SPIE **4006**, 310–320 (2000).

Very Large Telescope Interferometer (VLTI)

<http://www.eso.org/projects/vlti/>

A. Glindemann, A. Abuter, F. Carbognani, F. Delplancke, F. Dérie, M. Ferrari, A. Gennai, P.B. Gitton, P. Kervella, B. Koehler, S. A. Lévêque, S. Menardi, A. Michel, F. Paresce, T.P. Duc, A. Richichi, M. Schöller, M. Tarenghi, A. Wallander, and R. Wilhelm, “VLT interferometer: a unique instrument for high-resolution astronomy,” in *Interferometry in Optical Astronomy*, P.J. Léna and A. Quirrenbach, eds., Proc. SPIE **4006**, 2–12 (2000).

Space Interferometry Mission (SIM)

<http://sim.jpl.nasa.gov/>

SIM: Taking the Measure of the Universe, R. Danner and S. Unwin, eds., JPL 400-811 3/99 (Jet Propulsion Laboratory: Pasadena, California, 1999)

Terrestrial Planet Finder (TPF)

<http://tpf.jpl.nasa.gov/>

Terrestrial Planet Finder, C.A. Beichman, N.J. Woolf, and C.A. Lindensmith, eds., JPL Publication 99-3 (Jet Propulsion Laboratory: Pasadena, California, 1999)