

Two roads to planet detection

Thesis by
Michael Bottom

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

California Institute of Technology
Pasadena, California

2016
Defended May 10, 2016

*To my parents,
who worked very hard to give me a good life.*

Acknowledgments

It was, gentlemen, after a long absence—seven years to be exact, during which time I was studying in Europe—that I returned to my people. I learnt much and much passed me by—but that’s another story. The important thing is that I returned with a great yearning for my people in that small village at the bend of the Nile. For seven years I had longed for them, had dreamed of them, and it was an extraordinary moment when I at last found myself standing amongst them. They rejoiced at having me back and made a great fuss, and it was not long before I felt as though a piece of ice were melting inside of me, as though I were some frozen substance on which the sun had shone—that life warmth of the tribe which I had lost for a time in a land “whose fishes die of the cold.” My ears had become used to their voices, my eyes grown accustomed to their forms. Because of having thought so much about them during my absence, something rather like fog rose up between them and me the first instant I saw them. But the fog cleared and I awoke on the second day of my arrival, in my familiar bed in the room whose walls had witnessed the trivial incidents of my life in childhood and the onset of adolescence. I listened intently to the wind: that indeed was a sound well known to me, a sound which in our village possessed a merry whispering—the sound of the wind passing through palm trees is different from when it passes through fields of corn. I heard the cooing of the turtle-dove, and I looked through the window at the palm tree standing in the courtyard of our house and I knew that all was still well with life. I looked at its strong straight trunk, at its roots that strike down into the ground, at the green branches hanging down loosely over its top, and I experienced a feeling of assurance. I felt not like a storm-swept feather but like that palm tree, a being with a background, with roots, with a purpose.

-Tayeb Salih, *Season of Migration to the North*

It is proper and right to begin by thanking Professors Shri Kulkarni and John Johnson, my advisors, who mostly left me alone to make my own progress and learn from my own mistakes, as I preferred. I respect that. Similarly, I must thank Dr. Eugene Serabyn, my supervisor at the Jet Propulsion Lab, for his insight and leadership. I owe special thanks to Professor Dimitri Mawet, who was a patient teacher, collaborator, and mentor through my last years as a graduate student. His ideas are all over this thesis, and it’s so much the better for that.

I’d like to thank the folks at NASA, whose names I don’t even know, who saw fit to support me with the Space Technology Research Fellowship. It has made an enormously positive impact in the course of my research.

I owe very much to postdocs Phil Muirhead and Jon Swift. Phil taught me an enormous amount, and was a mentor and friend to me. Jon Swift was a great project manager and is a true Renaissance man, which made the time we spent together so enjoyable. I also owe much to Dr. Peter Plavchan,

who let me get knee deep in a fun second-year project, which gave me many skills I continue to draw on to this day.

I must thank Chris Shelton and Kent Wallace, brilliant engineers from JPL, the best in the business, who taught me so much about optics and instrumentation. I hope to be more like them some day.

I thank the exceptional staff at Palomar Observatory, especially Steve Kunsman—may he have a long and happy retirement. I will remember the many pleasant days and evenings at the magnificent Hale telescope for the rest of my life.

I'd like to thank all my teachers here at Caltech, both the good and the bad. Teachers don't get enough credit: when I came in here, I barely knew what a "magnitude" was; now, I can impress the hell out of anyone sitting next to me on an airplane, especially in economy class.

I owe much to my friends who made this journey so enjoyable: Brandon Dotson, the first person I met at Caltech and the most badass; Brett and Huilin Cornell, constant companions, pundits, and the first (and as of yet, only) two members of my ministerial flock; Abhilash Mishra and Io Kleiser, brilliant and fierce members of *la Résistance*; Leon Harding, a great Irish import and high flyer; Sirio Belli, for (or despite?) sharing an office for the better part of five years.

One more friend: Professor Wallace Sargent. Wal was an extraordinary person, with a great sense of humor, warm personality, and sharp mind. I loved chatting in his office about everything under the sun, and occasionally things beyond it. His passing left a huge hole in the department that can never be filled. I miss him.

I owe the most to Heather Duckworth, who has been my partner and best friend during this whole circus, and who I hope to partially repay in the coming decades. Here's to getting on with life together.

Abstract

One of the most exciting discoveries in astrophysics of the last last decade is of the sheer diversity of planetary systems. These include “hot Jupiters”, giant planets so close to their host stars that they orbit once every few days; “Super-Earths”, planets with sizes intermediate to those of Earth and Neptune, of which no analogs exist in our own solar system; multi-planet systems with planets smaller than Mars to larger than Jupiter; planets orbiting binary stars; free-floating planets flying through the emptiness of space without any star; even planets orbiting pulsars. Despite these remarkable discoveries, the field is still young, and there are many areas about which precious little is known. In particular, we don’t know the planets orbiting Sun-like stars nearest to our own solar system, and we know very little about the compositions of extrasolar planets. This thesis provides developments in those directions, through two instrumentation projects.

The first chapter of this thesis concerns detecting planets in the Solar neighborhood using precision stellar radial velocities, also known as the Doppler technique. We present an analysis determining the most efficient way to detect planets considering factors such as spectral type, wavelengths of observation, spectrograph resolution, observing time, and instrumental sensitivity. We show that G and K dwarfs observed at 400-600 nm are the best targets for surveys complete down to a given planet mass and out to a specified orbital period. Overall we find that M dwarfs observed at 700-800 nm are the best targets for habitable-zone planets, particularly when including the effects of systematic noise floors caused by instrumental imperfections. Somewhat surprisingly, we demonstrate that a modestly sized observatory, with a dedicated observing program, is up to the task of discovering such planets.

We present just such an observatory in the second chapter, called the “MINiature Exoplanet Radial Velocity Array,” or MINERVA. We describe the design, which uses a novel multi-aperture approach to increase stability and performance through lower system etendue, as well as keeping costs and time to deployment down. We present calculations of the expected planet yield, and data showing the system performance from our testing and development of the system at Caltech’s campus. We also present the motivation, design, and performance of a fiber coupling system for the array, critical for efficiently and reliably bringing light from the telescopes to the spectrograph. We finish by presenting the current status of MINERVA, operational at Mt. Hopkins observatory

in Arizona.

The second part of this thesis concerns a very different method of planet detection, direct imaging, which involves discovery and characterization of planets by collecting and analyzing their light. Directly analyzing planetary light is the most promising way to study their atmospheres, formation histories, and compositions. Direct imaging is extremely challenging, as it requires a high performance adaptive optics system to unblur the point-spread function of the parent star through the atmosphere, a coronagraph to suppress stellar diffraction, and image post-processing to remove non-common path “speckle” aberrations that can overwhelm any planetary companions.

To this end, we present the “Stellar Double Coronagraph,” or SDC, a flexible coronagraphic platform for use with the 200” Hale telescope. It has two focal and pupil planes, allowing for a number of different observing modes, including multiple vortex phase masks in series for improved contrast and inner working angle behind the obscured aperture of the telescope. We present the motivation, design, performance, and data reduction pipeline of the instrument. In the following chapter, we present some early science results, including the first image of a companion to the star delta Andromeda, which had been previously hypothesized but never seen.

A further chapter presents a wavefront control code developed for the instrument, using the technique of “speckle nulling,” which can remove optical aberrations from the system using the deformable mirror of the adaptive optics system. This code allows for improved contrast and inner working angles, and was written in a modular style so as to be portable to other high contrast imaging platforms. We present its performance on optical, near-infrared, and thermal infrared instruments on the Palomar and Keck telescopes, showing how it can improve contrasts by a factor of a few in less than ten iterations.

One of the large challenges in direct imaging is sensing and correcting the electric field in the focal plane to remove scattered light that can be much brighter than any planets. In the last chapter, we present a new method of focal-plane wavefront sensing, combining a coronagraph with a simple phase-shifting interferometer. We present its design and implementation on the Stellar Double Coronagraph, demonstrating its ability to create regions of high contrast by measuring and correcting for optical aberrations in the focal plane. Finally, we derive how it is possible to use the same hardware to distinguish companions from speckle errors using the principles of optical coherence. We present results observing the brown dwarf HD 49197b, demonstrating the ability to detect it despite it being buried in the speckle noise floor. We believe this is the first detection of a substellar companion using the coherence properties of light.

Published content and contributions

(See Section 1.5.2 for more details)

1. Bottom, Michael, Philip S. Muirhead, John Asher Johnson, and Cullen H. Blake. “Optimizing Doppler Surveys for Planet Yield.” *Publications of the Astronomical Society of the Pacific* 125, no. 925 (2013): 240-251.

M.B. wrote and performed the numerical simulations, derived the different figures of merit, produced all the figures, and prepared the manuscript.

2. Swift, Jonathan J., Michael Bottom, John A. Johnson, Jason T. Wright, Nate McCrady, Robert A. Wittenmyer, Peter Plavchan et al. “Miniature Exoplanet Radial Velocity Array (MINERVA) I. Design, Commissioning, and First Science Results.” *Journal of Astronomical Telescopes, Instruments, and Systems*, Volume 1, Issue 2 (21 April 2015)

M.B. selected and tested much of the hardware, built and tested the pointing control system, wrote the telescope controls software, performed many of the commissioning tests and observations, and contributed to the manuscript preparation

3. Bottom, Michael, Philip S. Muirhead, Jonathan J. Swift, Ming Zhao, Paul Gardner, Peter P. Plavchan, Reed L. Riddle et al. “Design, motivation, and on-sky tests of an efficient fiber coupling unit for 1-meter class telescopes.” In *SPIE Astronomical Telescopes+ Instrumentation*, pp. 91472E-91472E. International Society for Optics and Photonics, 2014.

M.B. contributed to the design of the unit, constructed it, wrote the control software and simulations, and designed and performed the verification tests.

4. Bottom, M., J. C. Shelton, J. K. Wallace, R. Bartos, J. Kuhn, D. Mawet, B. Mennesson, R. Burruss, E. Serabyn. “Stellar Double Coronagraph: a multistage coronagraphic platform at Palomar observatory” *Publications of the Astronomical Society of the Pacific*, in press

M.B. contributed to the construction of the instrument and initial alignment and commissioning, wrote the pointing control/adaptive optics system software, performed the observations, wrote and used the data reduction pipeline, and prepared the manuscript.

5. Bottom, Michael, Jonas Kuhn, Bertrand Mennesson, Dimitri Mawet, Jean C. Shelton, J. Kent Wallace, Eugene Serabyn. “Resolving the delta Andromedae spectroscopic binary with direct imaging”. *The Astrophysical Journal*, 809, 11 (June 2015)

M.B. performed the observations, reduced and analyzed the data, and prepared the manuscript.

6. Bottom, M., Bruno Femenia, Elsa Huby, Dimitri Mawet, Eugene Serabyn. “Speckle nulling wavefront control for Palomar and Keck” In *SPIE Astronomical Telescopes+ Instrumentation*, International Society for Optics and Photonics, 2016

M.B. wrote the wavefront control code, tested the performance, and prepared the manuscript.

7. Bottom, M., James K. Wallace, Randall D. Bartos, J. Chris Shelton, Eugene Serabyn “Speckle suppression and companion detection using coronagraphic phase-shifting interferometry”. Submitted to *Monthly Notices of the Royal Astronomical Society*.

M.B. installed and tested the interferometer in question, developed and tested the hardware and analysis algorithms, performed the observations, analyzed the data, and prepared the manuscript.

Contents

Acknowledgments	v
Abstract	vii
Published content and contributions	ix
1 Introduction	1
1.1 Overview of methods of exoplanet detection	1
1.2 Astrometry, Transits, and Microlensing	5
1.2.1 Astrometry	5
1.2.2 Transits	6
1.2.3 Gravitational microlensing	9
1.3 Radial velocities	12
1.3.1 Overview	12
1.3.2 Physical principles	13
1.3.2.1 Kepler’s Laws	13
1.3.2.2 Radial velocity determination	15
1.3.2.3 Radial velocity measurement	18
1.3.3 Instrumentation and instrumental error sources	22
1.3.3.1 Stability control	22
1.3.3.2 Calibration methods	24
1.3.4 Stellar error sources	25
1.4 Direct imaging of exoplanets	27
1.4.1 Overview	27
1.4.2 Physical principles	28
1.4.2.1 Detectability in reflected and emitted light	28
1.4.2.2 Detectability at different ages	31
1.4.2.3 Takeaways	35
1.4.3 Instrumentation	36

1.4.3.1	Coronagraphs	36
1.4.4	Major error sources	42
1.4.4.1	Exozodiacal light	42
1.4.4.2	Speckles	44
1.4.5	Observing strategies and data analysis	45
1.4.5.1	Hardware discrimination techniques	46
1.4.5.2	Data reduction techniques	47
1.5	Overview of thesis and original contributions	49
1.5.1	Thesis overview	49
1.5.2	Previously published work and individual contributions	50
2	Part I—Optimizing Doppler Surveys for Planet Yield	54
	Abstract	55
2.1	Overview	56
2.2	Introduction	56
2.3	Effects of Stellar type	58
2.4	Maximizing radial velocity survey yields	61
2.4.1	Surveys complete to limits in planet mass and orbital period	63
2.4.2	Habitable zone planet surveys	65
2.5	Instrumental effects and systematic noise floors	68
2.5.1	Effects of Spectrometer Resolution and Sampling	68
2.5.2	Effects of incompletely recovered instrumental profiles	70
2.5.3	Effects of noise floors on survey yields	72
2.6	Conclusion/Discussion	74
2.7	Appendix	76
2.8	Simulation Parameters	76
2.9	Spectral class conversions	77
3	Part I—Minerva: a robotic telescope array for exoplanet discovery	78
	Abstract	79
3.1	Introduction	80
3.2	Design considerations and survey yield	81
3.2.1	Doppler precision, spectral resolution, multi-dish arrays	84
3.3	Testing of MINERVA	85
3.3.1	Enclosure	87
3.3.2	Telescope system	88

3.3.3	Telescope system verification	88
3.3.3.1	Acquisition and tracking	88
3.3.3.2	Throughput	91
3.3.3.3	Telescope issues	93
3.3.4	Fiber guiding system	94
3.3.4.1	Introduction	94
3.3.4.2	Optomechanical design	94
3.3.4.3	Controller design	94
3.3.5	Fiber guiding system verification	96
3.3.5.1	Expected Performance	96
3.3.5.2	Measured Performance	99
3.3.6	Imaging cameras and filters	101
3.3.6.1	Camera requirements	101
3.3.6.2	Photometry optics and detector	102
3.3.6.3	Performance Validation	103
3.3.7	Site choice and commissioning	104
3.4	Conclusions and future work	105
4	Part II–SDC: a multistage coronagraphic platform for Palomar Observatory	112
	Abstract	113
4.1	Introduction and Motivation	114
4.2	Background	115
4.2.1	Optical Vortices	116
4.2.2	Single vortex coronagraph	116
4.2.3	Ring-apodized vortex coronagraph	117
4.2.4	Multistage vortex coronagraph	118
4.3	Design of the Stellar Double Coronagraph	119
4.4	Observing sequence and data reduction	122
4.4.1	On-sky calibrations and observing strategy	122
4.4.1.1	Correction of non-common path low-order aberrations: MGS	122
4.4.1.2	Correction of non-common path high-order aberrations: speckle nulling	123
4.4.2	Observing strategy	125
4.4.3	Data reduction pipeline	126
4.5	Coronagraph configurations and laboratory performance	127
4.5.1	Single and dual vortex observing modes	127
4.5.2	Ring-apodized vortex observing mode	128

4.6	On-sky performance	129
4.6.1	Confirmation of physical association of epsilon Cephei b	129
4.6.2	Identification of the “compact object” companion to delta Andromeda	131
4.7	Conclusion and future work	133
4.8	Acknowledgements	133
4.9	Appendix	133
5	Part II—Resolving the delta Andromedae spectroscopic binary with direct imaging	135
	Abstract	136
5.1	Introduction	137
5.2	Instrumentation, Observations and Data Analysis	138
5.2.1	Instrumentation	138
5.2.2	Observations	138
5.2.2.1	Data Analysis	138
5.3	Results and Conclusions	140
6	Part II—Speckle nulling wavefront control for Palomar and Keck	144
	Abstract	145
6.1	Introduction	146
6.2	Background principles	147
6.2.1	Speckles in high contrast imaging	147
6.2.2	Principles of speckle nulling	151
6.3	Overview of code	153
6.3.1	Calibration	153
6.3.1.1	Dark frames, flat fields, and bad pixel corrections	153
6.3.1.2	Deformable mirror - image plane registration	154
6.3.1.3	Deformable mirror intensity calibration	155
6.3.1.4	Control region definition	156
6.3.2	Controls	156
6.4	Performance on selected instruments	158
6.4.1	PHARO and P3K	158
6.4.2	TMAS and P3K	160
6.4.3	NIRC2 and Keck AO	160
6.4.3.1	L-band Vortex coronagraph	162
6.5	Discussion and future work	162

7	Part II—Speckle suppression and companion detection using coronagraphic phase-shifting interferometry	165
	Abstract	166
7.1	Introduction	167
7.1.1	Focal plane wavefront sensing techniques	167
7.1.2	Coronagraphic phase-shifting interferometry	169
7.2	Theory	170
7.2.1	Phase Shifting Interferometry	170
7.2.2	Coronagraphic phase shifting interferometer design	172
7.3	Optomechanical implementation	173
7.3.1	Stellar Double Coronagraph	173
7.3.2	Initial setup and alignment	176
7.4	Focal plane wavefront sensing via phase shifting interferometry	176
7.4.1	Focal-plane phase measurements and dark hole generation	176
7.5	Coherence-based companion detection	180
7.5.1	Principles	180
7.5.2	Combination with image post-processing	183
7.5.3	Demonstration: HD 49197b	184
7.6	Discussion	186
7.6.1	Potential as a method of speckle suppression	186
7.6.2	Potential as a method of planet-speckle discrimination	188
7.7	Conclusion	188
7.8	Acknowledgements	189
7.9	Appendix	189
8	Summary and future work	193

List of Figures

1.1	A plot of all known exoplanets, shown in mass and separation. The points are colored by discovery method: red is transits, blue is radial velocities, green is microlensing, and yellow is direct imaging. (Astrometry has not discovered any planets.) Solar system planets Venus, Earth, Jupiter, Saturn, Uranus, and Neptune are shown as letters. Error bars are omitted for clarity. This plot was generated from the exoplanet orbit database at www.exoplanets.org	2
1.2	The apparent path on the sky of a star orbited by a 15 Jupiter mass planet in an elliptical ($e = 0.2$) orbit of semi-major axis 0.6 AU, observed from the Earth. The straight dashed line is the path of the system's barycenter, with a proper motion of 50 mas/yr. The dotted line is the parallactic path of the system, with the looping behavior due the apparent motion of the star against background stars due to the orbit of the Earth around the Sun. The solid line shows the photocenter motion due to the orbiting planet, with the perturbation increased by a factor of 30 to make the effect visible. This figure is taken from Perryman (2011).	3
1.3	The transit probability can be calculated as the solid angle of the stellar diameter rotated about the orbital axis. See the text for details.	7
1.4	The relation between various angles relevant in gravitational microlensing. See the text for details. This figure is taken from Gaudi (2010).	10
1.5	An example of a microlensing light curve. The large "main" magnification event of ~ 2 magnitudes, or a factor of 6.3 in brightness, corresponds to the lens star, and is well described by Equation 1.15. The small perturbation at the 5% level corresponds to the planet around the M-dwarf lens star, a $5.5 M_{\oplus}$ cool Super-Earth at a ~ 2.6 AU separation. Different colored points correspond to different observatories. This figure is taken from Beaulieu et al. (2006)	11
1.6	A Keplerian elliptical orbit (see Table 1.3.2.1). The reference direction is generally north (i.e., towards the orbital axis of the Earth), the reference plane is the plane of the sky, tangent to the celestial sphere. This figure is taken from Perryman (2011). . .	14

- 1.7 An example of a radial velocity measurement at about the current limit of precision of 1-2 m/s, taken from the HARPS spectrograph (Mortier et al., 2016). (a) Radial velocities of the star HD175607 are measured over a significant time period. Note the x-axis of the top curve is in days, corresponding to nearly ten years of observations. The red curve shows the model fit, with the rapid variations due to the inner planet and the slow variations to a potential outer companion (b) The phase-folded light curves and fits of the signals, corresponding to periods of 29 days (top) and 1400 days (bottom). 17
- 1.8 A fake spectrum of a star, with a deep absorption line and a shallow, broad absorption line. The x-axis is in wavelength, and the y-axis is in flux, both in arbitrary units. The pixels of the spectrum corresponding to the deep line have high quality Doppler information (red shading); the pixels of the shallow line have relatively poor quality Doppler content. The pixels of the continuum have no Doppler information whatsoever. This is equivalent to saying that it is impossible to determine the translation of a function of the type $y = \text{constant}$ 20
- 1.9 Modal noise: The end of a fiber has two slightly different intensity distributions for two different wavelengths (top). The addition of a slit exacerbates the differences (top middle). The two slits are imaged as resolution elements onto the detector by the spectrograph. The detector spectrum has spatial variations in the cross dispersion direction *as a function of wavelength* (bottom middle). The extracted spectrum (summed in the vertical direction) has quasi-periodic variations in the continuum intensity (bottom) . 23
- 1.10 (Left) An image of the Sun's surface, showing spots and convective granulation, with the Earth to scale.¹(Right) An example of a radial velocity signal induced by a starspot. In this figure, the left part of the star is rotating towards the reader. As the spot rotates into view, following the stellar rotation period, a redshift is recorded. After the star crosses the center, a blueshift is recorded. For more complicated starspot paths, the signal would change in complex ways. This figure is taken from Dumusque et al. (2011a) 25
- 1.11 The flux of light from a sun-like star compared to various planets in the solar system, and a hypothetical hot Jupiter. All these are blackbody curves; real planets and stars have more complicated spectra, but the approximations are quite accurate for the continuum levels. Note that the shorter, visible wavelengths are all reflected light and track the spectrum of the star, while the longer infrared wavelengths are thermal emission and are somewhat larger. The contrast between the Earth and the Sun is ten orders of magnitude in visible light. This figure is taken from Seager & Deming (2010). 30

1.12	A simulation of the formation of Jupiter, showing the growth in mass and radius as a function of time. Mass grows slowly until the core reaches about 10 Earth masses, at which point runaway gas accretion occurs and the radius contracts rapidly. This figure is taken from Lissauer et al. (2009)	32
1.13	A numerical hydrodynamics simulation showing the fragmentation of a protoplanetary disk very far from the central star. The figure is taken from Boley (2009).	33
1.14	The evolution of effective temperature and radius of a planet as a function of age, plotted for different mass planets. Note the much higher initial temperature and radii due to the disk instability formation scenario (“hot start”) vs the core accretion (“cold start”) scenario. The effects of planet formation mechanism vanish after a few hundred million years. This figure is taken from Spiegel & Burrows (2012).	34
1.15	The evolution of flux in mJy as a function of wavelength, for different ages, for the disk instability (“hot start”, red curves) scenario, and the core accretion (“cold start”, blue curves) formation mechanism. The figure on the right is for a 1 Jupiter mass planet, the one on the left for a 10 Jupiter mass planet. Each curve is an isochrone corresponding to 1, 3, 10, 30, and 100 million year ages, with the younger ages having brighter fluxes. Note the different vertical scales on the left and right figure. The black lines at the top correspond to the J, H, K, L, M, and N infrared bands. This figure is taken from Spiegel & Burrows (2012).	34
1.16	The contrast of various planets compared to a Sun-like star, as a function of wavelength. The red and light blue curves show models similar to those presented in Figure 1.15; three known exoplanets are shown at the top. Jupiter, in pink, is shown at the bottom. Note that the relative contrast seems to reach a peak at around 3.5 - 5 microns, corresponding to the infrared L' and M' bands. This figure is taken from Skemer et al. (2014)	35
1.17	Basic operation of a Lyot coronagraph. The star is brought to focus and blocked by a focal plane mask. A second stop in a pupil plane (the Lyot stop, slightly smaller than the pupil) blocks starlight that is diffracted around the focal plane mask, and a camera brings the remaining light to a focus. For a planet near the star (red light, bottom panel), the planet is focused near, but not on, the first disk, since it is separated on the sky, and is not affected by the Lyot stop. In this design, a simple Lyot coronagraph can block about 99% of the stars light while allowing the planets light to pass through unaffected. (This image was generated from a combination of images from the www.Lyot.org website)	37
1.18	A one dimensional coronagraph. See the text for details. This figure is taken from Sivaramakrishnan et al. (2001)	39

1.19	A plot of the top contrasts achieved by various modern coronagraph designs, plotted in inner working angle—bandwidth space. Note some coronagraphs show up more than once if measurements were performed at more than one bandwidth. Contrast levels correspond to the size of the circles. For a description of the coronagraphs, see Table 1.4.3.1. These are empirical results, and are slightly out of date as of this writing. This figure is taken from Mawet et al. (2012)	42
1.20	The planetary and zodiacal light levels for a Solar system twin at 10 pc, as viewed with a 2m and 8m class telescope. An Earth twin is shown on the left, with a Jupiter twin on the right. The zodiacal light can dominate the light at the infrared, but visible is more amenable, particularly with a larger telescope. Compare the spectral features to Figure 1.11. This figure is taken from Kasting et al. (2009), in the Decadal review of Astronomy.	43
1.21	A sinusoidal phase ripple in the pupil plane will form two speckles in the image plane. This is real image data, not simulated, and was taken from (Martinache et al., 2014) .	45
1.22	Raw images from Project 1640, the Gemini Planet Imager, and SPHERE, all showing large amounts of speckle noise. Images from the respective project websites.	45
1.23	A cartoon of angular differential imaging. The planet (red) appears at different positions in the focal plane at different observing times due to sky rotation. A median frame contains the speckle light, but not the planet. This median is subtracted from all the individual data frames, which are then derotated and combined to reveal the planet. This image was taken from www.astrobites.org	46
2.1	Doppler precision as a function of wavelength range and star temperature for a <i>fixed amount of observing time</i> (60 s). The stellar spectra are derived from rotationally-broadened main-sequence templates from 2600-6200K, stepped in 200K increments, and the wavelength range is stepped in 100 nm increments. The contours indicate the velocity precision in m/s. From the perspective of velocity precision, the best result is achieved in the range of 400-600 nm. The hashed regions correspond to wavelengths where the infrared absorption is too high for ground-based observations to be effective. This simulation assumes a 1.28 m ² telescope dish, a spectrograph with $R=75000$, and sky-to-detector throughput of 10% (the full simulation parameters, including stellar parameters, are given in the Appendix).	59

- 2.2 The time (seconds) to detect ($\sigma_v = K$) a planet with $M_{pl}a_{pl}^{-1/2} = 5M_{\oplus}(1\text{AU})^{-1/2}$, 10 parsecs away, for a range of observing wavelengths and stellar effective temperatures. The hashed regions correspond to wavelengths where the infrared absorption is too high for ground-based observations to be effective. This simulation assumes a 1.28 m² telescope dish, a spectrograph with $R=75000$, and sky-to-detector throughput of 10% (the full simulation parameters, including stellar parameters, are given in the Appendix). 64
- 2.3 The maximum number of observable stars each night, with the goal of achieving the velocity precision necessary for a detection of a planet with $M_{pl}a_{pl}^{1/2} = 5M_{\oplus}\text{AU}^{1/2}$ on each star. This is plotted as a function of observing wavelengths, assuming 9 hours of observing time per night and 2 minute acquisition time between targets. The survey was simulated using the RECONS 7 pc sample and the present-day mass function (Reid et al., 2002), and extends to 20 pc (going out to 300 pc makes little difference, as bright and nearby stars are the most time-efficient targets). 64
- 2.4 The time (seconds) to detect ($\sigma_v = K$) a 5 M_{\oplus} planet in the habitable zone of its parent star, 10 parsecs away, for a range of observing wavelengths and stellar effective temperatures. The hashed regions correspond to wavelengths where the infrared absorption is too high for ground-based observations to be effective. This simulation assumes a 1.28 m² telescope dish, a spectrograph with $R=75000$, and sky-to-detector throughput of 10% (The full simulation parameters, including stellar parameters, are given in the Appendix). 66
- 2.5 The maximum number of observable stars each night, with the goal of achieving the velocity precision necessary for a detection of a 5 M_{\oplus} habitable-zone planet on each star. This is plotted as a function of observing wavelengths, assuming 9 hours of observing time per night and 2 minute acquisition time between targets. The survey was simulated using the RECONS 7 pc sample and the present-day mass function (Reid et al., 2002), and extends to 20 pc (going out to 300 pc makes little difference, as bright and nearby stars are the most time-efficient targets). 67
- 2.6 The resulting velocity precision and signal-to-noise ratio for an observation of a Sun-like star with varying resolution, (3.0 samples/resolution element). Note that the units on the vertical axis are m/s for the blue curve (velocity precision), and unitless for the red curve (signal-to-noise ratio). The exposure time is held fixed, and the resulting signal-to-noise ratio *decreases* as resolution increases, since less photons are incident per pixel. Increasing the resolution always improves the velocity precision, but the point of diminishing returns is reached at about $R = 45,000$, which corresponds to the point where almost all the spectral lines are fully resolved. 69

2.7	The resulting velocity precision for observation of a Sun-like star with $R=75000$ (3.0 samples/resolution element), and skewness varying from 10^{-1} to 10^{-4} . For reference, the inset shows a gaussian with skewness (not α) of 0.3, an order of magnitude higher than the maximum value considered (none of the skew-normal distributions simulated have skews large enough to be visually distinct from a normal distribution.) The skewness sets a signal-to-noise floor when it is greater than a part in 100, weakly dependent on signal to noise. The flattening out of the curves occurs where the signal-to-noise ratio limits the velocity precision.	72
2.8	The resulting velocity precision for observation of a Sun-like star with $R=75000$ (3.0 samples/resolution element), and perturbation amplitude varying from 10^{-4} to 10^{-1} of the peak amplitude of the LSF. For comparison, the inset shows a gaussian with a perturbation of 10%, equal to the maximum value considered. It is clear that for sub meter/sec precision, it is important that the perturbation amplitude of the distribution does not exceed 0.1 %, a value weakly dependent on the signal-to-noise ratio. The flattening out of the curves occurs where the signal-to-noise ratio limits the velocity precision.	73
2.9	The maximum number of observable stars as a function of observing wavelengths, assuming 9 hours of observing time per night and 2 minute acquisition time between targets. The survey was simulated using the RECONS 7 pc sample and the present-day mass function (Reid et al., 2002), and extends to 20 pc (going out to 300 pc makes little difference, as bright and nearby stars are the most time-efficient targets). This graph assumes a survey targeting habitable zone planets, with velocity precisions limited to 0.5 (top), 1, 3, and 5 m/s. Furthermore, the region of 0.7-0.8 μ m is the best area to observe overall.	74
3.1	The required integration time per night for the MINERVA array to detect 3 Earth-mass planets within the habitable zones of each star in the η_{\oplus} sample according to a photon limited noise model as a function of V magnitude. Data points are colored according to their effective temperatures, determined using stellar masses from Howard et al. (2010b). This figure appears in Swift et al. (2015).	82
3.2	The differences in cost of telescopes of different aperture diameters. Commercial (amateur) telescopes are shown in red, professional telescopes are shown in blue. The CDK-700s (the telescopes we eventually chose) are shown as green points. This figure appears in Swift et al. (2015).	86

3.3	The location of our test site at Caltech, with the first two telescopes visible. The astronomy department is in the orange building across California boulevard at the top left.	86
3.4	Engineering drawing of a modified Aqawan with two CDK-700 telescopes inside (see next section, compare to Figure 3.3).	87
3.5	CDK700 telescope in our Aqawan enclosure in Pasadena. The telescope is positioned in its neutral position, pointing south at about 30 degrees. Also visible is the Andor iKonL camera, guider, and filter wheel on one of the ports.	89
3.6	A single measurement of the pointing precision of the CDK-700 telescope. Each data-point corresponds to the radial offset of the center pixel taken at 45 second intervals. The telescope holds the star position to 2 arc-seconds over a 50 minute interval, better than the claimed level from the manufacturer. For reference, the fiber size on the sky and required level of pointing precision for negligible throughput loss are shown above. It is clear that while excellent, the pointing precision is not good enough for open-loop control.	91
3.7	Left: Manufacturer-supplied transmission curves for the telescope mirrors, lenses, filters, and camera quantum efficiency. Right: throughput measurements of the telescope. The high level of variability is primarily due to the high level of variability in the atmospheric conditions in the areas where the tests were performed. The derived efficiency of 70% is in line with expectations, especially as the derived extinction of 0.25 mags/airmass is a reasonable value in the V band. These figures appear in Swift et al. (2015).	93
3.8	Left: Optical layout of the fiber acquisition unit. The primary beam path (top to bottom) couples light from the telescope into the fiber. The second path (top to left) images the star onto a guide camera. The third path (bottom to right to left) images the fiber tip onto the detector. Right: mechanical drawing of the FAU; fiber tip to base is about 8 inches long.	95
3.9	Fiber acquisition unit mounted on a CDK-700 telescope in the Pasadena Aqawan. The different subcomponents are labeled, including the fiber used to make the throughput measurements. This figure appears in Swift et al. (2015).	96
3.10	Thorlabs BP108 pellicle transmission for different polarization states. This image was taken from the manufacturer website.	97
3.11	Left: Theoretical throughput including pellicle transmission (P_T), reflection (η_R), geometric fiber injection (η_{in}), and fiber transmission (η_{thru}) loss, for a range of seeing values. Right: Numerical simulation of geometric injection efficiency for a range of pointing accuracies. The different lines correspond to different values of seeing.	98

3.12	(Left) Raw pixel tracking data with the closed loop guiding on (shaded region) and off. The two curves correspond to the x and y pixel positions on the guider CCD. The pitch is 0.33 arc seconds per pixel. (Right) The amplitude spectral density of the guider error is smaller at essentially all frequencies when guiding, especially low frequencies corresponding to long-term drifts. The intersection of the green (guiding) curve with 0 at 0 Hz indicates there is no systematic error.	106
3.13	The focused image of a star on the guide camera, and the diverging output beam of the fiber. The region of relatively reduced flux in the center of the fiber image is due to the effects of focal ratio degradation.	107
3.14	A throughput measurement in typical Pasadena seeing conditions of $\sim 2''$. The efficiency is about 50% (45% throughput), in line with the expectations of Figure 3.11. The statistical deviation about the mean is 1.5% (absolute), which is smaller than our systematic error.	107
3.15	Photometric camera optical setup, with major subcomponents labeled. This figure appears in Swift et al. (2015).	108
3.16	Photometric time series of 16 Cygnus A and B; data binned to one-minute intervals. The photometry is quite stable with little evidence of residual correlated noise. This figure appears in Swift et al. (2015).	108
3.17	Allan deviation of the photometric light curve of 16 Cygnus A, showing a precision of 1 mmag reached in about 3 minutes, almost at the photon noise limit. This figure appears in Swift et al. (2015).	109
3.18	The final location of MINERVA, at Mt Hopkins, Arizona. The building to the immediate North of the two Aqawans is the now repurposed VERITAS observing station; the VERITAS dish has already been removed. The HATNet and MEarth observatories are labeled to the North. A scalebar is shown to the bottom right. The geographic coordinates of this location are $31^{\circ} 40' 49.1''\text{N}$, $110^{\circ} 52' 44.6''\text{W}$. This image was generated using Google Maps.	110
3.19	MINERVA installed at Mt. Hopkins. This photo was taken right after the four telescopes were installed. The clamshell dome in the bottom right is a sister project called MINERVA-RED (PI: Cullen Blake)	111
4.1	The basic vortex coronagraph. Light from a clear telescope pupil (bottom row) is focused on the vortex by the telescope optics. At the following pupil, the starlight is moved to outside the pupil, where it is blocked by a Lyot stop. The dark pupil is then reimaged by the camera, with the starlight removed. For a centrally obscured aperture (top row), there is a residual halo of starlight left in the pupil.	117

4.2	The ring-apodized vortex coronagraph apodizes the input pupil of the telescope, with the result being that in the pupil plane after the vortex, all the energy inside the pupil is localized to a ring (second square panel, top). A pupil stop can effectively block this light and thus have total starlight cancellation in principle.	118
4.3	Schematic of the dual-vortex coronagraph. The first vortex leaves a residual halo of light (4th panel from left) which is moved behind the pupil by the second vortex (2nd panel from right).	119
4.4	The SDC mounts between the P3K adaptive optics system (blue rectangle, left) and infrared imager PHARO (partially visible behind its red electronics box). The imager, coronagraph and adaptive optics system all attach to Cassegrain port of the telescope.	120
4.5	The optomechanical layout of the SDC; refer to the text for a more detailed description. Following the input beam from the top right of the figure: first fold mirror, dichroic beam splitter, linear coronagraphic slide, off-axis paraboloid, fold mirror, Lyot plane, fold mirror, off-axis paraboloid, linear coronagraphic slide, off-axis paraboloid, Lyot plane, off-axis paraboloid, fold mirror. The infrared tracker is the green square. The image and pupil viewing camera and lenses are shown on the left, directly below the first off-axis paraboloid. In this orientation, the output beam to the infrared imager PHARO exits downward into the page.	121
4.6	Coronagraphic speckle nulling in the dual-vortex mode using the internal white light source of the adaptive optics system; see 4.4.1.2. (a) The initial results of PSF correction using MGS (4.4.1.1) still leaves many residual speckles in the focal plane. (b) Four iterations of speckle nulling remove most of the residual wavefront errors (c) Nine iterations get to within a factor of two of the detector read noise from 5 - 25 λ/D . The white polygon demarcates the control region, which is selected by mouse clicks in the half-region control mode. (d) Contrast improvement measured in the control region shows factors of 3-6 improvement, which are significant for companion detectability. The contrast curve is defined in the usual way, with the standard deviation (ie, 1 σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The background limit is determined by the contrast in a region of the detector 100's of λ/D away. The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding.	124
4.7	Outline of the speckle nulling loop. See Section 4.4.1.2 for a description of the loop. .	125

- 4.8 Laboratory contrast measurement comparing single (dashed curve) and dual vortex mode (solid curve). The contrast curve is defined in the usual way, with the standard deviation (ie, 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding. 128
- 4.9 (a) Raw contrast measurement with the ring-apodized vortex coronagraph. The contrast curve is defined in the usual way, with the standard deviation (ie, 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding. (b) The coronagraphic (top) and non-coronagraphic PSF, shown on different logarithmic scales to enhance features. (c) The measurement of the output pupil intensity corresponds well to theoretical expectations, with the major discrepancy being outside the pupil. This is due to the presence of an chromium dot in the center of the vortex, reducing stellar leakage. The center of the PSF is the brightest, so light blocked there will not show up outside the pupil. 130
- 4.10 A reduced image of one of our target stars ($K=8$, $V=6$) with the associated 5σ contrast curve on the right. The reduction strategy used was a zonal principal components analysis (KLIP) algorithm (see Section 4.4.3), with the principal components generated from a calibrator star with similar brightness and $V - K$ color. The total open shutter time on this target was 14 minutes, with the same time on the calibrator star (backgrounds, flats, and non-coronagraphic PSF frames were recorded separately). This measurement did not involve speckle nulling, so contrast at small angles can likely be improved further in the future. 131
- 4.11 Epsilon Cephei b. (a) The original discovery image, from Mawet et al 2011 (Mawet et al., 2011a), using a 1.5 m well-corrected subaperture of the Hale telescope. (b) Raw (no reference subtraction) SDC image, dual vortex mode, 15s of 10 median combined frames. (c) Classic PSF subtraction of (b). In the SDC images, the first Airy ring is visible around the companion. 132
- 4.12 The PSF-subtracted coronagraphic image of delta Andromeda b, dual vortex mode. This image first appeared in Ref. 15. 132

- 5.1 Left: a) the background-subtracted target median image, b) the background-subtracted reference star median image, c) the background-subtracted point-spread function image, d) best-fitting model from the MCMC algorithm combining images b) and c) and attempting to match a) as explained above. The stretch is nonlinear to better show the companion and speckles. Right: All the one and two dimensional projections of the posterior probability distributions of the pixel shifts (x_c , y_c , the reference background scaling factor (R_a , and the PSF amplitude used to fit the companion P_a . The two-dimensional projections show very little covariance among any two parameters, and the marginal distribution histograms (along the diagonal) are nicely peaked. 140
- 5.2 The reduced, background-removed coronagraphic image of δ Andromedae. The first Airy ring is visible around the companion. The stretch in the image is linear. The colorbar shows the relative intensity (as a fraction) compared to the primary 141
- 5.3 Comparison of the approximate fluxes of δ And A, three K dwarfs, and a 50000K white dwarf. The spectral models are from Castelli & Kurucz (2004). The grey bar shows the span of the Bracket- γ filter. 142
- 6.1 Complex plane representation of speckle phasors. A speckle at positions λ/x_0 containing phase-type (orange arrows) and amplitude-type (red arrows) contributions with electric field phases $\pi/2 \pm \alpha$ and $\pi \pm \beta$, respectively, where α and β are the pupil phases. The dashed arrows refer to the components at $-\lambda/x_0$, and the solid arrows refer to the components at λ/x_0 . A speckle at λ/x_0 may be nulled out by the electric field phasor corresponding to the blue dashed arrow (as can be seen by computing the phasor sums when lying the arrows head to tail). This will not simultaneously null out the speckle at $-\lambda/x_0$, whose null solution is shown by the blue solid arrow. 150

- 6.2 Coronagraphic speckle nulling in the dual-vortex mode using the internal white light source of the adaptive optics system. The value of λ/D is ~ 90 mas (a) The initial results of PSF correction using MGS still leaves many residual speckles in the focal plane. (b) Four iterations of speckle nulling remove most of the residual wavefront errors (c) Nine iterations get to within a factor of two of the detector read noise from $5 - 25 \lambda/D$. The white polygon demarcates the control region, which is selected by mouse clicks in the half-region control mode. (d) Contrast improvement measured in the control region shows factors of 3-6 improvement, which are significant for companion detectability. The contrast curve is defined in the usual way, with the standard deviation (i.e., 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The background limit is determined by the contrast in a region of the detector 100's of λ/D away. The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding. 159
- 6.3 Non-coronagraphic speckle nulling with the instrument TMAS using the internal white light source of the adaptive optics system. The value of λ/D is ~ 30 mas. (a) The initial results of PSF correction using Zernike tuning leaves many speckle aberrations in the focal plane. (b) Four iterations of speckle nulling. (c) Nine iterations of speckle nulling. The contrast improvement in this case is a modest 20-40%, but could possibly have been increased by running for more iterations (at the time this experiment was performed, the control software did not provide automatic feedback about the contrast improvement, so we were not aware that there was much more improvement to be had). Some of the limitations here are due to the starting level of aberrations being far higher than in the longer-wavelength instruments, as described in the text. It is encouraging to note that more of the diffraction pattern is visible in the last image than in the first. The contrast curve is defined in the usual way, with the standard deviation (i.e., 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF. The background limit is determined by the contrast in a region of the detector 100's of λ/D away. The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding. 161

- 6.4 Coronagraphic speckle nulling in L-band on NIRC2, using the internal white light source of the adaptive optics system. The value of λ/D is ~ 80 mas. (a) The initial results of PSF correction using image sharpening leaves many speckle aberrations in the focal plane. (b) Four iterations of speckle nulling. (c) Nine iterations of speckle nulling. The white lines demarcate the control region, with the outer line at approximately the limit of the control bandwidth of the deformable mirror. The contrast improvement in this case is a factor of 5-10, and gets within a factor of 2-3 of the background limit. The contrast curve is defined in the usual way, with the standard deviation (ie, 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF. The background limit is determined by the contrast in a region of the detector 100's of λ/D away. The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding. 163
- 7.1 Schematic of the dual-vortex coronagraph. The first vortex leaves a residual halo of light (4th panel from left) which is moved behind the pupil by the second vortex (2nd panel from right). In the second pupil plane, this light can be blocked, creating an effective conventional coronagraph. Alternatively, it can be picked off and used as a "reference" beam, as we do here with a phase-shifting mirror, described below. This figure originally appeared in Bottom et al. 2016 173
- 7.2 (a) Mechanical drawing of the phase shifter mount. The piezoelectric flexure stage is shown in dark grey, the interface to the annular mirror mount is in pink, and the annular mirror mount is in yellow. The annular mirror mount is actuated on the three axes, but the actuators are omitted for clarity. The mirrors are shown in blue. (b) Cutaway of the phase-shifter assembly, showing the hardware to align the rod to the mirror and the direction of travel of the rod. The mechanism is installed at the "Lyot stop 2" location in Figure 7.4. 174
- 7.3 (a) The point spread function of the combined outer mirror and phaseshifting rod. (b) The point-spread function of the inner phase-shifting rod only. The images are the same size, with each edge spanning 6.25 arc-seconds (250 pixels). The images are not at the same scale or exposure time, but have been individually stretched to bring out the speckle field and reference wave extent. 174

7.4	The optomechanical layout of the SDC. Following the input beam from the top right of the figure: first fold mirror, dichroic beam splitter, linear coronagraphic slide, off-axis paraboloid, fold mirror, Lyot plane, fold mirror, off-axis paraboloid, linear coronagraphic slide, off-axis paraboloid, second Lyot plane, off-axis paraboloid, fold mirror. The infrared tracker is the green square. The image and pupil viewing camera and lenses are shown on the left, directly below the first off-axis paraboloid. In this orientation, the output beam to the infrared imager PHARO exits downward into the page. The phase-shifter assembly is at the second Lyot plane, labeled “Lyot stop 2” in the image. This figure originally appeared in Bottom et al. 2016	175
7.5	(a) An image is taken at each pistoning mirror position (z). The red line shows a cut through a particular pixel, at different mirror positions. (b) Raw data of intensity (in counts) vs mirror position for three different pixels of similar intensity in the image plane. The relative phases between the three different waves correspond to differences in electric field phase.	178
7.6	Phase sensing in the focal plane. (Top row) The left-hand plot shows the intensity measurements in the focal plane, with diffraction spikes clearly visible. This measure of intensity is a typical image shown on an camera. On the right, the corresponding phase map is shown. (Bottom row) The outlined region in the top plot, magnified to show detail. Speckles and their Airy rings are more easily identified in the phase plot, with a 180 degree phase shift between the cores and first Airy rings. Note that speckle chromatic effects such as elongation show themselves as phase gradients across an individual speckle; for example, in the central speckle in the image, indicated by the arrow, the measured phase is seen to increase from 240 to 300 degrees across it.	179
7.7	The derived map of the phase gradient over the same area as in Figure 7.6.	180
7.8	Laboratory demonstration of coronagraphic speckle-suppression using the phase-shifting interferometer. The white outline delineates the region of speckle control. Panels a, b, and c show the intensity after 0, 4, and 9 iterations of speckle removal by the deformable mirror. Panels d and e show the phase measurements of the initial and final light distribution, showing a clear difference in phase. Panel f shows the 1σ contrast curve measurement of the region after 1, 4, and 9 iterations. The contrast curve is defined in the usual way, with the standard deviation (ie, 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The preprocessing performed on the data was dark subtraction and flat-fielding.	191

- 7.9 Coherence modulated detection of HD 49197b. (a) The mean of the target frames. The substellar companion is not clearly visible in the image. The companion is present at an SNR of about 5. The stretch is logarithmic. (b) The incoherent intensity map (Equation 7.21). The companion is easily visible at a position angle of about 275 degrees, and has an SNR of about 7.5. The stretch is also logarithmic. We note that the same data frames were used to compute image (b) as in (a), except combined using the interferometer position to give extra information. No reference star is used. (c) Principle components analysis (KLIP) reduction of HD49197b, where the components are generated from only the coherent parts of the image data, as explained in Section 7.5.2. No reference star is used, and the SNR is 9. The stretch is linear. (d) A conventional PSF-subtracted image of the companion, using a nearby reference star, for comparison. All the data is used, and the SNR is 14. The stretch is linear. 192

List of Tables

1.1	Summary of coronagraphs presented in Figure 1.19.	41
3.1	CDK-700 optical, mechanical, electronic, and pointing parameters, as specified by the manufacturer Planewave Instruments.	90
3.2	Equation terms for determining telescope throughput	92
3.3	Andor iKon-L parameters and filters	103
4.1	The different on-sky observing modes of SDC; see section 4.2 for a description of each of these modes.	127
4.2	The actuated optics in the SDC, their degrees of freedom, and the optical fields they control. Refer to Figure 7.4 for the an optical layout	134
4.3	List of optics, electronics, and related information	134
5.1	Previously measured properties of δ And and newly measured properties of the companion	137
5.2	Summary of observations.	138
7.1	Summary of reductions. All observations of HD 49197 and reference star HD 48270 were on 22 November 2015.	186

Chapter 1

Introduction

1.1 Overview of methods of exoplanet detection

The major methods of detecting extrasolar planets are astrometry, transits, gravitational microlensing, radial velocities, and direct imaging. These are the only ways that have either detected many planets, or at least have the hope of detecting many planets. Each method has its advantages and disadvantages, and none can by itself explore the full diversity of planetary systems. A plot of all currently confirmed exoplanets is shown in Figure 1.1.

In the following sections, we present these different ways of detecting extrasolar planets, giving a basic introduction to the physical principles behind the methods, their sensitivities, and their range of applicability. We will take a “signals analysis” perspective, focusing on the frequency, amplitude, and noise sources of planet signals. We believe this is the most useful way to understand exoplanet science for experimental detection work. For a tabular comparison of methods, see Table 1.2.1.

We split the presentation into a few parts. First, we present a short overview of astrometry, transits, and microlensing, which are not very relevant to the work in this thesis, but useful for having a general picture of the current state of the field and its future. For these three methods, we draw heavily from Perryman (2011) and Seager (2010). We then devote one section each to radial velocities and direct imaging, the methods used in this thesis, in much more detail. The reader who is only interested in the work on radial velocities or direct imaging in this thesis might consider reading those respective sections, though background is provided in each chapter. The reader who is familiar with methods of planet detection may skip this introduction.

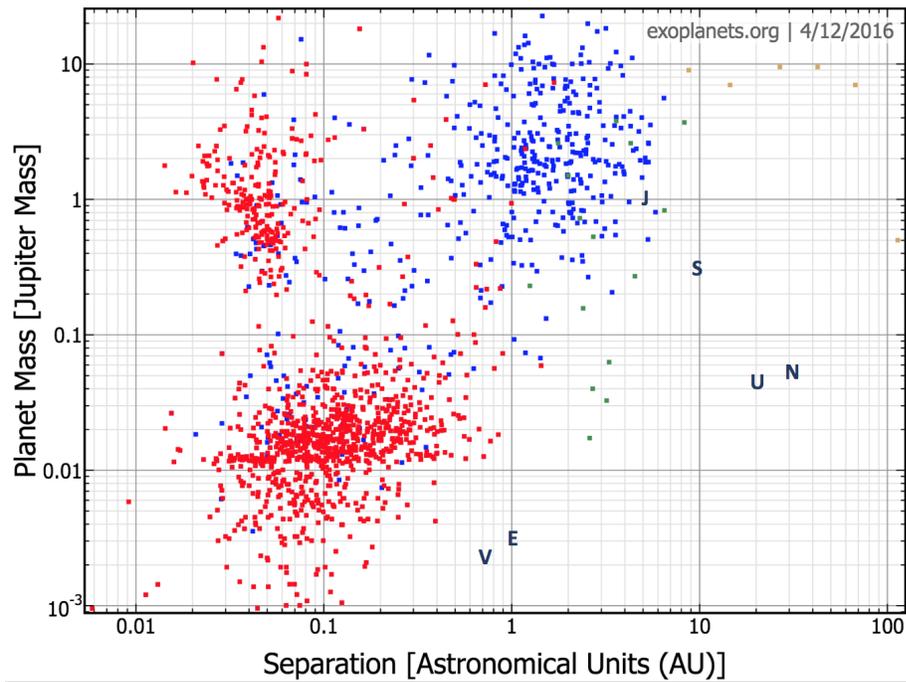


Figure 1.1: A plot of all known exoplanets, shown in mass and separation. The points are colored by discovery method: red is transits, blue is radial velocities, green is microlensing, and yellow is direct imaging. (Astrometry has not discovered any planets.) Solar system planets Venus, Earth, Jupiter, Saturn, Uranus, and Neptune are shown as letters. Error bars are omitted for clarity. This plot was generated from the exoplanet orbit database at www.exoplanets.org.

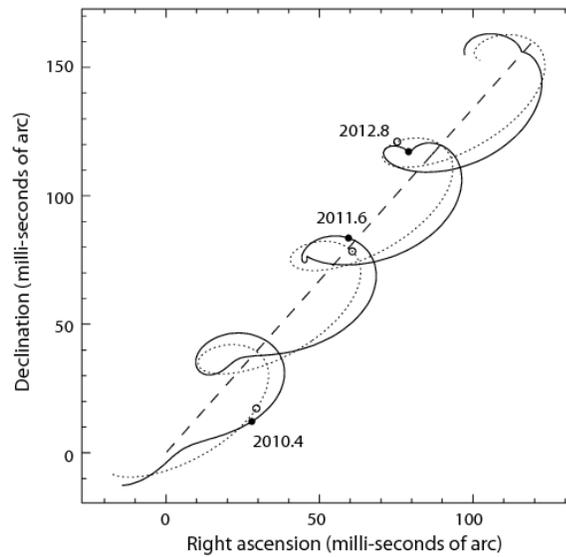


Figure 1.2: The apparent path on the sky of a star orbited by a 15 Jupiter mass planet in an elliptical ($e = 0.2$) orbit of semi-major axis 0.6 AU, observed from the Earth. The straight dashed line is the path of the system's barycenter, with a proper motion of 50 mas/yr. The dotted line is the parallactic path of the system, with the looping behavior due the apparent motion of the star against background stars due to the orbit of the Earth around the Sun. The solid line shows the photocenter motion due to the orbiting planet, with the perturbation increased by a factor of 30 to make the effect visible. This figure is taken from Perryman (2011).

Method	Signal	Scaling Law	Signal size	Fundamental frequency	Signal Duration	Detectability
Radial velocity	Stellar line shift/velocity	$v = 9.8 \left(\frac{P}{1\text{yr}} \right)^{-\frac{1}{3}} \left(\frac{M_p \sin i}{M_{\oplus}} \right) \left(\frac{M_s}{M_{\odot}} \right)^{-\frac{2}{3}} \text{cm/s}$ $\frac{\Delta\lambda}{\lambda} = \frac{v}{c}$	$\frac{\Delta\lambda}{\lambda} = 10^{-8}$ (Jupiter) $\frac{\Delta\lambda}{\lambda} = 10^{-10}$ (Earth)	Orbital period, days to decades	Always	Always, but high inclination angles problematic
Transit	Stellar intensity decrement	$\Delta f = \left(\frac{R_p}{R_*} \right)^2$	$\Delta f \approx 10^{-2}$ (Jupiter, Sun) $\Delta f \approx 10^{-4}$ (Earth, Sun)	Orbital period, days to decades	Hours	$P = R_*/a$, typically < 1%
Astrometry	Stellar photo-center shift	$\alpha = \left(\frac{M_p}{M_*} \right) \left(\frac{a}{1\text{AU}} \right) \left(\frac{d}{1\text{pc}} \right)^{-1} \text{arcsecond}$	$\alpha \approx 5 \cdot 10^{-4''}$ (Jupiter, Sun, d=10 pc) $\alpha \approx 3 \cdot 10^{-7''}$ (Earth, Sun, d=10 pc)	Orbital period, days to decades	Always	Always
Microlensing	Background star intensity perturbation	Complex, but very weakly dependent on planet mass(!)	$\frac{\Delta\text{Mag}}{\text{Mag}} > 0.05$	None, single-shot observation	Few hours to few days	10^{-6} events/star/yr near galactic bulge
Direct imaging	Planet light	$\frac{f_p(\phi, \lambda)}{f_*(\lambda)} = A(\lambda) \left(\frac{R_p}{a} \right)^2 g(\phi)$ (reflected) $f_p = \pi R_p^2 B(\lambda, T_p)$ (self luminous)	$\frac{f_p}{f_*} \approx 10^{-10}$ (Earth, Sun visible to near-IR) $\frac{f_p}{f_*} \approx 10^{-8}$ (Jupiter, Sun visible to near-IR)	0, Non-temporal signal	Always	Always

1.2 Astrometry, Transits, and Microlensing

1.2.1 Astrometry

The astrometric method of discovering planets is one of the oldest, and was the most heavily pursued methods up until the relatively recent past, likely due to its prolific results with binary stars. The principle of the astrometric discovery method is to look for small, periodic changes in the position of a star caused by unseen orbiting companions. By measuring many points of the angular deviation of the star against the plane of the sky, in particular nearby reference stars, the presence of a companion may be inferred, and all 7 Keplerian orbital elements can be determined. Since the center of mass in a system governed by gravitational forces is stationary, a planet and its host star will both orbit their (fixed) barycenter. In that case, the center of mass of the star will trace out an ellipse of semi-major axis

$$a_* = \frac{M_p}{M_*} a_p \quad (1.1)$$

where a refers to the semi-major axis of the orbit. The units of a_* are physical lengths; in reality, what is measured is an angular motion in the plane of the sky, which depends on the distance to the star. These two quantities may be related via the definition of the parsec, as

$$\alpha = \frac{M_p}{M_*} \frac{a_p}{d} = \left(\frac{G}{4\pi^2} \right)^{\frac{1}{3}} \frac{M_p}{M_*^{2/3}} \frac{P^{2/3}}{d} \quad (1.2)$$

$$= 3\mu\text{as} \frac{M_p}{M_{\oplus}} \left(\frac{M_*}{M_{\odot}} \right)^{-2/3} \left(\frac{P}{1\text{yr}} \right)^{2/3} \left(\frac{d}{1\text{pc}} \right)^{-1} \quad (1.3)$$

where the second equality follows from Kepler's laws (see Section 1.3.2.1), relating the semi-major axis of the planet's orbit to the period, with the approximation that $M_p \ll M_*$. From this equation, it can be seen that an Earth-like planet at 10 parsecs away would give a signal of about $3 \cdot 10^{-7}$ arcseconds, or $0.3 \mu\text{as}$. In order to measure this signal, or course, multiple epochs would have to be observed. The fundamental, shot-noise limited accuracy as to how well a telescope can measure a position is determined by the angular resolution and signal to noise ratio in a simple way (Lindgren, 1978):

$$\sigma_{ph} = \frac{\lambda}{4\pi D} \frac{1}{\text{SNR}} \quad (1.4)$$

where D is the dish diameter, but almost no telescopes on the ground approach this limit due to the deleterious effects of atmospheric turbulence. It should also be pointed out that Equation 1.4 scales, at best, as the square root of the exposure time, as given by the Poisson statistics of photon shot noise, so getting a tenfold improvement in precision requires over one hundred times

as long an observation. The limit from the ground achieved so far appears to be around $100 \mu\text{as}$ in optical and near-infrared frequencies (Cameron et al., 2009), with some groups claiming a precision of $30 \mu\text{as}$. These observations all use adaptive optics to sharpen the point-spread functions of the stars, and are thus limited in reference stars by the size of the “isoplanatic patch”; that is, the area of sky which can be well corrected by the adaptive optics system. Space telescopes have better accuracy, with Hubble being able to achieve $\sim 30 \mu\text{as}$ in drift scanning mode (J. MacKenty, private communication). Radio interferometers currently lead, with long-baseline setups achieving about $10 \mu\text{as}$ (Reid & Honma, 2014). (The D term in the equation above is replaced by the baseline of the interferometer separation, and the scale factor is slightly modified as well.)

The astrometric method has some useful features. It is one of the few methods where more distant planets give a stronger signal, though the time needed to make such a measurement is on the order of the orbital period, which mostly negates this advantage. Also, it has no fundamental “blind” spots; with a sufficiently accurate instrument, and enough observing time, a planet cannot “hide,” which is not the case for transits, microlensing, or sometimes radial velocities. It is the only method that can, even in principle, directly measure masses of Earth-type planets around nearby stars. Furthermore, it measures all the Keplerian orbital parameters directly, and may be used to derive a dynamical mass for the planet. Astrometric measurements may also be used to efficiently break the degeneracy between mass and inclination angle that the radial velocity method gives.

Drawbacks of the astrometric method are its reliance on the orbital period for a signal detection, the linear decrease in signal strength with stellar distance, and the substantial technical challenges in measuring such small photocenter shifts over long periods of time. It is currently the least successful method of exoplanet discovery, and has no confirmed detections, and many notorious false claims being later ruled out by follow-up observations. However, the astrometric method has a bright future. The *Gaia* mission, a dedicated space telescope optimized for high-precision astrometry, is expected to discover thousands of planetary systems. The data set from this mission will enhance our knowledge of massive, long-period planets like Jupiter and Saturn, of which almost no analogs are currently known.

1.2.2 Transits

The transit method detects planets by measuring the dip in starlight when a planet passes in front of its parent star. It is by far the most successful method used to discover planets, with the *Kepler* spacecraft having found thousands of planets, ranging from sizes of smaller than Earth to larger than Jupiter. Transiting exoplanets have revolutionized the field, allowing detailed statistical analyses of exoplanet populations and demographics, including discovering “Super Earths,” planets with sizes intermediate to Earth and Neptune, which are possibly the most common type of planet in the universe.

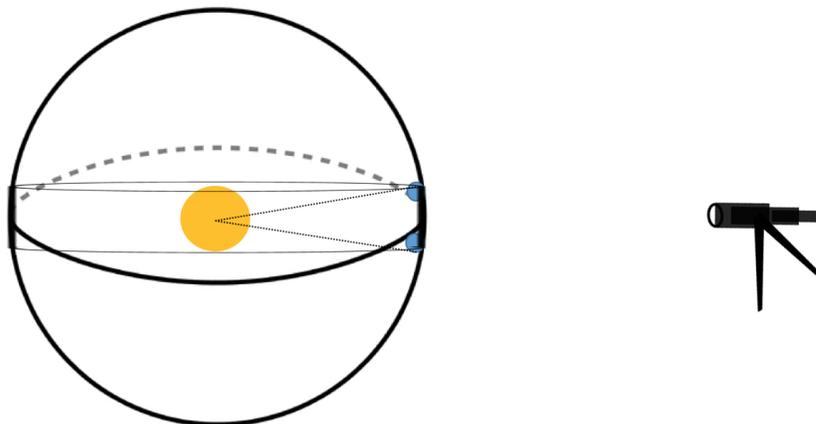


Figure 1.3: The transit probability can be calculated as the solid angle of the stellar diameter rotated about the orbital axis. See the text for details.

The fundamental physics of the transit method are quite straightforward to describe. The probability of a transit event occurring can be calculated by

$$P(\text{transit}) \approx \frac{(2\pi a)(2R_*)}{4\pi a^2} \quad (1.5)$$

$$= \frac{R_*}{a} \quad (1.6)$$

$$= 0.005 \left(\frac{R_*}{R_\odot} \right) \left(\frac{a}{1\text{AU}} \right)^{-1} \quad (1.7)$$

where R_* is the stellar radius and a the semi-major axis. The first line of the equation follows from imagining the extreme case where a planet just touches the top or bottom of the stellar limb (see Figure 1.3). The top and bottom of the star define the height of a cylinder of radius a when rotated about all possible viewing angles. The solid angle of this cylinder is A_{cyl}/a^2 , so the probability is just this solid angle divided by the total solid angle of a sphere, 4π . (Here we are ignoring effects like orbital eccentricity, which can somewhat boost the transit probability.) The equation shows that the probability of detecting an Earth-like planet is about one half of one percent, and that this does not depend on the distance to the star.

The transit depth is also possible to approximate fairly accurately. Assuming a stellar disk of uniform brightness, the fractional depth will be

$$\frac{\Delta F}{F} \approx \left(\frac{R_p}{R_*}\right)^2 \quad (1.8)$$

$$= 84 \cdot 10^{-6} \left(\frac{R_p}{R_{\oplus}}\right)^2 \left(\frac{R_*}{R_{\odot}}\right)^{-2} \quad (1.9)$$

where the first line just expresses the ratio of the sizes of the planet and star. We note that the transit method cannot measure the radius of the planet, only its relative size. Here, effects of limb darkening (the edges of the star being darker than the center) are ignored. From the ground, photometric accuracy is mostly limited by the atmosphere, with transparency variations and scintillation dominating the error budget. Generally, as will be described in Chapter 3, the accuracy appears to be limited to about $\Delta F/F = 10^{-3}$, making transit work from the ground only sensitive to planets that are about a thirtieth the radius of the star (of course, photon shot noise can affect this). From space, the main limitations are photometric stellar surface variations (at the 20×10^{-6} level) and pointing jitter of the telescope.

The duration of a transit depends on the impact parameter of the planet, that is, the height above or below the center of the star that the planet crosses. For an impact parameter of zero, the orbit can be approximated by

$$t_T \approx P \frac{2R_*}{2\pi a} \quad (1.10)$$

$$= \frac{2R_*}{\sqrt{GM_*}} a^{0.5} \quad (1.11)$$

$$= 13 \text{hours} \cdot \left(\frac{M_*}{M_{\odot}}\right)^{-0.5} \left(\frac{a}{1\text{AU}}\right)^{0.5} \left(\frac{R_*}{R_{\odot}}\right) \quad (1.12)$$

t_T is the transit duration and P is the orbital period. The first line simply states that the transit duration is equal to the period of the orbit multiplied by the fractional amount of the orbital length the star diameter subtends. The second line is derived from using Kepler's law to substitute the semi-major axis for the period (see Section 1.3.2.1). Substituting physical units on the third line shows that the transit is a very sparse signal which occurs during a tiny fraction of the orbital period. This makes transit work absolutely require dedicated, continuous surveys.

The advantages of the transit method are that the signals are relatively large (10^{-2} to 10^{-5}) compared to other methods like radial velocities, astrometry, and direct imaging. The signals also contain a very nice amount of information about stellar and planetary parameters, including limb darkening coefficients. One other second-order effect that is remarkably sensitive to low-mass planets is the measurement of transit time variations, which are caused by other bodies in the system gravitationally interacting with the transiting planet.

Disadvantages of the transit method are that it is completely blind to most planetary systems, and the transit signal is at the same period as the orbital frequency, and extremely sparse, requiring constant monitoring. This makes it exceptionally insensitive to outer planets, more so than any other method. Furthermore, it does not directly measure the radius of the planet, requiring follow-up on the parent star, and it also has no way of measuring the mass, even in principle. This requires investigation either from radial velocity or astrometry.

1.2.3 Gravitational microlensing

Gravitational microlensing is a method of discovering planets by observing magnifications of background stars when a planet-hosting star passes in front of them. It is a bizarre method, where neither the planet light nor the light from the host star is detected. The physics behind gravitational microlensing is complex, and the probability of a lensing event is tiny, and followup observations to characterize any detections are impossible. Yet microlensing holds a unique promise in being able to discover a population of planets from well below Mars size to Jupiter size at separations beyond 1 AU, even free-floating planets. So far, it has discovered about 20 planets, including a 2-Earth mass planet at 1 astronomical unit separation (Gould et al., 2014). Unlike other methods of planet detection, requiring factors like 10^{-6} or 10^{-10} to describe the planetary signal, the signals from lensing events may be magnifications of stars by factors of >100 , with planetary perturbations modifying these signals by tens of percent, over the course of minutes to hours. This is incredibly powerful.

Properly deriving the microlensing signal of a planet requires pages of algebra, and extracting planetary parameters from magnification light curves is extremely challenging, with many degenerate parameters confounding easy interpretation. Given that this thesis contains no microlensing work, and that the author of this thesis has a weak grasp of the technique, we declare defeat, content ourselves with providing a basic overview of the effect following Gaudi (2010), and direct the reader to that source for more detail.

The microlensing technique relies on the magnification of a background, or “source” star, by a foreground, or “lens” star. The magnification occurs because the lens star mass bends the spacetime around it, and photons of the source star will be deflected, causing a focusing effect which depends on the lens-source-observer position. In principle, this focusing distortion would be observable as two binary ovals of light around the source star, or an “Einstein ring” in case of perfect alignment, but this would require a telescope with a diameter of about a kilometer to resolve. Instead, the excess flux is detected, and the temporal evolution of the magnification due to the source-lens separation (caused by their relative motions through the galaxy) is the microlensing signal. For source stars near the galactic center, the typical timescale of a lensing event is one month, though it may last from days to years.

Referring to Figure 1.4, the lens is at position L , the source at S , and the image of the source at

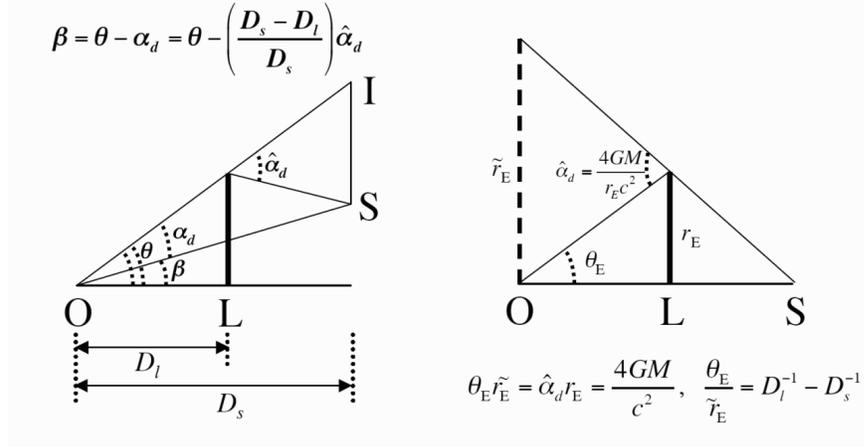


Figure 1.4: The relation between various angles relevant in gravitational microlensing. See the text for details. This figure is taken from Gaudi (2010).

I. The lens-observer distance is D_l , the source-observer distance is D_s . The angle between the true position of the source and the true position of the lens is β , the angle between the position of the lens and the image of the source is θ , and the angles between the position of the image of the source and the real position of the source is α_d . Finally, the bending angle $\hat{\alpha}_d$ gives the amount the source light is bent by gravity. We note that all these angles are tiny, and so neither the source, image, or lens are resolvable.

Since the angles are so small, the line from S to where it bends has length $\approx D_s - D_l$, so the physical distance from I to S is $\hat{\alpha}_d(D_s - D_l)$. Similarly, the S to I distance is given by $\alpha_d D_s$. With this equality, we have $\beta = \theta - \alpha_d = \theta - (D_s - D_l)/D_s \hat{\alpha}_d$. At this point, this has all been trigonometry, with no physics. The physics comes in with solving for $\hat{\alpha}_d$ in terms of the mass of the lens L and separation D_l , given by general relativity as $\hat{\alpha}_d = 4GM/(c^2 D_l \theta)$, where M is the lens mass. Thus,

$$\beta = \theta - \frac{4GM}{c^2 D_l \theta} \frac{D_s - D_l}{D_s D_l} \quad (1.13)$$

We note that we reproduce the famous ‘‘Einstein ring’’ for perfect alignment, setting $\beta = 0$ and solving for θ . This gives

$$\theta_E = 550 \mu\text{as} \left(\frac{M}{0.3 M_\odot} \right)^{0.5} \left(\frac{D_l - D_s}{D_s D_l} \right)^{0.5} \quad (1.14)$$

If we normalize all angles by the Einstein angle θ_E , as $u \equiv \beta/\theta_E$, $y \equiv \theta/\theta_E$ (from Figure 1.4), we find that Equation 1.13 reduces simply to $u = y - 1/y$, a quadratic equation. This means that there will, in general, be two images of the source on the sky, which tend to look like elongated ellipses.

It turns out that the total magnification of the source is simply related to u as

$$A(u) = \frac{u^2 + 2}{u\sqrt{u^2 + 4}} \quad (1.15)$$

where for very close alignments, u (also β) goes to zero and A is approximately $1/u$. This shows that magnification can go to very large values due to a microlensing event. The actual detected magnification is time dependent, due to the motion of the source and lens in the galaxy, and reaches a maximum when the source is at minimal angular separation from the lens. However, when a planet is present around the lens star, the curve will show a deviation from the function presented above, with a characteristic deviation size being on the order of a few percent to more than 100%, depending on the mass ratio and orbital parameters. An example microlensing curve with a planetary deviation is shown in Figure 1.5.

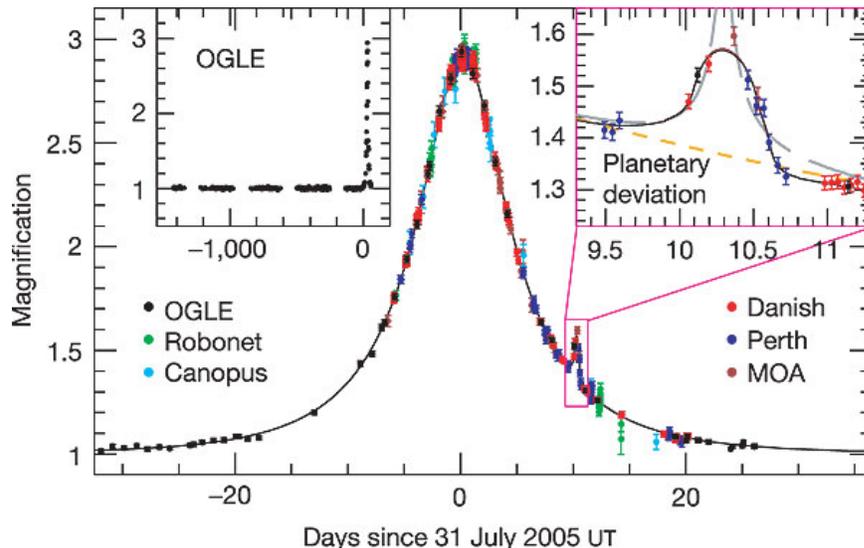


Figure 1.5: An example of a microlensing light curve. The large “main” magnification event of ~ 2 magnitudes, or a factor of 6.3 in brightness, corresponds to the lens star, and is well described by Equation 1.15. The small perturbation at the 5% level corresponds to the planet around the M-dwarf lens star, a $5.5 M_{\oplus}$ cool Super-Earth at a ~ 2.6 AU separation. Different colored points correspond to different observatories. This figure is taken from Beaulieu et al. (2006)

The minimum information that can be gained from a planetary microlensing event is the mass ratio between the planet and star and the projected separation in units of the Einstein radius. In many cases, it is possible to measure the Einstein ring in angular units, giving a mass-distance relation. Parallax measurements, detecting light from the lens itself, and proper motion can also break other degeneracies among the source-lens system, including direct measurements of the planet mass, and occasionally full orbital solutions.

In summary, microlensing method has many advantages. It has the most sensitivity to low-mass

planets on few-AU orbits, a region of parameter space that is mostly unknown and cannot be easily accessed by other methods. Furthermore, the characteristic signal preceded by a large magnification of a star, easily measurable by a small telescope. Even the small planetary deviation is easily measurable. Furthermore, the signal does not depend on the planetary orbital period. Disadvantages include the low probability of microlensing events, the highly nonintuitive and technical data analysis required, and the lack of information about the system gained in the “minimal information” case. We do not consider the “single-shot” measurement and lack of followup opportunities to be a disadvantage, as it is highly appealing to be able to publish a paper after a single observation. It has a very bright future in the WFIRST-AFTA mission, and will likely detect over a thousand planets at low masses and reasonable separations.

1.3 Radial velocities

1.3.1 Overview

The idea that planets might be detectable using spectroscopy seems to have first been put forward by Struve (1952) in a remarkable paper predicting both the existence of hot Jupiters (“...*it is not unreasonable that a planet might exist at an orbit of 1/50 astronomical unit...*”), the necessary technology to detect them (“...*the observed radial velocity...might just be detectable with the most powerful Coude spectrographs...*”) and the ability to do transit followup (“*There would, of course, be eclipses...ascertainable by modern photoelectric methods*”).

Struve’s far-reaching insight seems not to have had much effect for at least 20 years. High resolution spectrographs capable of measuring velocities up a few meters per second had been around for some time, but the accuracy seemed limited to a few hundred meters per second for reasons that were not understood until Griffin (1973). Griffin and Griffin explained that changes in the way the calibration lamp and stellar light were illuminating the spectrograph caused dramatic systematic offsets, and proposed a method to reduce radial velocity errors using absorption lines in the Earth’s atmosphere as fixed wavelength references, suggesting that precision on the order of 10 m/s might be reached, and *demonstrating* precisions of 30 m/s and 50 m/s on the stars Arcturus and Procyon. This may rightly be called the birth of precision radial velocities.

The first precision radial velocity survey was undertaken by Bruce Campbell and Gordon Walker from 1980 to 1992, which also had another key development, the use of absorption gas cells as common-path calibrators (Campbell & Walker, 1979).¹ Using hydrogen fluoride, they were able to achieve 15 m/s on a sample of 21 stars (Walker et al., 1995). While they did not detect any planets, the unprecedented long-term precision and monitoring set a standard for other groups to follow.

¹see Section 1.3.3.2 for more detail

There is an interesting afterthought to the Campbell and Walker study. In their sample, the star ϵ Eri actually has a planet of high eccentricity, on a seven year orbit, with an amplitude sufficiently high to have been detectable ($>19\text{m/s}$), finally discovered many years later by Hatzes et al. (2000). Why was it not detected? Perhaps there is a simple answer: in the text, as they stated that “On the other hand, with the Solar System satellites, and the pulsar...planets all being in nearly circular orbits, it is not unreasonable to limit our search to single giant planets in circular orbits.” One wonders if history would have taken a different track had eccentric orbits been considered.

Later in the same year that Walker published the results of their survey, the modern field of exoplanets was born with the discovery of 51 Pegasi b, a planet smaller than Jupiter on an extremely tight, 4-day orbit around a main-sequence star. (The orbit was about a factor of 2 off from Struve’s original prediction 43 years before.) Swiss astronomers Michel Mayor and Didier Queloz’ discovery was quickly confirmed by an American group led by Geoffrey Marcy and Paul Butler, who had a similarly sensitive instrument. Marcy and Butler followed up with two more discoveries: 70 Virginis b (Marcy & Butler, 1996) and 47 Ursae Majoris b (Butler & Marcy, 1996a), planets with periods of 100 days and 2.5 years.

The precision with which radial velocities could be measured continued to improve, with the Swiss and American teams leading the push down to about 1 m/s, discovering hundreds of planets in the process, including planets down to nearly the mass of Earth, though in very tight orbits. The radial velocity method of detecting exoplanets is now a mature and successful technique, with the current instruments now so sensitive that they are mostly limited by astrophysical noise sources like stellar surface variations.

In the following sections, we will discuss the physical principles behind precision radial velocity detection of exoplanets. First, we will review the physics of planetary orbital dynamics, and the expected effects of planetary motion on stellar spectra. We will discuss how these effects drive instrument requirements, and the major sources of noise encountered in precision radial velocity work, including instrumental and astrophysical sources.

1.3.2 Physical principles

1.3.2.1 Kepler’s Laws

Kepler’s three laws of solar system dynamics are that (1) the orbit of a planet is an ellipse, with the Sun at one focus; (2) equal areas are swept out by the line joining the planet and the Sun in equal times; (3) the squares of the orbital period are proportional to the cubes of the semi-major axes.

In the case of the first law, a Keplerian ellipse is specified by seven parameters, listed in Table 1.3.2.1 and sketched in Figure 1.6. This equation may appear in slightly different forms and with slightly different parameters depending on the reference consulted.

sition, measured from the center of the ellipse. This variable will be useful later in deriving the instantaneous radial velocity of the orbit.

The period, semi-major axis, eccentricity, and epoch are “fundamental” parameters expressing the physics of the orbit and do not depend on the specifics of the observer. The inclination, longitude of ascending node, and argument of pericenter are projection effects depending on where in the Galaxy the observer is located with respect to the system under study; they do not contain any physically relevant information.

The third law specifies a relationship between the period of the orbital ellipse and the semi-major axis, which depends only on the masses of the star and planet. In the case of a more general two-body problem, without reference to a particular star or planet, Kepler’s laws are still true, though we note that both the star and planet will also orbit their (fixed) center of mass in elliptical orbits. Also, the “relative” orbit of the planet with respect to the star is *still* an ellipse, with semi-major axis equal to the *sums* of the semi-major axes of the planet and star about the center of mass. In absolute terms, it is found that:

$$P^2 = \begin{cases} \frac{4\pi^2}{G(M_*+M_p)} a_{rel}^3 & \text{(relative orbit)} \\ \frac{4\pi^2}{G \frac{M_p^3}{(M_*+M_p)^2}} a_*^3 & \text{(stellar orbit about c.o.m.)} \\ \frac{4\pi^2}{G \frac{M_*^3}{(M_*+M_p)^2}} a_p^3 & \text{(planetary orbit about c.o.m.)} \end{cases}$$

where the period, P , is the same for all cases, M_i, a_i refers to the masses and semi-major axes of the stars and planets, and $a_{rel} = a_* + a_p$ is the relative orbit of the planet around the star. G is Newton’s gravitational constant. Kepler and other solar system dynamicists concern themselves with the first form, people trying to discover planets with astrometry with the second, and direct imaging people with the third. It should also be noted that in most cases of interest, $M_p \ll M_*$, and so $M_* + M_p \approx M_*$, to high accuracy, leading to a simplification of the equations above.

1.3.2.2 Radial velocity determination

Referring to Figure 1.6, we see that the radial velocity of the star—the velocity in the direction of the observer—is given by

$$v_r(t) = \dot{z}(t) = \frac{d}{dt} [r(t) \sin i \sin(\omega + \nu(t))] \quad (1.17)$$

$$= \sin i [\dot{r}(t) \sin(\omega + \nu(t)) + r(t) \dot{\nu}(t) \cos(\omega + \nu(t))] \quad (1.18)$$

Note that r refers to the polar coordinate of the ellipse, with respect to the focus, as in Equation

1.16. (To derive the first equality, it is helpful to project r along the descending node (see Figure 1.6); this is equal to $r \sin(\pi - \omega - \nu(t))$, then project along the vertical direction by multiplying by $\sin i$ to get z .) Some work is needed to express $r(t)$ and $\nu(t)$ as a function of $\nu(t)$. We skip the details, but the key step requires Kepler's second law, which shows that $r(t)^2 \dot{\nu}(t)$ is a constant that can be connected to the eclipse parameters. Finally, we obtain that

$$v_r(t) = \left(\frac{2\pi}{P} \frac{a_* \sin i}{\sqrt{1-e^2}} \right) \left(\cos(\omega + \nu(t)) + e \cos(\omega) \right) \quad (1.19)$$

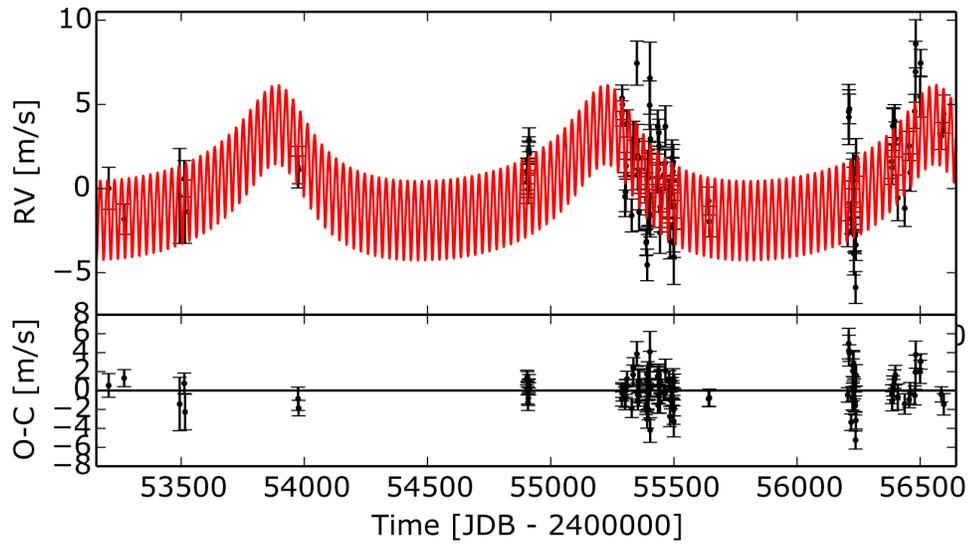
Equation 1.19 is important. The time series $v_r(t)$ is what is measured from astronomical spectrographs (see Figure 1.7). The first term in parentheses is called the “radial velocity semi-amplitude,” usually denoted as K , and the extreme values of the velocity are given by $K(1 + e \cos \omega)$ and $K(e \cos \omega - 1)$. The shape of the curve is determined by the values of e , a physical parameter, and ω , a parameter dependent on the viewing angle. The equation above is still implicitly in the time domain via the mean anomaly, but may be explicitly transformed into the time domain in a few steps by solving a transcendental formula known as Kepler's equation.² It is also apparent that not everything about the orbit can be measured. In fact, of the seven Kepler orbital parameters, only four can be determined: eccentricity e , argument of periastron ω , time of periastron t_p , and period P . The longitude of the ascending node Ω (not ω) cannot be measured at all, as any rotation of the orbital ellipse about a vector normal to the plane of the sky results in the same radial velocity. K can also be determined, but it is a combination of the period, eccentricity, semi-major axis a and orbital inclination i , with the latter two being degenerate. This directly translates into a degeneracy in the mass, by substitution of Kepler's third law:

$$\left(\frac{2\pi G}{P} \right)^{1/3} \frac{M_p \sin i}{(M_* + M_p)^{2/3}} \frac{1}{\sqrt{1-e^2}} \left(\cos(\omega + \nu(t)) + e \cos(\omega) \right) \quad (1.20)$$

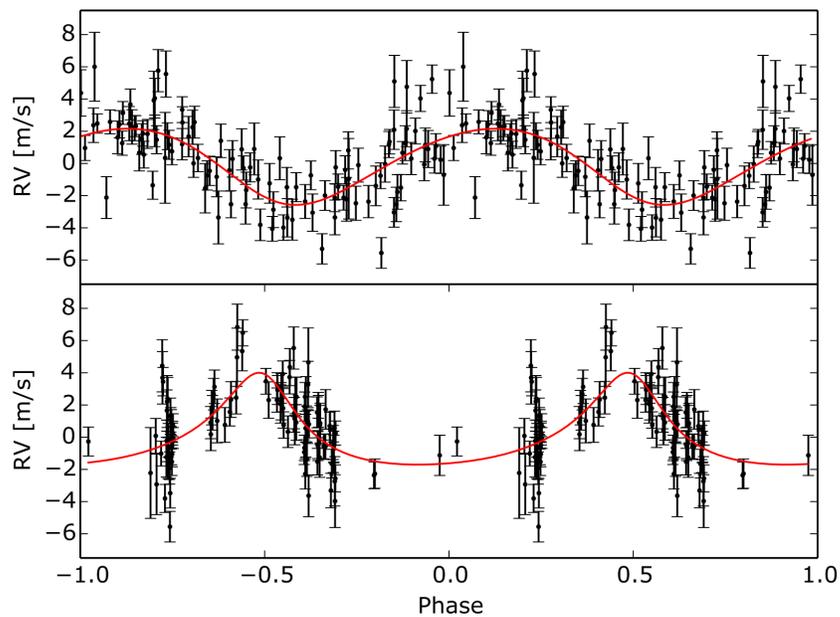
where Equation 1.3.2.1 is used to remove the dependence on a_* . Here it can be seen that only the quantity $M_p \sin i / (M_p + M_*)^{2/3}$ may be measured.

To get a better handle on the planetary and stellar masses of the system, a few things may be done. The most common one is to perform spectroscopic analysis of the star. Stellar spectral typing is a mature technique, and by determining the depth of different elemental lines in stars, the temperature, luminosity, and mass may be estimated. Knowing the mass of the star allows one to determine the mass of the planet up to the inclination factor, that is, $M_p \sin i$. After that

²The steps are: (0) Find the time since periastron passage, t ; (1) Compute the mean anomaly $M = \frac{P}{2\pi} t$; (2) Compute the eccentric anomaly E by solving Kepler's equation $M = E - e \sin E$; (3) Compute the true anomaly θ by solving $(1-e) \tan^2(\theta/2) = (1+e) \tan^2(E/2)$; (4) Compute the distance from the star $r = a(1 - e \cos E)$. This gives everything in polar coordinates r, ν .



(a) Raw time series with fit



(b) Phase-folded time series

Figure 1.7: An example of a radial velocity measurement at about the current limit of precision of 1-2 m/s, taken from the HARPS spectrograph (Mortier et al., 2016). (a) Radial velocities of the star HD175607 are measured over a significant time period. Note the x-axis of the top curve is in days, corresponding to nearly ten years of observations. The red curve shows the model fit, with the rapid variations due to the inner planet and the slow variations to a potential outer companion (b) The phase-folded light curves and fits of the signals, corresponding to periods of 29 days (top) and 1400 days (bottom).

point, nothing may be done unless the planet is fortuitous enough to transit, which sets $\sin i \approx 1$. Otherwise, the radial velocity method only returns a minimum mass. However, the mass-inclination

degeneracy is only problematic for individual targets, not surveys, as the average value of $\sin i$ is $\pi/4$. As a final step, we re-express the semi-amplitude in Equation 1.20 in more amenable physical units, rescaling to planetary semi-major axis and assuming a circular orbit and $M_p \ll M_*$:

$$K = 9\text{cm/s} \left(\frac{M_p \sin i}{M_\oplus} \right) \left(\frac{M_*}{M_\odot} \right)^{-1/2} \left(\frac{a}{1\text{AU}} \right)^{-1/2} \quad (1.21)$$

For reference, the values of the radial velocities of known and hypothetical planets are shown below in Table 1.3.2.2:

Name	$M_p [M_\oplus]$	$M_* [M_\odot]$	$a[\text{AU}]$	$P[\text{yr}]$	$K [\text{m/s}]$	Note
Earth	1	1	1	0.09	1	
Jupiter	318	1	5.2	12.7	12	
Neptune	17.1	1	30.1	164.8	0.3	Beyond purview of RV method
51 Pegasi b	150	1.1	0.052	0.01	56	First planet found around main-sequence star
Hot "Super Earth"	5	1	0.1	0.1	1.5	
Habitable "Super Earth"	5	1	1	1	0.5	Just beyond precision attainable today
Habitable "Super Earth" around M dwarf	5	0.3	0.1	0.06	2.5	Habitable zone estimate

1.3.2.3 Radial velocity measurement

We have so far discussed how an orbiting planet will impart a periodic radial velocity shift on its parent star, and quantified how large that shift would be. In this section, we will explain how the velocity shift leads to a shift in the wavelength, and how the latter can be measured.

The key to understanding how the motion of the star is observable is through the relativistic Doppler shift, first derived by Einstein in 1905, and given by the following formula:

$$\lambda_{\text{obs}} = \lambda_{\text{em}} \frac{1 + \frac{v}{c} \cos \theta}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (1.22)$$

where λ_{obs} , λ_{em} are the observed and emitted wavelengths of light, θ is the relative angle between the source and observer, v is the velocity of the source, and c is the speed of light. In practice, what is measured are the shifts of different spectral lines, with respect to a reference. When a star is moving toward the observer, the spectral lines will shift to bluer wavelengths; when the star is moving away, they will shift to redder wavelengths. In the limit of small speeds such that $v/c \ll 1$, and referring only to radial velocities now (absorbing the $\cos \theta$ term into v_r), we may transform the above expression into a more amenable form:

$$\lambda_{\text{obs}} = \lambda_{\text{em}} \left(1 + \frac{v_r}{c} \right) \quad (1.23)$$

$$\Rightarrow \frac{\lambda_{\text{obs}} - \lambda_{\text{em}}}{\lambda_{\text{em}}} = \frac{v_r}{c} \quad (1.24)$$

$$\frac{\Delta\lambda}{\lambda_{\text{em}}} = \frac{v_r}{c} \quad (1.25)$$

This is the fundamental equation of the precision radial velocity method.

A few points are in order, as this is not the full story, and there are a number of complicated effects in astrophysical situations that will effect the true wavelength measured by the observer. The starlight has to climb out of the potential well of the star, which causes a redshift described by a general relativistic formula depending on the mass of the star. Additionally, convective motions in the star preferentially blueshift the light. These kinds of effects, however, tend to contribute a fixed offset, and are not entirely relevant to measuring periodic variations. Similarly, the motion of the star through the galaxy will contribute a linear radial velocity offset. Other terms, such as the gravitational potential of the solar system, the orbital motion and rotation of the Earth, latitude of the telescope, etc, are absolutely relevant to precise radial velocities, contributing offsets of hundreds to many thousands of meters per second, and must be corrected for. However, these effects generally dependent on objects in the Solar system, whose positions and velocities are very well established, and thus do not cause any significant problems. Wright & Eastman (2014) characterize these systematic effects to a level well below 1 cm/s.

As a practical example, consider the H-alpha emission line at a wavelength of 656.3 nm. For a star moving at 50 m/s (such as one orbited by a hot Jupiter), the expected change in wavelength using the above formula would be about one part in ten million. How well such shifts can be measured, depending on the kind of star, the noise in the spectrum, and various errors, and how they affect survey yield, are the subject of the next chapter. However, we can derive a reasonably good approximation.

Consider a spectrum given by $I(\lambda)$, where I is a function of wavelength (see Figure 1.8). The flux measurement I has some uncertainty associated with it, whether due to photon shot noise statistics, read noise, or some other noise source. Call this σ_I . We are interested in the associated uncertainty in wavelength shift (and hence, velocity) associated with that noise. By the error propagation equation³, it can be seen that, for each pixel i ,

³If $y = f(x)$ then $\sigma_y = \sigma_x |df/dx|$

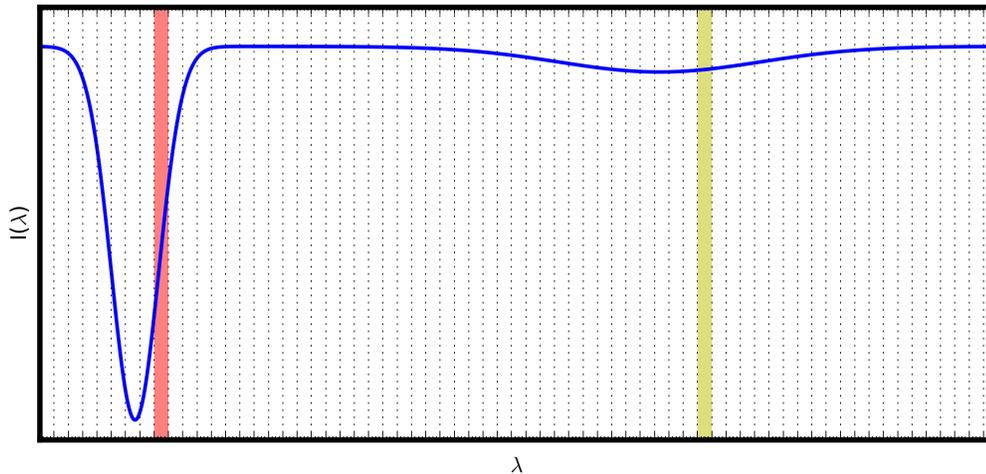


Figure 1.8: A fake spectrum of a star, with a deep absorption line and a shallow, broad absorption line. The x-axis is in wavelength, and the y-axis is in flux, both in arbitrary units. The pixels of the spectrum corresponding to the deep line have high quality Doppler information (red shading); the pixels of the shallow line have relatively poor quality Doppler content. The pixels of the continuum have no Doppler information whatsoever. This is equivalent to saying that it is impossible to determine the translation of a function of the type $y = \text{constant}$.

$$\sigma_{\lambda}[i] = \frac{\sigma_I[i]}{\left| \frac{dI}{d\lambda} \right|} \quad (1.26)$$

$$\Rightarrow \sigma_v[i] = \frac{c}{\lambda[i]} \frac{\sigma_I[i]}{\left| \frac{dI}{d\lambda} \right|} \quad (1.27)$$

where the derivative $dI/d\lambda$ is evaluated at each wavelength λ . Note that if I is measured in photons, and there is no detector noise, $\sigma_I = \sqrt{I}$ due to Poisson noise. For a larger chunk of spectrum, not just one pixel, we can evaluate the velocity precision as

$$\sigma_{\text{tot}} = \frac{1}{\sqrt{\sum \left(\frac{1}{\sigma_v} \right)^2}} \quad (1.28)$$

The takeaway from this is that high signal to noise is required (the σ_I term), spectra with dense, sharp lines are better (the $dI/d\lambda$ sum), and large spectral grasp is needed to reduce the random error contributed by each pixel.

What determines the slope term? First, instrumental effects will contribute to it: a spectrograph has an intrinsic resolution which corresponds to how much it will spread out an infinitely sharp spectral line. Another way of looking at this is that the resolution of the spectrograph corresponds to how close together two lines may be before they appear distinct on the detector. The resolution

of an instrument is usually expressed by the formula $R = \lambda/\delta\lambda$, a dimensionless number. To discover planets, we recognize that we should be operating at a resolution where the stellar lines are completely resolved, so that the instrument does not appreciably degrade the fundamental limit of velocity precision.

Stellar absorption lines are not infinitely narrow. At one extreme, the energy-time uncertainty principle contributes a “natural” line width, which is very small. Also, atoms in the stellar photosphere are constantly colliding with each other, shortening the lifetime of the emission process (and hence increasing the energy uncertainty). The hot temperatures also lead to high velocities, and via the Doppler effect, broadened lines. The bulk motion of the star also is a serious contributor. For many stars, the rotational broadening (via the Doppler shift) will dominate the width of the spectral line. Stellar (surface) rotation rates range from 2 km/s to dozens of km/s, and depend on things like spectral type and age. Therefore, the optimal targets are slow-rotating, bright stars with dense, sharp lines.

Since stellar rotation contributes the bulk of the line broadening, we can determine how much resolution is required for planet detection. For a Sun-like star, with a 2km/s rotation rate, we have, on either side of the star

$$\frac{\Delta\lambda}{\lambda} = \frac{\pm 2\text{km/s}}{c} \quad (1.29)$$

We approximate the “broadening” of the spectral line by assuming that the line shifts to either side by +2 and -2 km/s. (This is not quite right, as projection effects and the nonuniformity of the stellar surface velocities will affect the radial velocity measured; see Gray (2008), Chapter 17, for details on how to do this properly.) An instrument that completely resolves these lines would thus need a resolution of about

$$R \equiv \frac{\lambda}{\delta\lambda} = \frac{c}{4\text{km/s}} \quad (1.30)$$

$$\approx 75000 \quad (1.31)$$

Since a spectral line is resolved at 4km/s or $R = 75000$, it is easy to compute the actual shift in a spectral line due to a planet. For a 5 m/s planetary signal, the equivalent shift is $5/4000 \approx 0.001$ resolution elements or linewidths. The Nyquist-Shannon sampling theorem says that to properly extract the information of a line, we need to sample at least 2 pixels per resolution element, and grossly oversampling the line will not extract more information. This means that the shift of a line on the spectrograph detector is only a few ten thousandths of a pixel!

The previous two sections have demonstrated, from first principles, the magnitude of the effect needed to be measured, the type of instrumentation required to measure it, and the most amenable targets for such work. Of course, there are tradeoffs and optimizations, and much detail has been omitted, which is explored in Chapter 2. The discussion thus far may have given the impression that all that is needed is a high resolution spectrograph coupled to a reasonably sized telescope, and diligent observing over a few orbital periods. This is false. There are many noise sources in radial velocity studies, which contribute at various levels and timescales, which can completely swamp the small signals of interest. In the following section, we will discuss the major noise sources in radial velocity measurements, both instrumental and astrophysical.

1.3.3 Instrumentation and instrumental error sources

The previous section demonstrated that the spectral shifts induced by orbiting planets are quite small, on the order of 1/1000th of a pixel on a conventional CCD. These tiny shifts need to be tracked over periods corresponding to the planet's orbit, which may range from days to years. This requires both very high instrumental stability, and very stable calibration sources. While these used to be the main source of uncertainty in radial velocity programs, in the recent past, the problem of wavelength calibration seems to have mostly been solved, with laser frequency combs demonstrating intrinsic stabilities of better than a cm/s.

1.3.3.1 Stability control

Thermal, barometric (pressure), and mechanical control of spectrographs is a substantial challenge, especially as instruments on large telescopes must be commensurately sized to preserve high resolution (see Chapter 2). In order to determine the radial velocity effects of different environmental changes, finite-element analysis is required; however, simple calculations can show that in terms of the index of refraction of the air, a change of 0.01K in temperature or 0.01 Torr can be responsible for shifts of about 1 m/s. Stability of 0.001K and 10^{-7} Torr results in a system that does not have any substantial internal molecular transfer. This is the current goal for the next generation of instruments, though the time constant for such systems may be weeks to months.

While thermal, barometric, and mechanical changes may be tracked during the course of observations, at least in principle, errors like illumination control at the entrance to the spectrograph are much more problematic. For example, in a slit-fed spectrograph, it can be shown that a photocenter shift of only a percent in the star image can induce a shift of line centroid of tens of meters per second. Even with perfect guiding and tracking from the telescope, seeing variations and other atmospheric effects can contribute in ways that are difficult to simulate and impossible to track during an observation.

One important development to address this problem has been to move towards fiber-fed instruments, where the image of the star is injected into a length of fiber and then relayed to the input of the spectrograph. Fiber output illumination is quite stable to input illumination changes in the case of multimode fibers, with some caveats explained below. Single-mode fibers would provide the best possible illumination control, as the input image to the spectrograph would never change, except in brightness. However, the substantial technical challenge in efficiently coupling starlight into single-mode fibers makes them currently impractical for precision radial velocity work.

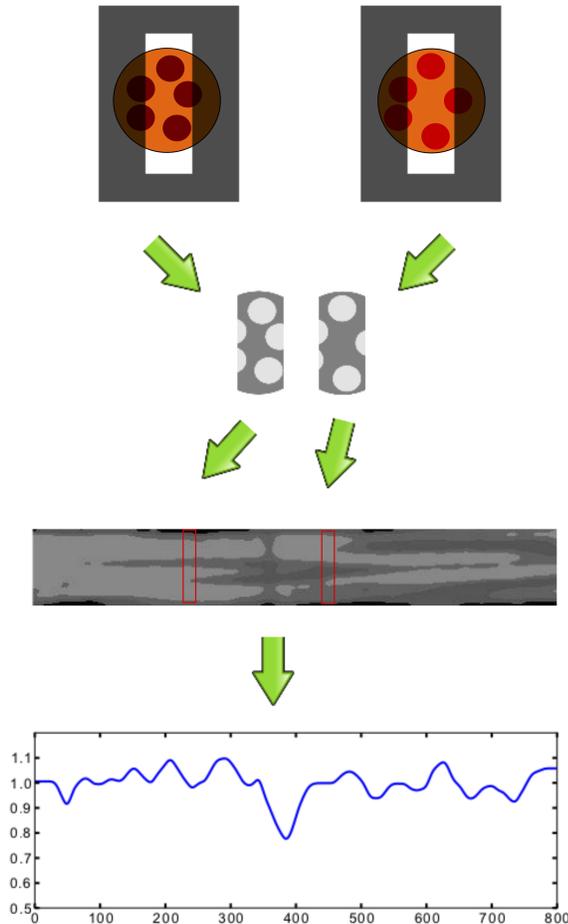


Figure 1.9: Modal noise: The end of a fiber has two slightly different intensity distributions for two different wavelengths (top). The addition of a slit exacerbates the differences (top middle). The two slits are imaged as resolution elements onto the detector by the spectrograph. The detector spectrum has spatial variations in the cross dispersion direction *as a function of wavelength* (bottom middle). The extracted spectrum (summed in the vertical direction) has quasi-periodic variations in the continuum intensity (bottom)

Instead, multimode fibers are used, typically those that support hundreds to thousands of modes at the wavelengths of starlight relevant for precision radial velocities. These have uniform output illumination, especially in the case of non-circular core fibers, for reasons not well understood. However, the finite numbers of modes interfering at the output of a fiber tip lead to quasi-static

illumination variations, which are wavelength dependent. This is called “modal noise” (see Figure 1.9). Modal noise is particularly bad in the infrared, where fibers support less modes, and hence less averaging at the fiber output tip. There are a number of solutions to this, including fiber agitation (or shaking, in common language), which causes the modal pattern to quickly change due to the variation of optical paths through the fiber, and double-scrambling systems (Barnes & MacQueen, 2010) which use ball lenses to invert the near and far-field of the fiber, as the far-field tends to be much more stable.

1.3.3.2 Calibration methods

Even with all due diligence on instrumental stabilization, it is not possible to measure radial velocities to close to the precision necessary to detect planets without using some kind of reference wavelength source. The use of a reference source changes an absolute wavelength measurement into a differential one, with very large increases in precision. There are two main techniques that are used: “self-referencing” via absorption gas cells, and “simultaneous referencing,” using a gas emission lamp or, more recently, a laser frequency comb.

The self-referencing technique passes starlight through a cell filled with absorptive gas in front of the spectrograph. The stellar continuum and lines are absorbed by the gas cell, and the detector measures the combined star and gas lines. The gas line depths and locations need to be characterized to very high resolution, usually through a Fourier transform spectrograph. The precise knowledge of this reference spectrum can be used to simultaneously model the instabilities in the instrument and the stellar spectrum shift. This method has been successful in improving radial velocity precision for slit-fed instruments (Butler et al., 1996), where the illumination control is challenging and changes to the line-spread function of the spectrograph would otherwise cause wild swings in the measured radial velocities. Disadvantages to the gas cell technique are the loss in stellar flux due to absorption of the cell (usually around 25% for molecular iodine, the most commonly used gas for optical instruments), the lack of good reference gases for infrared radial velocity work, and the complex data analysis necessary to disentangle the gas and stellar contributions.

The “simultaneous referencing” technique uses emission lamps to track changes to the instrument profile over the course of observations (these may be concurrent with observations, or more likely, once or twice an evening). The instrumental state must be unchanged between reference and science measurements, making this technique only appropriate for highly stable instruments. The advantage is that the starlight is measured “as-is”, without absorption and loss by dense gas cell lines. The most commonly used emission gas lamp used is Thorium Argon, which has some drawbacks, as different lines vary dramatically in brightness, and the lamp spectra change over long time periods. The measured “single-shot” stability is about 10-20 cm/s (Li et al., 2008); very high, but not high enough to detect Earths, which have reflex velocities of the same level.

A recent development which has dramatically improved the prospects of the simultaneous reference method is the laser frequency comb, a light source with highly stable ($<1\text{mm/s}$), bright, and uniformly spaced lines generated by a pulsed femtosecond laser (Li et al., 2008), with each line being separated by the repetition rate of the pulse. The one drawback of laser frequency combs is the need to suppress most of the lines using a Fabry-Perot cavity, as the pulse repetition rate leads to line spacings of 1 GHz, about a factor of 10 too close for a conventional spectrograph to resolve. The use of electro-optic modulators to control the pulse spacing can generate appropriately spaced lines without the cavity, using off-the-shelf equipment (Yi et al., 2016). Other recent advances include micro-resonator combs and lasers naturally pulsed at 10+ GHz. The sum total of these technologies will solve the calibration problem for the observable future.

1.3.4 Stellar error sources

Stars are complicated dynamical systems, which have a variety of kinds of “surface” motions. The poorly-characterized noise features in stellar radial velocity data sets are usually referred to as “stellar jitter”, and encompass a wide range of physics including asteroseismology, magnetic activity such as starspots and plages, granulation, and other effects. Stellar jitter is currently the most serious impediment to detecting low-mass planets in radial velocity datasets. Here we describe some forms of stellar jitter, their timescales, and their relative amplitudes. For a good exposition of these forms of stellar noise, see Dumusque et al. (2011c) and Dumusque et al. (2011a).

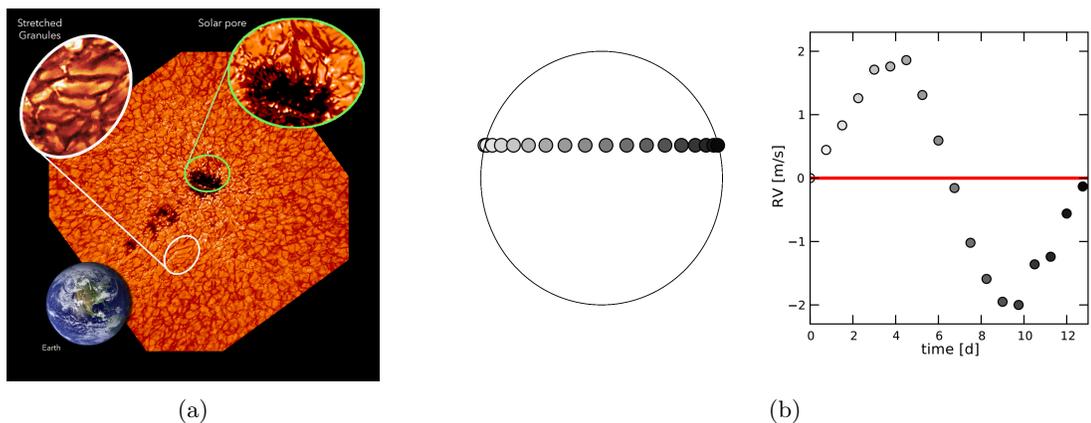


Figure 1.10: (Left) An image of the Sun’s surface, showing spots and convective granulation, with the Earth to scale.⁴(Right) An example of a radial velocity signal induced by a starspot. In this figure, the left part of the star is rotating towards the reader. As the spot rotates into view, following the stellar rotation period, a redshift is recorded. After the star crosses the center, a blueshift is recorded. For more complicated starspot paths, the signal would change in complex ways. This figure is taken from Dumusque et al. (2011a)

Magnetic activity jitter is a common form of noise, which occurs on the timescales of the stellar

⁴Image from <http://phys.org/news/2014-06-hidden-source-solar-surges.html>

rotation period, typically tens of days, as well as very long term cycles on the order of years. In the short term, as a star rotates, the radial velocity measured from the stellar photosphere is composed of blueshifted light from the part of the star rotating towards the observer, and redshifted light from the part rotating away. In a smooth, featureless photosphere, these motions would cancel out, resulting in a zero radial velocity signal. However, any differential brightness between the approaching and receding parts of the stars will create a net radial velocity signal. As an example, a starspot comprising a small fraction of the stellar surface will cause a radial velocity signal of a few m/s (see Figure 1.10). This starspot may last for months or longer, move around the star, and eventually disappear. Multiple, high cadence observations over significant periods of time would be required to disentangle this sort of signal from true underlying planetary signals. Long-term activity cycles, like the 11 year solar starspot cycle, can cause the mean level of noise to change significantly, but are not understood well enough to characterize their net effect on precision.

Magnetic activity can be countered in a few ways. First, target selection is important. Old, quiet, solar-type stars with less active dynamos have generally been the targets of choice for radial velocity surveys. Young stars, which are extremely active, make radial velocity investigation of low-mass planets essentially impossible. An empirical measurement of the activity-induced jitter by Hillenbrand et al. (2015) found that for stars younger than a few hundred million years, the jitter noise is tens to hundreds of meters per second, corresponding to masses of a third to a tenth of Jupiter’s. Observations of particular spectral emission lines, such as Ca II, known to correlate with activity, can flag data that is problematic. Another popular method is “bisector” analysis, where the exact shape of the spectral line is analyzed for tell-tale signs of starspots and plages (bright regions), which tend to skew the lines to one direction or the other, while planets shift the lines uniformly (Huerta et al., 2008). Finally, starspots and plages are less extreme at infrared wavelengths, so observations over independent bands (for example, H and V) can be used to distinguish activity from true planet signals, as the measured velocity at each band should be consistent with a planet, and different for a starspot.

Another surface effect is granulation, arising from the convective upwelling and downwelling of material in the photosphere (see Figure 1.10). The typical velocities of these effects are many km/s per granule, but the millions of granules on the surface of the star tend to average out the radial velocities to a much lower level over 15 minute timescales. These kinds of effects also have much larger scales, called “meso-” and “super-” granulation, with longer timescales of hours to a day, and scales of up to 30,000 km in the Sun. The presence of these phenomena require detailed, high cadence observing over at least a few days to successfully characterize and average over.

The final major form of jitter is stellar oscillations, which are caused by fast moving pressure waves in stars. These oscillations typically have bulk amplitudes of tens of cm/s, and periods of a few minutes. The stellar p -modes are such an example, with 30 cm/s amplitudes at 5-minute

periods. These may be averaged away by observing the star over multiple periods, up to several days, though the contributions of oscillation modes at very long periods is not perfectly understood.

The lack of control of stellar jitter at intermediate and long periods has seemed to impose a current velocity noise floor of slightly less than 1 m/s on the radial velocity method. No program has ever managed to exceed this level of precision over periods of time longer than a year, even with thousands of observations over many years. As an example of the difficulty of extracting small planetary signals at the instrumental precision, in the presence of many noise sources, a spectacular discovery of an Earth-mass planet orbiting alpha Centauri B at 0.5m/s amplitude (Dumusque et al., 2012), with an instrumental precision of about 0.7m/s, was recently disproven as an artifact of the data preprocessing (Rajpaul et al., 2016). It remains to be seen how the next generation of instruments, with precisions of about 10 cm/s, will be able to improve the planet yield in the presence of these noise sources.

1.4 Direct imaging of exoplanets

1.4.1 Overview

Direct imaging is a method of exoplanet science that seeks to analyze light from the planet itself. It is the only method that does so. Since direct imaging takes pictures of planets at a specific point in time, it does not rely on the planetary orbital periods for the signal, and so is sensitive to distant planets as well as those close to the parent star, and at all viewing angles.

The goals of direct imaging are lofty, and include studying the atmospheres of exoplanets and analyzing their chemical compositions, understanding the formation histories of planets and their relationships to protoplanetary disks, and perhaps one day, detecting the presence of chemical signatures in planetary spectra caused by life. It is the only method⁵ that can hope to answer these questions.

Direct imaging is also highly technically challenging, and despite the substantial investment in the field, the planetary yield has been decidedly meager. Less than 20 planets have been imaged, and some of those may be brown dwarfs. The first directly imaged exoplanet was found in 2004 by Chauvin et al. (2004), ten years after the first detection via the radial velocity method. This hot, super-Jupiter sized planet was found orbiting the brown dwarf 2M1207 at a distance of about the orbit of Pluto from the Sun. No other planets were found until 2008, when three were discovered orbiting the star HR 8799 (Marois et al., 2008), and two⁶ more were found around beta Pictoris (Lagrange et al., 2009) and Fomalhaut (Kalas et al., 2008). Since then, the number of planets

⁵transmission spectroscopy of transiting planets can extract some atmospheric information for a small subset of known exoplanets, but is by no means a general method.

⁶There was an additional companion thought to be a planet found around ROXs 42B (Currie et al., 2014), but it is almost certainly a brown dwarf.

discovered has trickled onwards, at the rate of two to three per year. Breathless predictions of a new planet per night for the second-generation instruments GPI (Macintosh et al., 2014) and SPHERE (Beuzit et al., 2008) have not materialized; between them a single planet has been published in over a year of operations (Macintosh et al., 2015). It is a hard business.

In the following sections, the expected characteristics of exoplanets will be derived, including their brightnesses, separations, and contrasts with respect to their host stars. These parameters will motivate a discussion of the instrumentation required for extrasolar planet imaging, which is complex and relatively unique. We will discuss the main error sources contributing to direct imaging, including astrophysical and instrumental errors. Finally, we will conclude with a view towards the future of the field.

1.4.2 Physical principles

In this section we will discuss the expected brightness, separation, and contrast of extrasolar planets, and motivate the instrumental requirements given in the next section. The focus will be on the general physics; detailed simulations and modeling results will be presented without in-depth derivations.

1.4.2.1 Detectability in reflected and emitted light

A simple calculation to define the parameters of the problem is the expected separation between the planet and star, in angular terms. For a planet located at a distance a from its host star, and the host star a distance d from the Earth, the expected angular separation is:

$$\theta = \frac{a}{d} \tag{1.32}$$

$$= 100\text{mas} \left(\frac{a}{1\text{AU}} \right) \left(\frac{d}{10\text{pc}} \right)^{-1} \tag{1.33}$$

where the implicit definition of the parsec is used to generate the second expression. For context, 100 mas is not a large angular separation; it is about the width of a human hair viewed from 100 meters away.

To understand how bright planets are, and in particular how bright they are compared to stars, we recognize that planets generate their own light from internal heat, but also reflect light from their parent stars. Both the reflected and emitted light may be approximated by blackbody spectra, at the temperature of the star and planet, respectively. The blackbody law is given by

$$B(\lambda, T) = \frac{2hc^2}{\lambda^5 \exp[hc/(\lambda k_B T)] - 1} \tag{1.34}$$

where the units of B are in specific intensity, that is, ergs/s/cm²/cm/sr. The law describes the amount of power emitted (ergs/s), per unit wavelength (cm), per unit surface area (cm²), per unit solid angle (sr) by a surface at temperature T at wavelength λ . In this formula, c is the speed of light, h is Planck's constant, and k_B is Boltmann's constant. At a distance d from a star of uniform brightness, we have

$$F = \int B \cos \theta d\Omega = \pi B(\lambda, T) \left(\frac{R_*}{d} \right)^2 \quad (1.35)$$

and hence the contrast between a planet and star is

$$\frac{F_p}{F_*} = \frac{R_p^2 B(\lambda, T_p)}{R_*^2 B(\lambda, T_*)} \quad (1.36)$$

Putting some numbers into this expression, we find that for a planet like Earth, the contrast at a wavelength of 10 microns is about 10^{-7} , about the same as Jupiter (the smaller radius is offset by the higher temperature of 275K vs 150K for Jupiter). At 1 micron, both are more than 20 orders of magnitude fainter. It is clear detecting planetary thermal emission at optical and near-infrared wavelengths is not feasible for solar system analogs.

The previous expressions made no explicit use of the planet-star separation distance. There is an implicit dependence, as the effective temperature of the planet will depend on the distance from the star, to some degree, though internal sources of heat and residual heat from formation are important caveats to this statement. For reflected light, the distance to the star plays an explicit role. As stated before, the stellar flux density at a distance d from the star is

$$F(\lambda, d) = \pi B(\lambda, T_*) \left(\frac{R_*}{d} \right)^2 \quad (1.37)$$

The factor of π comes from integrating the solid angle along the line of sight. The planet will reflect some of this light, and the amount reflected depends on the radius of the planet, R_p , its distance from the star a , and the reflectivity of the atmosphere or surface, called the albedo $A(\lambda)$:

$$F_p(\lambda, d) = \pi B(\lambda, T_*) \left(\frac{R_*}{a} \right)^2 A(\lambda) \Phi(\alpha) \left(\frac{R_p}{a} \right)^2 \quad (1.38)$$

There is also a phase function $\Phi(\alpha)$ which depends on the inclination of the orbit, and the portion of the orbit the planet is in. This is exactly analogous to the phases of the moon, and runs from 0 (such as when the planet transits a star) to 1. That shows that the planet to star flux ratio is

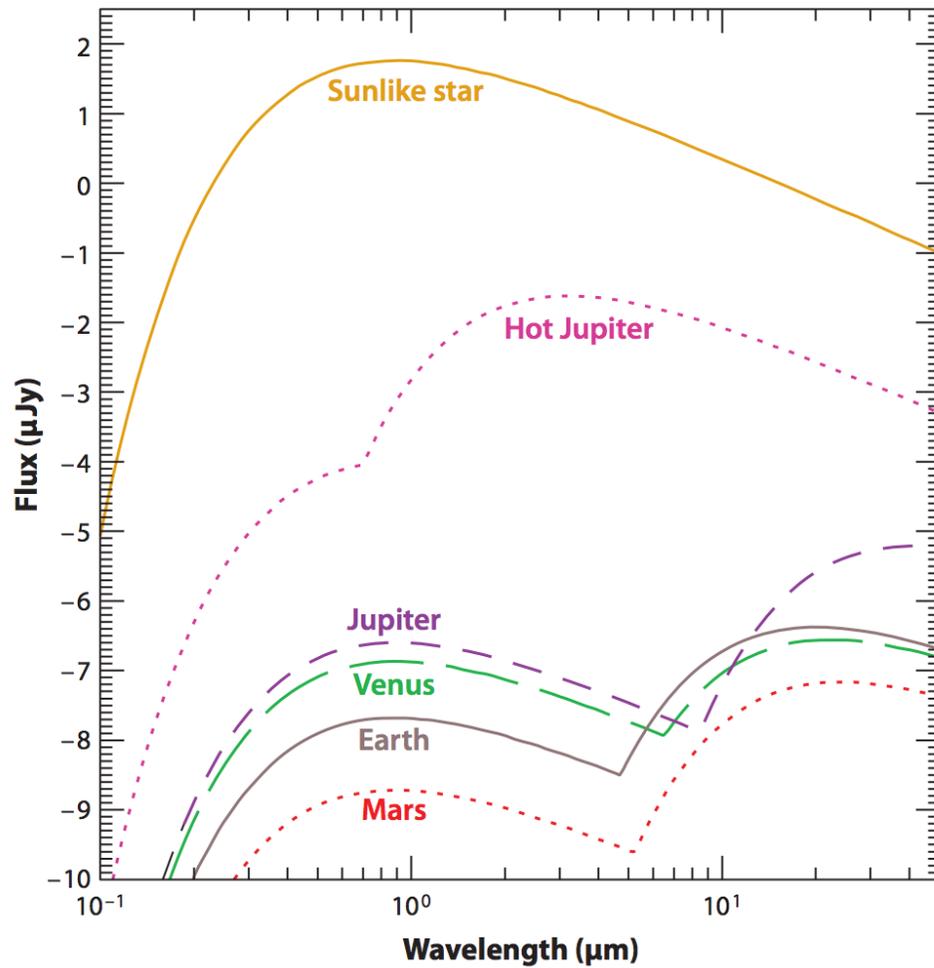


Figure 1.11: The flux of light from a sun-like star compared to various planets in the solar system, and a hypothetical hot Jupiter. All these are blackbody curves; real planets and stars have more complicated spectra, but the approximations are quite accurate for the continuum levels. Note that the shorter, visible wavelengths are all reflected light and track the spectrum of the star, while the longer infrared wavelengths are thermal emission and are somewhat larger. The contrast between the Earth and the Sun is ten orders of magnitude in visible light. This figure is taken from Seager & Deming (2010).

$$\frac{F_p}{F_*} = A(\lambda)\Phi(\alpha)\left(\frac{R_p}{a}\right)^2 \quad (1.39)$$

where we note that the wavelength dependence cancels out, except for the albedo. For general numbers of the Earth and Jupiter, and assuming an albedo of 0.3, we find reflected light contrasts of about 5×10^{-10} and 2×10^{-9} . The wavelength dependent reflected and emitted contrasts are presented in Figure 1.11.

The previous discussion defines two requirements for high contrast imaging: contrast and angular

resolution. We now give some context as to the challenges associated with these requirements. For angular resolution, the diffraction limit of a 2m class telescope operating at a visible wavelength of 500 nm is⁷ 50 mas. Of course, looking at nearby stars will help offset this, but there are only 250 star systems closer than 10 parsecs.⁸ Large telescopes are hence preferred to improve angular resolution, but on the ground, they need adaptive optics systems to sharpen the starlight and reduce the scattering effects of the turbulent atmosphere. In terms of contrast, even with a perfect telescope, diffraction will cause the image of the star to be surrounded by “Airy” rings; the first Airy ring is at about 1.5 resolution elements (λ/D) from the core, and is reduced in intensity by a factor of 100. This is still thousands to millions of times brighter than any putative planet at that location. This drives the need to develop diffraction control systems, such as coronagraphs or interferometers. Even with such systems, optical errors in the telescope and instruments can destroy most of their performance, driving the need to develop sophisticated focal plane wavefront control systems and post-processing algorithms to deal with this scattered light. Finally, with all this extra optical complexity, the throughput of the system has to remain high. A 2 m space telescope will receive about 100 million photons per second from a $V=7.5$ magnitude star; for a Jupiter-type planetary companion, this reduces to about 1 photon per second. If the throughput of the system is low, exposure times could stretch into days to accumulate enough signal, even in the best cases.

1.4.2.2 Detectability at different ages

Despite these challenges, there is one thing that can potentially improve the situation: age. Young planets are self-luminous, so instead of reflected light dominating the emission, internal heat will account for the vast majority of light from the planet. The thermal emission of young planets depends on on many factors, but primarily on the formation scenario.

For giant planets, there are two main theories about how they form, referred to as “core accretion” and “disk instability.” We provide a brief description of these scenarios and how they can lead to large boosts in the expected brightness of young planets. Both of models rely heavily on numerical simulations, and have many steps where different physics takes place, governing micron-sized objects up to objects the size of planets. For that reason, we will provide qualitative explanations and summaries of results rather than detailed derivations. For a review of the core accretion and disk instability models see Lissauer & Stevenson (2007) and Durisen et al. (2007).

In the core accretion scenario, dust and small grains stick together in the protoplanetary disk, eventually forming—through some poorly understood process—kilometer sized objects known as planetesimals. Collisions between planetesimals cause them to grow larger, eventually forming a planetary nucleus. Gas from the surrounding disk will also start to coalesce onto the nucleus once its

⁷ $\theta \approx \lambda/D$

⁸<http://www.recons.org/census.posted.htm>

escape velocity is larger than the average velocity of the gas in the disk. Once the planetesimal reaches 10 Earth masses, it is suspected that a runaway gas accretion occurs, with the planet gaining gas as fast as the disk can supply it (see Figure 1.12). This halts once the planet has cleared all the gas within its gravitational reach, forming a (possibly observable) gap in the disk. Core accretion is a “bottom-up” approach, with the entire process taking on the order of a few million years. This process is thought to have led to the formation of planets in the solar system, both the inner ones and the giants, with the inner ones remaining small primarily due to the influence of solar radiation on the formation process.

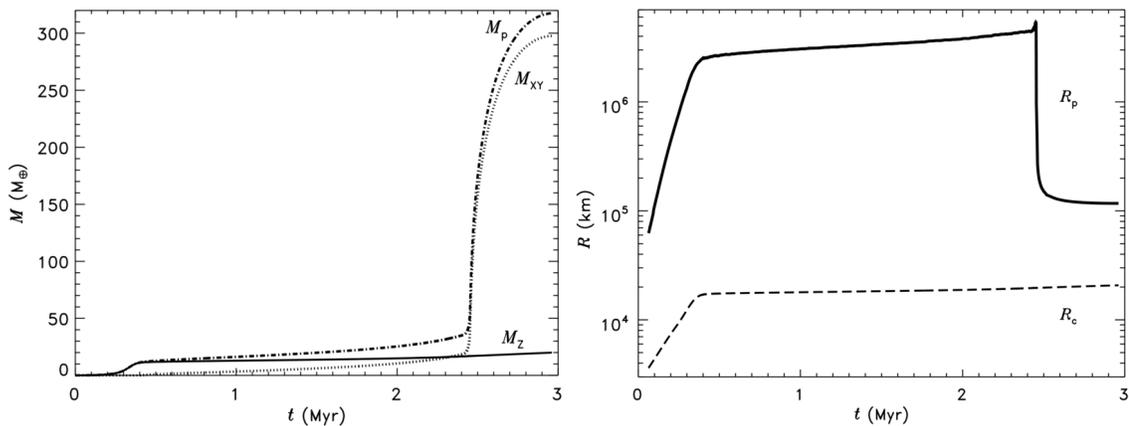


Figure 1.12: A simulation of the formation of Jupiter, showing the growth in mass and radius as a function of time. Mass grows slowly until the core reaches about 10 Earth masses, at which point runaway gas accretion occurs and the radius contracts rapidly. This figure is taken from Lissauer et al. (2009)

The main competitor to the core accretion theory is formation of giant planets via gravitational instabilities (Boss, 1997). At some phase of evolution, the protoplanetary disk may develop density perturbations causing the formation of spiral structures in the disk. As these spiral structures grow in amplitude and cause supersonic shocks, the disk energy will be converted into heat. If the heat dissipates slowly, the sound speed will rise in the disk and regulate the instability, eventually leading to quasi-stable cycles of heating and cooling. On the other hand, if the heat dissipates very rapidly, the instability will grow and cause fragmentation of the disk into self-gravitating clumps, which will quickly sweep up any gas, dust, and planetesimals nearby. Rather than a slow buildup via core accretion, this process is a “top-down” approach which only takes a few orbits to form a protoplanet, on the order of hundreds to thousands of years, not millions. A key issue is whether these density perturbations are stable or whether they lead to fragmentation in the disk, and this depends on how quickly the disk cools. Numerical hydrodynamics codes are employed to investigate these issues (see Figure 1.13), and consensus is that gravitational instabilities cannot form planets within a few tens of AU of the star, though formation may be possible at larger separations.

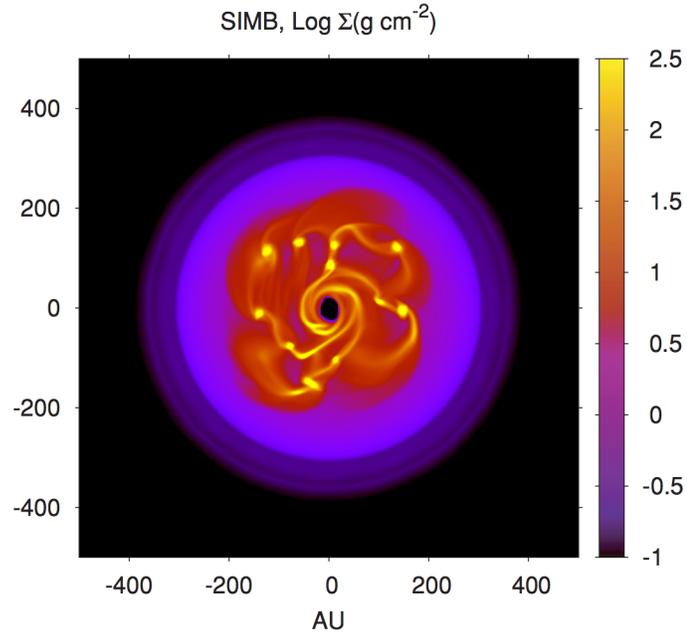


Figure 1.13: A numerical hydrodynamics simulation showing the fragmentation of a protoplanetary disk very far from the central star. The figure is taken from Boley (2009).

It may be that these two distinct formation modes are responsible for the formation of giant planets, with core accretion responsible for planets within 100 AU of the star and gravitational instability responsible for very distant giant planets (Boley, 2009). In this case, there should be two separate populations as well, and there is insufficient observational evidence to conclusively verify or refute this. It may also be the case that gravitational instabilities and core accretion work together to accelerate planet formation.

From the perspective of planet detection, there are observable consequences of planet formation, with very different outcomes depending on the formation mechanism (see Figure 1.14). In general, heat from planet formation can amplify the observable self-luminous signature of the planet (Equation 1.36). The disk instability formation mechanism creates a much hotter initial planet, so this is referred to as a “hot-start” model. The core accretion model leads to a colder initial temperature, so this is called a “cold-start” model. The temperatures of the planets equilibrate to the same level after a few hundred million years, causing no observable model-dependent temperature discrepancies.

In terms of flux difference, the hot start scenario is much more effective at making the planet detectable (see Figure 1.15), causing a maximum brightness boost of a factor of 100-1000 for young (1 - 5 Myr) planets, regardless of the mass. For the cold start scenario, the boost is more modest, about a factor of 10, and there is hardly any boost for a very massive $10 M_J$ planet.

What is really important is detectability, which does not simply depend on the flux of the planet.

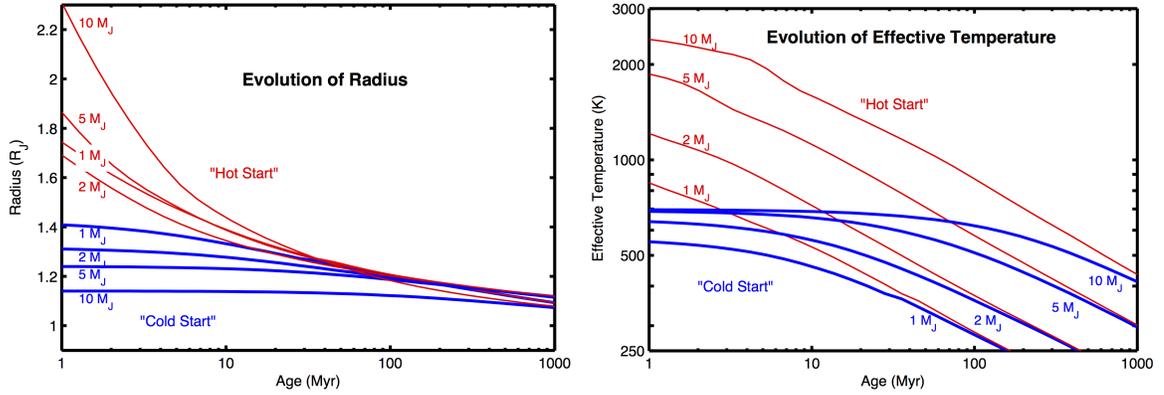


Figure 1.14: The evolution of effective temperature and radius of a planet as a function of age, plotted for different mass planets. Note the much higher initial temperature and radii due to the disk instability formation scenario (“hot start”) vs the core accretion (“cold start”) scenario. The effects of planet formation mechanism vanish after a few hundred million years. This figure is taken from Spiegel & Burrows (2012).

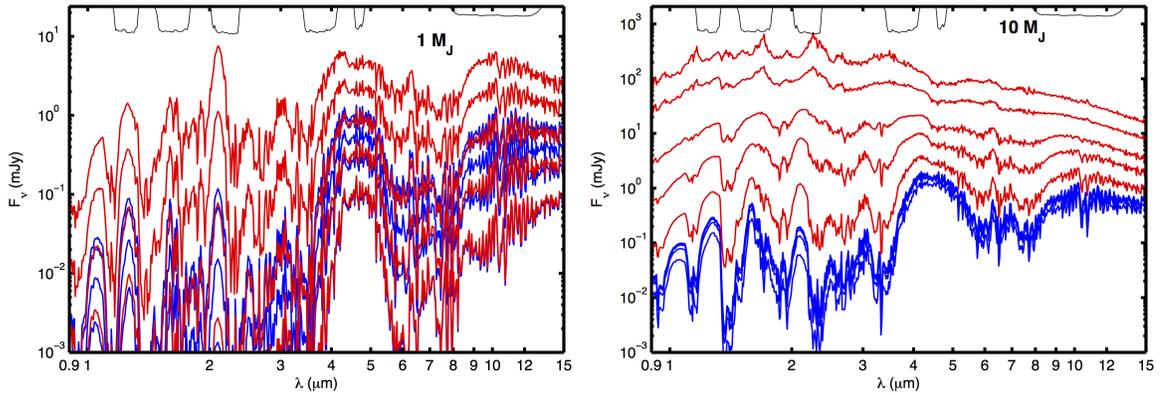


Figure 1.15: The evolution of flux in mJy as a function of wavelength, for different ages, for the disk instability (“hot start”, red curves) scenario, and the core accretion (“cold start”, blue curves) formation mechanism. The figure on the left is for a 1 Jupiter mass planet, the one on the right for a 10 Jupiter mass planet. Each curve is an isochrone corresponding to 1, 3, 10, 30, and 100 million year ages, with the younger ages having brighter fluxes. Note the different vertical scales on the left and right figure. The black lines at the top correspond to the J, H, K, L, M, and N infrared bands. This figure is taken from Spiegel & Burrows (2012).

The brightness of the star at different wavelengths, or more precisely the spectrum as a function of age, also affects F_p/F_* (Equation 1.36). For example, the Hubble extreme deep field had a (5σ) sensitivity of ~ 31 magnitudes⁹; an Earth-like planet around a Sun-like star 10 pc away would have a brightness of ~ 30 magnitudes, 2.5 times brighter. Jupiter would be hundreds of times above the noise limit in such an observation, but neither of these planets would be remotely detectable by Hubble. The reason is that the main error source in detecting planets, scattered starlight caused by diffraction and optical errors, scales with the stellar brightness (see Section 1.4.4), hence the

⁹<http://xdf.ucolick.org/>

focus on flux *ratio*. A plot of the wavelength-dependent star-to-planet contrast for various cold- and hot-start models is shown in Figure 1.16.

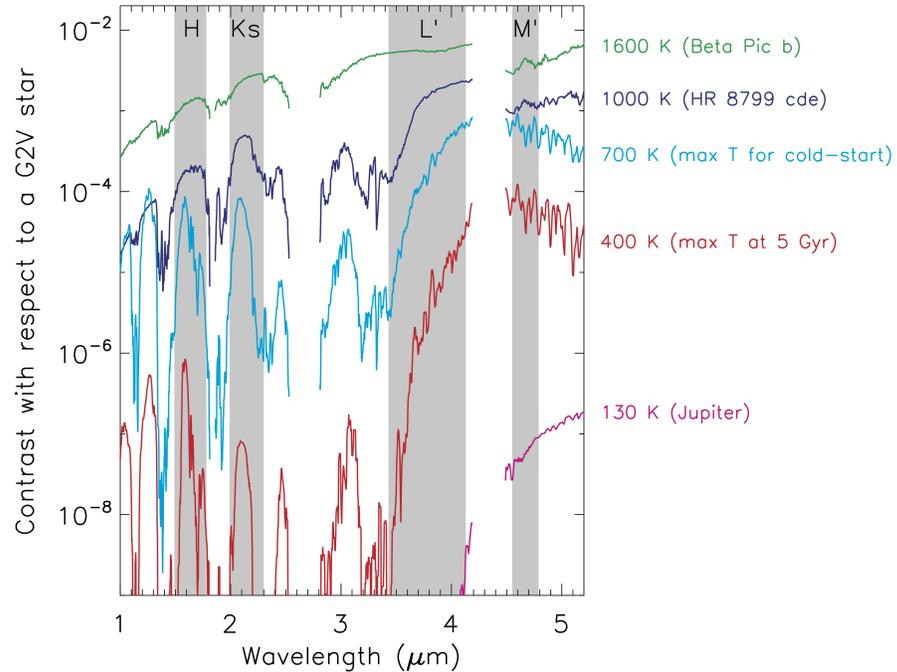


Figure 1.16: The contrast of various planets compared to a Sun-like star, as a function of wavelength. The red and light blue curves show models similar to those presented in Figure 1.15; three known exoplanets are shown at the top. Jupiter, in pink, is shown at the bottom. Note that the relative contrast seems to reach a peak at around 3.5 - 5 microns, corresponding to the infrared L' and M' bands. This figure is taken from Skemer et al. (2014)

1.4.2.3 Takeaways

The previous section has presented the expected brightness and separations of light from extrasolar planets, derived from physical principles. Typical stellar distances and planetary orbits cause star-to-planet separations of the order of 100 milli-arcseconds, comparable to the diffraction limit of most space and ground based telescopes. Planets both reflect their host star's light and generate their own light from internal heat. For reflected light, the relative flux compared to the star is on the order of 10^{-8} for a Jupiter-sized planet, and 10^{-10} for an Earth-sized planet. For thermal planetary emission, at far infrared wavelengths ($10 \mu\text{m}$), the contrast ratios are on the order of 10^{-8} , and at visible wavelengths they are greater than 10^{-20} .

However, these thermal emission contrasts may be greatly improved by observing young and newly-formed planets. Both core accretion and gravitational instability formation scenarios predict an excess of brightness from formation, though much more for gravitational instability. At young

ages, this internal heat has not yet radiated away, so the planet-to-star contrasts are much more agreeable, ranging up to 10^{-2} for some cases, especially when observed at mid infrared wavelengths. Imaging young planets around the nearest stars is the current strategy of most groups in the field.

So far, we have barely mentioned the required instrumentation, only focusing on the expected observables due to astrophysics. It turns out that attaining the high contrasts at small angles necessary to image planets is extremely challenging, and is limited both by physics and by engineering. This will be the subject of the next two sections.

1.4.3 Instrumentation

While the instrumentation required for radial velocity detection of exoplanets (telescopes and high resolution spectrographs), would have been recognizable to an astronomer working in the early 1900s, the instrumentation for direct imaging is decidedly modern. High performance adaptive optics systems are a recent invention, and most stellar coronagraph designs are so new that they are not mentioned in modern optical textbooks, even those focusing on astronomical instrumentation.

1.4.3.1 Coronagraphs

Diffraction of light through an optical system causes light to spread out over a large area around the center of the point-spread function, where the characteristic surface brightness of this light—typically in the shape of concentric rings—is many orders of magnitude larger than any planets which might be found there. The purpose of a coronagraph is to reduce the amount of diffracted starlight, while maximizing the amount of planet light let through.¹⁰

The basis of many modern designs is the “Lyot coronagraph,” first introduced in 1931 to image the solar corona (see Figure 1.17). The starlight is brought to a focus on a mask, which blocks the core of the light. The effect of the mask is also to diffract the Airy rings to the outer part of the pupil, where a “Lyot stop” will block them and prevent them from propagating to the final image plane. Any companion light will come in at an angle and focus off the mask, and so will only be slightly affected.

A conventional imaging camera may have two planes: the input pupil and the detector plane. Coronagraphs typically (but not always) have two intermediate planes between the input pupil and the detector plane, which are another image plane and a pupil plane. At each of these planes, the electric field of the light may be manipulated via optical elements such as masks and stops. Sometimes these optical elements modify the amplitude of the electric field, sometimes the phase,

¹⁰A common misconception is that coronagraphs are used to suppress starlight at the center of the point spread function. Suppressing the core of the PSF would have no effect whatsoever on improving planet detection capabilities, as no planets are found there. An equivalent operation would be to place a piece of dark tape on the pixels of the CCD corresponding to the PSF center.

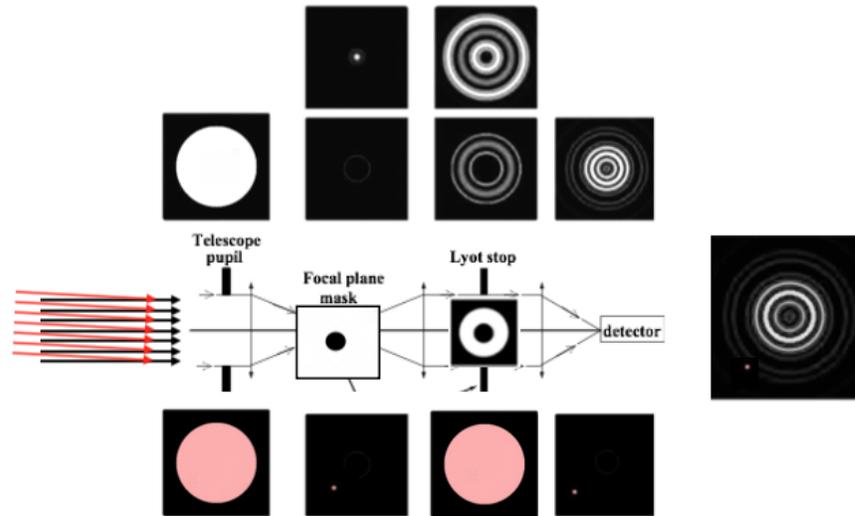


Figure 1.17: Basic operation of a Lyot coronagraph. The star is brought to focus and blocked by a focal plane mask. A second stop in a pupil plane (the Lyot stop, slightly smaller than the pupil) blocks starlight that is diffracted around the focal plane mask, and a camera brings the remaining light to a focus. For a planet near the star (red light, bottom panel), the planet is focused near, but not on, the first disk, since it is separated on the sky, and is not affected by the Lyot stop. In this design, a simple Lyot coronagraph can block about 99% of the stars light while allowing the planets light to pass through unaffected. (This image was generated from a combination of images from the www.Lyot.org website)

and sometimes both. A general optimization problem for coronagraphs is to minimize the amount of on-axis light making it to the focal plane (the star), while maximizing the amount of off-axis light.

Some knowledge of physical optics, specifically diffraction, is required to understand coronagraphs. For a first-principles approach, one may start with Kirchoff's diffraction formula, which leads to the well-known Huygens-Fresnel principle that each point on an optical wave can be seen as the source of secondary spherical wavelets. The spherical wavelet approach combined with some simplifications then can be used to derive the Fraunhofer approximation, which states that the electric field in the far-field of an aperture is simply a Fourier transform of the electric field in the aperture, up to a multiplicative phase term. It can also be shown that the far-field approximation, which is usually prohibitively large (>1 km for visible light and conventionally sized optics), also applies in the case of a thin lens, where the Fourier transform relationship will hold between the front of the lens and the focus of the lens. We will skip this derivation, and direct the reader to Goodman (2005), for the details.

Instead, we will start from the main result, which is that there is a Fourier transform relationship between image and pupil planes in an imaging system, where the Fourier transforms operate on the *electric field* of the light, not the intensity. The final intensity (what is detected by the camera) is related to the magnitude-square of the electric field. More precisely, the probability of detecting a photon is proportional to the square magnitude of the electric field at that location in the image

plane.

We now proceed through a simple, one-dimensional coronagraphic system, first presented in Sivaramakrishnan et al. (2001), shown in Figure 1.18. We denote pupil plane coordinates by x and image plane coordinates by θ . The input electric field in plane “a” is assumed to be a flat wavefront, a wave with no phase or amplitude errors, which is defined in extent by the pupil of the telescope, in this case a clear rectangular aperture¹¹ of diameter “D,” given by $\Pi(x/D)$. This is what would be expected in space, or from a perfect adaptive optics system. Planes “b,” “c,” and “d” correspond to immediately before, at, and immediately after the coronagraph plane. At plane “b,” the electric field intensity is the Fourier transform of the pupil (a rectangle function), which is a sinc ($\sin x/x$) function. Plane “c” is the coronagraphic image plane, and the coronagraph in this case is a apodized spot with shape $w(D\theta/s)$, of characteristic width s resolution elements. Note that the stop transmission is 100% far from the coronagraphic spot. Immediately after the apodized spot, the electric field is simply the sinc function multiplied by the coronagraphic plane transmission function $1 - w(D\theta/s)$. Planes “e,” “f,” and “g” correspond to immediately before, at, and immediately after the “Lyot stop” plane. The electric field at “e” is the Fourier transform¹² of the electric field at “d,” which can be written as

$$E(\text{plane e}) = \mathcal{F} \left[\frac{\sin(D\theta)}{D\theta} (1 - w(D\theta/s)) \right] \quad (1.40)$$

$$= \mathcal{F} \left[\frac{\sin(D\theta)}{D\theta} \right] * \mathcal{F} [1 - w(D\theta/s)] \quad (1.41)$$

$$= \Pi(x/D) * \left(\delta(x) - \frac{s}{D} W(sx/D) \right) \quad (1.42)$$

where the fact that the Fourier transform of a product is the convolution of the two individual Fourier transforms is used. (Note that the y-scale in the plot is significantly smaller in plane “d.”) This electric field intensity has large peaks at the edges of the pupil. An undersized Lyot stop of shape $\Pi(x/D_L)$ at plane “f” will block most of this light, transforming the electric field into

$$E(\text{plane g}) = \Pi(x/D_L) \left(\Pi(x/D) * \left(\delta(x) - \frac{s}{D} W(sx/D) \right) \right) \quad (1.43)$$

Finally, the electric field in the detector plane “h” is the Fourier transform of the electric field in plane “g,” given by

¹¹the rectangle function is defined as $\Pi(x) = 1$ if $x < 1/2$, 0 elsewhere.

¹²actually, it should be the inverse Fourier transform, but since we are dealing with symmetric functions, it makes no difference.

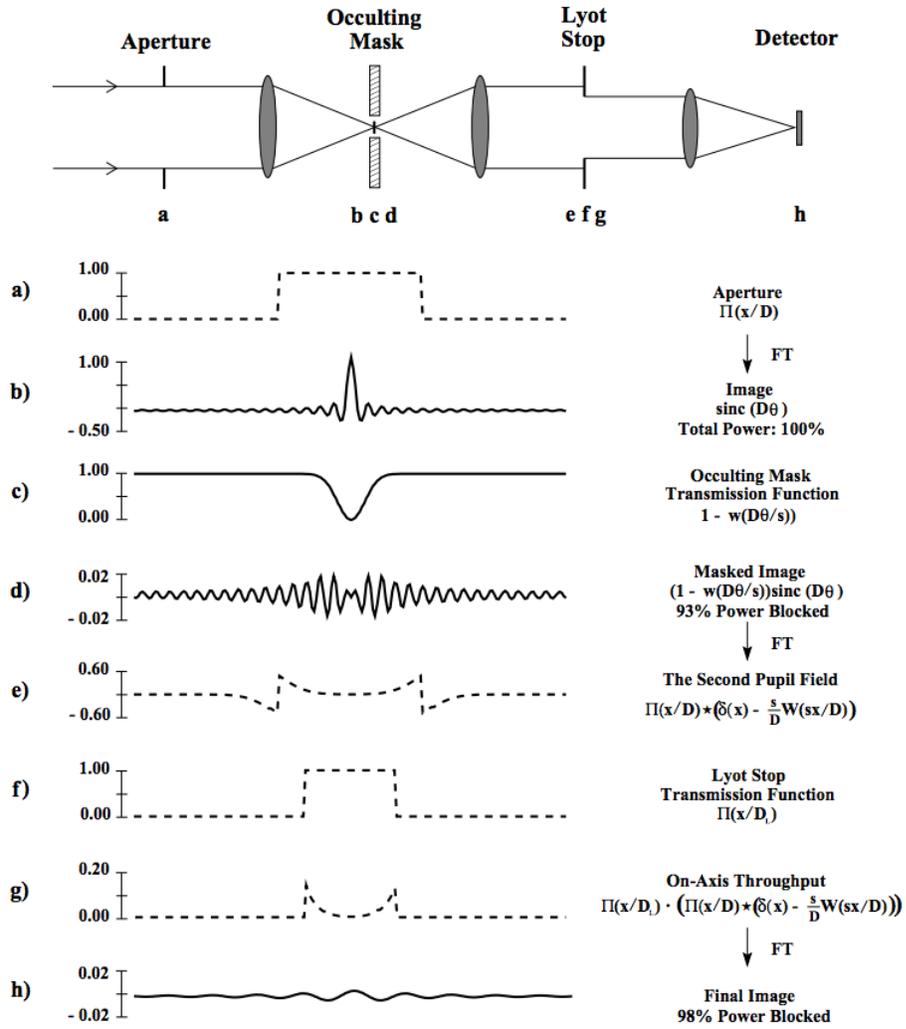


Figure 1.18: A one dimensional coronagraph. See the text for details. This figure is taken from Sivaramakrishnan et al. (2001)

$$E(\text{plane h}) = \mathcal{F} \left[\Pi(x/D_L) \left(\Pi(x/D) * \left(\delta(x) - \frac{s}{D} W(sx/D) \right) \right) \right] \quad (1.44)$$

$$= \mathcal{F} [\Pi(x/D_L)] * \mathcal{F} \left[\left(\Pi(x/D) * \left(\delta(x) - \frac{s}{D} W(sx/D) \right) \right) \right] \quad (1.45)$$

$$= \mathcal{F} [\Pi(x/D_L)] * \left(\mathcal{F} [\Pi(x/D)] \mathcal{F} \left[\delta(x) - \frac{s}{D} W(sx/D) \right] \right) \quad (1.46)$$

$$= \sin \left(\frac{D_L \theta}{D_L \theta} \right) * \left(\sin \left(\frac{D \theta}{D \theta} \right) (1 - w(sx/D)) \right) \quad (1.47)$$

which, as far as we can tell, has no simplified form. There are a few takeaways to be had from this example. First, notice that all the stops and masks in this simple coronagraph design are “amplitude” style, that is, they consist of real-valued functions that describe a transmissive intensity, which may be zero. Second, the central spot at plane “c” can attenuate a significant fraction of the image plane, depending on the value of “s,” which can hurt the ability to find planets near the star. Thirdly, the final angular resolution and throughput of the image is degraded somewhat by the use of the Lyot stop at plane “f,” which must be undersized to block the most extreme amplitudes of the electric field at plane “e.” Finally, there is still residual light left in the focal plane, though it has been attenuated by a significant amount. Balancing the width of the stop with throughput, resolution, and residual light is part of the optimization of a coronagraph design.

There are many other coronagraph designs. One important variant of the focal plane mask presented above is phase-type designs. Here, instead of using a partially opaque mask, a mask that imprints a *phase shift* on the light is used, such as by using transmissive glasses of different thicknesses. For example, a mask at plane “c” with

$$M(\theta) = \begin{cases} e^{+i\pi/2} & \theta > 0 \\ e^{-i\pi/2} & \theta < 0 \end{cases}$$

will create a pupil at plane “e” with electric field proportional to

$$E(\text{plane e}) \propto \ln \left(\frac{2x/D - 1}{2x/D + 1} \right)^2 \quad (1.48)$$

which is sharply peaked at $x = D$, showing much of the light is ejected to the outer edge of the pupil, so it may be more efficiently blocked. There is still light remaining inside the pupil, which means this is not a great coronagraph, but it shows how phase-masks can efficiently redistribute light around the pupil. One might ask whether there is some phase mask design that redistributes *all* the on-axis starlight to outside the pupil, for a circular input pupil. In fact, such a design does

exist. This is given by the two-dimensional mask function $M(r, \psi) = \exp[2n\pi i\psi]$, where ψ is the azimuthal coordinate, for values of $n = 2, 4, 6$, etc; and is called a vortex coronagraph of topological charge n (Mawet et al., 2005). This is the basis of the main kind of coronagraph used in this thesis, described at length in Part II.

When evaluating a coronagraph design, there are a few useful performance metrics to consider, including *contrast*, *inner working angle*, *throughput*, and *bandwidth*. For an unresolved, single star, contrast is the average brightness at a particular area of the image divided by the brightness of the star. Inner working angle is commonly defined as the angular separation where the flux of a putative companion is attenuated by the coronagraph to 50% of what it would be arbitrarily far away. Inner working angle is very important because planets are at usually at very small angular separations from their host stars. Throughput refers to the fraction of planet light making it through the coronagraphic system as a function of angular separation. Throughput is also very important, as accumulating enough companion photons even at far separations from a star can be a challenge with faint planetary sources. Large bandwidth is also needed for the same reason, as well as being necessary for detailed spectral characterization. Another metric is sensitivity to low-order aberrations, that is, how well the coronagraph works when there are slight imperfections like defocus or star misalignment on the input beam. Note that when comparing different coronagraph designs, many of these metrics are stated in units of diffraction beamwidths (λ/D) rather than absolute terms, since the merits of a design do not depend on the particular dish diameter.

Name	Abbreviation	Description
Regular Lyot	N/A	Binary amplitude focal plane mask using dark metallic spot, downstream Lyot stop
Fourth-order Hybrid Lyot	HBL4	Complex phase/amplitude focal plane mask with metallic and dielectric films, downstream Lyot stop
Four-quadrant mask	4Q	Focal plane mask with π phase shifts between quadrants, downstream Lyot stop
Eight-octant mask	4Q	Focal plane mask with $\pi/2$ phase shifts between octants, downstream Lyot stop
Charge-four Vector Vortex	VVC4	Focal plane mask with spiral phase ramp using sub-wavelength gratings, from 0 to 8π , downstream Lyot stop
Charge-two Vector Vortex	VVC2	Focal plane mask with spiral phase ramp using sub-wavelength gratings from 0 to 4π , downstream Lyot stop
Phase-induced amplitude apodization	PIAA	Aspheric pupil lenses/mirror losslessly remap beam into apodized shape
Binary shaped pupil	SP	Binary pupil mask apodizes pupil to create asymmetric PSF with dark/light regions

Table 1.1: Summary of coronagraphs presented in Figure 1.19.

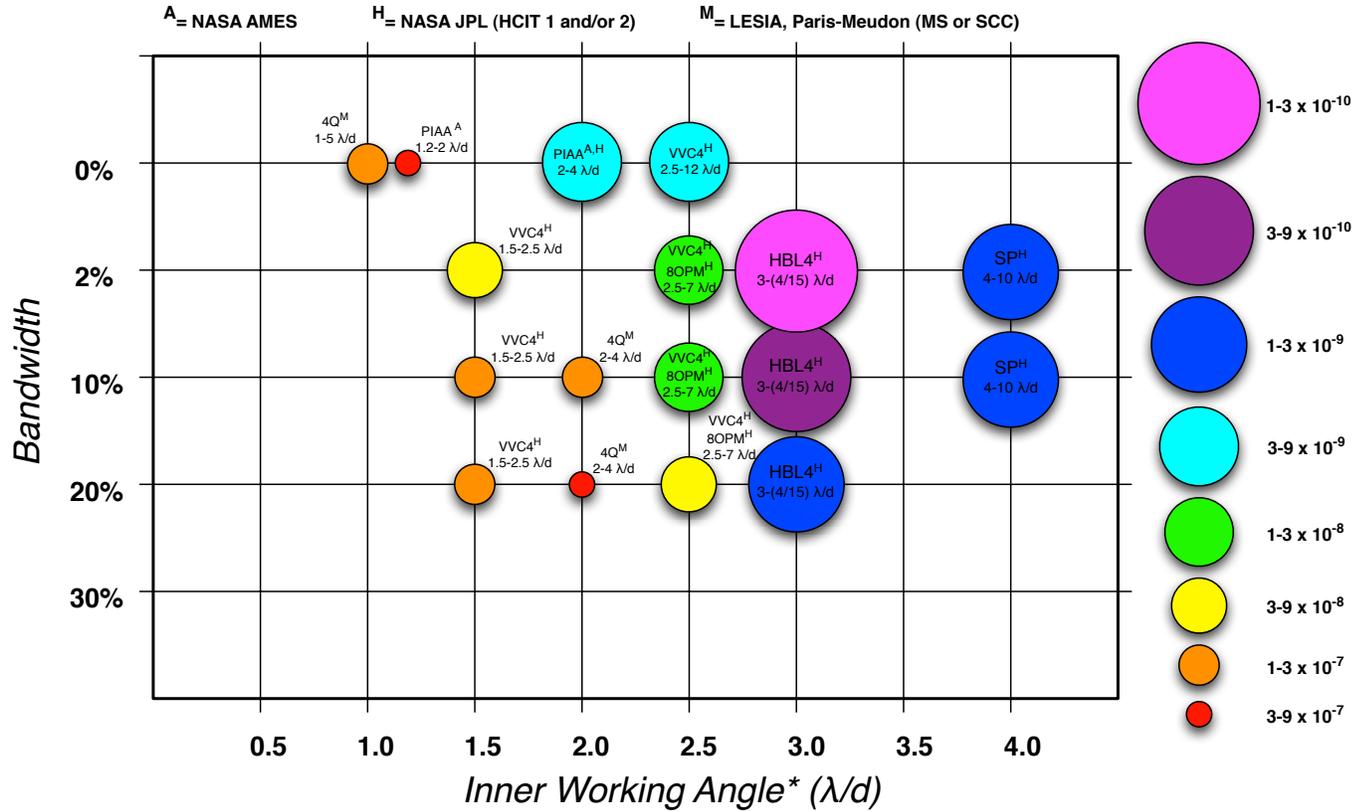


Figure 1.19: A plot of the top contrasts achieved by various modern coronagraph designs, plotted in inner working angle—bandwidth space. Note some coronagraphs show up more than once if measurements were performed at more than one bandwidth. Contrast levels correspond to the size of the circles. For a description of the coronagraphs, see Table 1.4.3.1. These are empirical results, and are slightly out of date as of this writing. This figure is taken from Mawet et al. (2012)

1.4.4 Major error sources

Like the radial velocity method, there are both astrophysical and instrumental error sources that degrade the precision of high contrast imaging systems. Unlike radial velocity studies, where instrumental control has well advanced to the point where the most interesting targets seem to be limited by astrophysical noise sources, in direct imaging, the instrumental errors completely dominate the error budget and have limited the field to the rather sad planet yield it has so far achieved.

In the following sections, we present a brief introduction to the primary astrophysical and instrumental error sources, exozodiacal light and speckle noise. The mathematics of speckles and speckle suppression is treated much more thoroughly in the penultimate chapter of the thesis.

1.4.4.1 Exozodiacal light

Dust in our own solar system scatters sunlight, and this may be seen in dark sky locations around the Earth, particularly after sundown or right before sunrise. This is called the “zodiacal light” or

“false dawn,” and is primarily composed of small ($<10\text{-}300\ \mu\text{m}$) particles from cometary evaporation and asteroid collisions, lying along the ecliptic plane.¹³ Zodiacal light exists in other star systems as well (Ertel et al., 2014), and the diffuse glow will be a major noise source, particularly as young protoplanetary disks are expected to have much higher levels of “zodi” than the around the Sun. Indeed, infrared interferometers have detected exozodiacal light hundreds to thousands of times brighter than the relative level in the solar system (Millan-Gabet et al., 2011). In direct images, this light will show up as a diffuse source in the image, which will dominate infrared and visible emission surrounding the star (see Figure 1.20) compared to the light expected from planets at those same locations.

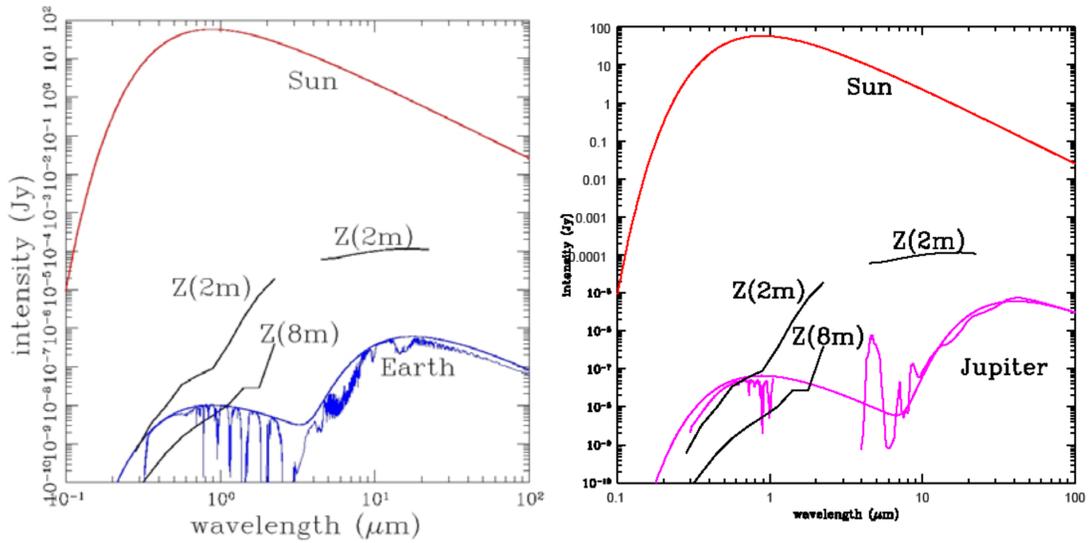


Figure 1.20: The planetary and zodiacal light levels for a Solar system twin at 10 pc, as viewed with a 2m and 8m class telescope. An Earth twin is shown on the left, with a Jupiter twin on the right. The zodiacal light can dominate the light at the infrared, but visible is more amenable, particularly with a larger telescope. Compare the spectral features to Figure 1.11. This figure is taken from Kasting et al. (2009), in the Decadal review of Astronomy.

The exozodi is a *surface brightness*, however, so the relative intensity compared to a planetary point source decreases quadratically with telescope diameter. Also, as diffuse emission, it can in principle be removed in post-processing if enough signal-to-noise is accumulated. Finally, in some configurations, the effects of a planet may be deduced from resonant structures in the exozodiacal light caused by the planet’s orbit (Kuchner & Holman, 2003). A better understanding of these issues will be needed in the coming decades as space-based missions become sensitive to planets at the exozodi locations.

¹³In the Northern Hemisphere, it is possible to see the zodiacal light, the Milky Way, and the North star at the same time. This allows one to get their bearings with respect to the plane of the solar system, the plane of the galaxy, and the Earth’s rotational axis in one shot.

1.4.4.2 Speckles

The previous discussion on coronagraph optics assumed the input beam to the telescope pupil was a perfectly flat wavefront. This is certainly not the case on the ground, with extreme distortion of the wavefront caused by its passage through the atmosphere, requiring an adaptive optics system. Even in space, telescope and instrument optics will distort the wavefront even in the best case, even in space. For a thorough discussion of speckles and speckle suppression, see the penultimate chapter of this thesis.

The electric field of a pure plane wave at a pupil of the imaging system can be written as

$$E(r, t) = A_0 e^{i(k \cdot r - 2\pi f t + \phi_0)} \quad (1.49)$$

Now, in the perfect case, A_0 is a constant as a function of pupil position x , as is ϕ . In the case of static phase and amplitude type errors at the pupil, we have

$$A(x) = A(x) e^{i\phi_p(x)} \quad (1.50)$$

here, we let $A(x) = A_0(1 - \phi_a(x)) \approx e^{\phi_a(x)}$ be a constant electric field amplitude plus some small perturbation, and let $\phi_p(x)$ be any phase aberrations in the beam. We then absorb the amplitude aberrations and phase aberrations into a single complex phase error $\phi(x)$ so that

$$A(x) = A_0 e^{i\phi(x)} \quad (1.51)$$

$$\phi(x) = \phi_p(x) + i\phi_a(x) \quad (1.52)$$

Now, consider a simple sinusoidal phase variation of spatial frequency k_x in pupil space. Since the Fourier transform of a sine function is two delta functions, two copies of the point-spread function will appear in the focal plane at locations $\pm\lambda/k_x$, with intensity A_0^2 (see Figure 1.21). A sinusoidal amplitude error, perhaps caused by a polishing nonuniformity, will cause a similar intensity distribution in the focal plane, with some subtle differences discussed in the speckle nulling chapter. Also, phase errors can convert to amplitude errors, and vice versa, through the Talbot effect.

More realistic figure and reflectivity errors may be decomposed into sines and cosines via standard Fourier analysis, and typical optics have error power spectra that scale as the inverse cube of the spatial frequency, that is, there are a lot more low-frequency errors than high frequency errors, so speckles will be brighter near the star. Separately, errors in the optics may change due to a varying thermal or gravity environment, such as after a telescope slew. Underlying diffraction can amplify speckles as well.

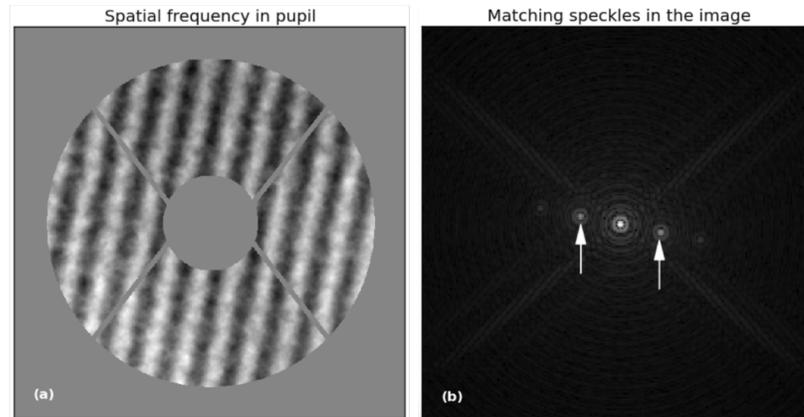


Figure 1.21: A sinusoidal phase ripple in the pupil plane will form two speckles in the image plane. This is real image data, not simulated, and was taken from (Martinache et al., 2014)

The overall effect is to create speckles in the image plane that may be many orders of magnitude brighter than any planetary companions. These are, by far, the most destructive noise source in high contrast imaging of exoplanets (see Figure 1.22). Speckles may be somewhat attenuated in post-processing through various techniques, and can also be partially removed optically through manipulations of the deformable mirror, if their electric field phases are discovered. The latter method is the subject of the last two chapters of this thesis.

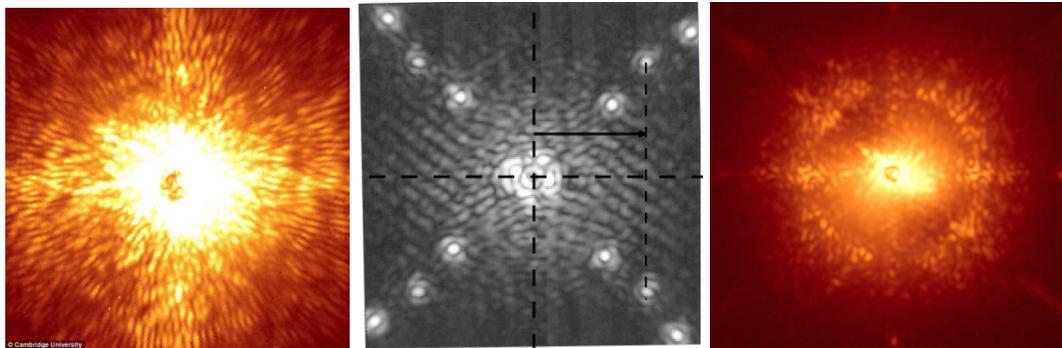


Figure 1.22: Raw images from Project 1640, the Gemini Planet Imager, and SPHERE, all showing large amounts of speckle noise. Images from the respective project websites.

1.4.5 Observing strategies and data analysis

High contrast imaging observing strategies and data analysis are structured around the need to remove speckle noise from the images. Many advances and discoveries in the field have occurred due to superior observing styles and reduction algorithms, without recourse to new instrumentation, and it is possible that more breakthroughs will allow archival datasets to be reanalyzed to reveal new planets and disks.

1.4.5.1 Hardware discrimination techniques

Speckles look very similar to planets, but there are some differences in the light that goes into a speckle and the light that goes into a planet. Speckles are generated by optical errors and composed of starlight, and planets are composed of planet light.

One possible exploit is that speckles look different at different wavelengths. More precisely, the position of a speckle is determined both by the spatial frequency of the error on the optical surface, and the wavelength of light. That means that when viewed at different wavelength, speckles will appear at different locations. Planets will stay fixed, since they are determined by angular position on the sky, not wavelength. Said another way, the angular separation of two objects does not depend on the wavelength used to view them. Using this wavelength-dependent information to separate speckle and planet light is called “simultaneous spectral differential imaging,” or SSDI (Marois et al., 2000; Racine et al., 1999). Of course, to use SSDI, observations at multiple wavelengths are required. It cannot be done at a single wavelength.

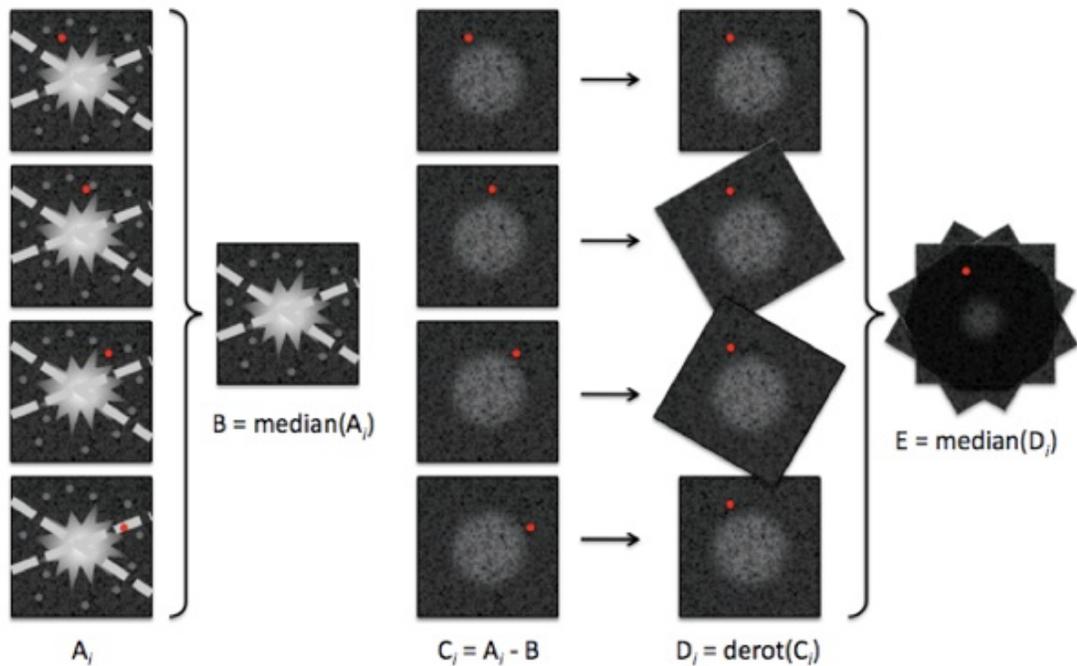


Figure 1.23: A cartoon of angular differential imaging. The planet (red) appears at different positions in the focal plane at different observing times due to sky rotation. A median frame contains the speckle light, but not the planet. This median is subtracted from all the individual data frames, which are then derotated and combined to reveal the planet. This image was taken from www.astrobites.org.

Another highly useful method is called “angular differential imaging” (Marois et al., 2006a) which exploits the fact that while speckles are fixed with respect to the telescope optical system, the sky image can rotate. In a space telescope, this can be accomplished by rolling the telescope; on the

ground, this is achievable in certain telescope configurations, notably the alt-azimuth mounts. In an alt-az system, the field of view rotates with respect to the sky, and usually this is compensated either by rotating the instrument being used or using a K-mirror to rotate the field of view before it enters the instrument. However, if this rotation is not done, then the sky image will rotate with respect to the fixed speckle pattern. This rotation information can be used to discriminate speckles from planets in post-processing.

Both ADI and SSDI are important techniques, but fail to work very well at locations close to the star. In ADI, the amount of rotation (in arc length) is proportional to the separation ($s = r\theta$), so for small separations little rotation is achieved in pixel space. For spectral differential imaging, the amount of motion is proportional to the spatial frequency of the error ($x = \lambda/k_x$, with x being the position on the detector and k_x the spatial frequency of the error in the pupil). Since most of the errors in optics are at low spatial frequencies, to get most of the speckle positions (x) to change significantly requires a prohibitively large bandwidth.

Another method is based on the fact that speckles are coherent with starlight—that is, speckle wavefronts have a fixed phase relationship with the stellar psf—and thus can be made to interfere with them using some kind of interferometric device. This is called “synchronous interferometric speckle subtraction” (Guyon, 2004). In this case, speckles will modulate with the interferometer position, while planet light will stay fixed. This method has never been implemented due to the complexity. We will have much more to say about interferometric speckle suppression in the final chapter.

1.4.5.2 Data reduction techniques

Spectral differential imaging and angular differential imaging are observing strategies, using either the telescope or an instrument with multiwavelength capabilities to disentangle the image of the planet from bright speckles. Another important strategy is to remove speckles in the image using a different reference star (or stars) and image processing techniques. These are techniques which may be combined with spectral and angular differential imaging when the telescope or instrument allows it, or used independently, as they are of general utility.

The crudest form of this process is classic point-spread function (PSF) subtraction, where an image of a nearby star is multiplied by an appropriate scale factor and subtracted from the target image. If the reference star is chosen nearby the target star, so that the gravity vector and hence optical state of the system is nearly identical, then the speckles between the two should correspond very well. In the subtraction step, the speckles will be preferentially removed, and the image will have an excess of flux at the position of a companion.

An improvement to classic PSF subtraction is called the “locally optimized combination of images” (LOCI) algorithm (Lafrenière et al., 2007). In this algorithm, rather than using means or

medians to combine the data frames and subtracting, a linear combination is used. Referring to the target images as T_i and the reference images as R_i , LOCI minimizes the quantity

$$\sigma_i^2 = \sum_p m_p (T_i - O_r)^2 \quad (1.53)$$

$$\text{where } O_r = \sum_i c_i R_i \quad (1.54)$$

where i refers to the image number, p refers to the pixel number, m_p is a binary mask to block out parts of the image not under consideration. In the second line, c_i are linear coefficients used to construct the optimal reference image for a particular mask m_i using the reference images R_i . In each target image T_i , at locations of the planet, the coefficients will not be able to well-fit the planet, since it does not exist in the reference images, so an excess of flux will remain. At the last step (not shown), each optimally-subtracted target image is combined with the others for the final result.

Efficiently calculating the coefficients c_i such that σ_i^2 is minimized relies on various matrix inversion schemes, which are beyond the scope of discussion. The main ideas behind LOCI’s success are the use of local segmentation of the image to more precisely capture noise variations, and the building up of linear combinations of basis vectors—in this case, the reference images—to create approximations of the speckles. LOCI and its derivatives, which use slightly different optimization parameters, are some of the most powerful methods of PSF subtraction currently used for high-contrast imaging datasets.

Another approach is using the Karhunen-Loeve eigenimage projection (KLIP) algorithm (Soumer et al., 2012). This is more conventionally called principal components analysis, or PCA. Here, the reference images are used to generate “eigenimages,” an orthonormal basis set which the first component captures the most variance in the reference images, the second captures the most variance after the first component has been removed, etc. These components in some sense define an “optimal” low-dimensional approximation of the image, where the first N principal components will capture the most variance in the reference images possible when using N or less orthogonal basis vectors. As such, rather than using the full reference set as in LOCI, the first few principal components form a separate optimal reference set, and each target image can be denoised with

$$T'_i = T_i - \sum_{j=1}^N \langle T_i \cdot Z_j \rangle \quad (1.55)$$

$$\text{where } \{Z\} = PCA[\{R\}] \quad (1.56)$$

where T'_i is the denoised image, the Z_j 's are the principal components generated from the reference set R , where the first N are used. N is a free parameter in the reduction, which may be experimented with. As usual, PCA may be performed on the full image or on a subset of an image, where the latter is preferable as PCA works best on images with uniform noise properties, and speckle noise varies as a function of distance from the star. There are many algorithms to calculate the principal components, which we do not delve into. We note that an advantage of the KLIP approach is that it is quite fast compared to LOCI, and can be used to efficiently forward-model any planets in the data or combined with Markov-chain Monte Carlo simulations.

Both of these algorithms have drawbacks. They will reduce the flux of any planetary signal, as the planetary signal will have a nonzero “overlap” with the basis vectors used in LOCI or PCA. This requires careful characterization for understanding the photometry of any detected companions. They also do not use information about the shape of a planet (that is, the size of a PSF), or any spatial structure in the image. For example, PCA will unwrap the image into a long one-dimensional vector, destroying information about proximity of nearby pixels, which will have similar noise properties.

There is much unexplored territory on image post-processing of exoplanetary data sets—the ultimate solution will likely involve incorporating all known physics of speckles and instrument telemetry in a full forward-model of the target images.

1.5 Overview of thesis and original contributions

1.5.1 Thesis overview

This thesis may be separated into two distinct projects, both dealing with the detection of planets orbiting other stars, but pursued through very different means.

The first project deals with the construction of a robotic telescope array for detecting extrasolar planets by the reflex motion of the star, using high-resolution spectroscopy to detect very small shifts in the absorption lines due to the presence of the orbiting companion. This is broken into two chapters; the first dealing with numerical simulations investigating the optimal design of such a system in order to maximize the survey yield and proving the feasibility of the project; the second dealing with the actual design, construction, testing, and verification of the observatory, called MINERVA. At the time of the work being done at Caltech, MINERVA's team leadership consisted of principal investigator John Johnson (Harvard, then at Caltech); co-investigators Jason Wright (Pennsylvania State University), Nate McCrady (University of Montana), and Robert Wittenmeyer (University of New South Wales). The Caltech contingent included postdocs Jonathan Swift and Philip Muirhead, software engineers Reed Riddle and Brian Lin, and Kristina Hogstrom and myself as graduate students.

The second project is also an instrument for exoplanet detection, this time for direct imaging of exoplanets. Direct imaging of exoplanets is challenging, requiring at minimum an adaptive optics system to sharpen the point spread function of the star after passing through the turbulent atmosphere; a coronagraph to suppress diffraction, allowing observation close to the star; and image post-processing to remove light scattered into the image plane from non-common path aberrations in the optical system. The latter half of this thesis deals primarily with the development of the Stellar Double Coronagraph (SDC), a multistage starlight suppression instrument at the Palomar telescope. This project is subdivided into four chapters. The first describes the SDC, including its philosophy, purpose, design, and performance. The second chapter describes an interesting companion discovery made using SDC. The third chapter describes a wavefront control code to remove optical errors that was used with SDC as well as with other high-contrast imaging systems at the Palomar and Keck telescopes. The last chapter details an optical subsystem of the instrument used to demonstrate a new way of detecting planets based on the coherence properties of the optical field, while simultaneously correcting for optical errors that can cause any image of a planet to be destroyed. The SDC collaboration is from NASA's Jet Propulsion Lab, led by Eugene Serabyn.

1.5.2 Previously published work and individual contributions

This thesis would not have been possible without the direct and indirect contributions of numerous people. This is true of all scientific research, but instrumentation development perhaps more than anything else. Most of this thesis is comprised of previously published work which benefited from strong contributions of different co-authors. As such, I would like to clarify individual contributions to the different chapters of this thesis. I apologize in advance for inadvertent omissions which may have occurred.

The two chapters of Part 1 are both based on previously published work. Chapter 2 was published essentially in its entirety as

- [1] Bottom, Michael, Philip S. Muirhead, John Asher Johnson, and Cullen H. Blake. "Optimizing Doppler Surveys for Planet Yield." *Publications of the Astronomical Society of the Pacific* 125, no. 925 (2013): 240-251.

Chapter 2 is a combination of a paper and a conference proceeding

- [2] Swift, Jonathan J., Michael Bottom, John A. Johnson, Jason T. Wright, Nate McCrady, Robert A. Wittenmyer, Peter Plavchan et al. "Miniature Exoplanet Radial Velocity Array (MINERVA) I. Design, Commissioning, and First Science Results." *Journal of Astronomical Telescopes, Instruments, and Systems*, Volume 1, Issue 2 (21 April 2015)

- [3] Bottom, Michael, Philip S. Muirhead, Jonathan J. Swift, Ming Zhao, Paul Gardner, Peter P.

Plavchan, Reed L. Riddle et al. “Design, motivation, and on-sky tests of an efficient fiber coupling unit for 1-meter class telescopes.” In *SPIE Astronomical Telescopes+ Instrumentation*, pp. 91472E-91472E. International Society for Optics and Photonics, 2014.

For Chapter 2, John Johnson defined the initial problem and gave guidance on how one might go about solving it. Phil Muirhead had many useful insights on measuring radial velocities, and provided patient help and insight on the astronomical tools available in the IDL programming language. Cullen Blake assisted with the statistical aspects of confirming the existence of planets based on multiple observations, and also had many useful comments on an early draft of the paper. I wrote and performed the numerical simulations, derived the different figures of merit, produced all the figures, and wrote up the work.

Chapter 3 is a large effort of a number of people from many universities, and many talented individuals contributed directly and indirectly. For the initial high-level system design, Philip Muirhead and I combined my results from Chapter 1 to calculate the survey yield and developed hardware requirements with these results, including things like telescope make and model, fiber type, spectrograph requirements, etc. (The survey yield calculation was subsequently improved by Nate McCrady and Chani Nava, who found results consistent with the initial calculation presented in this work to within the uncertainties.) Phil and I also both worked on the pointing control system, with Phil contributing the initial hardware design and myself responsible for simplification and improvements. The construction, testing, and validation of the pointing control system was done by myself with assistance from a high school summer student (now Caltech undergraduate) Erich Herzig. Kristina Hogstrom and myself set up the original computer control system of the telescope, with invaluable assistance from Kevin Ivarsen of Planewave Instruments. Kristina and Reed Riddle wrote the initial telescope and observatory control code in C++, which was eventually replaced by a Python version. I wrote the original Python control library for the telescopes with assistance from Kevin Ivarsen, which was also vastly expanded and improved by the efforts of Jason Eastman and many capable undergraduates.

The most significant person of all for the development of MINERVA detailed in Chapter 3 was Jonathan Swift, the project manager, who led the following efforts: telescope throughput validation, photometric validation, and site selection. On the throughput validation and photometric validation, we both performed the measurements; Jon analyzed the data and discovered the initial problem with the mirror coatings, and their solution. (I verified his results independently.) Jon led the published MINERVA paper [[2]], writing most of it. The text in Chapter 2 is all my own except for the sections I contributed in that publication. We participated equally in the site selection. Without Jon’s leadership and contributions, MINERVA would be far behind where it is today.

With respect to the last four chapters dealing with the Stellar Double Coronagraph and its tangents, all consist of previously published work except the last, which is currently in the process

of being published. The first three were published as

- [4] Bottom, M., J. C. Shelton, J. K. Wallace, R. Bartos, J. Kuhn, D. Mawet, B. Mennesson, R. Burruss, E. Serabyn. “Stellar Double Coronagraph: a multistage coronagraphic platform at Palomar observatory” Publications of the Astronomical Society of the Pacific, in press
- [5] Bottom, Michael, Jonas Kuhn, Bertrand Mennesson, Dimitri Mawet, Jean C. Shelton, J. Kent Wallace, Eugene Serabyn. “Resolving the delta Andromedae spectroscopic binary with direct imaging”. The Astrophysical Journal, 809, 11 (June 2015)
- [6] Bottom, M., Bruno Femenia, Elsa Huby, Dimitri Mawet, Eugene Serabyn. “Speckle nulling wavefront control for Palomar and Keck” In SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 2016
- [7] Bottom, M., James K. Wallace, Randall D. Bartos, J. Chris Shelton, Eugene Serabyn “Speckle suppression and companion detection using coronagraphic phase-shifting interferometry”. Submitted to Monthly Notices of the Royal Astronomical Society.

The Stellar Double Coronagraph, the paper comprising Chapter 4, benefited hugely from a very capable team from the Jet Propulsion Lab, led by Eugene Serabyn. Chris Shelton designed the optics and electronics of the SDC; Randy Bartos designed the optomechanics. System assembly and commissioning was done by myself, Chris Shelton, Kent Wallace, and Jonas Kuhn. I was primarily responsible for the initial assembly/alignment of the optics, electronics, and instrument control system, both low and high level, including the closed loop pointing control system to the adaptive optics system. In every single one of my responsibilities, the result was made much better due to input and guidance from Chris and Kent. I was also responsible for running all of the observations, writing the data analysis pipeline (with very much help from Jonas Kuhn), analyzing the data, and publishing all the results.

In Chapter 5, myself, Jonas Kuhn, and Eugene Serabyn performed the observation of the companion. Bertrand Mennesson and Eugene Serabyn provided useful guidance on how to interpret the data, and Jonas Kuhn had many comments which improved early drafts, as well as providing one of the final publication-quality images. I performed all the data analysis with code I wrote, and wrote all the text of the published work.

Chapter 6 involves wavefront control software I wrote, with the first version of the code was tested by Elsa Huby and Dimitri Mawet. For the results specific to the TMAS instrument, camera interface software was provided by Jennifer Milburn; Rich Dekany also contributed to the testing on TMAS. Bruno Femenia was immensely helpful in porting the software to the Keck Observatory, and without his resourcefulness, this wavefront control code would never have been able to make it

to the Keck big leagues. The general software design itself benefited greatly from discussions with Chris Shelton, Kent Wallace, Rick Burruss, and Eugene Serabyn.

The work described in Chapter 7 also benefited from many of the same contributors as before. Kent Wallace and Randy Bartos designed the optics and mechanics of the pistonning mirror. Kent and I installed and aligned the interferometer in the SDC. The initial commissioning and debugging was done by me, with help from Kent Wallace and Eugene Serabyn. The controls and calibration system, on-sky observations, data analysis, and interpretation were all done by me. Of course, the paper benefitted from insights of the other co-authors, which improved it and for which I am grateful.

Chapter 2

Part I—Optimizing Doppler Surveys for Planet Yield

Abstract

One of the most promising methods of discovering nearby, low-mass planets in the habitable zones of stars is the precision radial velocity technique. However, there are many challenges that must be overcome to efficiently detect low-amplitude Doppler signals. This is both due to the required instrumental sensitivity and the limited amount of observing time. In this paper, we examine statistical and instrumental effects on precision radial velocity detection of extrasolar planets, an approach by which we maximize the planet yield in a fixed amount of observing time available on a given telescope. From this perspective, we show that G and K dwarfs observed at 400-600 nm are the best targets for surveys complete down to a given planet mass and out to a specified orbital period. Overall we find that M dwarfs observed at 700-800 nm are the best targets for habitable-zone planets, particularly when including the effects of systematic noise floors. Also, we give quantitative specifications of the instrumental stability necessary to achieve the required velocity precision.

2.1 Overview

One of the most promising methods of discovering nearby, low-mass planets in the habitable zones of stars is the precision radial velocity technique. However, there are many challenges that must be overcome to efficiently detect low-amplitude Doppler signals. This is both due to the required instrumental sensitivity and the limited amount of observing time. In this section, we examine statistical and instrumental effects on precision radial velocity detection of extrasolar planets, an approach by which we maximize the planet yield in a fixed amount of observing time available on a given telescope. From this perspective, we show that G and K dwarfs observed at 400-600 nm are the best targets for surveys complete down to a given planet mass and out to a specified orbital period. Overall we find that M dwarfs observed at 700-800 nm are the best targets for habitable-zone planets, particularly when including the effects of systematic noise floors. Also, we give quantitative specifications of the instrumental stability necessary to achieve the required velocity precision.

2.2 Introduction

Over 400 extrasolar planets have been discovered by surveys using the precision radial velocity (PRV) method (Wright et al., 2011). The first discoveries of planets using the PRV method were mainly massive planets in close-in orbits, known as hot Jupiters, revealing a surprising diversity in planetary systems (Mayor & Queloz, 1995; Butler & Marcy, 1996b). More recently, technical advances have resulted in discoveries of planets with masses intermediate to terrestrial and gas giants, the “Super Earths” (McArthur et al., 2004; Rivera et al., 2005; Udry et al., 2007; Howard et al., 2011; Pepe et al., 2011). Future instrumentation promises the sensitivity needed to detect Earth analogs around Sun-like stars, but there are substantial challenges that must be overcome to attain this level of precision.

There are three kinds of limits to how well one can recover a radial velocity signal from a target star. First, there are statistical limits that come from the signal-to-noise ratio and the depth, density, and shape of the spectral lines. These effects may be further separated as having components that come from the physical properties of the star, such as the stellar luminosity and distance to the star, which affects the signal-to-noise ratio of the spectrum; the chemical composition, which sets the number and depth of spectral lines; and the rotation rate, which affects the broadening of the spectral lines. Additionally, there are effects that come from observing, such as the exposure time, telescope diameter, system throughput (sky to detector), and detector noise, which affect the overall signal-to-noise ratio; the resolution of the spectrometer and the sampling of the line spread function, which affect the width and clarity of the observed lines; and the decision of which wavelength range to observe. This latter point is notable as instruments are optimized to observe in a particular region

of the electromagnetic spectrum, and the line density and quality can vary significantly between various types of stars in different regions of their spectra. Furthermore, for ground-based studies, there are regions where telluric absorption makes observing impossible.

Even if all the negative effects above are minimized, there is still an important second class of problems which will hurt the velocity precision: the inability to properly control and/or characterize the changes in the instrumental profile of the spectrometer from night to night. Guiding errors, such as a star moving on a slit, can be a major component of this, as displacements in the center of light lead to skews in the instrumental profile. Additionally, small changes in the ambient pressure, temperature, etc. lead to changes in the optical path, which are degenerate with velocity shifts. This was realized as being a fundamental limit to velocity precision over 45 years ago (Griffin, 1967), and successful attempts to mitigate this require control of the environmental conditions and also active modeling of the fluctuating instrumental profile.

Finally, even in the limit of a perfect instrument and infinite observing time, there are wavelength-dependent stellar effects that can mimic the signal of a planet orbiting a star, such as starspots, stellar activity (jitter), and quasi-periodic oscillations of the stellar photosphere. These deleterious effects may be mitigated somewhat by clever observing strategies or modeling (Dumusque et al., 2011d,b), and the most successful programs are sometimes limited by this class of problems.

Many studies, notably Bouchy et al. (2001); Reiners et al. (2010); Wang & Ge (2011) have considered the best ways to detect planets around stars. However, we take a somewhat different approach, where our end goal is to optimize a radial velocity *survey* for sensitivity and detection efficiency. Optimizing an observing plan to detect a planet around a particular star has a different set of requirements than a plan that seeks to discover the maximum number of planets, or discover a particular class of planets. When considering how to optimize a survey, all the effects mentioned in the previous paragraphs must be considered together with the expected velocity signals caused by planets. Less massive stars are dimmer, but have higher velocity signals from similarly-sized planets. Additionally, there are more nearby low-mass stars than high-mass stars, given the present-day stellar mass function. In order to select targets for a survey, a balance must be struck, which will depend critically on the wavelengths of observation. As we will show, it will also depend on the particular kind of survey under consideration.

In this work, we will characterize the statistical limitations in radial velocity observations under the consideration of finite observing time. We will consider the most productive way to choose targets for a survey in order to maximize the recovery of *planets per unit of observing time*.

We arrive at our conclusions by the following chain of reasoning: in Section 2, we examine the precision achievable on different types of stars at a fixed distance and with fixed observing time, assuming a perfect instrument. In Section 3, we then relax the assumption of fixed observing time and discuss the time necessary to detect different types of planets. We remove the assumption of a

fixed stellar distance by considering nearby stars and the effects of the present-day mass function, and discuss the optimal way to select targets as a function of observing wavelengths. In Section 4, we discuss the effects of instrumental noise floors, how they arise, and how they affect the results in the previous sections.

2.3 Effects of Stellar type

Previous studies have sought to characterize the fundamental radial-velocity quality of different stellar spectra, considering the number density and depth of spectral lines. Additionally, these studies have examined the fundamental limitations set by the signal-to-noise of an observation (e.g. Bouchy et al. (2001)). These effects are not completely separate; they depend on the spectrum's shape and its specific intensity, modulated by the effect of stellar size and distance. If velocity precision is the ultimate goal, this requires additional consideration of at least one other thing, that being the observing wavelengths.

We choose to examine the effects of spectrum and wavelength range on velocity precision. We begin by using the latest spectral models (Allard et al., 2011) based on the PHOENIX code (Hauschildt et al., 1999) in the T_{eff} range of 2600 to 6200K, $[\text{Fe}/\text{H}]=0$, from 0.3 to 2.5 μm . For stars of $T_{\text{eff}} = 2600 - 3400$ K, we choose models where $\log g = 5.0$; for 3600-5800 K we use $\log g = 4.5$, and for 6000-6200 K we use $\log g = 4.0$. In order to simulate the effects of instrument resolution, we convolve the models with a generic Gaussian instrumental profile equivalent to a resolution of $\lambda/\Delta\lambda = 75000$ and a sampling of 3.0 pixels per resolution element. Next, we rotationally broaden the spectra of stars with $T_{\text{eff}} > 3600\text{K}$ to match a disk-integrated $v \sin i = 2.0$ km/s. For cooler stars, we use $v \sin i$ values taken from Reiners et al. (2010) (see Table 2.8 for the exact values). This can be considered the perfect observation, with the only degradation of spectral quality set by the spectrometer resolution and rotational broadening. To consider the effects of finite observing time, we then inject Poisson noise at a level equivalent to one minute of observing time, and Gaussian read noise at the level of 5 electrons/pixel spread over 10 pixels in the cross dispersion direction ¹. Note that fainter stars will have a lower S/N ratio, as will wavelength bands away from the peak of the spectral energy distribution.

Determining a radial velocity is equivalent to trying to recover the shift in a spectrum with respect to a reference spectrum, according to $\Delta\lambda/\lambda = v/c$. Algorithms to do this include cross-correlation maximization, least-squares minimization, forward modeling, and various simplex algorithms. For our simulated observations, the degradation of the spectrum with noise will introduce a spurious shift with respect to the convolved, non-noisy spectrum, which we recover using an algorithm where we fit the peak of the cross-correlation function with a polynomial. We checked that this converges to

¹for the typical (high) signal-to-noise levels in precision radial velocity observations, read noise does not contribute significantly to the error budget, a fact that we verified here

zero velocity shift linearly the signal-to-noise ratio approaches infinity, as expected. This precision is what would result from having a perfect reduction pipeline, regardless of the specifics of the calibration method.

As a measure of our velocity precision, we repeat this simulated observation three hundred times, and take the standard deviation of the radial velocities as a measure of the velocity precision, (σ_v). In order to assess the best wavelengths of interest and target spectral classes, we repeat this simulation for wavelength regions of size 100 nm and stellar effective temperatures from 2600-6200 K at steps of 200 K, from 0.3–2.5 μm , with an assumed distance to the star of 10 pc. The results are shown in Figure 2.1.

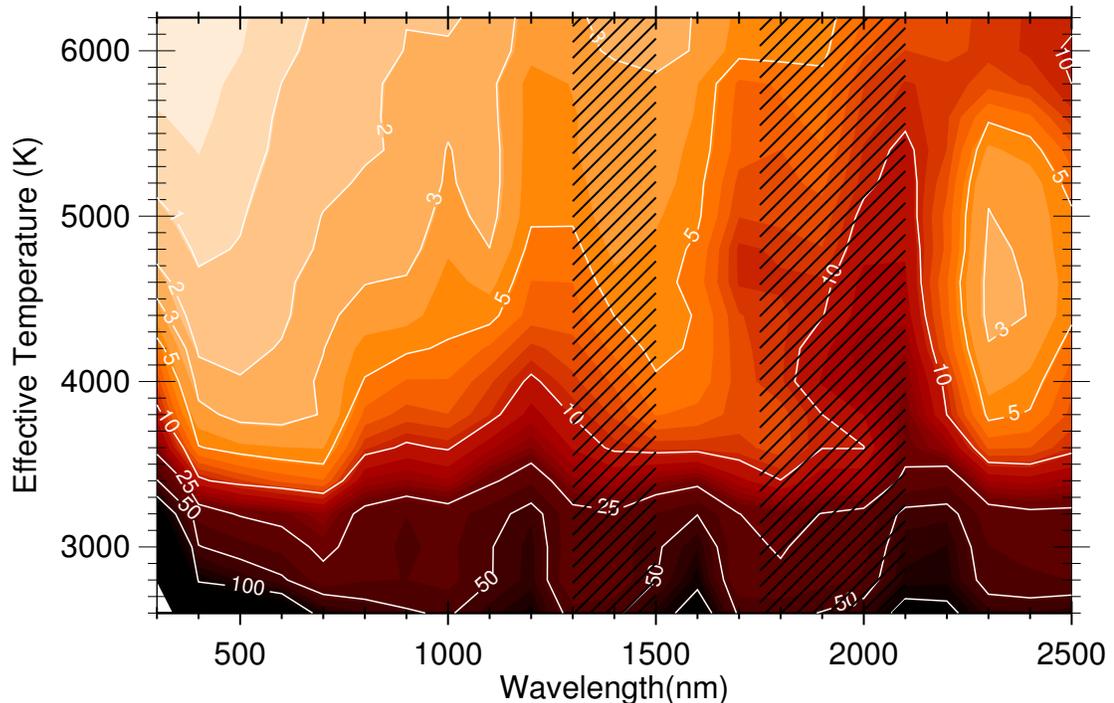


Figure 2.1: Doppler precision as a function of wavelength range and star temperature for a *fixed amount of observing time* (60 s). The stellar spectra are derived from rotationally-broadened main-sequence templates from 2600-6200K, stepped in 200K increments, and the wavelength range is stepped in 100 nm increments. The contours indicate the velocity precision in m/s. From the perspective of velocity precision, the best result is achieved in the range of 400-600 nm. The hashed regions correspond to wavelengths where the infrared absorption is too high for ground-based observations to be effective. This simulation assumes a 1.28 m² telescope dish, a spectrograph with $R=75000$, and sky-to-detector throughput of 10% (the full simulation parameters, including stellar parameters, are given in the Appendix).

The main conclusion of this numerical experiment is that in terms of best achievable *radial velocity precision* in a fixed amount of observing time for a star at a fixed distance, there is little advantage to moving redward of 600 nm, and the best overall area to observe for a large range of spectral

classes is the region between about 400 and 600 nm. This result is based on the convergence of a few physical effects. First, the spectral energy distribution peaks in this wavelength range for stars in the temperature range 4750-6000 K. This gives an advantage in terms of signal-to-noise. Second, high spectral quality occurs at these wavelengths, caused by deep atomic absorption features from metals.² While the cooler stars do share some of these features, their spectral energy distributions peak at redder wavelengths. This is evident in the contours moving rightward in the 2600-4250 K, 600-800 nm area. Additionally, the overall decrease in intensity of the stellar spectra at longer wavelengths leads to a general decrease in precision from left to right. Furthermore, the deep and rich molecular lines present in cooler stars partially offset the fact that the stars are smaller and dimmer, leading to bands of higher radial velocity precision from 1.4-1.6 μm and 2.3-2.6 μm , but the absolute level is still well below the precision from 400-600 nm. Any increase in the exposure time will only change the absolute scaling of the velocity precision; it will not affect the relative precision between different spectral types and wavelength bands.

There are a number of caveats to this analysis. In the infrared bandpass considered (longer than ~ 650 nm), there are significant absorption features by water and oxygen in the Earth's atmosphere, as well as OH emission lines, which make ground-based studies more challenging or even impossible. We have shaded the areas where infrared absorption is too high for effective observing from the ground. For the other regions, this simulation assumes that these features can be modeled and subtracted effectively and have no effect on the velocity precision. At high resolution, it is possible to identify telluric features and remove them, and recent results suggest that atmospheric calibration can subtract these lines effectively (stable to 10 m/s over 6 year timescales), and can even use these features as stable wavelength references to 2 m/s over short timescales (Figueira et al., 2010). A notable recent study that includes the effects of imperfect telluric subtraction can be found in Wang & Ge (2011). Even so, in our result, the photon errors are a few times higher, and dominate the error budget, despite the assumption of perfect sky subtraction.

Another important point, not obvious from the plot, is that the high average rotation rate (~ 9 km/s) of late M-dwarfs imposes a severe penalty on the velocity precision obtainable on these stars, as the lines are significantly blended at these speeds. Repeating this simulation with a fixed rotation rate of 2 km/s for these stars leads to about a factor of two higher in precision for the same amount of observing time. This is a major penalty, as a photon limited observation takes four times as long to get twice the velocity precision. Furthermore, the read noise is a proportionally larger part of the total signal at these lower photon counts.

The fixed integration time can reasonably be considered too severe a restriction, particularly for the cooler stars and at longer wavelengths, since similarly sized planet can cause a much larger reflex

²Coincidentally, this area happens to overlap the absorption lines of molecular iodine, a commonly used wavelength calibration source, which helps to explain the success of iodine cells in planet hunting.

velocity on a smaller star. Furthermore, the simulation considered different types of stars at a fixed distance from the Earth, which ignores the realities of the present day mass function. These are both fair points, and we consider the latter effect of detectability of planets in the next section.

2.4 Maximizing radial velocity survey yields

While it is clear that hotter stars at a fixed distance away are more amenable to high velocity precision, the goal of most radial velocity surveys is not solely high precision, but planet detection. Target selection is important. Therefore, there are two major revisions to the above analysis. The first is that planets of a given mass and orbital distance will cause a larger reflex velocity in lower mass stars. This means that less observing time will be needed to detect that planet, making lower mass stars more attractive targets, as their relative faintness is somewhat mitigated. Equally important is the distribution of stellar masses as a function of distance from the Sun—i.e., the probability that a star exists a certain distance from the Sun depends on how massive it is. Closer stars make better targets because because of the higher incident flux, though less massive stars have lower luminosities.

We first quantify the above statements. The reflex velocity on a star caused by an orbiting planet is given by³

$$v_* = 8.9 \text{ cm/s} \times \left(\frac{M_{pl}}{M_{\oplus}} \right) \left(\frac{M_*}{M_{\odot}} \right)^{-1/2} \left(\frac{a}{1\text{AU}} \right)^{-1/2} \quad (2.1)$$

where M_{pl} is the mass of the planet, M_* is the mass of the star, and a is the orbital distance of the planet (we consider only circular orbits for simplicity. Also, M_{pl} should always be considered as $M_{pl} \sin i$ to take into account inclination).

Secondly, we consider the amount of observing time per night it takes to detect a hypothetical planet which causes a given reflex velocity in its parent star. We consider a “detection” to be equivalent to a measurement with velocity precision equal to the reflex velocity of the star. A simpler way of saying this is that in the limit of many observations evenly spaced out over the planetary orbital phase, a single-point precision of 1 m/s is sufficient to detect a planet causing a 1 m/s modulation of its parent star’s velocity. We confirmed this using white-noise simulations of planet-induced stellar RV variations detected using both the Lomb-Scargle periodogram (Scargle, 1982) and a χ^2 test to reject the null hypothesis of constant stellar RV.

We repeat the calculation of Figure 2.1 for 10 minute’s worth of observing time, for a star at 10 pc, from Figure 2.1. For the more general case, we have a scaling:

³This expression can be derived quickly from conservation of momentum ($M_* v_* = M_{pl} v_{pl}$) and *Kepler’s* second law ($P^2 \propto a^3$), then scaling to the reflex velocity of Sun caused by the Earth, 8.9 cm/s.

$$t = 600 \text{ s} \times \left[\frac{v_{10\text{min},10\text{pc}}}{v_*(M_{pl}, M_*, a_{pl})} \times \frac{d}{10 \text{ pc}} \right]^2 \quad (2.2)$$

where v_* is the reflex velocity of the star; a function of the planet mass, stellar mass, and orbital distance, given in the previous equation. This expression can be derived in two steps. First, recall that the velocity precision scales as the inverse of the signal-to-noise ratio, which is proportional to the square of the observation duration. Of course, this scaling relation only applies in regimes where read noise is negligible, but this is true in the cases under consideration, where typical observing times are many minutes. Second, the signal-to-noise is equal to the square root of the number of photons, which scales as the inverse of the distance to the star ($\sqrt{N_{ph}} \propto d^{-1}$). Putting these two together gives the above expression, which is exact within our framework. This last result is interesting in of itself, because it is the inverse of *planets per unit observing time*, which we can evaluate as a function of wavelength of observation and stellar type.

Equation 2 depends on the variables M_*, M_{pl}, a_{pl} , which determine v_* ; on $\Delta\lambda_{obs}$ and T_{eff} , the bandwidth of observation and stellar effective temperature, which determine the velocity precision in ten minutes for a star at 10 pc, as well as the distance to the star, d . To simplify things and remove one parameter, we convert freely between stellar mass and effective temperature using the BCAH 98 isochrones (Baraffe et al., 1998) at 2 Gyrs ($[\text{Fe}/\text{H}]=0$), which are reasonably accurate for physical properties of low-mass stars. In this T_{eff} range, stellar properties are not very sensitive to evolution for the first few Gyrs on the main sequence.

In principle, all that is needed now is a complete list of stars of known spectral type and distance from the Sun (this gives d and M_* in the above expression), and the planet mass/orbital distance distribution as a function of spectral type. Then we can calculate the amount of time needed to detect a putative planet, as a function of observing wavelength range. We can then order the list in terms of observing time, which will show what wavelength range and spectral type maximizes the number of planet detections.

Unfortunately, the planet mass/orbital distance distribution function is not known, and there is hardly a complete list of stars in the galaxy with known distances and spectral types. Despite this, we can construct a reasonably accurate stellar census using the RECONS “100 nearest stars” sample (www.recons.org), which is complete out to ~ 7 parsecs. We populate the first 7 parsecs of our sample from RECONS. In extrapolating outwards, we assume a constant stellar number density per unit volume and proceeding outward in spherical shells, drawing from an estimate of the present day stellar mass function (Reid et al., 2002). As a check of this method, we found that this reasonably reproduces the number of stars of each spectral type in the RECONS 10 pc sample.⁴

⁴The full RECONS 10 pc sample has not been released as of the submission of this paper, though they have

with the deviation of our results at the few percent level, consistent with the statistical variability of the stellar neighborhood.

We do not attempt to guess the planet mass/orbital distance distribution as a function of spectral type; after all, discovering this is one of the goals of planet surveys. However, we can apply our analysis to cases where we can make reasonable assumptions. First, we consider surveys that are complete down to a certain planetary mass and orbital period. Second, we focus on searching for planets in stellar habitable zones.

2.4.1 Surveys complete to limits in planet mass and orbital period

Since RV surveys detect stellar accelerations, for a given stellar mass, the quantity $M_{pl}a^{-1/2}$ determines the necessary velocity precision. We can set this to a constant and then determine the necessary time to reach this precision as a function of stellar mass and distance:

$$t = 600 \text{ s} \times \left[\frac{v_{10\text{min},10\text{pc}} \times M_*}{(8.9 \text{ cm/s}) (M_{pl}a^{-1/2} = \text{const})} \times \frac{d}{10 \text{ pc}} \right]^2 \quad (2.3)$$

A plot of this result with the distance fixed at 10 pc and $M_{pl}a^{-1/2} = 5M_{\oplus}\text{AU}^{-1/2}$ is shown in Figure 2.2. The higher mass stars generally take less time in terms of planet detection, and the region of 400-600 nm is still the best for a range of stars.

To include the effects of real stellar populations, we simulate performing such a survey. We apply the results of Figure 2.2 to our simulated sample of stars out to 20 pc. For each star, we calculate the amount of observing time necessary to get a radial velocity precision lower than the reflex velocity caused by the planet.

In order to maximize the survey yield, we order these times in increasing order and select the stars until we exceed the amount of observing time per night (9 hrs, with 2 minutes of acquisition per target). We plot this result in Figure 2.3 as a function of wavelength range and spectral type (for our conversions from mass to spectral type, see the appendix).

The results in Figure 2.3 show that the best place to observe is around 400-500 nm, with the primary targets being F, G, and K stars. Redward of 600 nm, early M dwarfs become the primary targets, but the number of stars that can be observed from night to night is lower by a factor of two to four. Repeating this experiment for a distance out to 300 pc gives roughly the same results, as nearby stars are the most efficient, though there is a slight bias towards hotter stars in the visible wavelengths, because their luminosities make them observable over a larger volume.

The results here do not consider the effects of stellar activity, but it should be pointed out that F stars are known to be quite jittery and don't make ideal targets. Regardless, they make up a released the number of stars of each spectral class in the 10 pc sample.

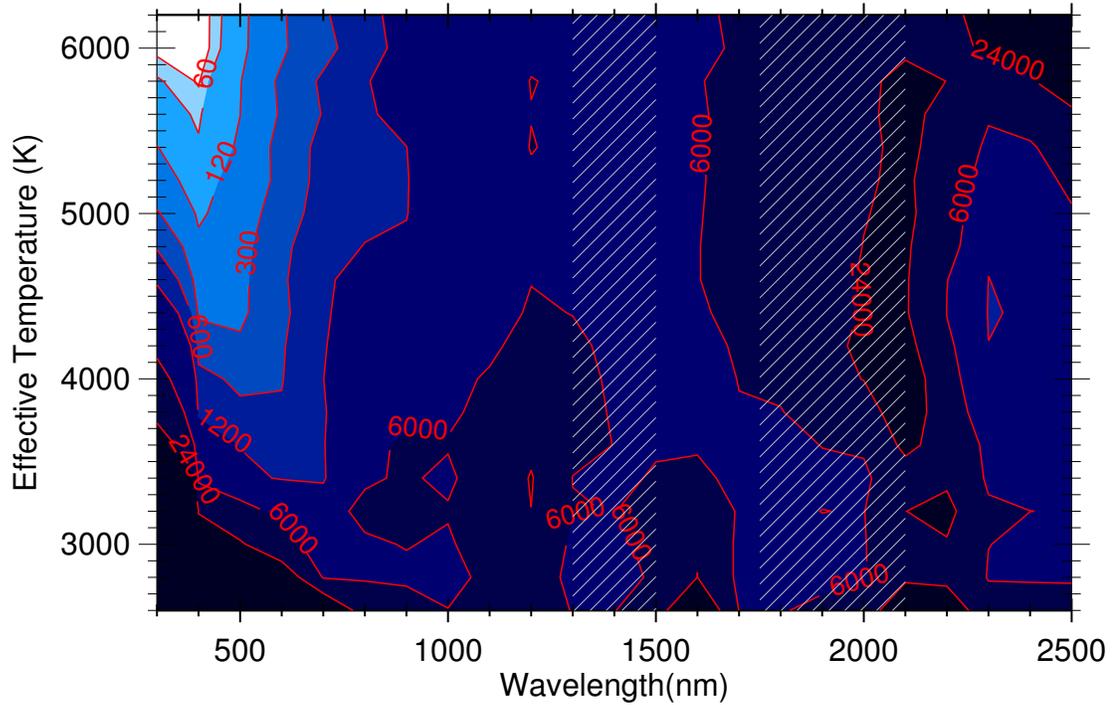


Figure 2.2: The time (seconds) to detect ($\sigma_v = K$) a planet with $M_{pl}a_{pl}^{-1/2} = 5M_{\oplus}(1\text{AU})^{-1/2}$, 10 parsecs away, for a range of observing wavelengths and stellar effective temperatures. The hashed regions correspond to wavelengths where the infrared absorption is too high for ground-based observations to be effective. This simulation assumes a 1.28 m^2 telescope dish, a spectrograph with $R=75000$, and sky-to-detector throughput of 10% (the full simulation parameters, including stellar parameters, are given in the Appendix).

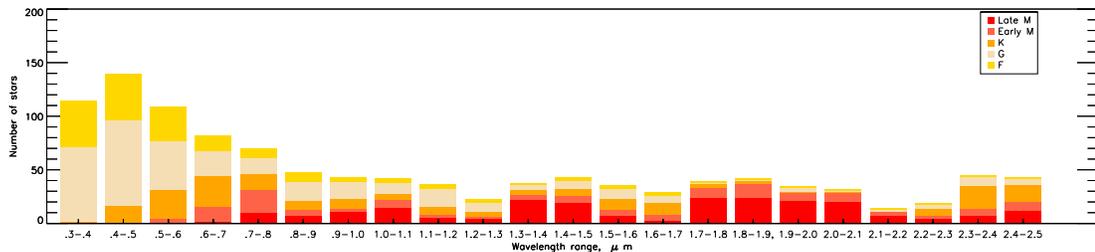


Figure 2.3: The maximum number of observable stars each night, with the goal of achieving the velocity precision necessary for a detection of a planet with $M_{pl}a_{pl}^{1/2} = 5M_{\oplus}\text{AU}^{1/2}$ on each star. This is plotted as a function of observing wavelengths, assuming 9 hours of observing time per night and 2 minute acquisition time between targets. The survey was simulated using the RECONS 7 pc sample and the present-day mass function (Reid et al., 2002), and extends to 20 pc (going out to 300 pc makes little difference, as bright and nearby stars are the most time-efficient targets).

relatively small percentage of the total stars, so the results do not change very much.

2.4.2 Habitable zone planet surveys

Next, we consider the case of planets in the habitable zones of their parent stars. Here we can avoid the question of planet distribution as a function of orbital distance and spectral type. For a star of a given T_{eff} , we can calculate the inner and outer habitable zone, where we define the habitable zone to be the region where the equilibrium temperature of the planet is between 175 and 275 K (Kaltenegger & Sasselov, 2011):

$$T_{eq} = T_{\text{eff}} \left[\frac{(1-A)R_*^2}{4\beta a^2(1-e^2)} \right]^{1/4} \quad (2.4)$$

where we set the albedo, $A = 0.5$, the eccentricity $e = 0$, and the planetary re-radiation fraction β to 1 for non-tidally locked planets and 0.5 for tidally locked planets (Peale, 1977). We set the radius to be a function of the effective temperature of the star, and use values generated from the BCAH 98 isochrones (Baraffe et al., 1998). (For main sequence stars in this T_{eff} range, the physical properties of the stars (mass, radius, etc) are not greatly affected by stellar evolution over a reasonably long timescale, so we simply use the values from the 2 Gyr isochrones.) With these assumptions in place, we can solve for the two values of a , the inner and outer habitable zones. We then calculate the reflex velocity of the stars caused by these planets, given by

$$v_* = 8.9 \text{ cm/s} \times \left(\frac{M_{pl}}{M_{\oplus}} \right) \left(\frac{M_*}{M_{\odot}} \right)^{-1/2} \left(\frac{a_{HZ}}{1 \text{ AU}} \right)^{-1/2} \quad (2.5)$$

once again using the BCAH 98 isochrones to derive a consistent mass estimate for the star from the effective temperatures, and placing the planet at the edge of the inner HZ.

We repeat the calculation of the observing time necessary to detect a $5 M_{\oplus}$ planet in the center of the habitable zone of its parent star, given by

$$t = 600 \text{ s} \times \left[\frac{v_{10\text{min},10\text{pc}}}{v_*(5M_{\oplus}, M_*, a_{HZ})} \times \frac{d}{10 \text{ pc}} \right]^2 \quad (2.6)$$

This result gives a notion of *habitable planets per unit observing time*, even if it depends on the distance to the star. A plot of this result, for a fixed distance of 10 parsecs, is shown in Figure 2.4.

Here, the M-dwarfs easily make the most attractive targets, as it takes much less observing time to recover a habitable zone planets around them. Furthermore, this result ignored systematic noise floors, which will make some of the brighter stars problematic as targets for habitable zone planets, as their reflex velocities can be below the noise floor.

Again, for the simulated stellar neighborhood of 20 pc, we calculate the observing time for each

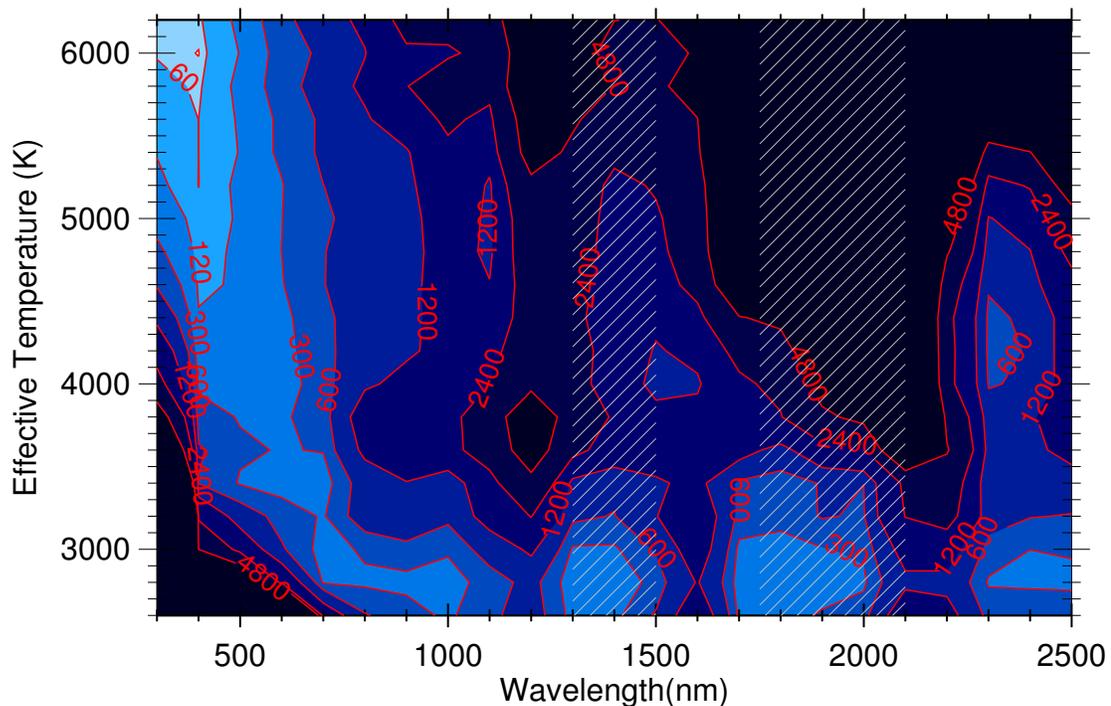


Figure 2.4: The time (seconds) to detect ($\sigma_v = K$) a $5 M_{\oplus}$ planet in the habitable zone of its parent star, 10 parsecs away, for a range of observing wavelengths and stellar effective temperatures. The hashed regions correspond to wavelengths where the infrared absorption is too high for ground-based observations to be effective. This simulation assumes a 1.28 m^2 telescope dish, a spectrograph with $R=75000$, and sky-to-detector throughput of 10% (The full simulation parameters, including stellar parameters, are given in the Appendix).

star necessary to detect a habitable zone planet according to the results of Figure 2.4. We assume the same 9 hours of observing time per night at 2 minute acquisition time between targets, and plot the results in Figure 2.5. Here, there are dramatically different results compared to the previous case. Because the habitable zone orbital distance decreases so rapidly for lower mass stars, the corresponding radial velocity signal is much larger for the same mass planet. This advantage is so pronounced that late M dwarfs become the primary targets for surveys operating at essentially all wavelength ranges except the bluest. The best wavelength to observe now becomes 700-800 nm, but this is somewhat constant for the range of 400-800 nm. Note that the absolute number of targets are somewhat higher for the habitable zone survey. This is an effect of the observing strategy; the observing time for each target is set by the *required* velocity precision, not the *ultimate* velocity precision of the instrument. If one can detect a habitable-zone planet at 5 m/s in 1 minute, getting twice the velocity precision in four times the observing time is not worth it, even if the spectrometer is able to reach the precision easily. Figure 5 may give the impression that there is little point to moving to the near-infrared, but this is somewhat an artifact of the choice of a $5 M_{\oplus}$ planet as the

target and the fact that systematic noise floors are not considered yet. For a lower mass planets and more massive stars, the radial velocities of the stars are so small that they would almost certainly be below the systematic noise floor of the instrument, meaning that the *most* promising targets would be in the near infrared, where the stellar radial velocities would be higher. See Section 4.3 and Figure 9 for further discussion of these effects.

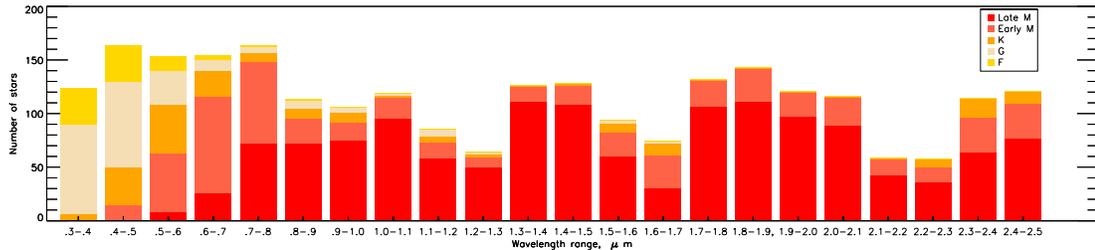


Figure 2.5: The maximum number of observable stars each night, with the goal of achieving the velocity precision necessary for a detection of a $5 M_{\oplus}$ habitable-zone planet on each star. This is plotted as a function of observing wavelengths, assuming 9 hours of observing time per night and 2 minute acquisition time between targets. The survey was simulated using the RECONS 7 pc sample and the present-day mass function (Reid et al., 2002), and extends to 20 pc (going out to 300 pc makes little difference, as bright and nearby stars are the most time-efficient targets).

As previously mentioned, we did not consider the “contamination” of stellar spectra by telluric lines in the Earth’s atmosphere, effects which are wavelength-dependent. In the case of the ideal survey complete to a mass-orbital distance limit, telluric lines are not relevant, because observations will be taking place in the visible wavelengths at 400-600 nm, where the atmosphere is mainly transparent. For the habitable-zone survey, they will certainly be present in many of the infrared regions. However, many telluric lines are stable at the the few m/s level, which is often *lower* than the reflex velocities of the low mass stars, the primary targets of the survey. This means telluric lines could be used as broadband wavelength references. Due to the potential of M-dwarf infrared surveys, efforts put towards improving telluric referencing and calibration are highly important (Blake & Shaw, 2011).

Notably, we did not consider stellar jitter caused by oscillations, granulation, or activity such as spots. For the interested reader, a thorough exploration of these subjects can be found in Dumusque et al. (2011d,b) where these effects are analyzed in detail for solar-type stars, and observing strategies/corrections to mitigate different kinds of jitter are explored. Of these three effects, the latter is considered the most troublesome in terms of planet detection, as the characteristic timescales of spot-related jitter are similar to planetary periods. The simulations in this paper deal primarily with statistical errors and their dependence on observing wavelengths and stellar effective temperatures; the effects of stellar jitter are categorically different, as they are real radial velocity signals with the potential to confuse the actual planetary signal. Furthermore, the spectral-type dependence of spot number, size, and shape are not well constrained, making meaningful simulation of these

effects difficult within our framework. It is possible that spot jitter would skew the results above; for example, if one spectral type typically has spot distributions that are extremely stable and similar to planetary signals, whereas another has spots that are easily distinguishable from planetary signals, then the latter would be preferable to the former. Of course, this would be dependent on the ability of the data reduction and observing strategy to distinguish spot jitter from true planetary signals, whereas the results above show more fundamental limitations. With this in mind, recent advances (Lanza et al., 2011; Aigrain et al., 2012) have shown a promising ability to subtract out spot jitter with the combination of high-precision photometry. Perhaps most encouragingly, by modeling stellar effects, Dumusque and collaborators were able to discover an Earth mass planet around α Centauri B, a star with a “stellar noise” level many times higher than the planetary signal (Dumusque et al., 2012).

2.5 Instrumental effects and systematic noise floors

The most unrealistic assumption so far is the assumption of a perfect instrument in our hypothetical survey, as all past, present, and future instruments have a limiting precision. In practice, this means that after a certain point, increasing the exposure time does not lead to an increase in velocity precision.

Up to this point, we have assumed that we have a particular instrumental configuration; a resolution 75000 spectrometer with 3 pixel sampling and a velocity precision limited by the signal-to-noise ratio and read noise (which is negligible for most cases). We first consider our choices of resolving power and sampling, and how they contribute to the velocity precision. Furthermore, we consider the effects of instrumental instabilities and show how they can lead to systematic noise floors. Finally, we assess the effects of these noise floors on our hypothetical survey, and show how they can substantially affect the optimal target selection.

2.5.1 Effects of Spectrometer Resolution and Sampling

The choice of resolving power of a spectrograph is important, as more sharply resolved lines lead to higher Doppler precision. A simple analytic calculation for one line suggests that the accuracy should scale as $\sigma_v \propto R^{-1}$, which was approximately reproduced in an early study (Hatzes & Cochran, 1992).

We examine the effect of spectrometer resolution on velocity precision at a fixed exposure time on a Sun-like (Teff=5800 K, log g = 4.5 and [Fe/H]=0) star. We simulate one thousand measurements at resolving powers ranging from 10,000 to 150,000, in the wavelength range of 500-600 nm, and as before, take the standard deviation of the velocity measurements as our velocity precision. The results are shown in Figure 2.6.

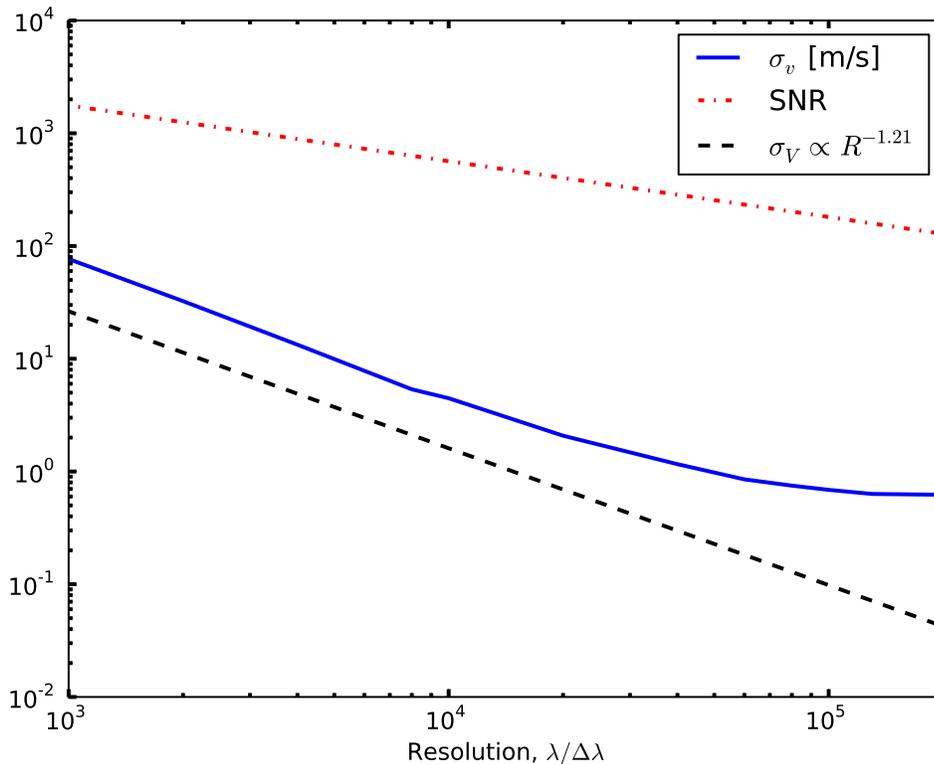


Figure 2.6: The resulting velocity precision and signal-to-noise ratio for an observation of a Sun-like star with varying resolution, (3.0 samples/resolution element). Note that the units on the vertical axis are m/s for the blue curve (velocity precision), and unitless for the red curve (signal-to-noise ratio). The exposure time is held fixed, and the resulting signal-to-noise ratio *decreases* as resolution increases, since less photons are incident per pixel. Increasing the resolution always improves the velocity precision, but the point of diminishing returns is reached at about $R = 45,000$, which corresponds to the point where almost all the spectral lines are fully resolved.

Our results show that the best-fitting power law is $\sigma_v \propto R^{-1.2}$, and that diminishing returns appear at about $R=45000$, where the spectral lines are resolved. It should be mentioned that the power in the proportionality is a function of the spectral quality (for example, the line density), but we were unable to find much deviation from the value of 1.2 in interesting spectral regions. We were not able to find any region where the power was as high as 1.5. Additionally, although precision improves with higher resolution, resolving powers above 100,000 yields little additional benefit, as essentially all the lines are resolved.

Satisfying the Nyquist sampling theorem requires that at least 2 pixels cover each resolution element, but more pixels may be used. Increasing the sampling turns out to have *no effect on velocity precision, assuming the exposure time stays constant*. The reason for this may be seen in that the number of photons per pixel is reduced by a factor of N , where N is the number of pixels per resolution element, and the number of pixels in the data product is increased by the same factor.

Since the radial velocity precision is inversely proportional to the signal-to-noise ratio ($\propto N^{1/2}$) and to the square root of the bandwidth, these effects cancel out. We verified this (non)effect with numerical simulations. We point out that in the limit of extreme (cm/s) velocity precision, it is not safe to ignore pixel topology effects, as the pixels may vary in efficiency over the center to the edge. This can become a problem when the line-spread function is minimally sampled.

Taken together, these two results show that our choice of resolution 75000 and sampling of 3 pixels in our simulated instrument was reasonable, and did not affect the recovery of radial velocities adversely.

2.5.2 Effects of incompletely recovered instrumental profiles

In general, the measured output from a spectrograph is the intrinsic spectrum of the object convolved with the instrumental profile (IP) or line spread function (LSF) of the instrument:

$$m(\lambda) = \int_{-\infty}^{\infty} s(\lambda') \text{IP}(\lambda - \lambda') d\lambda'$$

In this equation, $m(\lambda)$ is the measured spectrum, $\text{IP}(\lambda)$ is the instrumental profile (which is normalized to unity by conservation of flux), and $s(\lambda)$ is the “true” stellar spectrum. The instrumental profile is fiducially a gaussian with a full-width-half-max equal to the resolution of the spectrograph.

A typical extraction of a radial velocity datapoint involves consideration of the entire spectral region. First, the instrumental profile is extracted from a wavelength reference source, and after the observation, this instrumental profile is deconvolved from the stellar spectrum. In cases of simultaneous calibration (as with an iodine cell), the full transmission spectrum is modeled. Finally, the radial velocity datapoint is extracted from the shift in the spectrum with respect to the wavelength solution.

Properly characterizing the IP of the spectrometer is a challenging task, but is essential to recovering radial velocities accurately (Valenti et al., 1995), especially as all further steps depend on its characterization. It is a function of the optical path, and hence depends on environmental parameters like temperature and pressure, as well as slit illumination and focus. These parameters can change during and between observations, and thus the IP must be recalculated for each observation; that is, the IP varies with time. Furthermore, since optical elements have wavelength-dependent properties, the IP is wavelength-dependent. This means that in practice it is necessary to model an IP varying with wavelength, rather than a constant one. It is clear that any change in the IP that can be accounted for and modeled is not relevant. However, any changes not accounted for will be interpreted as radial velocity shifts—this is due to having two different IPs; the physical instrument profile, and the approximation that is deconvolved from the observation.

To model this, we convolve our model spectrum with a perturbed IP. While it is obviously not

feasible to examine every possible perturbation, we can derive useful rules of thumb from characterizing simple cases. We begin by restricting ourselves to a gaussian LSF with equivalent resolution of 75000 and 3.0 samples/resolution element as our ideal IP. We simulate the effects of LSF mismatch by convolving with a stellar spectrum ($T_{\text{eff}}=5800$ K, $\log g=4.5$ [Fe/H]= 0) with the perturbed spread function and trying to recover the velocity shift (which should be zero) with respect to the spectrum convolved with the ideal IP. As before, we repeat this many times and take the standard deviation of our derived velocities as our velocity precision.

In the first case, we consider a gaussian LSF with some skew added. Skewness ($\hat{\gamma}_3$) is a measure of asymmetry of the distribution, and is a property that can be straightforwardly calculated for a particular instrumental profile. Physically, skewness in the IP results when a source moves perpendicular to the slit direction.⁵

Mathematically, the skewness of a function is given by:

$$\hat{\gamma}_3 = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] \quad (2.7)$$

This equation is general and applies to any distribution. For our purposes, we consider the skew-normal distribution with parameter α , given by the function

$$f(x) = 2\phi(x)\Phi(\alpha x) \quad (2.8)$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \Phi(x) = \int_{-\infty}^x \phi(t) dt \quad (2.9)$$

Note that the skew parameter α is not actually equal to the skewness of the distribution, which is a complicated function of α :

$$\frac{\alpha}{\sqrt{1 + \alpha^2}} = \sqrt{\frac{\pi}{2} \frac{|\hat{\gamma}_3|^{\frac{2}{3}}}{|\hat{\gamma}_3|^{\frac{2}{3}} + ((4 - \pi)/2)^{\frac{2}{3}}}} \quad (2.10)$$

where the skewness ($\hat{\gamma}_3$) is recovered by inverting the equation above for α . Despite the rather opaque equation above, the usefulness of this parametrization is that it is, in a sense, a “simple” way of adding skewness in a distribution, and for $\alpha = 0$ one recovers the normal distribution, which is our fiducial way of representing an instrumental profile.

It is apparent from Figure 2.7 how a velocity floor can arise from an uncorrected skewed IP. For example, at a skewness of 0.05, it will be impossible to do better than 5 m/s in precision, regardless of the signal-to-noise ratio.

In the second case, we consider an IP with a small gaussian perturbation with varying amplitude.

⁵A real example of this effect can be found in the *Herschel* (Pilbratt et al., 2010) observer’s manual. <http://herschel.esac.esa.int/Docs/PACS/html/ch04s07.html>, particularly figure 4.17.

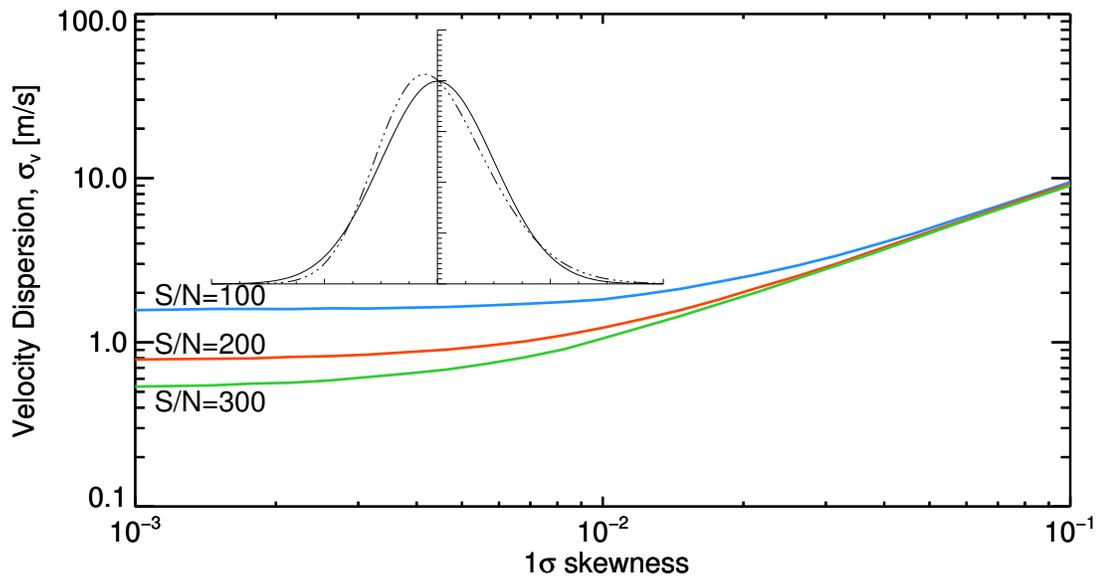


Figure 2.7: The resulting velocity precision for observation of a Sun-like star with $R=75000$ (3.0 samples/resolution element), and skewness varying from 10^{-1} to 10^{-4} . For reference, the inset shows a gaussian with skewness (not α) of 0.3, an order of magnitude higher than the maximum value considered (none of the skew-normal distributions simulated have skews large enough to be visually distinct from a normal distribution.) The skewness sets a signal-to-noise floor when it is greater than a part in 100, weakly dependent on signal to noise. The flattening out of the curves occurs where the signal-to-noise ratio limits the velocity precision.

The choice of this form of perturbation is due to the fact that a common practical way to represent an imperfect IP is through many small gaussian functions added together on top of the main gaussian IP (Butler & Marcy, 1996b; Endl et al., 2000; Kambe et al., 2002). The position of the perturbation is set to vary normally with a standard deviation equal to the standard deviation of the ideal IP (1). Also, the width of the perturbation is fixed to be of a characteristic size of one of the mini-gaussians used to model the IP. We examine the effects of the perturbation amplitude on the velocity precision under the same conditions as the previous test.

The result of this simulation, in Figure 2.8, reinforces how little tolerance there is in terms of characterizing the instrumental profile. A perturbation as small as 3 % IP peak is able to set a velocity floor of 5 m/s, independent of the signal-to-noise ratio.

2.5.3 Effects of noise floors on survey yields

The preceding section demonstrates the need of a stiff combination of stabilizing the IP through temperature, pressure, and illumination control and immediately capturing any changes that occur. However, it is reasonable to assume that there will be some velocity floor in every survey. We examined the effect by putting arbitrary noise floors at different velocities. We repeated our simulated

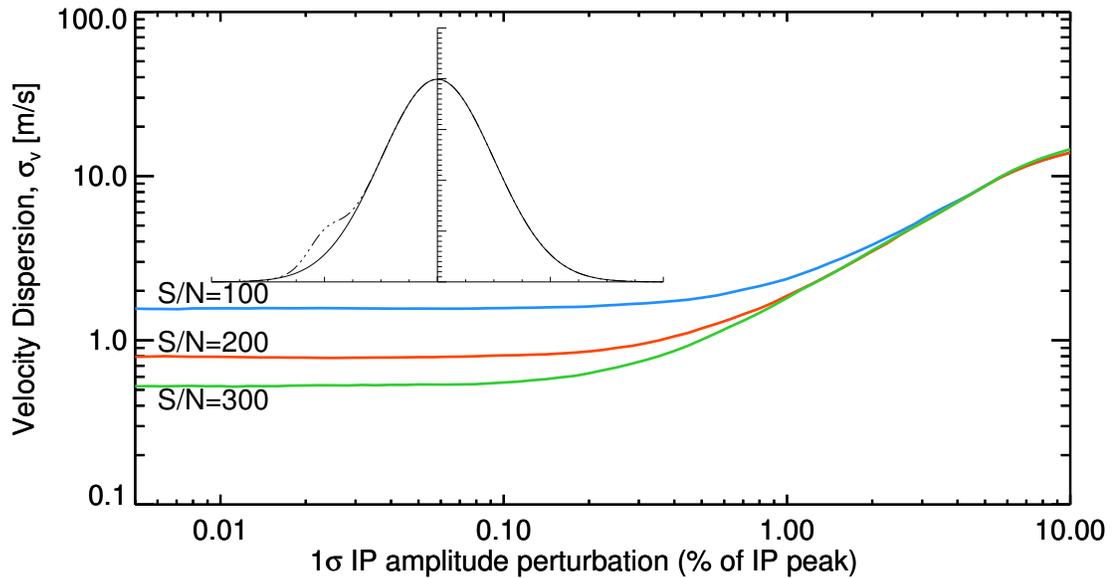


Figure 2.8: The resulting velocity precision for observation of a Sun-like star with $R=75000$ (3.0 samples/resolution element), and perturbation amplitude varying from 10^{-4} to 10^{-1} of the peak amplitude of the LSF. For comparison, the inset shows a gaussian with a perturbation of 10%, equal to the maximum value considered. It is clear that for sub meter/sec precision, it is important that the perturbation amplitude of the distribution does not exceed 0.1 %, a value weakly dependent on the signal-to-noise ratio. The flattening out of the curves occurs where the signal-to-noise ratio limits the velocity precision.

habitable-zone survey, except that we removed stars from the target list if their predicted planetary signal was below the noise floor (for the other survey, the very existence of noise floors negates its completeness). This makes massive stars become inaccessible targets, and they are progressively replaced by the next less massive stars as the noise floor increases, and observations go to the limiting precision. There are less targets per night as well, though the decrease is not as severe as one would think, as removing stars frees up available observing time for other targets within the detection limits: G stars replace F stars, K stars replace G stars, and so on.

We simulated this effects for noise floors of 0.5, 1, 3, and 5 m/s. The results are shown in Figure 2.9, which demonstrates that mid and late-M dwarfs become better targets as precision decreases, and that the number of stars observable per night does not decrease substantially—after about 800 nm, the amount of observable stars in a given night is set primarily by the duty cycle. Additionally, the wavelength range of 700-800 nm is the best overall in terms of total number of targets.

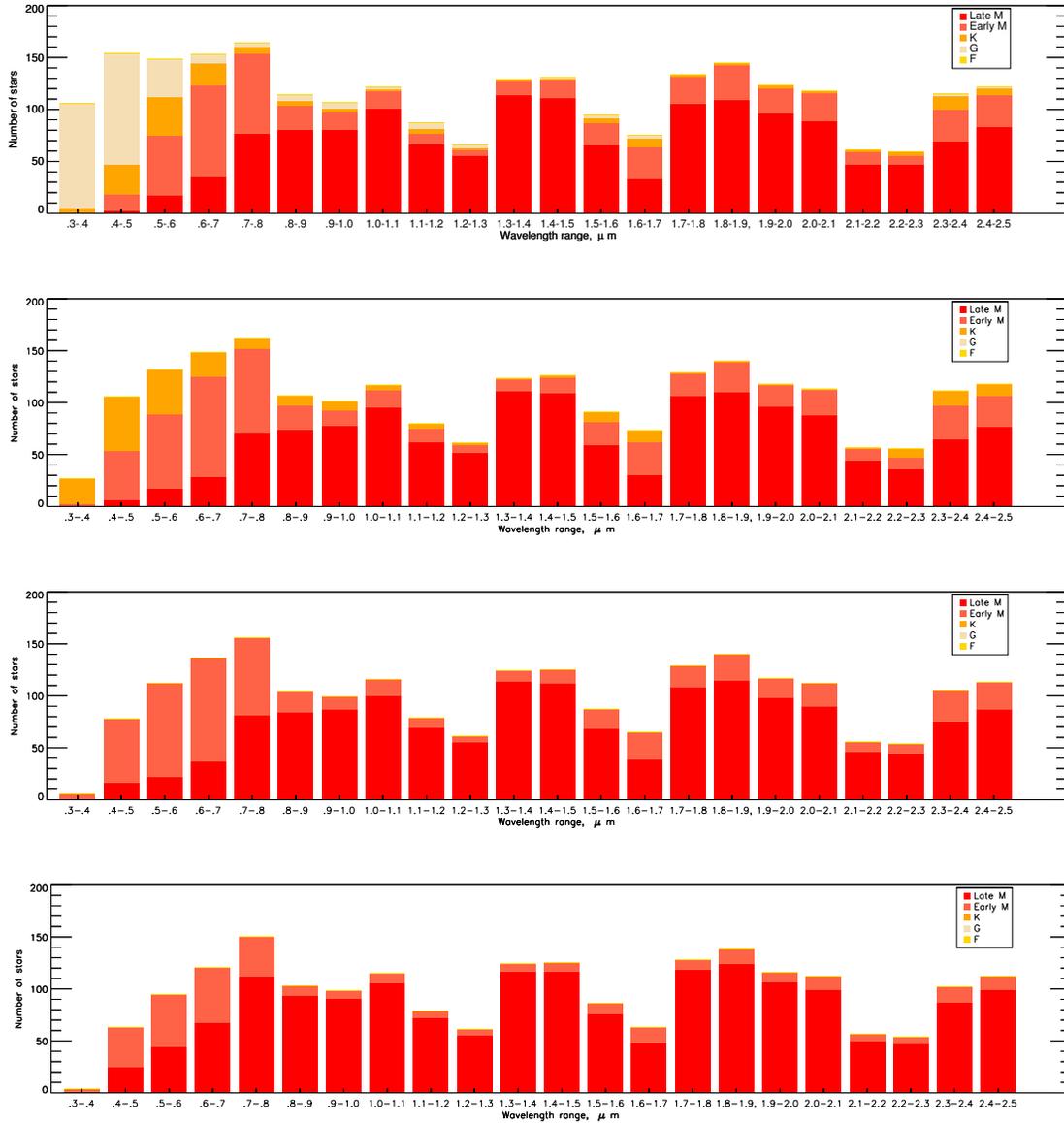


Figure 2.9: The maximum number of observable stars as a function of observing wavelengths, assuming 9 hours of observing time per night and 2 minute acquisition time between targets. The survey was simulated using the RECONS 7 pc sample and the present-day mass function (Reid et al., 2002), and extends to 20 pc (going out to 300 pc makes little difference, as bright and nearby stars are the most time-efficient targets). This graph assumes a survey targeting habitable zone planets, with velocity precisions limited to 0.5 (top), 1, 3, and 5 m/s. Furthermore, the region of 0.7-0.8 μm is the best area to observe overall.

2.6 Conclusion/Discussion

We have investigated the design requirements and performance expectations for radial velocity surveys, deriving the best wavelength bands and targets for different survey goals. For surveys targeting completeness out to a particular planetary mass-orbital distance product, the best targets are F, G,

and K dwarfs observed at wavelengths spanning 400-600 nm. For surveys of habitable-zone planets, the best targets are late M dwarfs in the wavelength range of 700-800 nm, though the number of possible targets stays flat from 400-800 nm. Of the two survey methods, those searching for habitable-zone planets are more productive, as the larger expected radial-velocity signals lead to more targets, and hence more detections. These results are based on the consideration of a number of input parameters, within the framework of maximizing planet detections in a fixed amount of observing time per night. In particular, we considered how the spectral quality, stellar flux, photon noise, stellar and planetary mass, and stellar mass distribution in the galaxy play against each other to produce different ideal targets.

For surveys that are complete to a constant $M_{pl}a^{-1/2}$, brightness of targets is paramount: the best targets are hotter stars, and the visible wavelengths where these stars' spectral energy distributions peak, coincident with where the density of absorption lines is highest. The target selection result holds even when considering the present-day stellar mass function. Generally, the observing strategy consistent with this analysis is to target as many bright, non-jittery stars as possible out to a limiting magnitude (or volume), then move to lower masses.

In the case of habitable zone planet surveys, our results demonstrate the potential of infrared surveys of M dwarfs. For an observing program that has relatively modest velocity precision, it is possible to have a survey of M dwarfs that is complete within the limits of the habitable zone. Note that this requires a different sort of observing strategy than one where each observation is taken to the noise floor of the velocity precision. This would make sense for F, G, and K stars, as the reflex velocities for habitable-zone candidates are very low. However, for M-dwarfs, this is inefficient, as habitable-zone velocity precisions are higher than systematic instrumental limitations. Since the velocity precision scales approximately as the *square* of the observing time, getting excessive precision will dominate the nightly time budget. This directly penalizes the number of target stars in the limit of complete phase coverage. A better choice is to observe only until the *target* precision is reached, and then move on to the next target.

We demonstrated how instrumental imperfections can lead to systematic noise floors, gave quantitative prescriptions for the level of stability needed in a restricted number of cases, and explored how these effects would change the scope of a habitable-zone planet survey. Significantly, the number of potential targets is highly dependent on the noise floor for shorter wavelengths, but is basically unchanged for longer wavelengths. The main change is that the target stars become later and later, demonstrating the high potential of infrared surveys.

2.7 Appendix

2.8 Simulation Parameters

The following table gives the parameters in our simulation of velocity precision in Figure 2.1. It is stated in the text when any of these parameters are changed.

Simulation parameters		
Property	Default Value	Unit
Stellar atmosphere models	PHOENIX	BT-Settl 2009
	[Fe/H]	0.0
	log g	
	$2600 \text{ K} \leq T_{\text{eff}} \leq 3400 \text{ K}$	5.0
	$3600 \text{ K} < T_{\text{eff}} \leq 5800 \text{ K}$	4.5
	$5800 \text{ K} < T_{\text{eff}}$	4.0
	α -enhancement	0.0
Stellar isochrone models	BCAH (1998)	2 Gyr
Stellar rotation rate, $v \sin i$		
$2600 \text{ K} \leq T_{\text{eff}} < 2800 \text{ K}$	9.0	km/s
$2800 \text{ K} \leq T_{\text{eff}} < 3200 \text{ K}$	6.0	km/s
$3200 \text{ K} \leq T_{\text{eff}} < 3800 \text{ K}$	3.0	km/s
$3800 \text{ K} \leq T_{\text{eff}}$	2.0	km/s
Distance	10	pc
Telescope area	1.28	m ²
Observing time	60	s
Spectrograph resolution	75000	
Spectrograph sampling	3.0	pixels per resolution element
Throughput (sky to detector)	10	%
Read noise	5	electrons
Cross-dispersion	10	pixels

2.9 Spectral class conversions

Spectral Class	Mass (M_{\odot})
Late M	0.08–0.23
Early M	0.23–0.51
K	0.51–0.79
G	0.79–1.05
F	1.05–1.6

Chapter 3

Part I—Minerva: a robotic telescope array for exoplanet discovery

Abstract

MINERVA (MINiature Exoplanet Radial Velocity Array) is a robotic observatory for the discovery and characterization of exoplanets. The system is designed for the sensitivity necessary to detect super-Earth type exoplanets around bright, nearby stars. The primary science goal of MINERVA is to be pursued by radial velocity operations; the secondary goal is photometric transit followup of planets. The primary advantage of MINERVA is its modular design and automated exoplanet observations every night, allowing for very high orbital phase coverage and quick turnaround on interesting targets of opportunity. We describe the science goals, design, construction, testing, and verification of MINERVA from its conception as project to its installation as a functioning observatory at Mt. Hopkins in Arizona.

The work and results described in this chapter can be viewed as the first stage of the project, up to installation and verification on Mt. Hopkins. This is not the end of the project, as the spectrograph installation, verification, and scientific operations are currently ongoing, with the upcoming spectrograph verification being the most critical part.

3.1 Introduction

The two most remarkable findings to come out of the young field of extrasolar planet science are that i) the Galaxy is full of planets; in fact, there are likely more planets than stars, and ii) the frequency of planets increases with decreasing planet size; that is, there are more small, rocky planets than large, gaseous ones (Howard et al., 2011). The latter result is particularly exciting, as recent work has shown that for the case of Earth-sized planets at orbital distances similar to Earth’s, the prevalence is on the order of 10% (Petigura et al., 2013). For smaller M-dwarfs, the most common stars in the Galaxy, the prevalence of habitable-zone Earth-sized planets is similarly calculated to be approximately 15% (Dressing & Charbonneau, 2013). These results are even more impressive in light of the fact that just 25 years ago, there were no known planets outside of our own solar system. Now, there are hundreds of confirmed planets, ranging in size from smaller than Mars to tens of times the mass of Jupiter; orbiting normal sun-like stars, tiny M dwarfs, pulsars (Wolszczan & Frail, 1992), even binary star systems (Doyle et al., 2011). To a large degree, this progress reflects the achievements of the two most successful methods of planet detection. The first is the Doppler method, by which the tiny shift of starlight caused by the gravitational tug of an orbiting planet is measured with respect to a stationary reference source. The second, exemplified by NASA’s *Kepler* mission, is the transit method, by which the dip in intensity of starlight caused by a planet passing in front of its host star is measured.

The two results point to the conclusion that the nearest stars to the Sun should be teeming with planets, particularly low-mass planets of sizes comparable to the Earth and so-called “Super-Earths,” a recently discovered class of planets larger than Earth but smaller than gas giants. However, few of these planetary systems in our solar neighborhood have been discovered. Doppler studies are the most promising way of finding these planets, but the precision required for detecting short-period Earths and habitable-zone Super-Earths is beyond the range of almost all current instruments. Even more important is the tremendous cadence needed to get dense orbital phase coverage, which requires dedicated, repeated observing that is unrealistic within the framework of shared telescope time allocation. This issue is unappreciated, especially as many groups are trying to develop “extreme” precision instruments with the goal of reaching a few cm/s of performance. This level of single-point precision requires long integration times, and without phase coverage does not necessarily increase detection efficiency. Even dedicated exoplanet programs typically are allocated a few dozen telescope nights per year (Sousa et al., 2008; Howard et al., 2010a). The net result is a very poor planetary census of the solar neighborhood.

To fill this scientific need, we are building a dedicated, multi-telescope robotic observatory for detection of extrasolar planets around nearby stars, called the MINiature Exoplanet Radial Velocity Array, or MINERVA . The observatory will pursue exoplanet science every possible night, with the

flexibility of using both precision Doppler spectroscopy or transit photometry. It will consist of four small telescopes which can observe different targets or work together to synthesize a larger effective aperture. It will observe every night, giving an unprecedented level of orbital phase coverage. Finally, MINERVA will be built in a modular fashion, using tested, commercially available technologies, and incorporating the most important lessons from fifteen years worth of ground-based exoplanet instrumentation.

The primary goal of MINERVA will be to discover Earth-like planets in close-in (<50 day) orbits around nearby stars, and super-Earths ($M > 3 M_{\text{Earth}}$) in the habitable zones of sun-like stars. This will lead to new, potentially exotic planetary systems, testing planet formation theories, finding targets for the next-generation of space-based missions, and placing the Earth in a broader context among planets in the Solar neighborhood. The target precision of MINERVA is 0.8 m/s, a factor of two better than Keck/HIRES (Howard et al. 2010), which has discovered the most planets by the Doppler method, and comparable to HARPS, the leading European spectrometer (Lovis et al. 2006). The secondary science goal stems from the fast and accurate instrument switching capabilities of the dual Nasmyth ports of our telescopes, allowing us to pursue photometric transit work of our Doppler-detected planets. While the Doppler method measures the mass of a planet, a transit measurement gives the radius. Measuring both the mass and radius can help characterize the interior structures of planets in a mass range not found in our solar system, 3-15 Earth masses.

We performed extensive simulations of the capabilities of MINERVA, considering both instrumental limitations (spectrograph resolution, dish area, optical efficiency, optical errors, integration time, statistical errors, etc) and fundamental physical limitations (stars spectra, luminosities, distances, spots, rotation rates, etc). We used this performance model along with published Kepler data on planet frequencies to simulate observations over a three-year timescale, including overheads like duty cycle, weather, and other confounding factors. Our simulations indicate that MINERVA will discover 15 ± 4 planets, where 3.2 ± 1.3 will reside within the habitable zone and 1.0 ± 0.17 will transit.

3.2 Design considerations and survey yield

MINERVA is a medium-budget project, and design considerations were driven by a combination of cost and performance. The design choices in the following section were informed primarily by the conclusions drawn from the previous chapter.

The survey target list is drawn from the *NASA/UC* η_{\oplus} sample made up of 166 nearby, chromospherically inactive stars currently monitored by Keck/HIRES for orbiting exoplanets (Howard et al., 2009). The projected yield from this target list dictates the minimum effective aperture required for MINERVA. We choose from this list the maximum number of targets that can be observed from

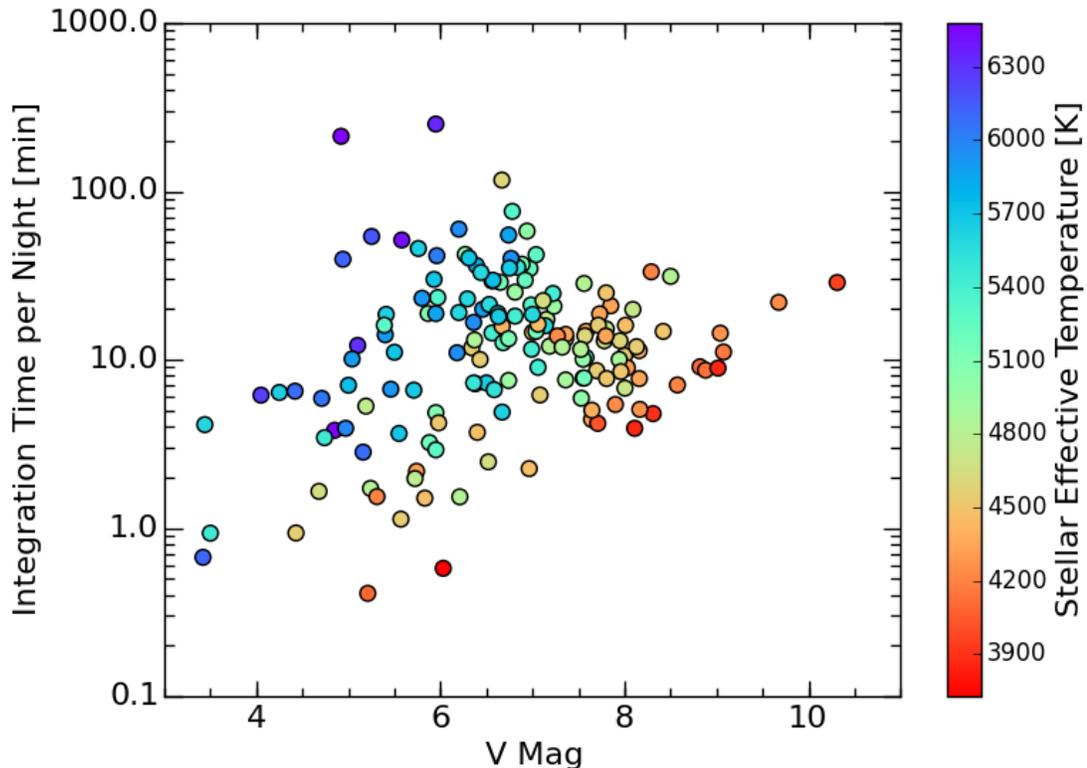


Figure 3.1: The required integration time per night for the MINERVA array to detect 3 Earth-mass planets within the habitable zones of each star in the η_{\oplus} sample according to a photon limited noise model as a function of V magnitude. Data points are colored according to their effective temperatures, determined using stellar masses from Howard et al. (2010b). This figure appears in Swift et al. (2015).

the final location of the array in southern Arizona to the precision necessary to detect planets with $m \sin i = 3$ Earth masses in their respective habitable zones with 3 years of observations (McCrary & Nava, 2014). Figure 3.1 shows the η_{\oplus} target list with the required integration time per night. As we do not have commissioning data for our spectrograph in hand, we perform simulations assuming a hard RV precision limit of 0.8m/s and a photon noise model. We assume a total system throughput of 10%, as stated in the previous chapter.

Our simulations account for stellar jitter, target observability, and weather losses based on historical weather records for southern Arizona. The radial velocity stellar jitter is modeled to match the spot-induced photometric variations observed in quiet G and K dwarfs by *Kepler* (Basri et al., 2011). We find that a spot model with 4 to 7 spot pairs per star with sizes ranging from 1.4% to 1.8% of the stellar radius reproduce well the observed photometric variations of between 3×10^{-4} to 7×10^{-4} . We populate the surfaces of our simulated stars with spots having lifetimes and latitudes

following the Solar “butterfly diagram” and assign rotation periods based on the period distribution seen by *Kepler* (McQuillan et al., 2013). The spot-induced pseudo-Doppler shifts are added to the simulated dynamical shifts caused by the planet. We then add white noise to the simulated radial velocities, recording the reduction in detectability of the simulated planet. The maximum amount of white noise added such that the planet was detected in 99% of the realizations was taken as the required per-night precision per star. Our final target list is chosen in consideration of the length of nights, declination of target stars, calibration observations, 10% overhead for secondary science and other programs, and a 25 s telescope slew and source acquisition time.

The projected exoplanet yield from the MINERVA Doppler survey is estimated using the statistical results from the *Kepler* Mission (Howard et al., 2012). We extrapolated the reported occurrence rates out to periods of 400 days for planets above 2 Earth radii and assume the same frequencies per log bin as those for 85-day periods. The latter is likely a conservative assumption, as the frequency of planets appears to rise beyond 50-day periods. For each target in the MINERVA sample, we randomly drew planets in the radius-period grid based on the extrapolated frequency surface and then converted the exoplanet radius to a mass using a density relation, $M \propto R^{2.29}$ (Howard et al., 2012). Signals larger than 3 m/s would have already been detected by the η_{\oplus} program, and are thus not included in the yield. Multi-planet extractions were not performed in this simulation and may delay confirmation of the largest RV signals in multi-planet systems.

The final yield results are obtained from the results of repeating the simulation 1000 times. Using an effective aperture of ~ 1.3 m (4 CDK-700 telescopes) we are able to observe the 82 brightest stars from the η_{\oplus} list, and we find a mean yield of 15 ± 4 planets with amplitudes between 0.8m/s and 3m/s. Improvements to these estimates are currently being pursued with the use of a more realistic stellar activity model and optimized observing strategies based on our recovery methods. Modeling of the stellar activity cycles will be necessary for the lowest mass planets and we recognize that the efficacy of these algorithms, which have yet to be quantified, will effect our estimated yield.

However, the ability to measure Doppler shifts of our targets at this precision every night (weather permitting) is unique to MINERVA and will be a key factor in recovering the RV signals of low-mass planets in their respective habitable zones. The observing cadence achievable with MINERVA allows us to account for the stellar variability of our sample of stars in a way that current facilities cannot. Based on our simulations and our expectations from a more refined treatment of stellar variability we take a 1.3m effective aperture to be the minimum required for ensured success of the project. We therefore design MINERVA around the use of four 0.7 m telescopes.

Built in to the design of MINERVA is the capability for flexible scheduling and simultaneous science and education programs. Of the 15 ± 4 simulated detections described above, 1.0 ± 0.8 are expected to transit their host star. Currently there are 16 RV detected planets with declinations $\delta > -20^{\circ}$, periods less than 30 days, $M \sin i < 50$ Earth masses, and $V < 10$. With an additional ~ 10

from MINERVA, the total transit yield is expected to exceed unity. This motivates the secondary science objective of MINERVA to search for transits of super-Earths among its RV discovered planets and to further characterize known transiting planets with multi-band light curves. This requires a broadband photometry precision of < 1 mmag in the optical. We demonstrate a comparable level of photometric precision from our commissioning site on the Caltech campus in Pasadena, CA (see Figure 3.3). On the educational front, students in lab courses can use one of the telescopes to conduct their course assignment and gain valuable observing experience—a useful community service—while the other telescopes simultaneously conduct the primary science program.

During the early stages of the project MINERVA will be used to follow up newly found Jupiter- and Neptune-sized planets from surveys like HATNet (Bakos et al., 2002), WASP (Pollacco et al., 2006) and NGTS (Wheatley et al., 2013) in addition to providing long-term monitoring for some TERMS long period planets (Dragomir et al., 2011). MINERVA photometry can also be used to follow up space-based discoveries. The *Kepler* K2 mission (Howell et al., 2014) will produce thousands of transit discoveries, but will only monitor each target field for approximately 75 days. At later times the number of potential targets will grow considerably as the TESS Mission (Ricker et al., 2014), set to launch in 2017, will yield many more detections over the whole sky, but with continuous monitoring of only 27 days for each non-overlapping field.

3.2.1 Doppler precision, spectral resolution, multi-dish arrays

There is a geometrical relationship between spectrograph resolution, spectrograph size, and telescope diameter, given by

$$R = \frac{\lambda}{\delta\lambda} = \frac{d\beta}{d\lambda} \frac{\lambda}{r\phi} \frac{d}{D} \quad (3.1)$$

where $\frac{d\beta}{d\lambda}$ is the angular dispersion, r is the anamorphic magnification, ϕ is the angle the slit (or fiber) subtends on the sky, and d, D are the spectrograph collimator and telescope diameters, respectively. It is clear from the previous chapter that Doppler velocimetry requires a relatively high resolution of $>50,000$; this sets R . For a constant R , and absent changes to the dispersion (typically set by the physical parameters of the grating) and slit/fiber geometry (which are usually matched to atmospheric seeing conditions), the collimator diameter must grow as the telescope diameter. In other words, doubling the telescope diameter quadruples the collecting area but also requires a spectrograph volume about eight times as large. For some practical examples, CSHELL ($R=30000$) on the 3 meter IRTF is about the size of an old CRT television set, while HIRES ($R=60000$) on the 10 meter Keck telescope is the size of a small car. From the perspective of ultra-precise Doppler velocimetry, a large spectrograph is harder to atmospherically stabilize, increasing the cost and complexity of the design.

One way around the tyranny of the spectrograph equation is to use smaller dishes to synthesize a larger effective aperture. (Of course, the telescopes are not actually co-phased as in an interferometer, but used as “light-buckets”). In this case, while the collecting area can be made large, the spectrograph size can stay small. The tradeoff exists in the detector size, which must be able to accommodate the spectra from multiple telescopes, but the prevalence of large-format, high quantum efficiency, low noise detectors makes this an acceptable price to pay.

Another advantage of having multiple telescopes is on the effects of output slit illumination. The output illumination pattern of the slit or fiber depends on telescope pointing. In the former case, the line spread function will have skew asymmetry when there is a pointing error; in the latter case, pointing jitter is one factor affecting the coupling into the fiber, particularly what modes are excited at the input end. Thus, different pointing errors will illuminate slightly different parts of the spectrograph optics, particularly the grating, where each groove has different optical properties due to imperfect construction. Having multiple fibers mitigates these kinds of errors, as the telescope pointing and fiber coupling errors are completely independent.

A final advantage in using multiple telescopes instead of a larger aperture is in the cost savings. It turns out that there is a sharp break in price at around the 1 meter aperture mark, corresponding to the difference between commercial and professional telescope designs (see Figure 3.2). All other things being equal, one can save factors of 4-6 in cost by using high-end amateur telescopes rather than professional telescopes for the number of photons/second collected.

3.3 Testing of MINERVA

Installation and commissioning of MINERVA at the test site in Pasadena took place from late 2012 to late 2014. Concrete was poured in late 2012, with cable junctions and connectors to route power and data installed shortly after. The computer systems and remote access systems were installed in addition to an uninterruptible power supply in early 2013. The Aqawan enclosure in February 2013, followed by the first telescope in April 2013. The second telescope was installed in March 2014; before this, various levels of instrumentation such as cameras, filter wheels, and the fiber acquisition unit were installed and tested on the first telescope. The two years the telescopes, enclosure and instrumentation were present on the Caltech campus allowed us to retire significant risk and streamline development while finding and preparing an appropriate location for the eventual installation of the project. In the following sections, we will describe the design and validation of each of the MINERVA subsystems. It should be noted that currently, the spectrograph is the only component not to have been deployed and tested on campus.

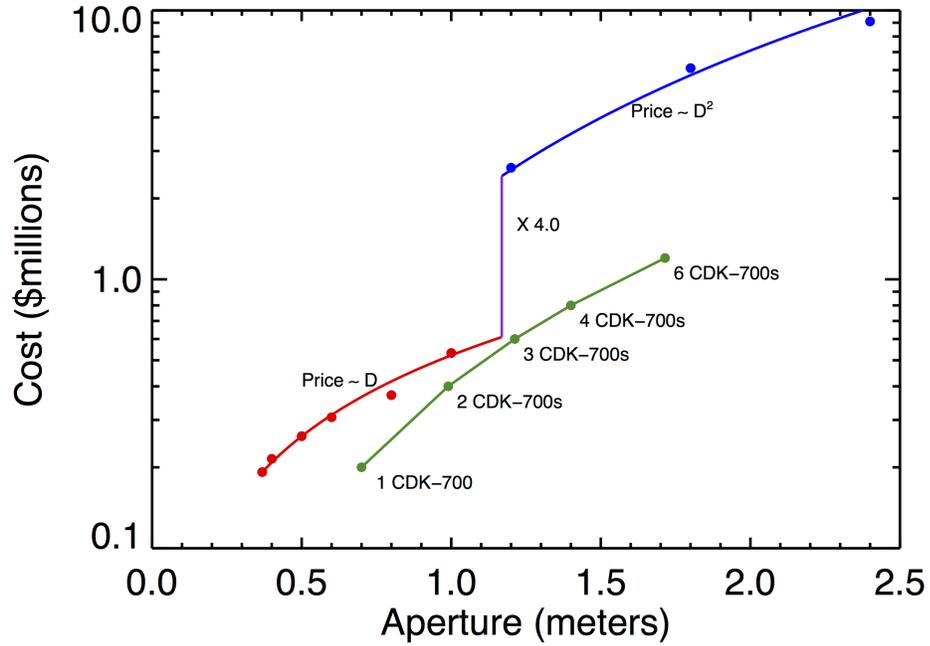


Figure 3.2: The differences in cost of telescopes of different aperture diameters. Commercial (amateur) telescopes are shown in red, professional telescopes are shown in blue. The CDK-700s (the telescopes we eventually chose) are shown as green points. This figure appears in Swift et al. (2015).



Figure 3.3: The location of our test site at Caltech, with the first two telescopes visible. The astronomy department is in the orange building across California boulevard at the top left.

3.3.1 Enclosure

The enclosure we selected for MINERVA is a modified “Aqawan,” designed by Annie Hjelstrom of LCOGT and assembled by Harmon Construction. The design is largely identical to those built at LCOGT but somewhat enlarged to accommodate our telescopes, which have dishes about twice as large as those used by LCOGT. The Aqawan is a cylindrical clamshell type design which can open and close in about 20 seconds, is watertight, and can be overpressurized to keep out dust and other undesirables. It has multiple useful features inherent in its design, including an internal steerable web camera, backup power capabilities, temperature and humidity sensors, reliable firmware with TCP/IP based communications that automatically close the dome if a “heartbeat” command is not received each minute, and easily accessible analog backups for emergencies.

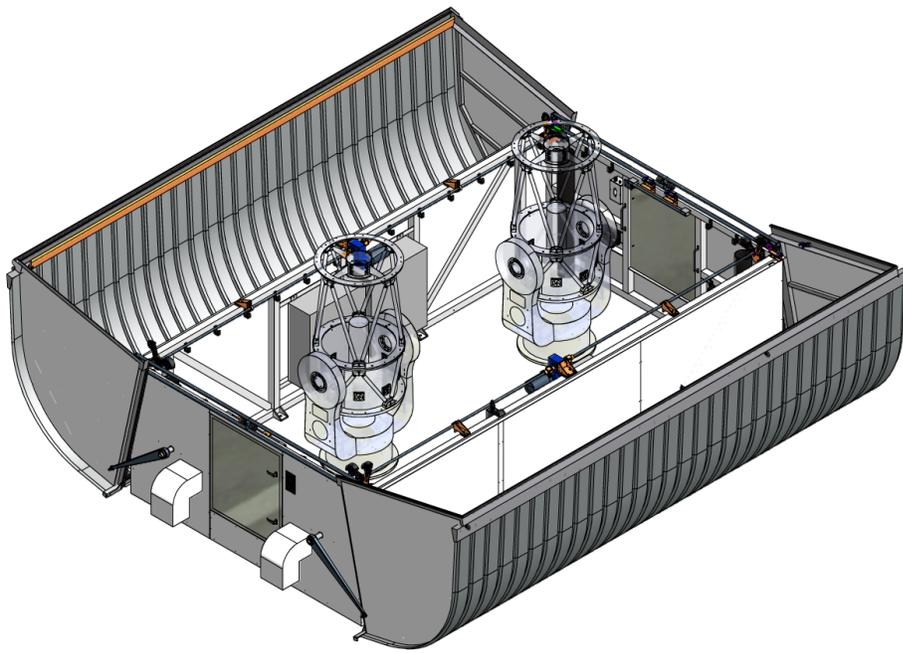


Figure 3.4: Engineering drawing of a modified Aqawan with two CDK-700 telescopes inside (see next section, compare to Figure 3.3).

During the testing of the Aqawans in Pasadena, we did not have serious issues with them, but overlooked a problem that became more serious down the line. Due to a miscommunication between the designer and the manufacturer, some seals and brushes in the clamshell-type roof were not installed. In the mild weather of Pasadena, even during heavy rain, there were no issues and the enclosures kept the interior dry. In the more serious weather at our eventually chosen site, where heavy rain combined with wind gusts of over 50 mph, the rain was able to eventually find a way into the enclosure. We discovered this, thankfully, during our initial commissioning run on the mountain, and sorted out the miscommunication and had the seals installed properly, at which

point the Aqawans became watertight.

3.3.2 Telescope system

The PlaneWave CDK-700 is a 0.7 meter, alt-azimuth mounted telescope system (Hedrick et al. 2010). It has a compact design, standing just under 8 feet tall when pointing at the zenith, with a 5 foot radius of maximum extent when pointing horizontally. The telescopes use a corrected Dall-Kirkham¹, optical setup, a two-mirror setup with a pair of correcting lenses to remove off-axis coma, astigmatism, and field curvature. This results in a flatter, more coma and astigmatism- field than the Ritchey-Chretien design, with the added benefit that the spherical secondary mirror makes alignment forgiving compared to the hyperbolic secondary of the Ritchey-Chretien. The CDK-700 has dual Nasmyth port outputs at f/6.5, with an image scale of 22 microns per arcsecond. The telescope pointing is controlled by two direct drive motors with highspeed encoders, resulting in a pointing accuracy of 10 arcseconds RMS, a pointing precision of 2-3 arcseconds, and a tracking accuracy of 1 arcsecond over a three-minute period. Additionally, the focuser is also motor- controlled and can be remotely adjusted, useful for defocusing the telescope when doing photometric work. Cooling fans and temperature sensors are used to equilibrate the primary mirror, and the control software is built to automatically correct for wind gusts and other perturbations, increasing the stability of the overall system. A detailed list of parameters is found in Table 3.1.

The two main criteria for evaluating the telescope system are pointing performance and throughput. The pointing performance can further be broken down into acquisition (that is, how accurately the telescope can acquire a star after a large slew) and guiding (that is, after acquiring a star, how well the telescope can follow the star position). Pointing and throughput are both important to photometric and spectroscopic performance.

3.3.3 Telescope system verification

3.3.3.1 Acquisition and tracking

The acquisition performance of the telescope depends on the internal motor and fork quality, the encoder resolution and repeatability, and the quality of the pointing model, that is, the map between sky position and encoder count. The method of measuring acquisition performance is relatively straightforward. First, known stars are slewed to and imaged, and a model is fit to these data points, consisting of considerations like zero point offset, sinusoidal type flexure and misalignment terms, etc. Many points are added to the model until a convergent solution with low scatter is found (typically, the first few stars are discarded from the model as the pointing errors for the first

¹This is distinct from the “Modified Dall-Kirkham” design, pioneered by Wynne and Rosin after World War II. The “Corrected Dall-Kirkham was developed by David Rowe of Planewave Instruments.

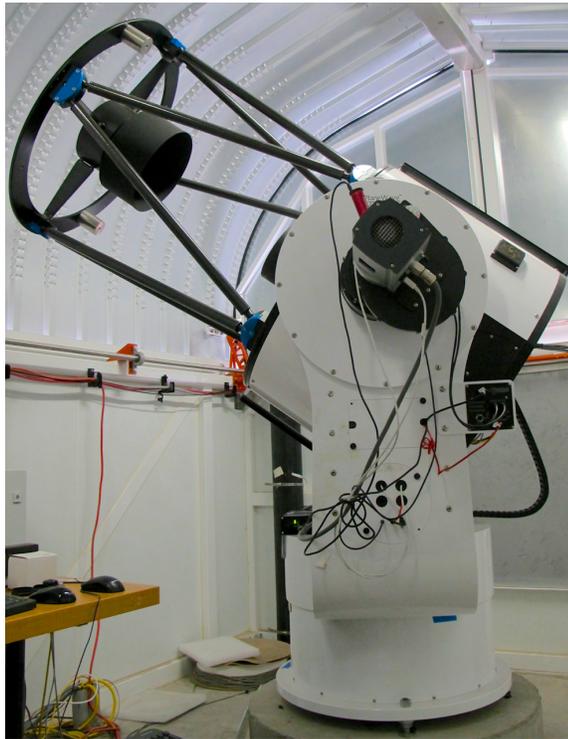


Figure 3.5: CDK700 telescope in our Aqawan enclosure in Pasadena. The telescope is positioned in its neutral position, pointing south at about 30 degrees. Also visible is the Andor iKonL camera, guider, and filter wheel on one of the ports.

CDK-700 parameters	
Optical Design	Corrected Dall-Kirkham (CDK)
Aperture	700 mm (27.56 in)
Focal Length	4540 mm
Focal Ratio	6.5
Central Obscuration	47% primary diameter
Back Focus	305 mm from mounting surface
Focus Position	Nasmyth (dual)
Dimensions	93.73" H × 43.25" W × 39" D
Weight	1200 lbs
Optical Performance	1.8 μm RMS spot size on axis
Image Scale	22 μm per arcsecond
Optimal Field of View	70 mm (0.86 degrees)
Fully Baffled Field	60 mm
Mount	Altitude-Azimuth
Fork	Monolithic U-shaped fork arm
Azimuth Bearing	20 in diameter thrust bearing
Altitude Bearing	2 × 8.5 in OD ball bearings
Optical Tube	Dual truss structure
Motors	Direct drive, 3 phase motor
Encoders	Stainless steel, 81 mas resolution
Motor Torque	~ 35 ft-lbs
Slew Rate	15° s ⁻¹ (max)
Pointing Accuracy	10" RMS
Pointing Precision	2" RMS
Tracking Accuracy	1" RMS over 3 minutes
Field De-Rotator	3 μm (p-p), 35 mm off axis (1hr)

Table 3.1: CDK-700 optical, mechanical, electronic, and pointing parameters, as specified by the manufacturer Planewave Instruments.

few points are too large to accurately centroid). The entire procedure typically takes 20-40 minutes. The model is saved and then tested on subsequent nights by slewing to different stars and measuring their positional offsets from the center of the detector. Our tests found that a newly built pointing could achieve a scatter of about 3 arcseconds, which degraded slightly over time, but stayed within the manufacturer specified tolerances of 10 arcseconds for a period of a month. Alternatively to rebuilding a pointing model each month, another way to proceed would be to generate pointing model updates “on the fly” every time a star is observed, with more weight given to more recent points. This would allow for changes in the pointing model due to slow changes in the optical system to be tracked over time. We are currently working with Planewave Instruments to implement such a system.

The tracking performance is also very important, as it determines the starting point for an active pointing control system. We measured the tracking performance of the telescope by slewing to different stars and calculating the center pixel offset as a function of time tracking on the source, tracking for much longer than our expected exposure times on stars. An example of such a measurement is shown in Figure 3.6. The pointing performance of the telescope was found to be very good,

even better than the manufacturer specifications. However, the tracking was not good enough for open-loop control, though this was expected in advance.

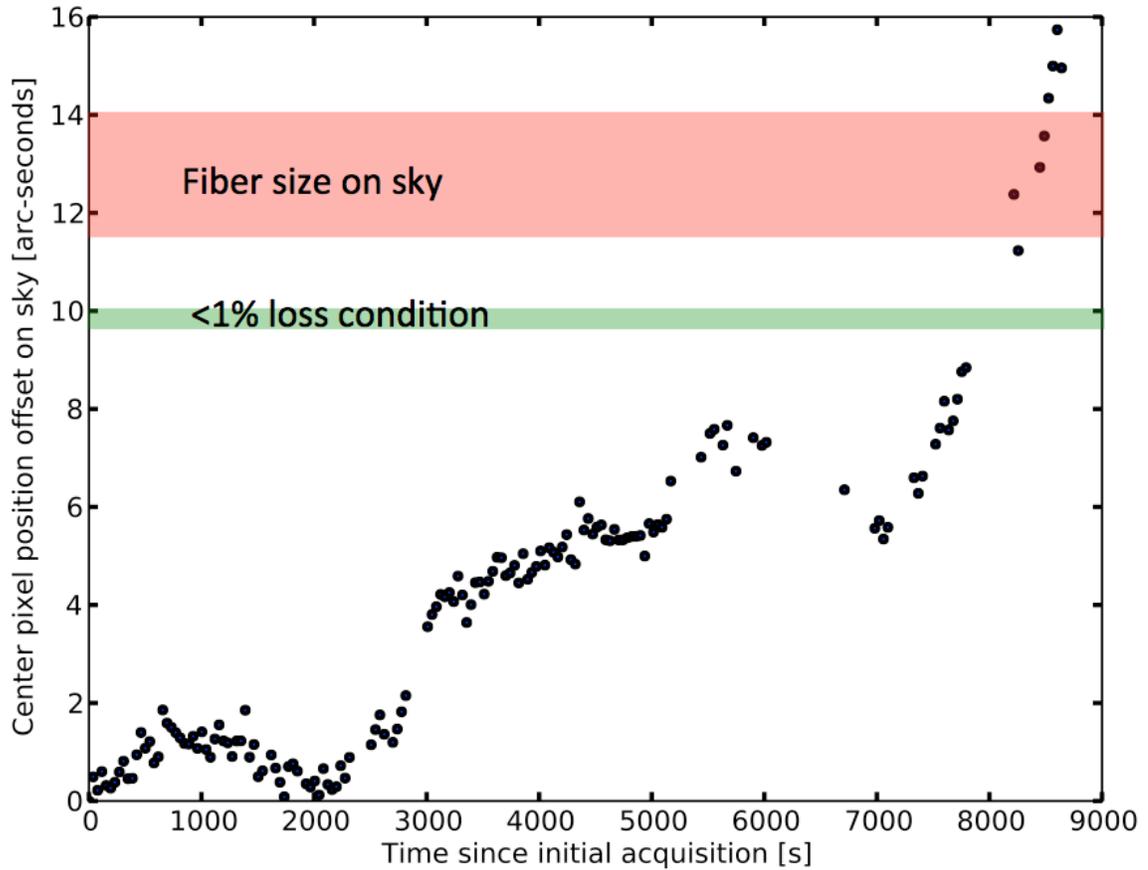


Figure 3.6: A single measurement of the pointing precision of the CDK-700 telescope. Each datapoint corresponds to the radial offset of the center pixel taken at 45 second intervals. The telescope holds the star position to 2 arc-seconds over a 50 minute interval, better than the claimed level from the manufacturer. For reference, the fiber size on the sky and required level of pointing precision for negligible throughput loss are shown above. It is clear that while excellent, the pointing precision is not good enough for open-loop control.

3.3.3.2 Throughput

Telescope throughput is a bit more challenging to measure, and requires an accurate characterization as it effects every science program on the observatory. The standard way to measure throughput is to measure the photoelectron production rate (counts/sec) of a source, and compare that to expectations given factors like the area of the telescope dish, reflectivities, filters used, camera parameters (gain, quantum efficiency), atmospheric attenuation (which depends on airmass), and the expected photon count from the star. Throughput tests were performed in Pasadena and Rancho Dominguez, which was further complicated by the highly variable atmospheric conditions in the area,

particularly with respect to dust levels and other aerosols.

The equation used to determine telescope throughput is (McLean (2008), chapter 9):

$$S = \int_{\lambda_{\text{low}}}^{\lambda_{\text{high}}} A_{\text{eff}} \tau(\lambda) f(\lambda) \eta(\lambda) F_{\lambda}(0) \times 10^{-0.4m} \exp[-\alpha(\lambda)X] \frac{\lambda}{hcg} d\lambda \quad (3.2)$$

with each term described in Table 3.2.

Term	Parameters in throughput determination equation	Unit/Comment
	Description	
S	Photoelectron detection rate	ADU/s
λ	Wavelength	Ang
A_{eff}	Effective Area	2998.33 cm ² (22% obscuration)
$\tau(\lambda)$	Other attenuation factors	Unitless (mirror reflectivity, etc)
$f(\lambda)$	Filter transmission function	Unitless
$\eta(\lambda)$	Camera quantum efficiency	Unitless
$F_{\lambda}(0)$	Photometric zeropoint at airmass X	ergs/cm ² /s/Ang
m	magnitude relative to photometric scale	Unitless
$\alpha(\lambda)$	0.4ln 10· extinction coefficient	Unitless, mags/airmass
X	sec(z), z =airmass	Unitless, accurate at low airmass
g	Camera gain	electrons/ADU
h	Planck's constant	6.626·10 ⁻²⁷ erg s
c	speed of light	2.99·10 ¹⁰ cm s ⁻¹

Table 3.2: Equation terms for determining telescope throughput

The quantities we measured in the equation above were the photoelectron rate S , from CCD counts and exposure time; the camera gain g , from flatfields; and airmass X , from telescope angle and location considerations. The quantities we determined from standard astronomical literature values were filter transmission f , target star magnitude m , and extinction coefficient α . Manufacturer data supplied effective dish area A_{eff} , reflective and transmissive efficiency η , and upper limits for other attenuation factors τ . Planck's constant and the speed of light h and c are fundamental and have well-known values. This completely determines the equation, as every term is accounted for.

However, the term τ covers essentially all of the interesting information on telescope throughput. Using the upper limits for $\tau(\lambda)$, the equation above can be used to derive a theoretical upper limit to the telescope throughput when combined with manufacturer-supplied transmission curves for mirror reflectivity, lens transmissivity, filter attenuation, and camera quantum efficiency. This theoretical upper limit is found to be 84 %, but the true throughput is expected to be somewhat lower. Pulling the mean value of τ out of the equation, we can see how this throughput can be measured:

$$\langle \tau \rangle_{V \exp[-\alpha X]} = \frac{Shcg}{\int_{\lambda_{\text{low}}}^{\lambda_{\text{high}}} A_{\text{eff}} f(\lambda) \eta(\lambda) F_{\lambda}(0) 10^{-0.4m} \lambda d\lambda} \quad (3.3)$$

Collecting data of known standards (known m) and airmasses ($z > X$), and measuring S allowed us to measure the efficiency τ as a function of airmass as well as fitting the extinction

coefficient $\alpha(\lambda)$; extrapolating to zero airmass (ie, above the atmosphere) gives the intrinsic telescope efficiency.

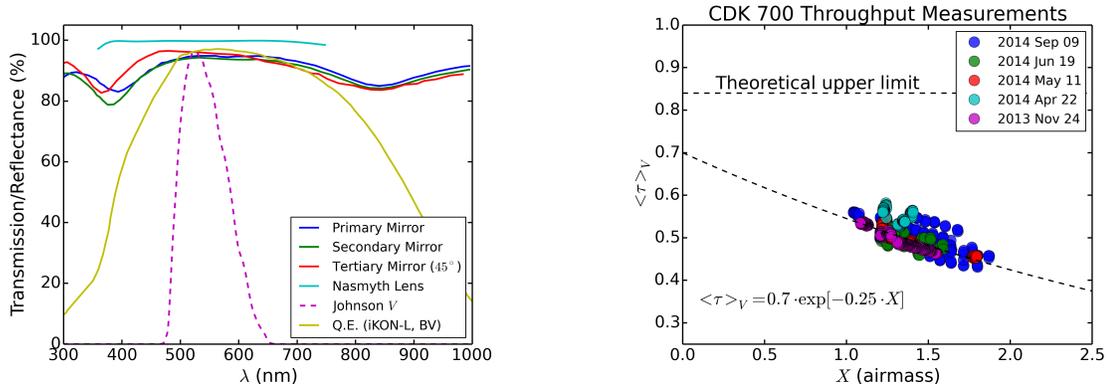


Figure 3.7: Left: Manufacturer-supplied transmission curves for the telescope mirrors, lenses, filters, and camera quantum efficiency. Right: throughput measurements of the telescope. The high level of variability is primarily due to the high level of variability in the atmospheric conditions in the areas where the tests were performed. The derived efficiency of 70% is in line with expectations, especially as the derived extinction of 0.25 mags/airmass is a reasonable value in the V band. These figures appear in Swift et al. (2015).

The measured telescope efficiency was 70%, about 10% less than the theoretical maximum efficiency of 84% derived from the filter/mirror transmission curves. The errors on this measurement were formally high (30%) due to covariance between the two parameters, but repeated measurements with different sites, telescopes, and cameras gave best fit values that always agreed within the error bars, building confidence in the answer. These measurements will be repeated in the future at Mt. Hopkins as needed to characterize the degradation of the mirrors over time.

3.3.3.3 Telescope issues

There were a number of problems encountered during commissioning in Pasadena which will be briefly mentioned here for the record. When performing throughput tests, we first discovered an unacceptably low throughput (45%), which we traced to improperly coated mirrors. The problems were brought to the attention of Planewave instruments, who found that their subcontractor coating the mirrors was lying about how well they performed. It should be noted that this inability to tell the truth is a common feature of optics companies. After a recoating downtime of some weeks, the performance improved dramatically. Throughput calculations and measurements described above refer to the recoated mirrors. Another optical problem we discovered was vignetting by the tertiary baffle. This was an easier fix; a new baffle was quickly obtained from Planewave.

Another serious problem we encountered during our tests was a catastrophic failure of the telescope drive motors on August 2013. The controls software applied maximum torque on the motor against one of the internal hard stops, which was not noticed as the telescope was still. The excess

current applied over the course of an evening caused the motors to overheat and melt, jamming the telescope in a particular configuration. The smell of the melted motor was remarkably similar to that of a deceased rodent, which caused us to initially suspect a short circuit in the motor electronics caused by such a creature. We had the motor drive replaced and Planewave modified their control software to immediately kill power to the motors when maximum torque/current is applied for more than 5 seconds; nothing like this has happened since.

3.3.4 Fiber guiding system

3.3.4.1 Introduction

As demonstrated in the previous section the intrinsic tracking of the CDK-700 telescopes is not good enough to efficiently feed a star into our fiber tip and hold it there over the course of an observation, on the order of 10-15 minutes. We designed a closed-loop guider, called the Fiber Acquisition Unit, or FAU, to accomplish this task. Such a system must be robust and efficient to reduce observing overhead and maximize signal. We describe below the optomechanical and controller design and tests to verify the performance of the system.

3.3.4.2 Optomechanical design

The FAU has three accessible optical paths. In the primary path, the telescope beam is fed directly into the fiber at the native $f/6.6$ of the telescope. Separately, there is an optical path which relays a portion of the light to a guiding camera via a pellicle. Finally, there is a set of relay lenses and a corner retroreflector. If the fiber is illuminated from the exit end near the spectrograph, the input end of the fiber tip will be imaged on the guide camera. The corner retroreflector guarantees that misalignments of the optics do not affect the image position; this allows determination of the pixel position on the guider that corresponds to the fiber tip, and hence a setpoint for guiding.

The fibers we use are 50 μm Ceramoptec octagonal fibers, which corresponds to a sky diameter of 2.3 arc seconds. The median seeing at Mt. Hopkins observatory location is about a factor of two better than this, ensuring that minimal flux is lost at the wings of the seeing disk.

3.3.4.3 Controller design

While the CDK-700 telescope has quite good open-loop tracking (accurate to a couple of arc-seconds over a ten minute period), this is not good enough for our purposes, as typical integration times are from 10-20 minutes and require keeping the star centered on a 2.3" diameter fiber tip. Therefore, we have implemented a positional PID-type controller to correct for drifts and other inaccuracies. Specifically, the correction term calculated at each step is given by:

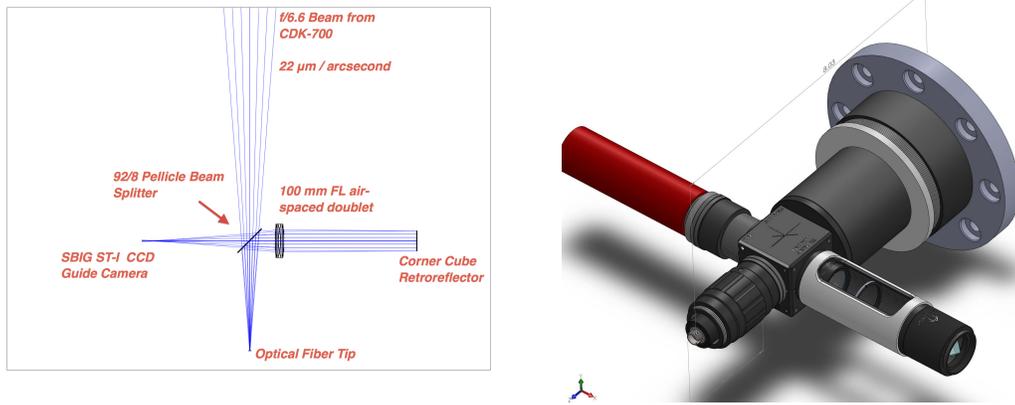


Figure 3.8: Left: Optical layout of the fiber acquisition unit. The primary beam path (top to bottom) couples light from the telescope into the fiber. The second path (top to left) images the star onto a guide camera. The third path (bottom to right to left) images the fiber tip onto the detector. Right: mechanical drawing of the FAU; fiber tip to base is about 8 inches long.

$$\vec{e}(t) = K_p \vec{e}(t) + K_i \int_0^t \vec{e}(\tau) d\tau + K_d \frac{d\vec{e}(t)}{dt} \quad (3.4)$$

where $e(t) = \vec{x} - \vec{x}_0$, \vec{x}_0 being the pixel position of the set point of the loop (in pixels). Good values for the gains K_p , K_i and K_d are determined by experimentation, as are appropriate iteration durations, which may be modified by the relative brightness of the target star and seeing conditions, to reduce centroid noise. Furthermore, derivative smoothing, a deadband option, and an integrator “clamp”—that is, a limit on the maximal size of the K_i term to prevent overshoot—are added to the controller to make it more robust. The controller does not need to send corrections more than once every few seconds, as too short corrections tend to “chase” the seeing. This will be described more in the following sections.

As the controller input is the star position on the camera, and the output is offsets (in arcseconds) to the telescope mount, a calibration of scale and rotation is needed to convert one unit into the other. We therefore have a “one-time” calibration of the rotation and scale, which is performed by making a fixed slew of the telescope (in arcseconds) and determining the angle and distance (in pixels) on the detector. This calibration takes about 15-20 seconds. Note that since the controller is closed loop, the exact rotation and scale parameters do not need to be precisely known, as these will affect the convergence *rate*, not the setpoint.

The controller pointing requirements will be established below, as a natural extension of the seeing jitter requirements, but we briefly mention that the limit of performance can be expected to be about 0.08 arcseconds, the resolution of the encoder tape in the telescope mount.

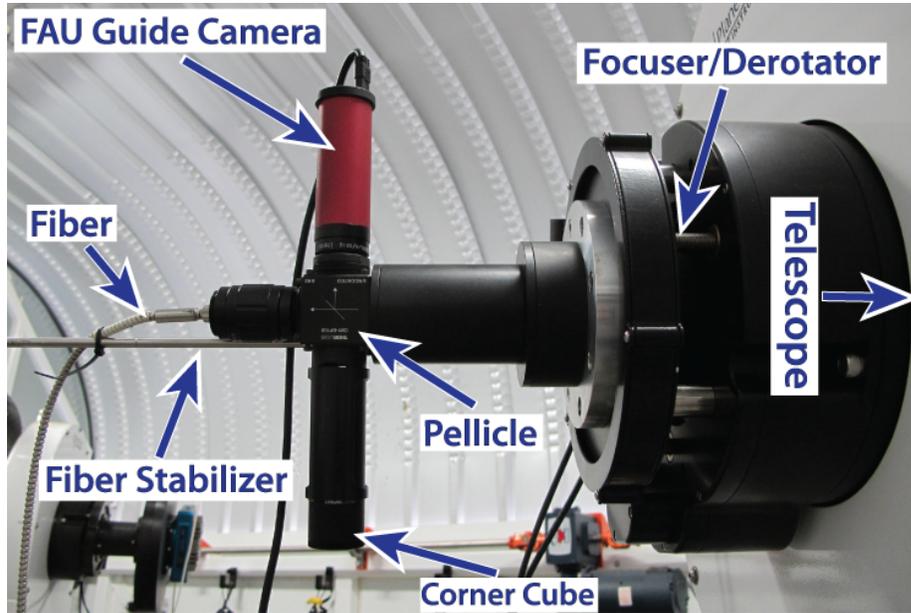


Figure 3.9: Fiber acquisition unit mounted on a CDK-700 telescope in the Pasadena Aqawan. The different subcomponents are labeled, including the fiber used to make the throughput measurements. This figure appears in Swift et al. (2015).

3.3.5 Fiber guiding system verification

3.3.5.1 Expected Performance

The throughput of the fiber system is defined as the ratio of output light from the fiber to incident light on the FAU. The throughput may be expressed in the following way:

$$\text{Throughput} = \text{Pellicle Transmission} * \text{Fiber Transmission} \quad (3.5)$$

$$= P\eta_{R_{in}}\eta_{in}\eta_{thru}\eta_{R_{out}} \quad (3.6)$$

where η_{in} is the input coupling efficiency of the fiber tip, η_{thru} is the throughput of the fiber (which depends on the fiber length and wavelength of light), and $\eta_{R_{out}}$ is the output transmission of the other fiber tip. Implicit in the above equation is the wavelength dependence on all these quantities, and in all that follows, the flux should be interpreted as an integral over the V-band. We will consider these terms one by one.

Pellicle transmission

The pellicle is a Thorlabs BP108 92% / 8% transmission/reflection model. (While these values are strongly polarization dependent, the expected polarization level of the light is low.) Assuming the unpolarized transmission curve, we calculate the transmission over the Johnson V-band and find

an expected transmission of 90.12 %. The pellicle transmission is presented in Figure 3.10, shown below.

$$\frac{\int P_T(\lambda)V(\lambda)d\lambda}{\int V(\lambda)d\lambda} = 90.12\% \quad (3.7)$$

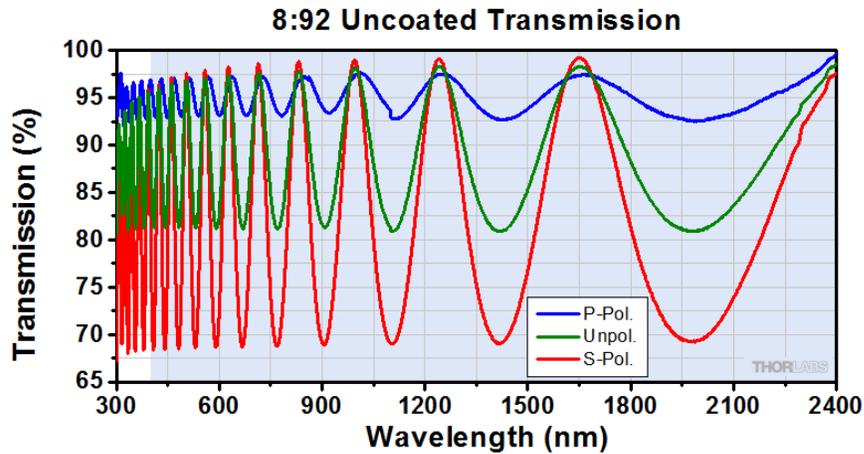


Figure 3.10: Thorlabs BP108 pellicle transmission for different polarization states. This image was taken from the manufacturer website.

Fiber reflection losses: $\eta_{R_{in}}$, $\eta_{R_{out}}$

The input coupling efficiency is dependent on two factors, the reflection from the fiber tip and the coupling into the fiber tip. The reflection loss can be determined from the Fresnel equation(s) for s - and p - polarized light. Both of these expressions can be approximated by

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2 \quad (3.8)$$

which gives a reflection loss of **5.3%** on input assuming an index of refraction of silicon of 1.6. Using the exact Fresnel equations does not change the answer very much (5.30% and 5.34%, respectively) even assuming all the rays are propagating at the most extreme angle of the telescope's $f/\#$. The same result holds on output, as the expression (the simplified one) is the same on interchange from $1 \rightarrow 2$.

Fiber input losses, η_{in}

The input efficiency depends on the coupling integral between the fiber and the image of the

star²:

$$C = \frac{\int_{-\infty}^{\infty} \text{Fiber}(\theta_x, \theta_y) \times \text{Star}(\theta_x, \theta_y) d\theta_x d\theta_y}{\int_{-\infty}^{\infty} \text{Star}(\theta_x, \theta_y) d\theta_x d\theta_y} \quad (3.9)$$

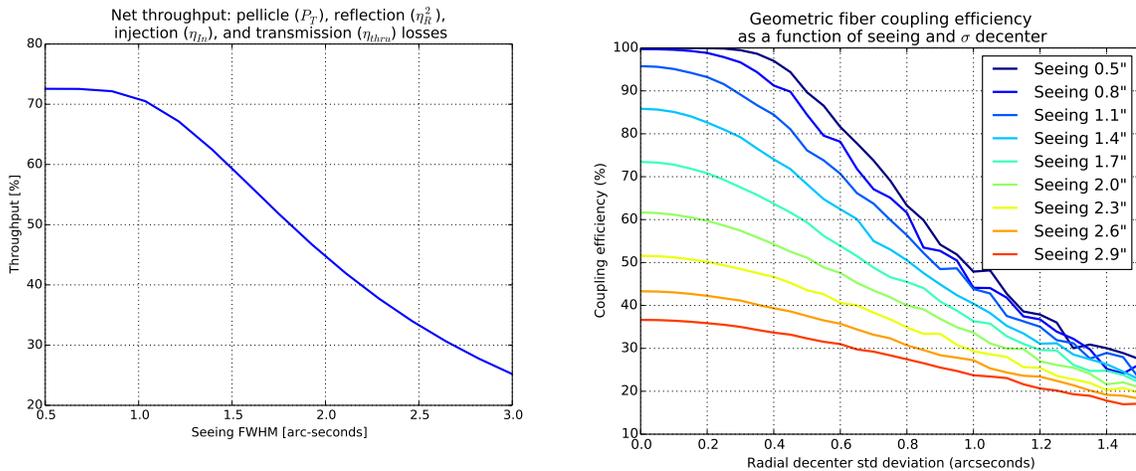


Figure 3.11: Left: Theoretical throughput including pellicle transmission (P_T), reflection (η_R), geometric fiber injection (η_{in}), and fiber transmission (η_{thru}) loss, for a range of seeing values. Right: Numerical simulation of geometric injection efficiency for a range of pointing accuracies. The different lines correspond to different values of seeing.

The full integral needs to be performed in 2 dimensions. The input seeing is assumed to be a symmetric gaussian with a full-width-half-max equal to the seeing; the fiber is assumed to be an octagon with width equal to the fiber image on the sky. Full transmission is assumed inside the octagon; no transmission is assumed outside. It can be seen in Figure 3.11 that losses depend heavily on the seeing, and in Pasadena (where the tests were performed), we are operating well below optimal conditions.

It is also possible to calculate how the geometric coupling efficiency depends on pointing accuracy. To do this, we run a Monte Carlo simulation of the coupling efficiency by treating the centroid of the seeing profile as a random variable with zero mean error and a standard deviation equal to the pointing precision, shown as well in Figure 3.11. It is interesting to note that for good pointing (0.1"-0.2"), the most severe losses occur at *good* but not *excellent* seeing. The reason for this is that at excellent seeing (0-0.5"), the starlight is smaller than and completely contained within the fiber tip and minor pointing errors do not change this. For good seeing (0.8-1.7"), the same pointing errors cause significant portions of the seeing disk to miss the fiber tip. For poor seeing (1.7+"), the seeing disk is already leaking out of the sides of the fiber tip, and small pointing errors only have a modest additional effect. This helps define an upper pointing tolerance; it can be seen that less

²this is a simplification; the full equation involves the mode field diameter of the each fiber mode and the electric field of the input star light

than 5% loss is achieved across any seeing conditions for 0.25 arcseconds of jitter. The best possible performance, as previously stated, is 0.08 arcseconds, corresponding to the encoder resolution.

Summary of expected transmission

A summary of the expected transmission for each optical element is shown below. The weakness in these values is that except for the pellicle, the transmissions were all derived from data sheets or calculated from numerical simulations. Therefore, the uncertainty in the final number are expected to be at the 10-20% level.

Item	Label	Transmission	Source	Comment
Pellicle	P	0.9	Thorlabs data, measured	Integrated over V-band
Input Reflection	$\eta_{R_{in}}$	0.947	Calculated	Fresnel equation, unpolarized
Input Coupling	η_{In}	0.45–0.55	Numerical simulations	Coupling integral, seeing=2–2.5"
Fiber Transmission	η_{thru}	0.90	Optran Data sheet	40 m fiber, calculated over V-band
Output Reflection	$\eta_{R_{out}}$	0.947	Calculated	Fresnel equation, unpolarized
Product		0.34–0.45		

3.3.5.2 Measured Performance

Controller performance To measure the controller performance, we slewed to different stars, then engaged the control system when the star was in the capture range. The controller locked on the target star and sent it to the commanded pixels. During development, we identified good gain values to use through manual tuning. We found that a positional gain value of 0.5 and an integral gain value of about 0.05 gave the best performance, though the integral term did not give enough benefit to justify the risk of having overshoot taking the star off the CCD camera, however unlikely this would be. We measured the centroid of the star over reasonable periods of time, then analyzed this data to determine pointing performance. The same test was repeated with the controls system off to determine the noise added to the system by the controller, through spectral analysis, and repeated for different stars, with stellar magnitudes chosen that correspond well to our targets, with $V < 9$. An example measurement of the star Alkaid is shown below.

Analysis of the data shows that the controller operates near the optimal level, showing an rms pointing precision of about 0.2 arc seconds, dominated by uncertainties in the measurement from seeing variations. Simulations of the required pointing accuracy indicate that very low coupling penalties are incurred for a pointing accuracy of 0.25" rms at any seeing from 0.5 - 2.5", typically from 0-5%. It is thus unlikely that the control system is contributing to any major loss of throughput in the system. Results from pointing tests are shown in Figure 3.12. The most convincing evidence

that the controller is not adding significant noise to the pointing system is from examining the amplitude spectral density of the pointing errors; the error is comparable or lower at all sensed frequencies when the telescope is guiding.

Throughput performance To assess the performance of the fiber acquisition unit, we used two CCD cameras to synchronously measure the reflected starlight and light transmitted out of the end of the fiber. The two cameras are identical SBIG ST-i monochrome models, which were both fully characterized to assess gain and dark current differences. One of the cameras was the actual camera used for guiding; the other was attached to the end of the fiber with an identical V-band filter. A typical output image from the test is shown below in Figure 3.13.

The fundamental quantity measured from these two cameras is ADUs, which can be related to the input flux and other optical constants:

$$\text{Guidecam[ADU]} = \frac{I[\text{photons/s}] \cdot P_R \cdot \text{QE}[e^-/\text{photon}]}{\text{Gain}_{\text{guider}}[e^-/\text{ADU}]} \cdot t_{\text{exp,guider}}[\text{s}] \quad (3.10)$$

And the integrated counts on the fiber are:

$$\text{Fibercam[ADU]} = \frac{I[\text{photons/s}] \cdot P_T \cdot \eta_R \eta_{\text{In}} \eta_{\text{thru}} \eta_R \cdot \text{QE}[e^-/\text{photon}]}{\text{Gain}_{\text{Fibercam}}[e^-/\text{ADU}]} \cdot t_{\text{exp,fiber}}[\text{s}] \quad (3.11)$$

Where P_R , P_T is the pellicle reflectance and transmission. Assuming the QE is the same for both cameras (which is reasonable, as they use the same sensor), we can take the ratio of the two expressions above to find

$$\frac{\text{Fibercam[ADU]}}{\text{Guidecam[ADU]}} = \frac{P_T}{P_R} \frac{\text{Gain}_{\text{guider}}}{\text{Gain}_{\text{Fibercam}}} \frac{t_{\text{exp,fiber}}}{t_{\text{exp,guider}}} \times \eta_R^2 \eta_{\text{In}} \eta_{\text{thru}} \quad (3.12)$$

or more usefully

$$\text{FAU efficiency} = \eta_R^2 \eta_{\text{In}} \eta_{\text{thru}} = \frac{\text{Fibercam[ADU]}}{\text{Guidecam[ADU]}} \frac{P_R}{P_T} \frac{\text{Gain}_{\text{Fibercam}}}{\text{Gain}_{\text{guider}}} \frac{t_{\text{exp,guider}}}{t_{\text{exp,fiber}}} \quad (3.13)$$

$$\text{FAU Transmittance} = P_T \eta_R^2 \eta_{\text{In}} \eta_{\text{thru}} = \frac{\text{Fibercam[ADU]}}{\text{Guidecam[ADU]}} P_R \frac{\text{Gain}_{\text{Fibercam}}}{\text{Gain}_{\text{guider}}} \frac{t_{\text{exp,guider}}}{t_{\text{exp,fiber}}} \quad (3.14)$$

We performed multiple tests in both open and closed loop to assess the performance of the

optical system and control system. The tests were performed over a number of stars of differing magnitudes and spectral types. We found that we typically measure throughputs of 35-45%, in line with the expectations from Figure 3.11 and consistent with independent seeing measurements performed using a wide-field camera on the opposite Nasmyth port of the telescope. An example of such a measurement can be seen in Figure 3.14.

3.3.6 Imaging cameras and filters

3.3.6.1 Camera requirements

The secondary science goal (and initial science results) for the MINERVA array is photometric followup to newly discovered or previously known exoplanets. Specifically for transit photometry, a super-Earth type planet (for example, 3 Earth radii) around a K-dwarf will produce a transit decrement of

$$\Delta m = 2.5 \log \left[1 - \left(\frac{3R_e}{0.8R_\odot} \right)^2 \right] = 1.3 \text{ mmag} \quad (3.15)$$

A “typical” transit time across a stellar surface can be estimated as

$$t_{\text{trans,max}} = \frac{2R_\odot}{v_p} = \frac{R_\odot}{2\pi a_\oplus / T_\oplus} \approx 12 \text{ hours} \quad (3.16)$$

and similarly, the ingress and egress time for the planet to cross the limb of the host star can be estimated as

$$t_{\text{in/egress}} = t_{\text{trans,max}} \frac{R_p}{R_\odot} \approx 20 \text{ minutes} \quad (3.17)$$

The transit time is more of an upper bound as the planet is unlikely to transit across the full diameter of the host star, and most planets that are found in closer-in orbits. The ingress and egress time are more insensitive; ingress and egress times of 10 to 40 minutes capture a very large range of planetary and stellar systems. Therefore, a sensitivity requirement can be stated as needing a photometric precision of about 1 mmag over a timescale less than 20 minutes, as multiple data points are needed to accurately characterize a transit.

For this level of transit precision, experience has shown that the atmosphere sets the most severe constraints on transit precision. For comparison, *Kepler*, a space mission optimized for transiting planets, had a photometric precision of 30 ppm, or equivalently about 30 μmag . From the ground, in good atmospheric seeing conditions, many groups have reported a limit to ground-based photometry of a bit less than 1 mmag; see for example López-Morales (2006), Mann et al. (2011). To reach this limit, differential measurements are absolutely required for ground-based work—that is, reference stars of a similar magnitude need to be measured as close to simultaneously with the target star as possible. This requirement leads to the need for a wide field of view, so that more stars can be

used as simultaneous references. Of course, field of view depends on both the optical performance of the telescope and the detector size. Independently of the size of the usable field, the detector needs to have acceptable dark current and readout noise. Additionally, detector nonuniformities can adversely affect long-term precision, so the need for active guiding to keep stars on the same pixels in measurements spanning many nights is necessary. This requires a high level of pointing control from the telescope, though not as high as in fiber guiding.

A tertiary soft design requirement for MINERVA is a camera suitable for outreach, particularly in the form of undergraduate-level lab courses which typically do not have access to telescopes. This requirement is derived from the demands of certain funding agencies that significant time be used for outreach, often at the 10% level. From this perspective, apart from the precision necessary to do basic eclipsing binary work, transit photometry, multi-filter spectral characterization, etc, it is useful to have a camera with a field of view large enough to image solar system objects, Messier objects, and other diverse celestial objects. For these kinds of observations, narrow-band filters can be very useful as well, as they can be used to both limit the flux from very bright objects and to reveal phenomena like H II regions in galaxies, which are frequently found in undergraduate astronomy curricula.

One advantage of MINERVA is that the telescopes are situated further apart than the typical coherence length of the atmosphere. This means that transit photometry datasets are expected to be weakly correlated, if at all, when measured simultaneously with the same filter from multiple telescopes. On a similar note, when measured simultaneously with different filters on each telescope, interesting constraints on the planetary atmosphere can be made.

3.3.6.2 Photometry optics and detector

The first camera we purchased (Apogee U230) had completely unacceptable transient noise properties which resulted in low photometric precision and an inability to properly flat field. Sending it back to the manufacturer, resulted in months long delays and radio silence, and eventually a new camera with a host of other problems. We gave up and moved away from Apogee in the end.

The model we eventually chose for the standard photometric camera was the Andor iKon-L with the BV chip for three out of four telescopes and the BEX2-DD chip for the other. The advantages of this camera are high efficiency, high dynamic range, and an efficient thermoelectric cooler without the need for liquid coolants, which can be less reliable. The BV chips have a quantum efficiency of greater than 95% from 500 to 650 nanometers; the BEX2-DD has a quantum efficiency that is lower but covers a much greater wavelength range, higher than 80% from 380 to 900 nanometers. Camera parameters are presented in Figure 3.3.

Our camera setup is identical for each telescope except for the choices of filters. Each camera is mounted via custom adapter to an Apogee AFW50-7S filter wheel, which have seven accessible slots.

(We also tested a dual-ten slot filter wheel on one of telescopes, but this proved to be unreliable with frequent wheel slip due to a fundamentally poor mechanical design.) The filter wheels are connected to an off-axis guider optic which sends light to an ST-i guide camera with a field of view of 4 x 6 arcminutes. (The guide camera is the same model as used in the fiber acquisition unit.) The entire unit is connected to the telescope focuser and derotator. An image of the setup can be seen in Figure 3.15. The current filters are standard Sloan g', r', i', z' , which are present on every filter wheel, as well as combinations of $U, B, V, R, I, H \alpha, [S II], [OIII]$. Typically, a slot is left open for luminance measurements. The filters were provided by Astrodon Photometrics.

Active Pixels	2048 x 2048
Sensor Size	27.6 x 27.6 mm
Pixel Size	13.5 μm x 13.5 μm
Well depth (typical)	100,000 e-
Max. Readout Rate	5 MHz
Read Noise	2.9 e- (min)
Maximum Cooling	-100 C
Frame Rate	0.95 fps(max)
Coating	BV (3x), BEX2-DD(1x)
Dark current (-70 C)	0.00013 (BV), 0.006 (BEX2-DD)
Usable field	21' x 21'
Johnson filters	U, B, V, R, I
Sloan filters	g', r', i', z'
Narrow filters	$H \alpha, [S II], [OIII]$

Table 3.3: Andor iKon-L parameters and filters

3.3.6.3 Performance Validation

We measured the single-telescope photometric precision at the MINERVA test site in Pasadena. The typical test procedure involved heavily defocusing the field to reduce the effects of interpixel variations, Poisson noise, and shutter effects which are annoying to accurately cancel with very short exposures. This slightly complicates the pointing calibration, as the guide camera must be operated in focus while the science camera is out of focus, and without an automatic adjustment a one-time manual adjustment is required. Exposures are taken every ~ 10 seconds.

One of our target stars was 16 Cyg, a visual binary with a separation of about $30''$. We measured this star for about one hour (about 200 images) while guiding on an off-axis star 20 arcminutes away. The guider was not functioning perfectly as there were some unwanted drifts over the course of the observation of a few pixels. Following the standard CCD reductions of bias, dark, and flatfield correction, we extracted aperture photometry using a simple annular method found in the program AstroImageJ, using a 30 pixel aperture for the defocused spot and 90 to 100 pixels for the sky annulus. There were five appropriate reference stars in the field for lightcurve detrending and systematics control.

The results of the above analysis showed a single-shot photometric precision of 2.7 mmags, with a binned precision of less than 1 mmag on 3-5 minute timescales, see Figures 3.16 and 3.17. This is well within the required precision characterized in the previous section. Given the unpleasant light conditions at the MINERVA test site in Pasadena (e.g. passing cars, nearby streetlights, etc), the underperformance of the guider in this particular test, and the simple analysis involved, it is encouraging to see that we achieve the required level of precision already. This result strongly suggests we will be able to achieve this level of photometric performance routinely at the observatory site.

3.3.7 Site choice and commissioning

Many locations were considered for the siting of MINERVA, and there were a number of factors that influenced the final decision. The sites that were considered were Palomar Observatory (Palomar Mountain, CA), Whipple observatory (Amado, AZ), McDonald Observatory (Jeff Davis County, TX), Mt. Wilson Observatory (Los Angeles, CA), and San Pedro Martír observatory (Baja California, Mexico). The criteria for evaluation were based on weather conditions and atmospheric seeing, existing infrastructure, commitment of support from the governing institutions, and annual financial commitment required to maintain and lease the site.

After visiting all the sites, we decided on Whipple Observatory as the best choice overall. Within Whipple, the site that we chose was the now-decommissioned site of the VERITAS 10m gamma ray Cherenkov telescope, located at an elevation of 7816 ft above sea level. Advantages to this site include the full use of the VERITAS control building, a low horizon (between 15 and 20 degrees), existing auxiliary facilities like lightning rods, ground conduits, access roads, etc. Furthermore, there are several sources of accurate, up-to-date weather and seeing conditions available from the MEarth Irwin et al. (2009) and HATNet projects Bakos et al. (2002), located a few hundred feet away. The data suggest about 90 days will be lost per year due to weather, occurring primarily during the Arizona monsoon. Median seeing for the remaining nights is expected to be about 1.1 arc-seconds, well within the tolerances proscribed in Section 3.3.5.1.

Despite the significant infrastructure present, there were many modifications of the VERITAS site that needed to be completed to make it appropriate for MINERVA. The first involved the removal of the 10m Cherenkov telescope, the installation of appropriate fiber, data, and power systems, including an uninterruptible power supply. These modifications were generally completed in the commissioning and second round of work in 2014 and early 2015. An important further item was the conversion of one of the existing rooms of the VERITAS control building into an temperature and humidity controlled “clean” room to house the spectrograph. This work commenced in late 2015 and has been mostly completed.

3.4 Conclusions and future work

We have presented the scientific motivation, design, and verification of the multi-aperture observatory MINERVA, built to perform spectroscopic and photometric discovery and characterization of planets orbiting nearby stars. We presented the expected survey yield, built off of simulations from the previous chapter combined with planet occurrence statistics informed by the *Kepler* space telescope. To achieve the desired survey yield, the use of a distributed aperture array was shown to have distinct advantages in terms of simplicity, stability, and cost. In such an arrangement, a much smaller spectrograph may be used with an equivalently high resolution, at the modest cost of detector space. We also discussed the choices of enclosure and telescope, and the testing and verification of these critical components, including the telescope throughput and pointing precision, at our test site in Pasadena.

We discussed the design, testing, and performance of a fiber coupling system for the telescopes, which performs the important role of getting starlight to the spectrograph. We analyzed the expected performance in different seeing conditions, and found that excellent coupling could be achieved for pointing jitter less than 0.25 arcseconds. After designing and testing the controller, we were able to achieve a pointing jitter of 0.2 arcseconds, within the optimal design parameters of the system. With this level of performance, we were able to measure throughputs consistent with the expectations from the numerical simulations, and exceeding the level required for efficient coupling to the spectrograph. Further tests will be done at the final location for the observatory to confirm consistent performance in high quality astronomical conditions not achievable from our test site in Pasadena. We also presented the other focal plane instrument, the wide field camera and filter set, for performing transit photometry and outreach observations. We demonstrated the required precision for detecting Super-Earth type transits from our test site in Pasadena, retiring the risk for that portion of the experiment.

After a search of several sites in North America, MINERVA was selected to be installed at Mt Hopkins observatory in Arizona, which had the combination of excellent astronomical observing conditions, existing infrastructure, and a large level of institutional support. MINERVA has now been commissioned at Mt Wilson, and the critical verification stage of the spectrograph is commencing. Unfortunately, the spectrograph commissioning and testing was not able to be completed by the time of this thesis completion, and thus must be omitted.

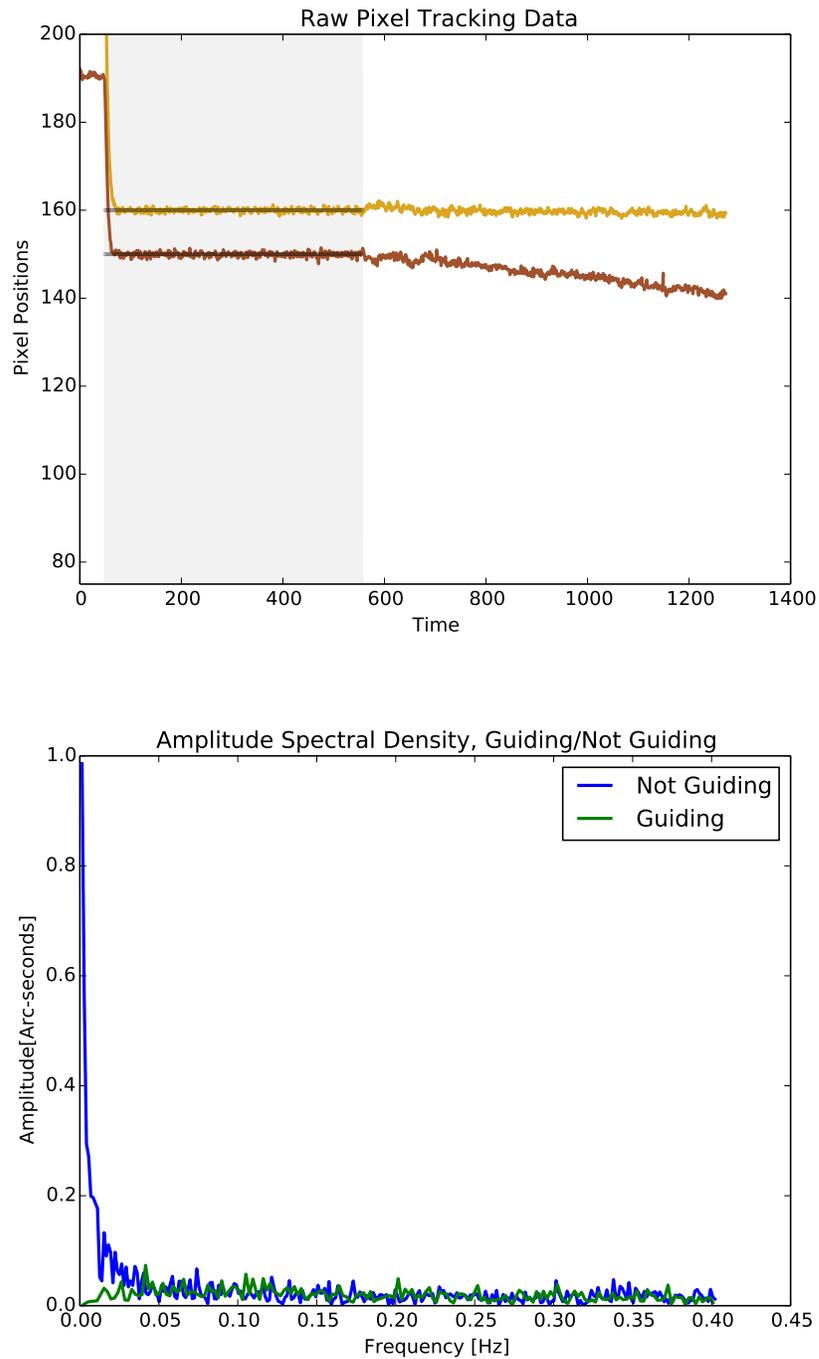


Figure 3.12: (Left) Raw pixel tracking data with the closed loop guiding on (shaded region) and off. The two curves correspond to the x and y pixel positions on the guider CCD. The pitch is 0.33 arc seconds per pixel. (Right) The amplitude spectral density of the guider error is smaller at essentially all frequencies when guiding, especially low frequencies corresponding to long-term drifts. The intersection of the green (guiding) curve with 0 at 0 Hz indicates there is no systematic error.

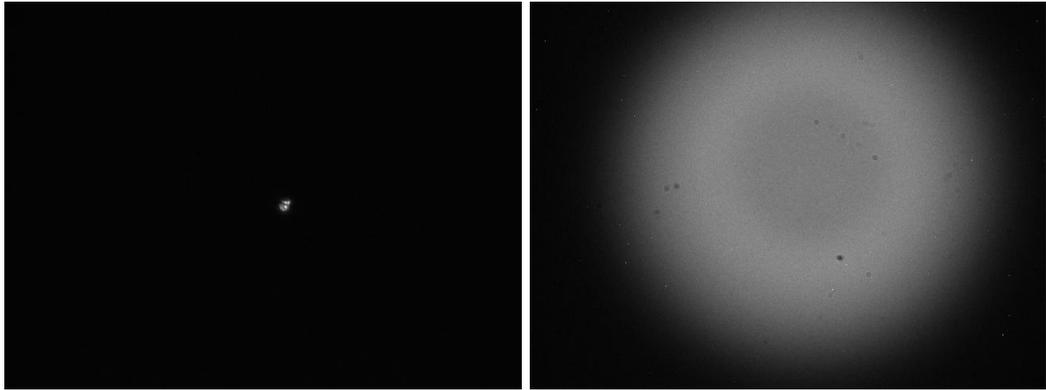


Figure 3.13: The focused image of a star on the guide camera, and the diverging output beam of the fiber. The region of relatively reduced flux in the center of the fiber image is due to the effects of focal ratio degradation.

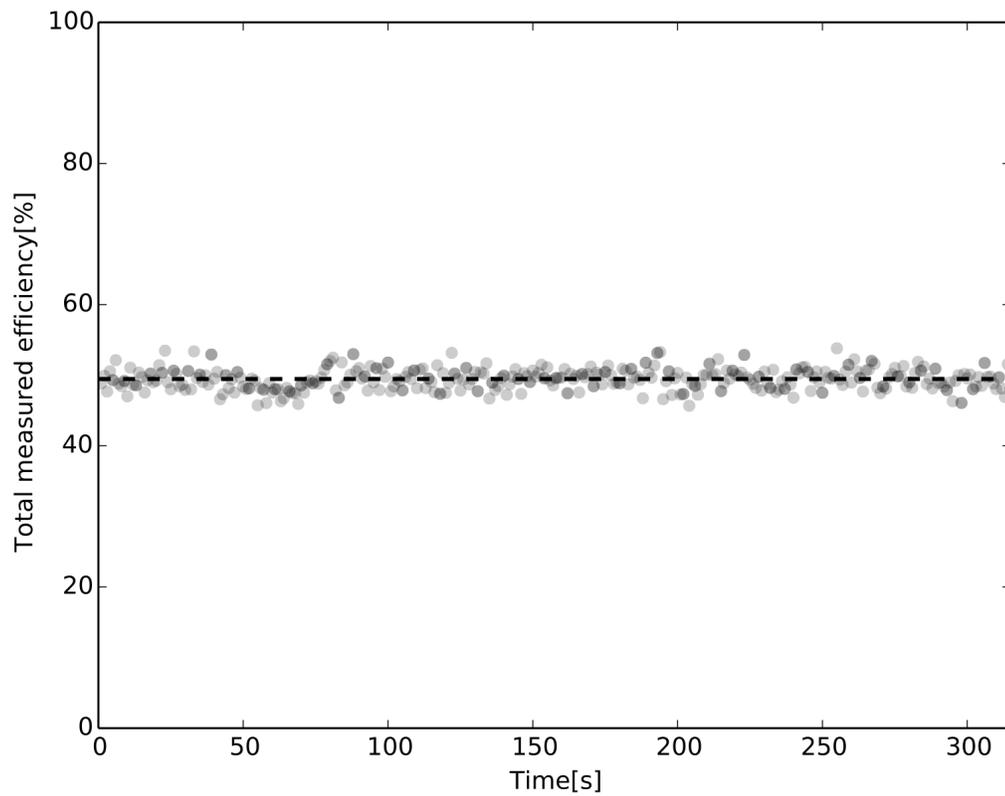


Figure 3.14: A throughput measurement in typical Pasadena seeing conditions of $\sim 2''$. The efficiency is about 50% (45% throughput), in line with the expectations of Figure 3.11. The statistical deviation about the mean is 1.5% (absolute), which is smaller than our systematic error.

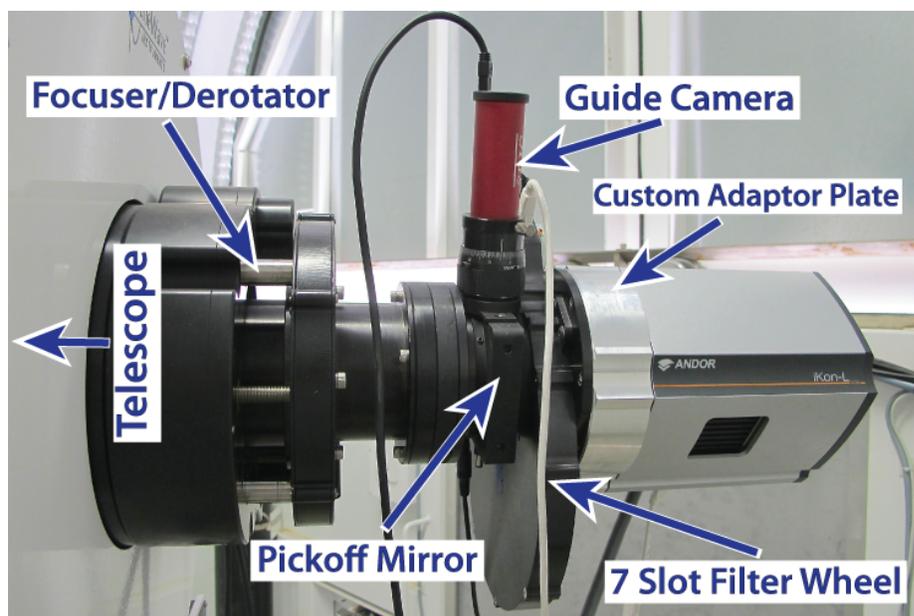


Figure 3.15: Photometric camera optical setup, with major subcomponents labeled. This figure appears in Swift et al. (2015).

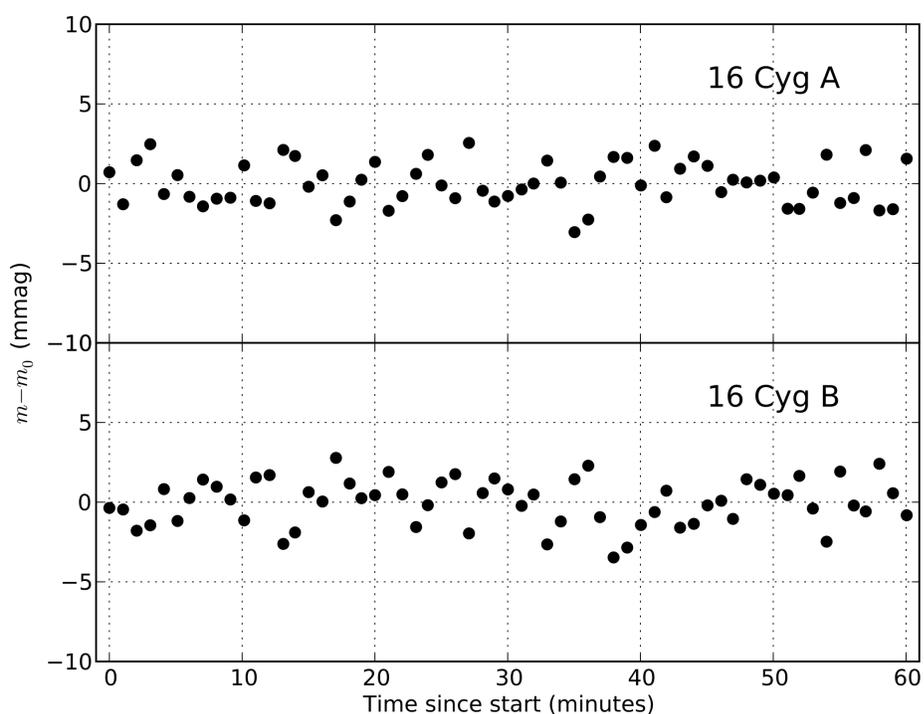


Figure 3.16: Photometric time series of 16 Cygnus A and B; data binned to one-minute intervals. The photometry is quite stable with little evidence of residual correlated noise. This figure appears in Swift et al. (2015).

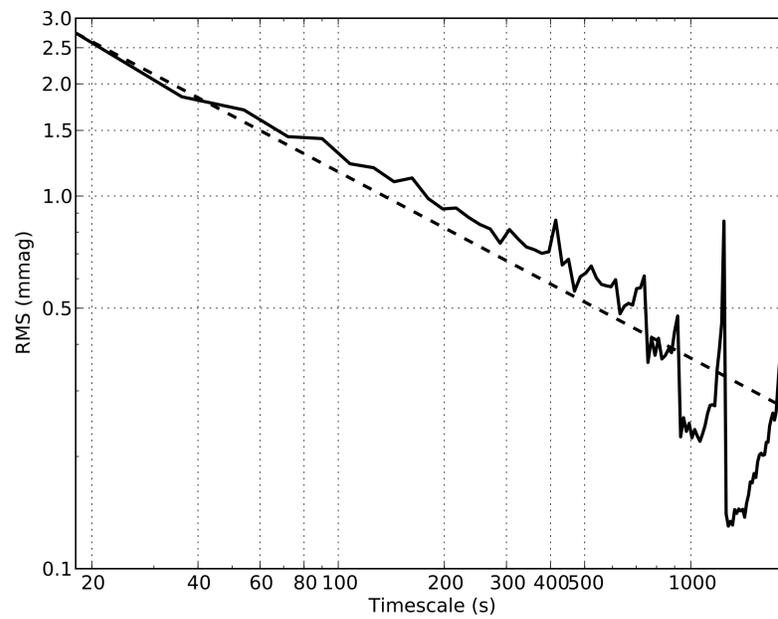


Figure 3.17: Allan deviation of the photometric light curve of 16 Cygnus A, showing a precision of 1 mmag reached in about 3 minutes, almost at the photon noise limit. This figure appears in Swift et al. (2015).

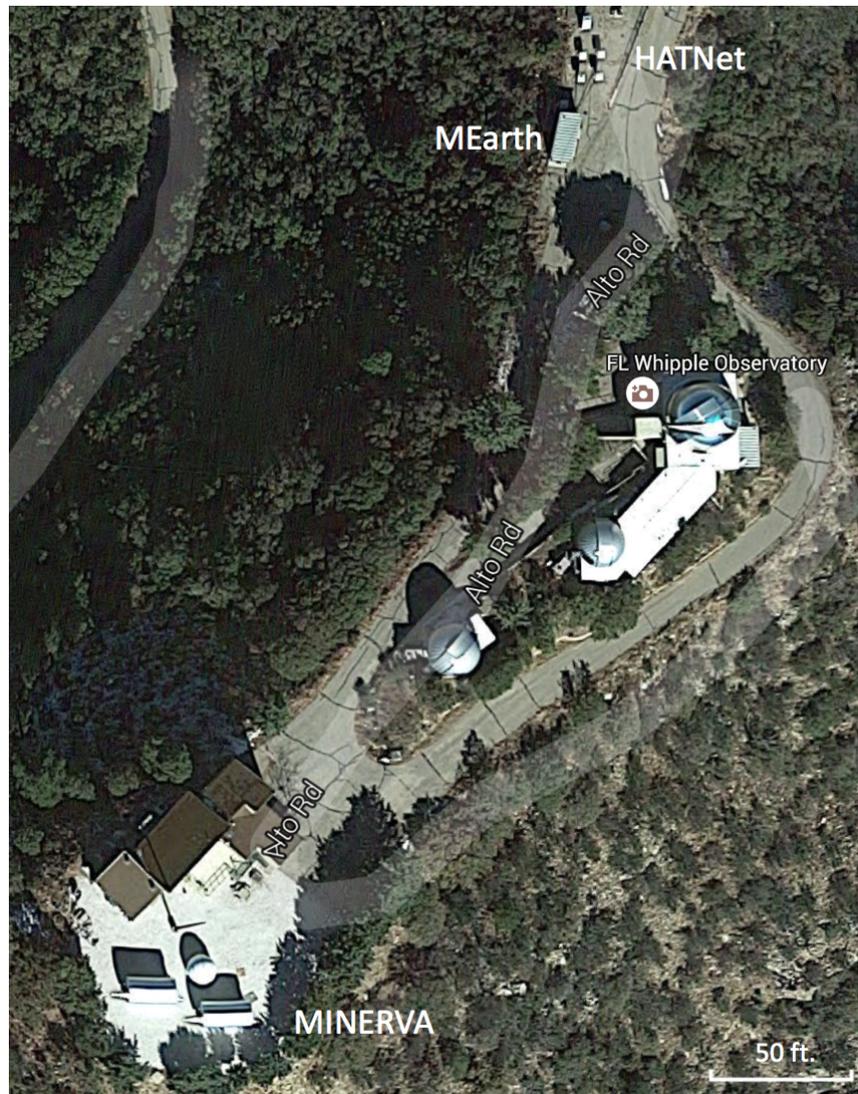


Figure 3.18: The final location of MINERVA, at Mt Hopkins, Arizona. The building to the immediate North of the two Aqawans is the now repurposed VERITAS observing station; the VERITAS dish has already been removed. The HATNet and MEarth observatories are labeled to the North. A scalebar is shown to the bottom right. The geographic coordinates of this location are $31^{\circ} 40' 49.1''\text{N}$, $110^{\circ} 52' 44.6''\text{W}$. This image was generated using Google Maps.



Figure 3.19: MINERVA installed at Mt. Hopkins. This photo was taken right after the four telescopes were installed. The clamshell dome in the bottom right is a sister project called MINERVA-RED (PI: Cullen Blake)

Chapter 4

Part II–SDC: a multistage coronagraphic platform for Palomar Observatory

Abstract

We present a new instrument, the “Stellar Double Coronagraph” (SDC), a flexible coronagraphic platform. Designed for Palomar Observatory’s 200” Hale telescope, its two focal and pupil planes allow for a number of different observing configurations, including multiple vortex coronagraphs in series for improved contrast at small angles. We describe the motivation, design, observing modes, wavefront control approaches, data reduction pipeline, and early science results. We also discuss future directions for the instrument.

4.1 Introduction and Motivation

High contrast imaging is a rapidly evolving field, offering one of the most promising ways of obtaining spectra of extrasolar planets. Direct imaging of planets is challenging for two main reasons: first, stars are brighter than their orbiting planets by many orders of magnitude, and second, planets and their host stars are close to each other in angular separation, often only a few telescope resolution elements. In order to practically overcome these issues, at least three things are needed: an extreme adaptive optics system (ExAO) to reduce scattered starlight, a coronagraph to attenuate the diffraction pattern and allow observations very close to the host star, and post-processing image analysis to reduce quasi-static speckles due to remaining non-common path spatial and temporal optical imperfections. The large amount of scientific and technical effort towards improving these three areas has led to rapid developments in the field. While direct imaging has only discovered a handful of planets so far, there are recently commissioned instruments such as GPI (Macintosh et al., 2014), SPHERE (Zurlo et al., 2014), and SCEXAO (Jovanovic et al., 2015a) that will expand the census of imaged planets around nearby stars.

When evaluating new coronagraphic systems and technologies, there are a few useful performance metrics to consider, including *contrast*, *inner working angle*, and *throughput*. For an unresolved, single star, contrast is the average brightness at a particular area of the image divided by the brightness of the star. Inner working angle is commonly defined as the angular separation where the flux of a putative companion is attenuated by the coronagraph to 50% of what it would be arbitrarily far away. The importance of inner working angle can be appreciated by noting that at a star 10 pc away, a planet at 1 AU will only be separated by 100 mas, just above one diffraction beamwidth for a 5 meter telescope operating at $2 \mu\text{m}$. Also, the areal discovery space around a star scales as the inverse square of the inner working angle, and the number of accessible stars in a survey complete down to an orbital radius goes as the inverse *cube* of the inner working angle (that is, halving the inner working angle allows one to observe a planet twice as far away at the same star-planet separation). Throughput refers to the fraction of planet light making it through the coronagraphic system as a function of angular separation. Throughput is also very important, as accumulating enough companion photons even at far separations from a star can be a challenge with faint planetary sources. Note that when comparing different coronagraph designs, many of these metrics are stated in units of diffraction beamwidths (λ/D) rather than absolute terms, since the merits of a design do not depend on the particular dish diameter.

One of the most promising technologies in the field is the vector vortex coronagraph (Mawet et al., 2005). Using a phase mask known as an optical vortex in an intermediate focal plane, for a clear aperture the coronagraph can provide very high contrast at an inner working angle of $0.9\lambda/D$, near the theoretical limit set by diffraction, with nearly 100% throughput at larger angles. Despite

its advantages, a vector vortex suffers the same fate that all coronagraphs do when operating behind a telescope with a secondary mirror, sharply reduced contrast and degraded inner working angle due to diffraction from the secondary and any support struts in the telescope pupil.

There are certain ways to get around the limitations set by the secondary mirror and assorted support spiders. One way is to use a mask to re-image only a clear, unobscured pupil subaperture. This leads to severely reduced throughput and resolution; regardless, this method holds the record for closest directly imaged planet in units of diffraction beamwidths for a conventional coronagraph (Serabyn et al., 2010), though interferometric methods relying on non-redundant masking have done better (Kraus & Ireland, 2012). Two other proposed designs also allow for improved contrast at low inner working angles when dealing with centrally obscured apertures. The first is based on a specially made pupil-plane apodizer upstream of the vortex (Mawet et al., 2013), which has the effect of redistributing the diffracted starlight in a way that it can be completely blocked in a post-coronagraphic Lyot plane, also at the cost of reduced throughput¹; the second is by introducing another vortex in series with the first, which moves the light diffracted by the secondary to the center of the pupil, where it can be blocked (Mawet et al., 2011b). The SDC was built by the Jet Propulsion Lab² to allow for flexible development, testing, and on-sky evaluation and useful observing with these designs, as well as other ideas in wavefront control and coronagraphy, as will be elaborated below.

The layout of this paper is as follows. The next section of this paper discusses the physics and optics behind the vector vortex coronagraph and its various observing modes. Section 4.3 presents the optomechanical design of the instrument. Section 4.4 discusses the observing approach and data reduction pipelines for the instrument. Laboratory measurements and predicted performance are presented in Section 7.4. First on-sky results are presented in Section 4.6.

4.2 Background

The initial coronagraphic elements in the SDC are K-s band optical vortices, and most of the observing modes implemented in the instrument thus far involve their use in some way. In this section, we will summarize the principles behind optical vortices and their implementation as focal-plane coronagraphic elements. We also provide a theoretical description for the different instrumental observing modes. Readers may consult the cited papers below for more detailed discussions. We expect the SDC to evolve with time, adding functionalities and improving performance. Thus, what is presented here is a snapshot of the current state of the instrument.

¹modern designs using pupil-remapping apodizers can recover most of this lost throughput, but are not used in this work

²Principal Investigator: Eugene Serabyn

4.2.1 Optical Vortices

An optical vortex is a device that creates an azimuthal phase ramp of $e^{il\theta}$, $l = 1, 2, 3, \dots$ in light passing through it, where the integer l is referred to as the “topological charge”. A simple way to picture it is as a piece of glass whose thickness increases like a spiral staircase. As light passes through the spiral glass plate, it accumulates a different phase delay depending on its azimuthal position. For a particular choice of glass index and thickness profile, a smoothly increasing phase delay of 0 to 4π can be constructed; this is a “charge-2 vortex”.

The spiral glass plate, also known as a “scalar vortex,” is not the actual optic used in the SDC, as the machining tolerances are difficult to achieve, and the chromaticity of glass limits the usable bandwidth. There are three methods more commonly used to create high quality optical vortices, subwavelength gratings (Delacroix et al., 2014), photonic crystals (Murakami et al., 2013) and liquid crystal polymers. All these methods generate the phase delays with a spatially varying half-wave plate, the fast axis rotates twice as fast as the azimuthal coordinate. These are “vector vortices” since they use polarization to generate a phase delay through the Pancharatnam-Berry phase, rather than creating it through thickness variations in optics. The liquid crystal polymer method, the one we use, is discussed in Mawet et al. (2009).

4.2.2 Single vortex coronagraph

The principle behind the vortex coronagraph is the following. Light brought to a focus at the center of an optical vortex, experiences a total phase discontinuity, which creates a dark hole. If the vortex is placed at the focus of a telescope, the dark hole will propagate and expand at the center of the optical axis. The result is that at the following pupil, the dark hole will encompass the entire pupil— all the starlight will lie outside the aperture. A conventional pupil stop can then block this light. Any planetary companions that come to a focus *off* the center of the vortex will not experience the phase discontinuity, and will thus propagate almost normally. Figure 4.1 shows a schematic of this optical system.

It is useful to consider the vortex coronagraph in the optical Fraunhofer approximation, where there is a Fourier transform relationship between subsequent image and pupil planes. The input pupil is Fourier transformed to the image plane, then multiplied by the vortex function ($e^{il\theta}$, where l is the topological charge), the *product* of which is Fourier transformed to the following pupil plane. In this view, the “perfect” performance of the unobscured vortex coronagraph is simply a statement that the Fourier transform of the product of an Airy disk and $e^{il\theta}$ (for even l) has no energy interior to the input pupil radius. In the case of a telescope with an on-axis secondary mirror (Mawet et al., 2011b), the electric field in the pupil following the vortex is

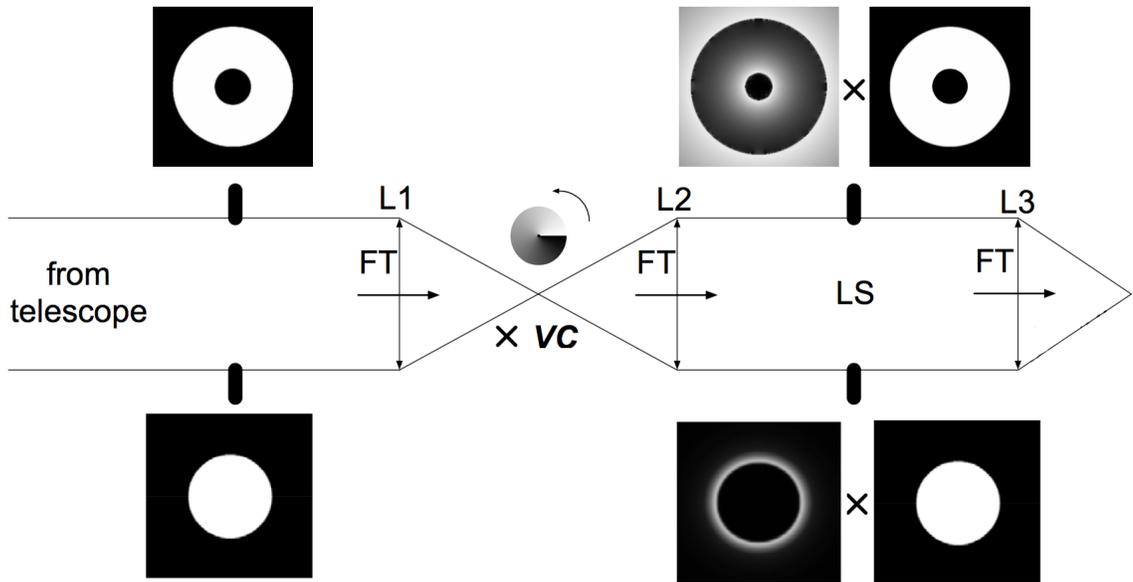


Figure 4.1: The basic vortex coronagraph. Light from a clear telescope pupil (bottom row) is focused on the vortex by the telescope optics. At the following pupil, the starlight is moved to outside the pupil, where it is blocked by a Lyot stop. The dark pupil is then reimaged by the camera, with the starlight removed. For a centrally obscured aperture (top row), there is a residual halo of starlight left in the pupil.

$$E_L(r) = \begin{cases} 0 & r < r_0 \\ -e^{i2\psi} \left(\frac{r_0}{r}\right)^2, & r_0 < r < R \\ e^{i2\psi} \left[\left(\frac{R}{r}\right)^2 - \left(\frac{r_0}{r}\right)^2\right] & r > R \end{cases}$$

where r specifies the radial coordinate, ψ specifies the angular coordinate, R is the outer pupil radius, and r_0 is the radius of the secondary mirror. In the case of $r_0 = 0$, the unobscured case, the cancellation is perfect ($E_L = 0$) in the regime $0 < r < R$.

The perfect cancellation is only true for an Airy function input, not arbitrary circularly symmetric distributions of light. This is important because in cases of obscured apertures, the vortex will leak at a net level of $(r_0/R)^2$; see Figure 4.1. This may be improved somewhat using an oversized central mask in the Lyot plane, but at the cost of degraded throughput. It should be noted that the issue of pupil obscurations is a problem that inhibits the performance of all coronagraphic designs, and is not unique to the vortex coronagraph.

4.2.3 Ring-apodized vortex coronagraph

One solution to the problem of obscured apertures can be provided by actually apodizing the *input* pupil with a partially transmissive ring (Mawet et al., 2013); see Figure 4.2. In the case of a

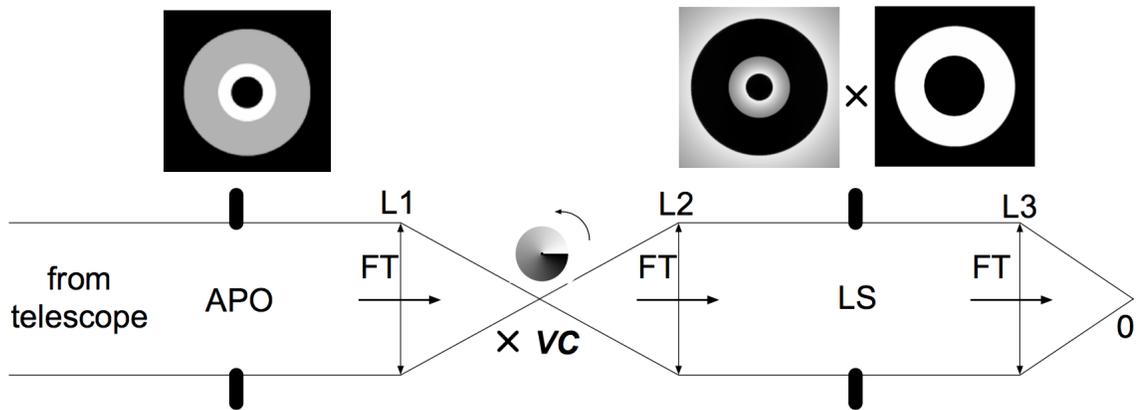


Figure 4.2: The ring-apodized vortex coronagraph apodizes the input pupil of the telescope, with the result being that in the pupil plane after the vortex, all the energy inside the pupil is localized to a ring (second square panel, top). A pupil stop can effectively block this light and thus have total starlight cancellation in principle.

transmissive ring apodizer, the electric field in the pupil following the vortex is

$$E_L(r) = \begin{cases} 0 & r < r_0 \\ -\left(\frac{r_0}{r}\right)^2, & r_0 < r < r_1 \\ (1-t)\left(\frac{r_1}{r}\right)^2 - \left(\frac{r_0}{r}\right)^2 & r_1 < r < R \end{cases}$$

where r is the radial coordinate; r_0 , r_1 , and R are the radius of the secondary, ring apodizer, and pupil, respectively, and t is the transmissivity of the apodizer. By building the apodizer such that $t = 1 - (r_0/r_1)^2$, it is possible to create a region of total cancellation in the following pupil plane from r_1 to R . A Lyot stop can then remove the remaining light, interior to r_1 . However, this comes at a cost of reduced throughput, as the size of the dark region is only a fraction of the pupil. For example, with a 35% central obscuration radius, the maximal throughput of the optical system can only be about 33%. Regardless, this approach can solve the problem of the secondary obscuration.

4.2.4 Multistage vortex coronagraph

Another approach to mitigating the problem of secondary obscurations is by having a second vortex in series with the first (Mawet et al., 2011b). The residual starlight from the first vortex is “folded back” by the second vortex to behind the secondary obscuration in the pupil plane, where it can then be blocked. In this case, the electric field after the final Lyot stop is

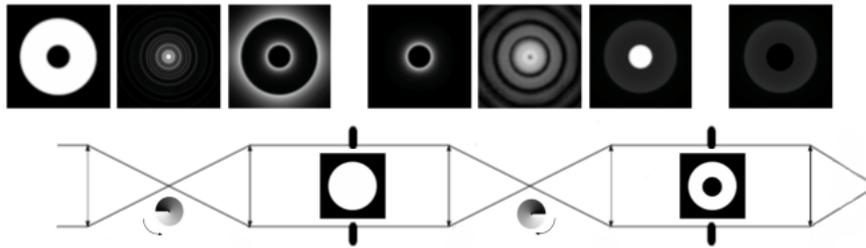


Figure 4.3: Schematic of the dual-vortex coronagraph. The first vortex leaves a residual halo of light (4th panel from left) which is moved behind the pupil by the second vortex (2nd panel from right).

$$E_L(r) = \begin{cases} 0 & r < r_0 \\ \left(\frac{r_0}{R}\right)^2, & r_0 < r < R \\ 0 & r > R \end{cases}$$

where r is the radial coordinate, r_0 is the radius of the secondary obscuration, and R is the radius of the pupil. The electric field is not completely nulled, but reduced in amplitude significantly, with the leakage being reduced to $(r_0/R)^4$. In addition to the imperfect cancellation, another disadvantage of this method is the precise alignment needed between the two vortices. The significant advantage of the multistage coronagraph, however, is the high throughput.

4.3 Design of the Stellar Double Coronagraph

The SDC was built to demonstrate and use these and other coronagraphic approaches on-sky. It operates between the P3K adaptive optics system and the near-IR imager PHARO on the Hale 200" telescope; see Figure 4.4. The dimensions of the instrument are 25.4cm x 45.7cm x 91.4cm, and the weight is about 70 kg.

The SDC accepts the f/15.7 output beam from the adaptive optics system. An input fold mirror (Figure 7.4, top right) at a compound angle steers the beam towards the coronagraph. Before the coronagraph, an infrared dichroic splits off J-band light towards a quad cell tracker (see next paragraph). After passing through the first vortex phase mask, the light hits an off-axis paraboloid (Figure 7.4, top left), where it is collimated and then bent towards the first Lyot stop at the first pupil plane (green rectangular base). The optics in the Lyot plane are designed to be mounted at a 5 degree angle, so that the reflected (rejected) light returns at a different angle to the previous fold mirror. An internal camera with a flip lens (Figure 7.4, bottom left) uses this broadband infrared light to image the pupil (or object), assisting with initial alignment. The light passing through the

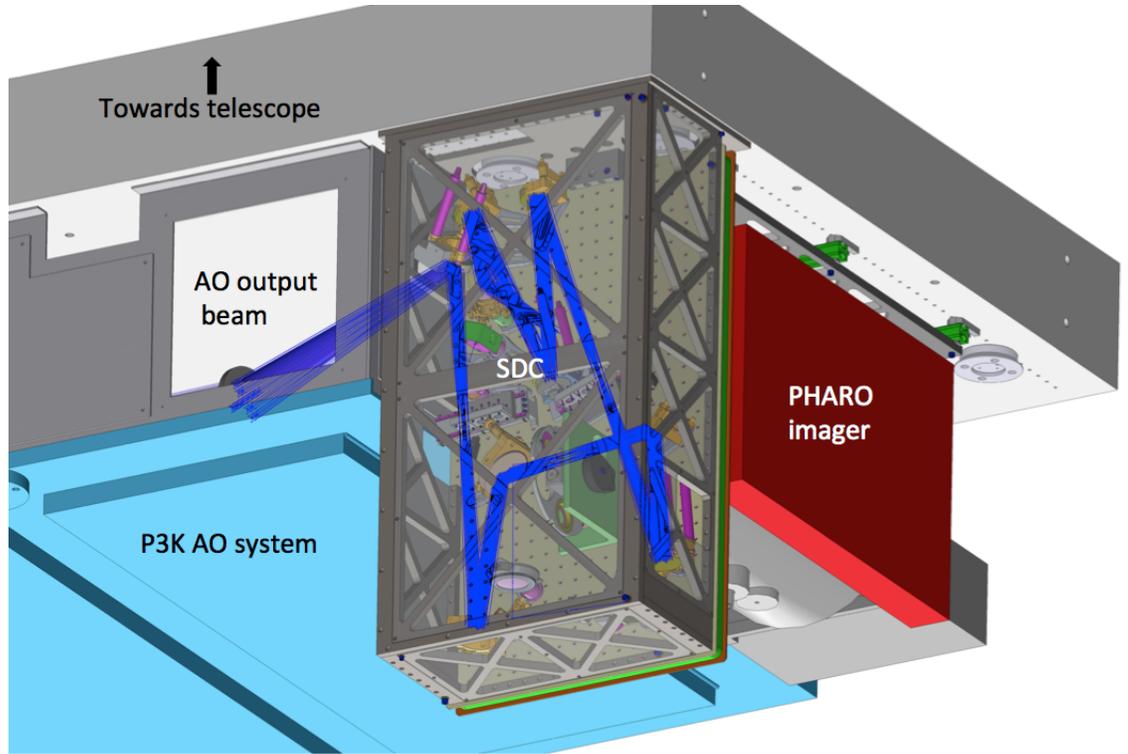


Figure 4.4: The SDC mounts between the P3K adaptive optics system (blue rectangle, left) and infrared imager PHARO (partially visible behind its red electronics box). The imager, coronagraph and adaptive optics system all attach to Cassegrain port of the telescope.

Lyot stop is folded to a second off-axis paraboloid, which focuses it onto the second vortex phase mask, mounted on an identical linear slide. The second off-axis paraboloid controls the focus and vertical image position on the second vortex mask, and the horizontal position is controlled by the slide itself—this avoids changing the off-axis angle of the paraboloid. After the second vortex, another off-axis paraboloid collimates the beam to a second Lyot plane, where a second (reflective) Lyot stop may be installed. The collimated beam hits a final off-axis paraboloid (Figure 7.4, top right), and the converging beam is bent towards the infrared imager PHARO by an output fold mirror. A summary of the degrees of freedom and controls is presented in Table 4.2 of the Appendix.

A very important source of error in a coronagraphic system is tip/tilt “leakage” error, i.e., misalignment of the input image with the coronagraph. In a Cassegrain instrument, this is especially challenging as the differential flexure between the coronagraph and wavefront sensor of the adaptive optics system means that the star will slowly drift off the coronagraph through the course of an observation. Our solution to this problem is a custom quad-cell infrared tracker. A dichroic splitter sends J-band ($1.1\text{--}1.4\ \mu\text{m}$) light to the infrared tracker (green box at top of Figure 7.4), and lets the science wavelengths (H and K band) pass to the coronagraph. The tracker is positioned to be very close to the coronagraphic optical element, so differential flexure is minimal. The output

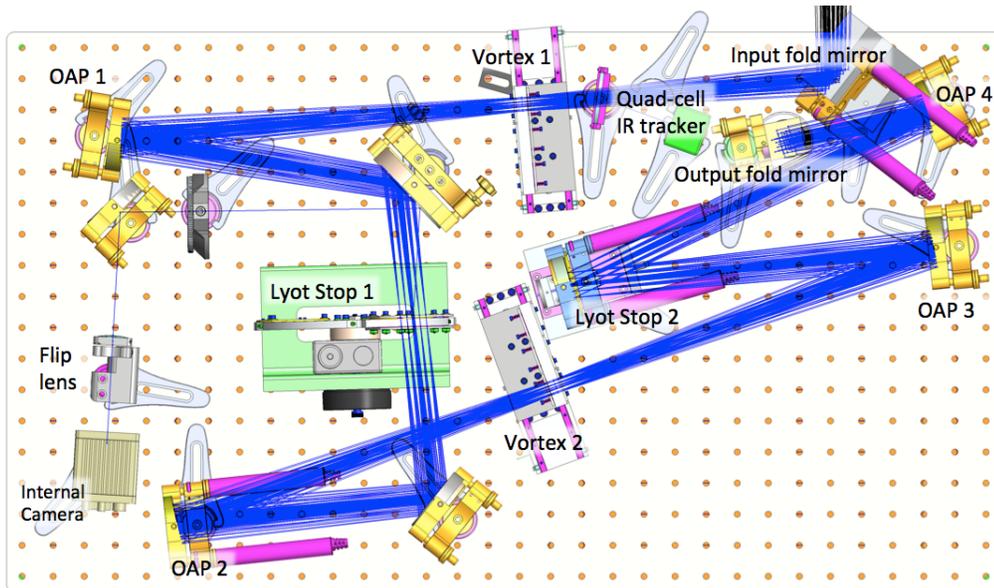


Figure 4.5: The optomechanical layout of the SDC; refer to the text for a more detailed description. Following the input beam from the top right of the figure: first fold mirror, dichroic beam splitter, linear coronagraphic slide, off-axis paraboloid, fold mirror, Lyot plane, fold mirror, off-axis paraboloid, linear coronagraphic slide, off-axis paraboloid, Lyot plane, off-axis paraboloid, fold mirror. The infrared tracker is the green square. The image and pupil viewing camera and lenses are shown on the left, directly below the first off-axis paraboloid. In this orientation, the output beam to the infrared imager PHARO exits downward into the page.

of the tracker is digitized, analyzed, and locked to the adaptive optics system in a proportional-integral-derivative type controller, with a variable update time. Typically, corrections are sent once every 5-20 seconds, depending on observing conditions, as the goal is to correct slow non-common path drifts. Using a Hamamatsu G6489-01 quad cell as the sensor, the tracker delivers a positional accuracy of better than 3 mas, and has a linear range of ± 250 mas. The tracker uses a variation of the denominator-free centroid technique (Shelton, 1997), allowing acquisition of targets well outside this linear range or the AO field stop size. Its dynamic range corresponds to stars between 0 and about 8.5 magnitudes, the fainter magnitude approximately coinciding with the high-Strehl ($>80\%$) cutoff of the P3K adaptive optics system.

There are a few design choices that make the instrument quite flexible for adapting new configurations. The mounts holding the coronagraphs are connected to the motorized linear stages by magnets, which can be quickly removed and replaced, with a positional repeatability of a few microns. Furthermore, each mount accommodates up to three separate coronagraphic elements, which can be slid in and out remotely, although one slot is typically occupied by a single-mode alignment fiber. The Lyot wheel has five slots, and the face of the wheel can be quickly removed for installation of new pupil masks. Finally, the second Lyot plane can also accept pupil masks, which can be positioned accurately via the use of magnets installed on the optical mount. Some of these optics are shown in Figure 7.4.

4.4 Observing sequence and data reduction

4.4.1 On-sky calibrations and observing strategy

The Hale telescope has an equatorial mount, so Cassegrain instrument gravity vectors vary with time and depending on the target. In practice, this changing vector leads to two important effects, which are linked. First, the AO system output Strehl ratio degrades after a large slew; second, the speckle pattern changes due to flexure. These two effects can be viewed as introduction of low-order and high-order aberrations, respectively. We compensate for these two effects separately in two calibration steps that we perform after executing a large slew of the telescope. These steps are performed using the internal white light source of the adaptive optics system.

4.4.1.1 Correction of non-common path low-order aberrations: MGS

The Strehl ratio at the first coronagraphic mask is important, as the total performance of that vortex depends sensitively on the amount of low-order aberration present. We compensate for the Strehl degradation with telescope pointing by performing the Modified Gertzberg-Saxton (MGS) phase retrieval algorithm (Burruss et al., 2010) in the detector plane after slewing to a new target, using

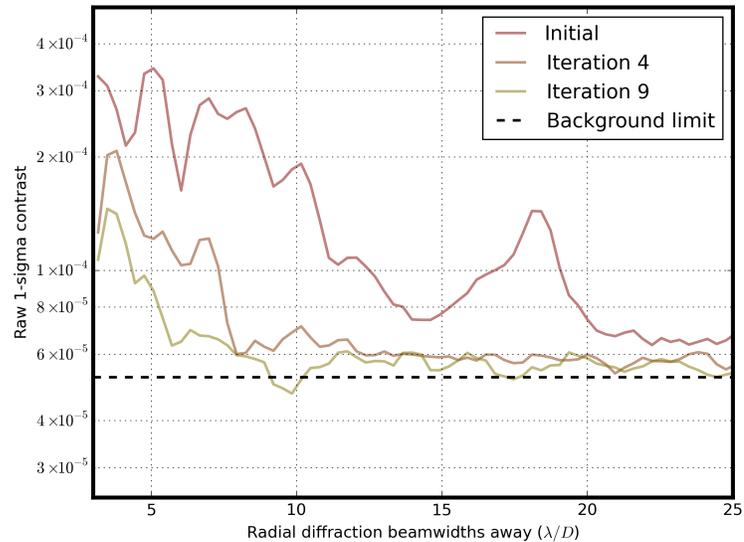
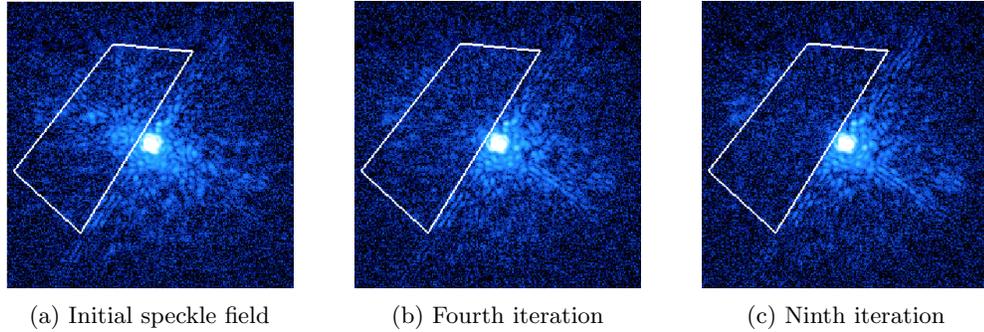
the internal AO white light source. This technique calculates phase errors using a set of defocused PSF, iteratively updating the shape of the deformable mirror to maximize the Strehl ratio. While the Strehl ratio at the detector is not the same as on the first vortex mask focal plane, the SDC internal Strehl ratio, i.e., that measured from the first coronagraphic focal plane to the detector, is typically about 95%. Therefore, it is safe to state that the MGS algorithm’s optimization of low-order aberration between the deformable mirror and the detector focal plane achieves a high Strehl PSF on the *coronagraphic mask* plane as well. Using MGS, K_s Strehl ratios are typically improved from about 70% to 95% from the input of the adaptive optics system to the detector after large slews or other disruptive events.

4.4.1.2 Correction of non-common path high-order aberrations: speckle nulling

The most important factor limiting contrast in high contrast imaging systems are speckles, quasi-static bright spots in the focal plane caused by non-common-path phase and amplitude aberrations in or after the beamsplitter that separates science light from light to the AO wavefront sensor. The total intensity of these speckles (that is, the speckle flux in the focal plane from $0-33\lambda/D$) is very low in absolute terms, as they correspond to wavefront errors with amplitudes of 5-25 nm, but they are typically much brighter than any planetary companions, at raw contrasts of 10^{-3} - 10^{-4} in K_s . They are also sensitive to gravitational flexure in the optical train that evolves on \sim minute timescales. The MGS algorithm has difficulty correcting these low intensity, high spatial frequency speckle aberrations, as the signals-to-noise of these errors are very low in the out-of-focus images that the procedure uses to determine wavefront quality. Despite the fact that the detector readout time is a few seconds for PHARO, these slowly-evolving static speckles may be reasonably tackled using the science camera as a sensor.

To lower speckle intensity, we have implemented a speckle nulling code (Savransky et al., 2012). The technique involves using the deformable mirror to generate spots at the exact locations of bright speckles in the image, then varying the phase of the electric field of the spots. By measuring the modulation in the intensity due to the interference of the spot and speckle, it is possible to calculate the phase of the speckle, and hence cancel it with the deformable mirror.

A typical iteration proceeds as follows. First, speckles are identified in the image using a local maximum filter, and aperture photometry is used to get their intensities. The position on the detector and intensity of an individual speckle is converted to an equivalent sinusoidal shape on the deformable mirror, with the spatial frequency and orientation determining the position, and the amplitude determining the brightness. The phase of the sinusoid on the deformable mirror is changed four times, with the intensity of the speckle measured at each step. These intensities are then used to determine the phase that would cancel the speckle completely, the “null phase.” In order to compensate for deformable mirror hysteresis, intensity calibration imperfections, and other



(d) Contrast in region

Figure 4.6: Coronagraphic speckle nulling in the dual-vortex mode using the internal white light source of the adaptive optics system; see 4.4.1.2. (a) The initial results of PSF correction using MGS (4.4.1.1) still leaves many residual speckles in the focal plane. (b) Four iterations of speckle nulling remove most of the residual wavefront errors (c) Nine iterations get to within a factor of two of the detector read noise from 5 - 25 λ/D . The white polygon demarcates the control region, which is selected by mouse clicks in the half-region control mode. (d) Contrast improvement measured in the control region shows factors of 3-6 improvement, which are significant for companion detectability. The contrast curve is defined in the usual way, with the standard deviation (ie, 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The background limit is determined by the contrast in a region of the detector 100's of λ/D away. The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding.

possible errors, a second loop then modulates the amplitude of each null sinusoid in four steps (0%, 33%, 66%, 110%), with the speckle intensity recorded once again and used to determine the optimal amplitude of the sinusoid on the deformable mirror. While the second loop is not strictly necessary, and is not used for the first few iterations, it is very useful as it prevents speckle amplification if the phase is incorrectly measured. It also helps define a stop condition—when the calculated optimal gains start dropping to zero, there is no benefit to continue running the loop. All in all, either four

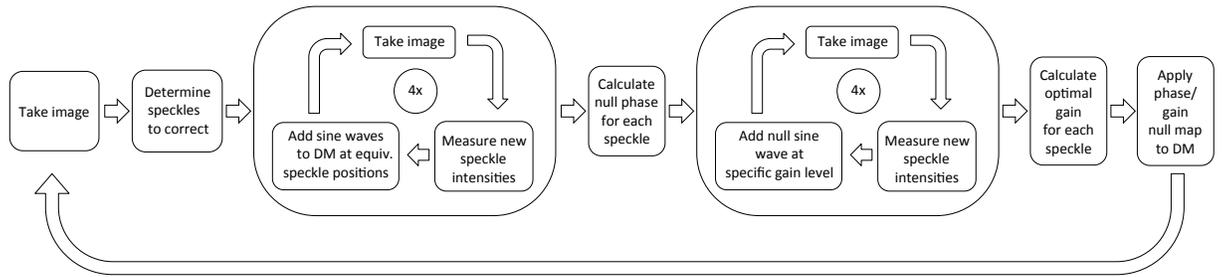


Figure 4.7: Outline of the speckle nulling loop. See Section 4.4.1.2 for a description of the loop.

or eight images are required per iteration.

After a few iterations, the internal instrumental contrast through the adaptive optics system improves typically by a factor of 3-6 when nulling half of the focal plane, and less when nulling a centro-symmetric region. The reason for the difference is that in the half aperture case, both phase and amplitude errors can be corrected with the single deformable mirror; in the full aperture case, only phase variations can be corrected. The system-level description of the speckle nulling code is shown in Figure 4.7.

4.4.2 Observing strategy

Similar to other high contrast imaging programs, our observing strategy is driven by the need to reduce residual speckles. There are many different methods to deal with speckles, including angular differential imaging (Marois et al., 2006a) and spectral differential imaging (Marois et al., 2006b). The former is precluded by the equatorial/Cassegrain mount configuration, and the latter because our imager is not an integral field spectrograph.

The strategy we use is the classical approach called “reference differential imaging” (Mawet et al., 2011a), and involves quickly alternating observations between our target star and a reference star. The reference star is selected to be nearby, and to have a similar visible magnitude and spectral type. The proximity of the star ensures that the gravity vector is similar, leading to a nearly identical speckle pattern. The visible magnitude and spectral type ensure that the AO correction and brightness in the detector plane are very similar. Dithering between the target and reference on a few-minute cadence means that the slow temporal evolution of the speckle pattern is thus tracked. With datacubes of the target and reference, it is possible to remove speckles in post-processing, as will be described below.

4.4.3 Data reduction pipeline

Data reduction is an integral part of high contrast imaging, and significant advances in performance have come from new algorithms and reduction methods (Lafrenière et al., 2007; Soummer et al., 2012). Understanding of reduction code behavior and performance have a direct bearing on observing strategies.

The SDC pipeline is an automated high contrast imaging reduction code written in Python. It performs the low-level preprocessing of the images, registration, and various forms of point spread function subtraction, as will be explained below.

The pipeline begins by building a database of each observing sequence from logsheet entries, taking as input exposure numbers. It checks for errors in the logsheet such as mislabeled filters, stops, etc, by examining file headers. A bad pixel mask is generated from flat fields of varying fluxes, and bad pixels are treated with a spatial median filter. Dark subtraction/flatfield normalization is performed after this point. Image registration is done to the subpixel level by a discrete Fourier-transform upsampling algorithm described in Guizar-Sicairos et al. (2008), with shifting done using linear interpolation. During observing, we generate astrometric reference spots by adding a checkerboard pattern to the deformable mirror, corresponding to the outer spatial frequency controlled by the mirror. In practice, this checkerboard is generated on-sky by changing the convergence points of the wavefront sensor, with an amplitude corresponding to 5% of the stroke of the deformable mirror.

Using the registered reference and target image cubes, three independent point-spread function subtractions are performed. The first is classic PSF subtraction, where an intensity-scaled median reference PSF is subtracted from the median target PSF. The second method is a full-frame principal components analysis, also known as the Karhunen-Loeve eigenimage decomposition (KLIP) algorithm, described in Soummer et al. (2012). Here, each target image is projected onto a low-dimensional subspace derived from the principal components of the *reference* image library. The low-dimensional projection is subtracted from the input data. If an off-axis source is present in the target, but not the reference, it should largely be unaffected after subtraction of the principal components, though modeling of the algorithm’s effects is necessary. One issue with using full frame KLIP is that the noise properties vary significantly as a function of radius from the star, and principal components analysis does not work well with this spatially varying noise. The way we compensate for this is to partition the image into small, overlapping zones and perform KLIP on each zone, then take the median to reconstruct the image. The zones are chosen to be small enough such that the noise is fairly uniform over each of them.

In addition to the reference cube, a “supercube” is constructed by choosing the most highly similar images to the target datacube over the entire observing run (not just the target and reference frames), selected via image correlation coefficients. Using the supercube as the reference, the data is

reduced again; this typically gives a 30-40% improvement in contrast at close inner working angles. Unsurprisingly, most of the frames selected by the supercube generating algorithm are from the calibrator star.

4.5 Coronagraph configurations and laboratory performance

Observing mode	1st coronagraph	First Lyot	2nd Coronagraph	Second Lyot
Open	Open	Open	Open	None
Single Vortex	Ks band vortex	Lyot stop 1	Open	PHARO pupil stop
Ring-apodized vortex	Open	Ring apodizer	Ks band vortex	Lyot stop 2
Dual Vortex	Ks band vortex	Lyot stop 1	Ks band vortex	PHARO pupil stop

Table 4.1: The different on-sky observing modes of SDC; see section 4.2 for a description of each of these modes.

The SDC currently has four operational on-sky observing modes, listed below in Table 4.1. The “open” observing mode is used for flat-fielding exposures and non-coronagraphic observing. The “single-vortex” mode is used for comparing contrast ratios between the two vortices, but is not generally used for science as it offers no observational advantage due to the effect of the secondary. The “ring-apodized mode” and “dual vortex mode” are used for science observations, with the majority of time thus far going to the dual vortex mode. The reasons for this are twofold. First, the sensitivity to tip/tilt errors is much higher in the ring-apodized mode compared to dual-vortex mode (θ^2 vs. θ^4). This problem is compounded by the fact that the tip/tilt pointing control is not as good, as the tracker is located further away from the vortex mask used (that being in the *second* SDC focal plane, after the first Lyot wheel housing the apodizer). Furthermore, despite having theoretically perfect rejection, the theoretically maximal throughput of the ring-apodized coronagraph is only about 33%, as mentioned earlier.

We first measured performance of the observing modes in controlled laboratory conditions. In this case, the same P3K adaptive optics system, coronagraph, and detector were used, but the light source was the internal single-mode fiber white light source from the AO system. Additionally, all the tests were performed at a stationary, vertical gravity vector.

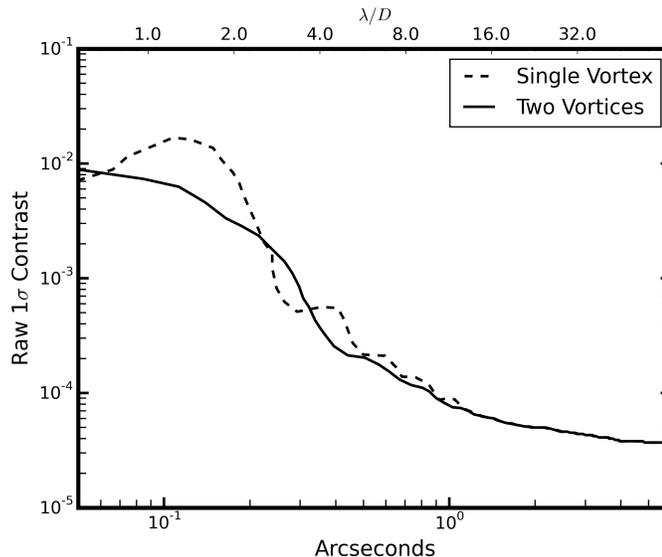
4.5.1 Single and dual vortex observing modes

The results comparing single and dual vortex modes are shown in Figure 4.8. The dual vortex mode provides a dramatic improvement in inner working angle, as shown by the large boost in contrast from $1 - 2 \lambda/D$. Contrast is also improved at $3 - 10 \lambda/D$, with the diffraction rings substantially removed. The point-spread functions are shown at the top; note that the dual vortex PSF closely reproduces the original Airy function PSF, at a reduced intensity, as expected from Figure 7.1, where

the output pupil is a copy of the input pupil but fainter. The measured peak rejection of about 100:1 is consistent with theoretical prediction of 80:1 for the Palomar secondary/primary mirror size ratio; the “better” than expected performance is mainly due to the fact that there is a small $25\ \mu$ chromium dot at the center of the vortex to compensate for manufacturing imperfections that can lead to stellar leakage. We note that radial contrast, not peak rejection, is a true measure of coronagraph performance, but peak rejection provides a quick way to check whether the coronagraph is working to design expectations.



(a) Zero, single, and dual vortex PSFs; same (linear) scale



(b) Raw contrast

Figure 4.8: Laboratory contrast measurement comparing single (dashed curve) and dual vortex mode (solid curve). The contrast curve is defined in the usual way, with the standard deviation (ie, $1\ \sigma$) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding.

4.5.2 Ring-apodized vortex observing mode

The ring-apodizer was installed and tested on-sky in February 2015. Lab tests indicated performance consistent with theoretical expectations (see Figure 4.9 (b)). The advantages are high starlight suppression at small inner working angles. (For a telescope design with a 20% secondary obscuration,

such as Keck, the throughput would be about twice as high, making this approach much more advantageous). Regardless, the lab tests demonstrate the validity of the concept, with a close match between theory and measurement. However, the minimum rejection is about 1000:1 at the peak of the psf, as opposed to infinite. There are a few explanations for the imperfect rejection, such as imperfect Strehl at the position of the second vortex. Additionally, the central chromium spot on the vortex is not taken into account in the idealized performance model. Similarly, bright light from defective pinned actuators in the AO system were clearly visible in the “null” region of the mask. Finally, interferometric testing revealed a 20 nm phase difference between the opaque and transmissive annuli in the apodizer, though at the operational wavelength of over 2 μm , this is unlikely to be significant.

4.6 On-sky performance

First light observations with SDC took place in February-March 2014, with full science observations in dual-vortex mode commencing in October 2014. Other observing runs were in Feb 2015 (nearly all lost to weather) and May 2015. This section will present some of the early engineering and science results of the instrument, some of which have been already published. It is not the intent of this paper to present the complete analyses of all our science targets, but a few preliminary results showcasing the performance of the instrument are summarized.

In general, sky performance (as measured by expected peak-to-peak rejection ratio) is within a factor of two of the lab-measured values. Much of the deviations from theory can be attributed to the imperfections in adaptive optics performance, such as imperfect Strehl ratio, bad actuators, etc. Internal rigidity of the instrument is generally very good, with the two focal-plane vortex masks staying co-aligned throughout observations despite large slews between different target stars. An example of a measurement of one of our survey targets is shown in Figure 4.10, with the image showing low residual diffraction with no evidence of companions.

4.6.1 Confirmation of physical association of epsilon Cephei b

Epsilon Cephei (HD 21136) had a previously reported (stellar) companion about 50 times fainter at 330 mas separation (Mawet et al., 2011a), but it was unknown whether it was physically associated or a background alignment. The close separation measured in 2010 ($3.6 \lambda/D$ for a 5 meter aperture) made it an attractive target for our first science observations. Additionally, it would allow more accurate measurement of position as the angular resolution of the full 5 meter dish would be used; the previous observation was performed in 2010 with a 1.5 m clear sub-aperture of the Hale telescope. We observed the star on 13 October 2014, and were able to see the companion in the raw coronagraphic image, with no post-processing (Figure 4.11). Classical PSF subtraction was all that was needed to

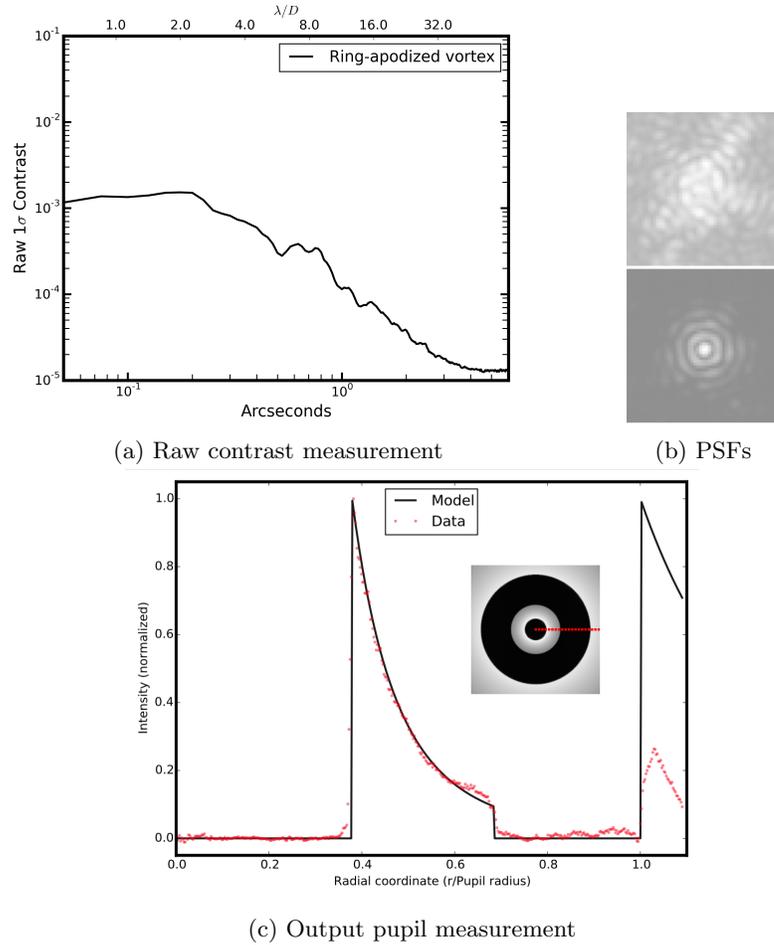
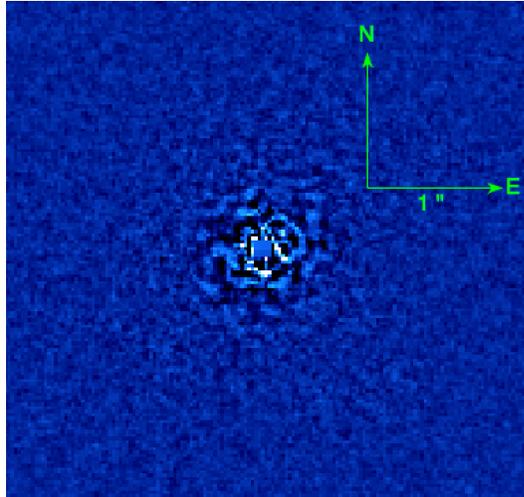
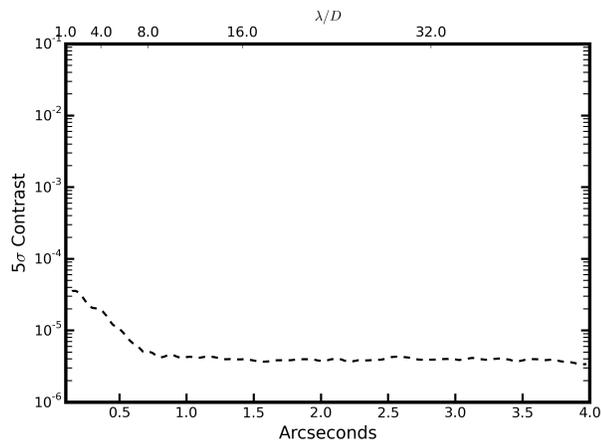


Figure 4.9: (a) Raw contrast measurement with the ring-apodized vortex coronagraph. The contrast curve is defined in the usual way, with the standard deviation (ie, 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding. (b) The coronagraphic (top) and non-coronagraphic PSF, shown on different logarithmic scales to enhance features. (c) The measurement of the output pupil intensity corresponds well to theoretical expectations, with the major discrepancy being outside the pupil. This is due to the presence of an chromium dot in the center of the vortex, reducing stellar leakage. The center of the PSF is the brightest, so light blocked there will not show up outside the pupil.

measure astrometry and photometry. Given that the proper motion of Epsilon Cephei is more than 400 mas/yr , and our measured companion separation is $216 \pm 6 \text{ mas}$ ($2.1 \lambda/D$) after more than four years, physical association is definitively confirmed. Significant orbital motion is also evident, as the position angle has changed from 90 ± 10 degrees to 66 ± 3 degrees, and the orbit separation decreased from $330 \pm 50 \text{ mas}$ to the currently measured value. The data confirms that the orbit is clearly far from edge-on, but the two data points do not allow for a detailed characterization.



(a)



(b)

Figure 4.10: A reduced image of one of our target stars ($K=8$, $V=6$) with the associated 5σ contrast curve on the right. The reduction strategy used was a zonal principal components analysis (KLIP) algorithm (see Section 4.4.3), with the principal components generated from a calibrator star with similar brightness and $V - K$ color. The total open shutter time on this target was 14 minutes, with the same time on the calibrator star (backgrounds, flats, and non-coronagraphic PSF frames were recorded separately). This measurement did not involve speckle nulling, so contrast at small angles can likely be improved further in the future.

4.6.2 Identification of the “compact object” companion to delta Andromeda

On the initial science run, we observed the star delta Andromeda, which is a spectroscopic binary with a period of about 58 years. The companion had been previously hypothesized to be a white dwarf (Gontcharov & Kiyeva, 2002; Judge et al., 1987), but had never been imaged due to its faintness and proximity to the primary. Thanks to the high contrast at low inner working angles, we were able to easily detect the companion in the raw image. The companion separation was found

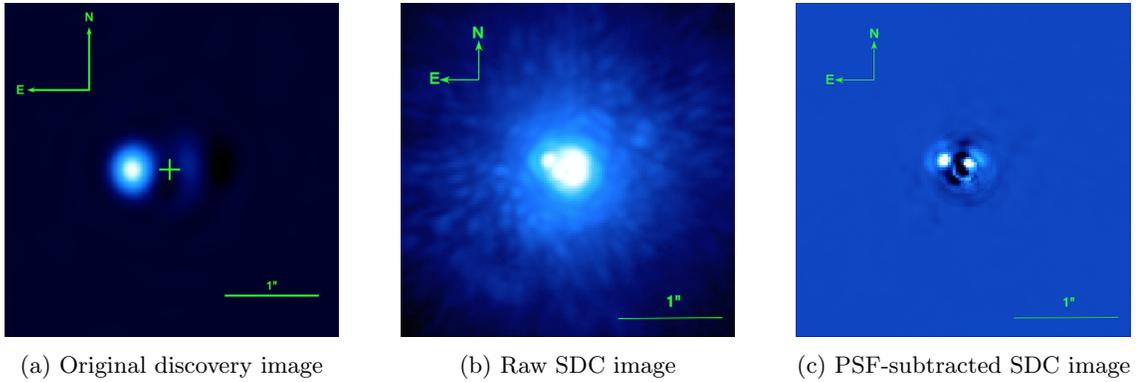


Figure 4.11: Epsilon Cephei b. (a) The original discovery image, from Mawet et al 2011 (Mawet et al., 2011a), using a 1.5 m well-corrected subaperture of the Hale telescope. (b) Raw (no reference subtraction) SDC image, dual vortex mode, 15s of 10 median combined frames. (c) Classic PSF subtraction of (b). In the SDC images, the first Airy ring is visible around the companion.

to be about 360 mas ($4 \lambda/D$) with a contrast of 6.2 magnitudes in K-band. Bottom et al. 2015 showed that the companion was much too bright to be a white dwarf, and was more likely a main sequence star of K-type. Again, the advantage of high contrast at low inner working angles allowed for a robust detection and characterization; see Figure 4.12.

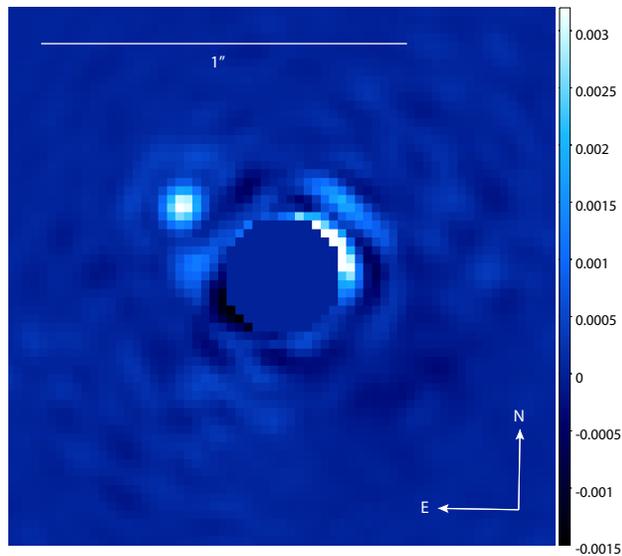


Figure 4.12: The PSF-subtracted coronagraphic image of delta Andromeda b, dual vortex mode. This image first appeared in Ref. 15.

4.7 Conclusion and future work

We have presented the motivation, design, and current performance of a new multistage coronagraphic instrument at Palomar observatory's 200" Hale telescope. The SDC is currently the only multistage coronagraph in operation, and has also successfully tested the ring-apodized vortex coronagraph concept, a promising way of pursuing high contrast at low inner working angles when behind obscured telescope pupils. The SDC is fully operational and actively pursuing both astronomical observations and new technical developments at the same time.

Other observational and wavefront sensing modes can be envisioned, and will be implemented as funding and teaming arrangements allow. Potential coronagraphic modes include other focal plane mask-based coronagraphs, such as the band-limited Lyot coronagraph (Kuchner & Traub, 2002), and also pupil-plane coronagraphs such as the shaped pupil (Kasdin et al., 2003) and phase-apodized coronagraph (Snik et al., 2012). Shorter-wavelength operation is also possible, as is a nulling interferometry mode. In the wavefront sensing area several steps are conceivable, including phase-shifting interferometry for direct measurement of speckle phases (Serabyn et al., 2011), Lyot-plane wavefront sensing (Singh et al., 2014), and speckle phase measurements with the self-coherent camera approach (Galicher et al., 2008). Finally, post-coronagraphic spectroscopy is also now enabled at Palomar, first with PHARO itself using the internal grisms, and potentially with other spectrometers, such as upcoming energy-resolving MKID detectors (Mazin et al., 2014).

4.8 Acknowledgements

We are pleased to acknowledge the Palomar Observatory staff for their enthusiastic and excellent support. We thank the referee for a careful and thorough read, and comments which improved the paper. MB is supported by a NASA Space Technology Research Fellowship, grant NNX13AN42H. Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA).

4.9 Appendix

Element	Degrees of Freedom	Controls
AO output beam	Tip, Tilt	X, Y on focal plane masks
Input fold mirror	X, X+Y	Lateral input pupil position
Focal plane mask 1 slide	X	Mask 1 X position
First Lyot Wheel	Θ	Lyot stop choice
OAP #2	X, Y, X+Y	Mask 2 Image Y position/focus
Focal plane mask 2 slide	X	Mask 2 X position
Lyot plane #2	X, Y, X+Y	Image position on detector
Output fold mirror	X, Y	Output pupil position
Flip lens	In/Out	Image/Pupil on internal camera

Table 4.2: The actuated optics in the SDC, their degrees of freedom, and the optical fields they control. Refer to Figure 7.4 for the an optical layout

Item	Vendor	Part	Notes
Vortices	JDSU		25 μm dot in center
Optics mounts	Newport/Thorlabs	Assorted 2 and 3"	
Actuators	Newport	TRA6	0.2 μm step, 1.5 μm abs.
Actuator controller	Galil Motion Control	DMC-4183	
Lyot wheel		Custom	5 slots
Lyot controller	Sigma-Koki Co.	PAT-001	250,000 cts/rev
Vortex stage controllers	Applied Motion Products	ST5	Had to filter PWM signal
Internal camera	Sensors Unlimited	InGaAs	J, H, K sensitivity
Piezo stage	Physik Instruemnte	P-752	0.2 nm resolution
Piezo controller	Physik Instrumente	E-516	Few nm resolution

Table 4.3: List of optics, electronics, and related information

Chapter 5

Part II—Resolving the delta Andromedae spectroscopic binary with direct imaging

Abstract

We present a direct image of the innermost companion to the red giant δ Andromedae using the Stellar Double Coronagraph at the Palomar Observatory. We use a Markov-chain Monte Carlo based algorithm to simultaneously reduce the data and perform astrometry and photometry of the companion. We determine that the companion is a main-sequence K-type star rather than the previously hypothesized white dwarf.

5.1 Introduction

δ Andromedae (K3 III) is red giant with a visual magnitude of 3.28. It has a UV excess which implies a hot, high-velocity wind and a 60 and 100 μm excess (Judge et al., 1987) which is most likely due to a debris disk (Decin et al., 2003). It is the brightest star in a quadruple system; of the outer companions (28.7 and 48 arc seconds), the first has been classified as an M2 V star with $V = 11.3$, probably physically associated with the primary as it shares the same proper motion (Bakos, 1976). The outer component does not share the proper motion of the system and is most likely a background object.

δ Andromedae is a spectroscopic binary with a rather long period of about 52 years; see Table 5.1 for a summary of its physical properties. The presence of the secondary has been confirmed both spectroscopically (ibid) and photocentrically (Gontcharov and Kiyaveva 2002). The companion has been conjectured to be either a main-sequence star later than G-type (Judge et al., 1987) or white dwarf near the Chandrasekhar limit (Gontcharov & Kiyaveva, 2002). It has never been directly imaged, however, due to the secondary’s faintness and proximity to the primary. In this work we image the companion for the first time and measure its magnitude and separation. We find that the secondary is at approximately the expected separation and determine that it is not a white dwarf, but a main sequence star of K-type. This work demonstrates the potential of high contrast imaging with low inner working angles applied to spectroscopic binaries.

Physical properties of δ And		
Mass (M_A+M_B)	$2.6 \pm 0.4 M_\odot$	Gontcharov & Kiyaveva (2002)
a_A+a_B	$0.62 \pm 0.04''$	Gontcharov & Kiyaveva (2002)
Radius	$13.6 \pm 0.3 R_\odot$	Piau et al. (2011)
Luminosity	$68 \pm 4 L_\odot$	Piau et al. (2011)
Surface gravity ($\log g$)	2.0 ± 0.3	Judge et al. (1987)
Temperature	$4315 \pm 9 \text{ K}$	Massarotti et al. (2008)
Metallicity [Fe/H]	0.04 dex	Massarotti et al. (2008)
Rotational velocity ($v \sin i$)	6.5 km/s	Massarotti et al. (2008)
Age	3.2 Gyr	Decin et al. (2003)
δ And b properties (this work)		
ΔM (Bracket- γ)	6.22 ± 0.05	
Angular Separation	$0.357 \pm 0.0035''$	
Position angle	$56 \pm 1^\circ$	
Physical Separation	$11.55 \pm 0.13 \text{ AU}$	parallax: Hipparcos (van Leeuwen, 2007)
Spectral type	$K4 \pm 2$	derived from ATLAS9 spectra

Table 5.1: Previously measured properties of δ And and newly measured properties of the companion

5.2 Instrumentation, Observations and Data Analysis

5.2.1 Instrumentation

The Stellar Double Coronagraph (SDC) is a JPL-developed instrument designed for high-contrast imaging of close-in companions to stars, particularly exoplanets. It uses two optical vortices in series to simultaneously diffract starlight out of the pupil of the instrument and partially correct for the secondary obscuration of the telescope (Mawet et al., 2011b). It has an inner working angle of approximately $1\lambda/D$, or 90 mas in K-band ($2.2\ \mu\text{m}$) behind the 5 m Hale telescope. It is installed between the P3K adaptive optics system (Dekany et al., 2013) and the near-IR imager PHARO (Hayward et al., 2001).

5.2.2 Observations

We observed δ And on October 8-9 2014 (UTC), during the course of normal science operations. The seeing at Palomar was $1.2''$, with the adaptive optics system delivering a Strehl ratio of about 85% at an airmass of 1.02-1.03. Our observing strategy involved frequently dithering between the target star and a reference star, then using post-processing to subtract the speckle pattern from the target images using the reference (see the following section for more detail). A Bracket- γ filter was used concurrently with neutral density filters to reduce the flux from the target when off the coronagraph to below detector saturation. Absolute transmissivity of the neutral density filters was measured separately, and found to be consistent with Metchev & Hillenbrand (2004). Sky backgrounds were interspersed with the target and reference star observations; sky flats were taken five days later. A summary of the observations is presented in Table 5.2.

Observing date: Oct. 9-10 2014, JD 2456939-10				
Target	Images	Filters	Exposure Time[s]	Purpose
δ And	29	Br- γ , ND2	9.91	Photometry, Astrometry
β And	100	Br- γ , ND2	2.83	Photometry
δ And	10	Br- γ , ND3	2.83	Non-coronagraphic, Photometry

Table 5.2: Summary of observations.

5.2.2.1 Data Analysis

One of the main challenges in high contrast imaging is trying to remove speckles due to aberrations in the optics after the wavefront sensor. There are a number of ways to tackle this contrast-limiting/quasi-static aberration problem; our strategy is sometimes called “reference differential imaging” (Mawet et al., 2011a). This involves dithering between the target and a nearby star of

similar visible magnitude, spectral type, and airmass. This leads to a similar AO correction gravity vector, ensuring a similar speckle pattern. It is then possible to remove some of the speckles by either subtracting the reference image or using a more advanced image processing technique such as the Karhunen-Loeve eigenimage decomposition (Soummer et al., 2012). The latter method gives better results than the former in terms of contrast, but has the unfortunate side effect of reducing the flux of any nearby companions that might be in the image, therefore rendering accurate photometry difficult. In this paper, we use a slightly different approach where we forward model the target image as a combination of a scaled reference image and a shifted, attenuated point-spread function image. This method has some advantages that will be explained below.

We acquired coronagraphic images of δ Andromedae and the reference star, β Andromedae. We aligned and median combined these images after flat-fielding, background subtraction and bad pixel removal. We derive a relative magnitude and offset between the star and companion using a Markov Chain Monte Carlo (MCMC) fitting algorithm (Foreman-Mackey et al., 2013). This is somewhat different than the usual approach to analyzing fluxes and positions, where one prioritizes maximising the signal to noise ratio of the companion, often performing astrometry and photometry separately. Here the image reduction, raw photometry, and astrometry are all performed at the same time by the MCMC algorithm. There are a few advantages to doing everything at once with MCMC. First, one can measure the precision of the reduction algorithm much more accurately: the per-pixel uncertainties are Poissonian and straightforward to propagate in the model above. Furthermore, the MCMC returns marginal likelihoods, which shows the precision in each parameter as well as any correlations. Finally, one does not need an analytic model of the PSF but can use images of the instrumental point spread function taken off the coronagraph. This reduces the number of parameters in the model, decreases degeneracy, and improves accuracy.

The generative model for the image data is

$$T[x, y] = R_a \cdot R[x, y] + P_a \cdot P[x - x_c, y - y_c] \quad (5.1)$$

where T is the coronagraphic image of δ And, R_a is a constant scale factor, R is the coronagraphic reference image of β And, P_a is another constant scale factor, and P is the point-spread function (ie, a unit-intensity normalized, non-coronagraphic image of a point source). Images T , R , and P are all single median images. The indices x, y refer to pixel coordinates, and the factors x_c, y_c are shifts in point-spread function imaging data (ie, $P[x-1, y-0.34]$ corresponds to a pixel shift of 1, 0.34). The constant R_a is to correct for the fact that the background speckle field in the science image is of a different mean intensity, due to differing magnitudes or exposure times. P_a is the intensity scaling prefactor of the point spread function of the companion. MCMC is used to solve for x_c, y_c, P_a , and R_a simultaneously; the results are presented in Figure 5.1. The “reference subtracted” image,

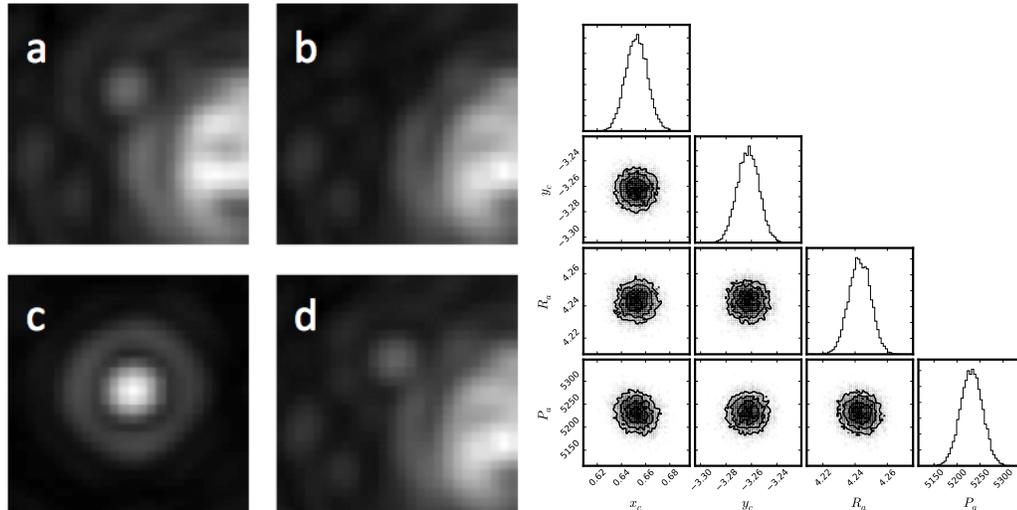


Figure 5.1: Left: a) the background-subtracted target median image, b) the background-subtracted reference star median image, c) the background-subtracted point-spread function image, d) best-fitting model from the MCMC algorithm combining images b) and c) and attempting to match a) as explained above. The stretch is nonlinear to better show the companion and speckles. Right: All the one and two dimensional projections of the posterior probability distributions of the pixel shifts (x_c , y_c , the reference background scaling factor (R_a , and the PSF amplitude used to fit the companion P_a . The two-dimensional projections show very little covariance among any two parameters, and the marginal distribution histograms (along the diagonal) are nicely peaked.

$T[x, y] - R_a \cdot R[x, y]$, is shown in Figure 5.2.

In order to determine the relative brightness, we similarly use the unit-intensity PSF model to fit a non-coronagraphic image of δ And, and the derived intensity allows us to establish a relative intensity in magnitudes. The uncertainty in relative intensity is dominated by the uncertainty on the neutral density filters' absolute extinctions. For the companion location, the typical error in this case for x_c and y_c was about 0.01 pixels, or less than a milliarcsecond at $0.025''/\text{pixel}$. However, this is not the true uncertainty in *separation* because the primary star's image is suppressed and distorted by the coronagraph and its true position is not obvious to calculate. In order to locate the position of the primary, we imposed a waffle pattern on the deformable mirror of the adaptive optics during observations. The waffle generates astrometric spots $3.9''$ away from the primary, which can be used to locate the position of the star, and we verified our result using the Radon transform (Pueyo et al., 2015). The waffle centration has an uncertainty of about 0.1 pixels, which dominates the total separation uncertainty.

5.3 Results and Conclusions

The results of the above analysis are shown in Table 5.1. Judge et al. suggest that the secondary companion is either a main sequence star later than G type or a white dwarf. Gontcharov and

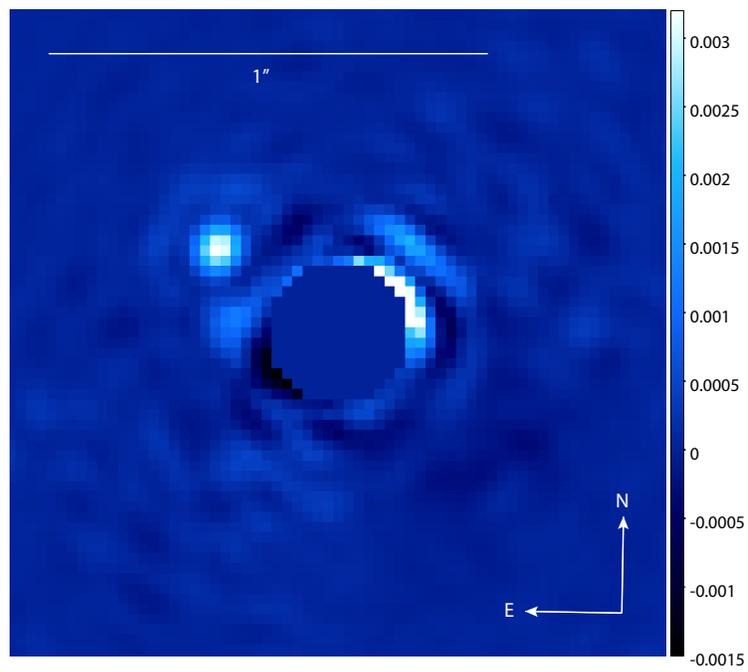


Figure 5.2: The reduced, background-removed coronagraphic image of δ Andromedae. The first Airy ring is visible around the companion. The stretch in the image is linear. The colorbar shows the relative intensity (as a fraction) compared to the primary

Kiyaeva measure a mass fraction $m_B/(m_A + m_B) = 0.5 \pm 0.1$ for the binary system and then favor the white dwarf assumption and suggest an even split of mass between primary and secondary, placing the white dwarf very near the Chandrasekhar mass limit.

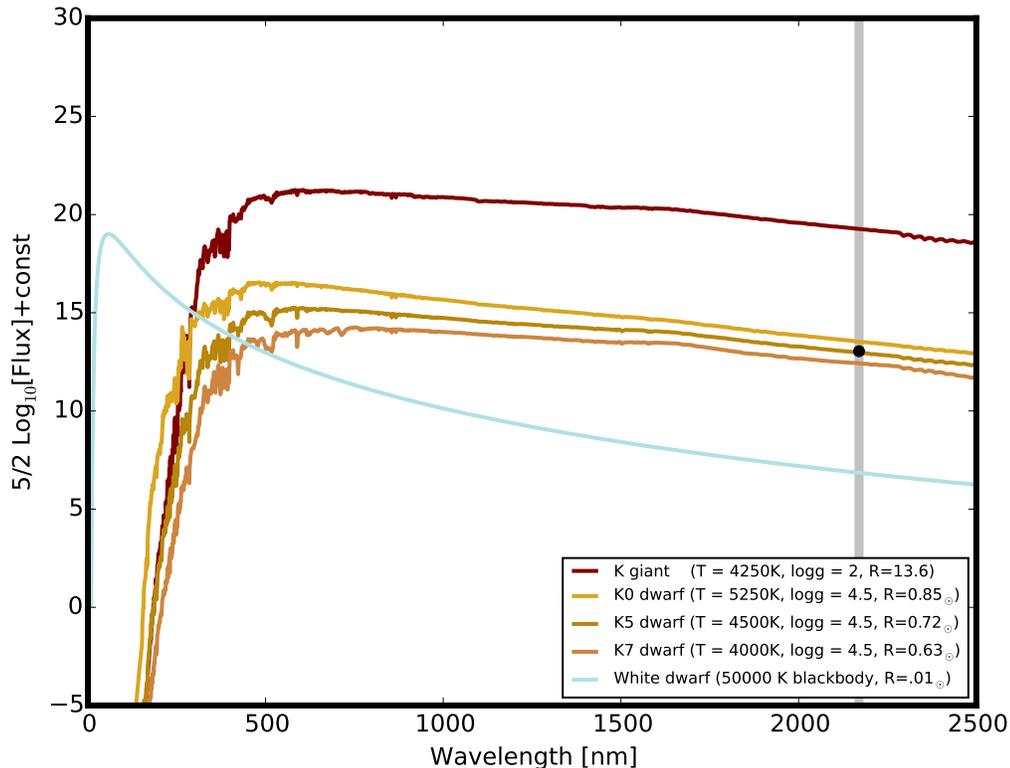


Figure 5.3: Comparison of the approximate fluxes of δ And A, three K dwarfs, and a 50000K white dwarf. The spectral models are from Castelli & Kurucz (2004). The grey bar shows the span of the Bracket- γ filter.

However, assuming that the companion is a white dwarf, its radius is constrained to be about $0.01 R_{\odot}$, as white dwarfs of $0.5\text{-}1.4 M_{\odot}$ span the radius range of 0.014 to $0.005 R_{\odot}$. Comparing the expected flux levels of a hot white dwarf to that of the primary (see Figure 5.3), the magnitude difference through the Bracket- γ filter would be about 12 magnitudes, not the measured 6, a discrepancy of greater than 100 times our photometric uncertainty. Furthermore, such a hot white dwarf would have a UV continuum that was not detected in Judge et al., who constrain the white dwarf's temperature to less than 10000K if it exists. This low temperature would make the magnitude difference even more extreme. The white dwarf possibility is thus definitively excluded.

On the other hand, the measured flux is quite consistent with a main-sequence K-type dwarf. Shown in Figure 5.3 are spectral models of K0, K5, and K7 stars, with our measured flux shown as a black dot. For the δ And primary, the effective temperature and radius is taken from published results

(Table 5.1; note that the radius is known accurately from interferometry (Piau et al., 2011)). For the secondary, the effective temperatures and surface gravities are taken from Castelli & Kurucz (2004) and the radii are taken from Cox (2000). While the formal photometric error is smaller than the size of the datapoint, lack of knowledge about the companion’s radius and temperature make it impossible to give a completely specific spectral classification; the best we can say is that the companion is most likely a K-type dwarf. Making a more accurate measurement of the secondary spectral type is possible in principle. The simplest way would likely be a similarly precise coronagraphic measurement in J band, as the J - K colors of K dwarfs change by about 200 millimag over the spectral type. Alternatively, an AO-fed integral field spectrograph might be able to measure the spectral type from the CO band at approximately $2.3 \mu\text{m}$.

The conclusion that the companion is K-type is mostly consistent with previous work. As mentioned before, Judge et al. concluded that a main sequence companion would have to be a star later than G-type. A K-dwarf has a mass of between 0.6 - $0.8 M_{\odot}$; taking values of 1.1 - $1.2M_{\odot}$ for δ And A gives 0.3 - 0.4 as the mass fraction, reasonably consistent with the value of Gontcharov & Kiyaveva (2002) of 0.5 ± 0.1 . The results presented here demonstrate the potential of high contrast imaging applied to medium to long-period spectroscopic binaries. In particular, the contrast differences between main sequence stars in binaries are readily accessible to a coronagraphic system, and the information gained can improve orbit characterization, or as in our case, distinguish quickly between different companion possibilities.

Chapter 6

Part II—Speckle nulling wavefront control for Palomar and Keck

Abstract

We present a speckle nulling code currently being used for high contrast imaging at the Palomar and Keck telescopes. The code can operate in open and closed loop and is self-calibrating, requiring no system model and minimal hand-coded parameters. Written in a modular fashion, it is straightforward to port to different instruments. It has been used with systems operating in the optical through thermal infrared, and can deliver nearly an order of magnitude improvement in raw contrast. We will be releasing this code to the public in the near future.

6.1 Introduction

The largest current barrier to imaging extrasolar planets is the presence of “speckle” aberrations, which arise from errors in figure or transmissiveness of optical elements after the wavefront sensor of the adaptive optics system. These speckles show up in the focal plane as bright points of light that look similar to the point-spread function, typically tens to thousands of times brighter than any planetary companions that might be present in the image. To a large degree, observing approaches on ground-based telescopes are oriented towards removing speckles in post-processing. This includes techniques such as angular differential imaging (Marois et al., 2006a), spectral differential imaging (Marois et al., 2006b), and reference star differential imaging.

Rather than only remove speckles in post processing, it is possible to remove them “optically.” This has the advantage that the fundamental photon shot noise level is reduced, so all post-processing algorithms will have better performance, as they are all eventually limited by shot noise. There are a number of different techniques to reduce speckles optically. The most powerful general approach is called “electric field conjugation” (EFC) (Give’on, 2009), where the deformable mirror response function is linearized, small perturbations are put on it to determine the phase over the focal plane; then the linearized response function is used to compute the corrections in a specified control region. EFC has been used in a number of stable high contrast imaging testbeds, with contrasts of better than 10^{-8} being achieved in some cases (Thomas et al., 2010).

Electric field conjugation has not found wide implementation on high contrast imaging instruments on telescopes, perhaps because it requires an accurate model of the imaging system to achieve good performance. In many instruments, system states can change quite rapidly due to changing gravity vectors, temperature, or other states. This motivates us to consider a different algorithm for robust performance on non-ideal systems. An alternative to EFC is speckle nulling, an algorithm which identifies bright points of light in the image, uses the deformable mirror to create an artificial “speckle” at the same location, and calculates the phase of the original speckle by changing the phase of the artificial speckle (Savransky et al., 2012; Martinache et al., 2014). Finally, it uses the deformable mirror to put a speckle of the opposite phase on the original offending speckle, which makes it disappear. Doing this to many speckles at a specified location in the image plane will create a region of lowered intensity and higher contrast.

In this work, we will first present the background principles of speckles: how they arise, how they propagate, and how it is possible to (partially) correct them. Then we will describe the main points behind speckle nulling, and discuss our specific algorithmic implementation. Finally, we will present results of our speckle nulling code on different instruments at the Palomar and Keck observatories.

6.2 Background principles

6.2.1 Speckles in high contrast imaging

Speckles are formed by optical errors in the imaging system, and can be decomposed into two different types; “phase-type” and “amplitude-type”. A phase-type speckle is formed by a phase error on an optical surface, such as an imperfect figure, while an amplitude-type speckle is formed by variations in reflectivity, such as an imperfect polish. Significantly, phase-type errors can turn into amplitude-type errors and vice versa, via the Talbot effect, with the conversion taking place over distances of approximately $2x_0^2/\lambda$, where x_0 is the spatial frequency of the error and λ is the wavelength; this “Talbot distance” corresponds to about 1 meter of optical propagation. Speckles in the focal plane generally consist of both phase-type and amplitude-type, and a single speckle may be a superposition of both types. Here, we follow the explanation in Traub & Oppenheimer (2010), mostly conserving their notation. Readers are encouraged to consult that source for a thorough and insightful exposition of these phenomena.

The electric field of a pure plane wave at a pupil of the imaging system can be written as

$$E(r, t) = A_0 e^{i(k \cdot r - 2\pi f t + \phi_0)} \quad (6.1)$$

In the following discussion, since we are concerned with the final intensity, not electric field, we ignore the quickly oscillating terms of type $e^{i2\pi f t}$, where f is the frequency of light, on the order of 100 THz for these systems. Similarly, we restrict the coordinates to a single variable, x , which corresponds to the coordinate at a one-dimensional pupil that runs from $-D/2$ to $D/2$, D being the diameter of the pupil. Let us now consider static phase and amplitude type errors at the pupil, so that the amplitude and phase of the wave is not constant, but has a dependence on the position in the pupil

$$A(x) = A_0 e^{i\phi_p(x)} \quad (6.2)$$

Here, we let $A(x) = A_0(1 - \phi_a(x)) \approx e^{\phi_a(x)}$ be a constant electric field amplitude plus some small perturbation, and let $\phi_p(x)$ be any phase aberrations in the beam. We then absorb the amplitude aberrations and phase aberrations into a single complex phase error $\phi(x)$ so that

$$A(x) = A_0 e^{i\phi(x)} \quad (6.3)$$

$$\phi(x) = \phi_p(x) + i\phi_a(x) \quad (6.4)$$

Consider the case where the error consists of a single phase error¹ (such as a ripple in a mirror) of amplitude a , spatial frequency x_0 and “pupil phase” α . By pupil phase, we refer to the pupil decentration of the perturbation—with the convention that a pure cosine with the peak at the center of the pupil has pupil phase 0. The single phase ripple here is described by

$$\phi(x) = a \cos(2\pi x/x_0 + \alpha) \quad (6.5)$$

To see the effect of this perturbation in the image plane (where the detector is located), we recall there is a Fourier transform between the electric field of the pupil and image plane in an optical system. (Even if a detector measures intensity and not electric field, it is the field that carries the phase information explicitly.) Therefore, the electric field at the following image plane will be given by the Fourier transform of this ripple ϕ , as follows:

$$A_{\text{image}} = \int_{-D/2}^{D/2} e^{i\phi(x)} e^{i2\pi\theta x/\lambda} dx \quad (6.6)$$

$$\approx \int_{-D/2}^{D/2} (1 + ia \cos(2\pi x/x_0 + \alpha)) e^{i2\pi\theta x/\lambda} dx \quad (6.7)$$

$$= \int_{-D/2}^{D/2} e^{i2\pi\theta x/\lambda} dx + \frac{ia}{2} \int_{-D/2}^{D/2} e^{i(2\pi\theta x/\lambda + \alpha + 2\pi x/x_0)} dx + \frac{ia}{2} \int_{-D/2}^{D/2} e^{i(2\pi\theta x/\lambda - \alpha - 2\pi x/x_0)} dx \quad (6.8)$$

where the approximation is justified by the phase ripple being a small error. The last line follows from the second by expanding $\cos x = (e^{ix} + e^{-ix})/2$. We also note that the dx here refers to the pupil coordinate.

Letting the Fourier transform of a function f be denoted by $\mathcal{F}[f(x)] = \hat{f}(\theta)$, we now recall the following Fourier transform relations

$$\mathcal{F}[e^{i2\pi g x} f(x)] = \hat{f}(\theta - g) \quad (6.9)$$

$$\mathcal{F}[f(x/c)] = c\hat{f}(c\theta) \quad (6.10)$$

$$\mathcal{F}[\text{rect}(x)] = \frac{\sin(\pi\theta)}{\pi\theta} \quad (6.11)$$

$$(6.12)$$

where

¹More realistic errors can be decomposed into linear superpositions of such errors via Fourier analysis.

$$\text{rect}(x) = \begin{cases} 0 & |x| > \frac{1}{2} \\ \frac{1}{2} & |x| = \frac{1}{2} \\ 1 & |x| < \frac{1}{2} \end{cases}$$

Here we use the variable θ to denote angle. If a ripple in the pupil plane has a spatial frequency of cycles/cm, or cm^{-1} , then θ should strictly be in units of cm, just as the Fourier transform of a time series has units of t^{-1} , or Hz. However, since an imaging system essentially converts angles into positions, we casually write the position coordinate on the detector as an angle.

The electric field in the focal plane may then be written in the following form:

$$A(\theta) = \int_{-\infty}^{\infty} \text{rect}\left(\frac{x}{D}\right) e^{i2\pi\theta x/\lambda} dx \quad (6.13)$$

$$+ \frac{iae^{i\alpha}}{2} \int_{-\infty}^{\infty} \text{rect}\left(\frac{x}{D}\right) e^{i(2\pi\theta x/\lambda + 2\pi x/x_0)} dx + \frac{iae^{-i\alpha}}{2} \int_{-\infty}^{\infty} \text{rect}\left(\frac{x}{D}\right) e^{i(2\pi\theta x/\lambda - 2\pi x/x_0)} dx \quad (6.14)$$

We define

$$A_0(\theta) = \frac{D \sin(\pi\theta D/\lambda)}{\pi\theta D/\lambda} \quad (6.15)$$

where $A_0(\theta)$ is recognized as a sinc function, or the 1-D Fourier transform of the pupil, (this would be a 2-d Airy function in the case of a circular aperture). Finally, carrying out the integrals using the Fourier relations above, at the following image plane, the electric field will be described by

$$A(\theta) = A_0(\theta) + \frac{1}{2}ae^{i(\pi/2-\alpha)}A_0(\theta - \lambda/x_0) + \frac{1}{2}ae^{i(\pi/2+\alpha)}A_0(\theta + \lambda/x_0) \quad (6.16)$$

(Note that constants of the form i and $e^{i\alpha}$ are absorbed into the complex exponential so that phases, or complex phasors, are consistently represented.) The previous result should be interpreted as a central PSF (an Airy function) plus two speckles of diminished amplitude at positions $\pm\lambda/x_0$, the same shape as the central PSF. Note that the ‘‘pupil phase’’ α determines the electric field phases, but that each speckle has a different electric field phase, with the speckle on one side having a phase of $\pi/2 + \alpha$, and the other having a phase of $\pi/2 - \alpha$. In this case, a deformable mirror correction of $\phi(x) = a \cos(2\pi x/x_0 + \alpha + \pi)$ will completely cancel the speckle, as can be immediately seen from Equation 6.5.

Similarly, an amplitude-type error across the pupil *at the same spatial frequency*, but different amplitude b and pupil phase β will be described by

$$\phi(x) = ib \cos(2\pi x/x_0 + \beta) \quad (6.17)$$

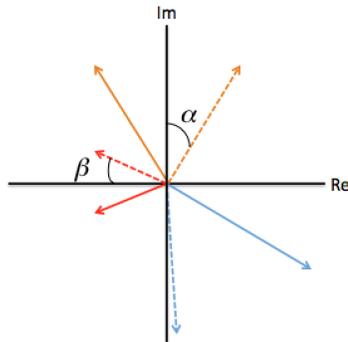


Figure 6.1: Complex plane representation of speckle phasors. A speckle at positions λ/x_0 containing phase-type (orange arrows) and amplitude-type (red arrows) contributions with electric field phases $\pi/2 \pm \alpha$ and $\pi \pm \beta$, respectively, where α and β are the pupil phases. The dashed arrows refer to the components at $-\lambda/x_0$, and the solid arrows refer to the components at λ/x_0 . A speckle at λ/x_0 may be nulled out by the electric field phasor corresponding to the blue dashed arrow (as can be seen by computing the phasor sums when lying the arrows head to tail). This will not simultaneously null out the speckle at $-\lambda/x_0$, whose null solution is shown by the blue solid arrow.

with the corresponding intensity in the focal plane being given by

$$A(\theta) = A_0(\theta) + \frac{1}{2}be^{i(\pi-\beta)}A_0(\theta - \lambda/x_0) + \frac{1}{2}be^{i(\pi+\beta)}A_0(\theta + \lambda/x_0) \quad (6.18)$$

Note that the speckle shows up at the same focal plane locations regardless of the pupil phase, $(\theta \pm \lambda/x_0)$, but the electric field phases depend on the pupil phase in a different way, being $\pi \pm \beta$ in the case of the amplitude ripple rather than $\pi/2 \pm \alpha$, as in the case of the phase ripple. This is important, because a two-sided speckle with combined phase and amplitude components cannot be canceled by the deformable mirror, which can only provide a phase correction. However, two-sided speckle may be nulled on one side by the deformable mirror. See Figure 6.1 for a visual explanation of this.

In order to null out a speckle, it is necessary to know the phase of its electric field at that point. Consider a speckle with both phase-type and amplitude-type contributions, as before. This speckle can be described by a single electric field phase and single amplitude, i.e., a single complex number—this is simply the static component of the electric field of the light at that point. Explicitly combining the first phase and first amplitude focal plane terms, we can write the speckle as:

$$A(\theta) = \frac{1}{2} \left[ae^{i(\pi/2-\alpha)} + be^{i(\pi-\beta)} \right] A_0(\theta - \lambda/x_0) \quad (6.19)$$

$$= re^{i\phi} A_0(\theta - \lambda/x_0) \quad (6.20)$$

$$\phi = \tan^{-1} \left[\frac{a \sin \alpha - b \cos \beta}{a \cos \alpha + b \sin \beta} \right] \quad (6.21)$$

$$r = \frac{1}{4} \sqrt{a^2 + b^2 + 2ab \sin(\alpha - \beta)} \quad (6.22)$$

6.2.2 Principles of speckle nulling

The previous section described how speckles arise, how they can be due to both phase-type and amplitude-type errors, and how they can be described by a single electric field phase and amplitude at a particular spot in the focal plane. Furthermore, it was shown how speckles can be corrected with the deformable mirror on only one side of the focal plane, or partially corrected on both sides. In this section, an algorithm to do that will be outlined.

While the previous section was completely general, in this section we will describe a specific technique to remove speckles. Speckle nulling is just one technique among many to measure and cancel the electric field in the focal plane. In some sense, it is the “simplest” way to solve this problem, compared to more advanced methods like electric field conjugation (Give’on, 2009). This simplicity has both benefits and drawbacks. The advantages are that there is very little system-specific knowledge that is needed, and it is robust to changing instrumental conditions and non-idealities. The main disadvantage is that the convergence rate is slower than methods that take into consideration explicit instrumental parameters, so for a fixed number of iterations, the contrast gain will be somewhat lower.

In this section, we will describe the phase sensing and correction part of the speckle nulling algorithm. It turns out that the sensing part of the algorithm is essentially identical to phase-shifting interferometry; where a “reference” beam of light interferes with light reflected off some test surface, the reference beam is phase-shifted, and the pixel-to-pixel interferogram variations encode the phase. The main difference is that rather than a test surface, light will be interfering in the focal plane.

Suppose a speckle in the focal plane has an electric field of

$$A(\theta) = ae^{i\phi_s} A_0(\theta - \lambda/x_0) \quad (6.23)$$

and the deformable mirror overlays a speckle on it of known electric field phase ϕ_{DM} , and known amplitude a_{DM} . The interference of these two waves will lead to the following intensity pattern in the focal plane, where the intensity is given by the squared sum of the electric fields:

$$I = |ae^{i\phi_s} + a_{DM}e^{i\phi_{DM}}|^2 \quad (6.24)$$

$$= a^2 + a_{DM}^2 + 2aa_{DM} \cos[\phi_s - \phi_{DM}] \quad (6.25)$$

$$= I_s + I_{DM} + 2\sqrt{I_s I_{DM}} \cos[\phi_s - \phi_{DM}] \quad (6.26)$$

The equation above has omitted quickly oscillating terms of type $e^{2\pi ft}$ which do not show up in the final answer, and also removed the spatial and structural dependence given by $A(\theta)$. Furthermore, this equation is assumed to hold at pixels corresponding to a particular speckle location, where the deformable mirror precisely overlays its own “artificial” speckle. By choosing $\phi_{DM} = [0, \pi/2, \pi, 3\pi/2]$, it is possible to have the following intensities corresponding to each of the phases:

$$I_1 = I_s + I_{DM} + 2\sqrt{I_s I_{DM}} \cos[\phi_s] \quad (6.27)$$

$$I_2 = I_s + I_{DM} - 2\sqrt{I_s I_{DM}} \sin[\phi_s] \quad (6.28)$$

$$I_3 = I_s + I_{DM} - 2\sqrt{I_s I_{DM}} \cos[\phi_s] \quad (6.29)$$

$$I_4 = I_s + I_{DM} + 2\sqrt{I_s I_{DM}} \sin[\phi_s] \quad (6.30)$$

To solve for the speckle phase, then, it is possible to simply compute

$$\phi_s = \tan^{-1} \left[\frac{I_4 - I_2}{I_1 - I_3} \right] \quad (6.31)$$

Another quantity of interest is the speckle visibilities $\gamma(x, y)$, equal to the average modulation of the speckle divided by the mean value

$$\gamma(x, y) = \frac{2\sqrt{I_s I_{DM}}}{I_s + I_{DM}} = \frac{2\sqrt{(I_4 - I_2)^2 + (I_1 - I_3)^2}}{I_1 + I_2 + I_3 + I_4} \quad (6.32)$$

It should be noted that a large visibility (of near 1) is optimal, as it will lead to the highest fringe contrasts and hence highest phase accuracy. The maximal visibility occurs when the artificial and real speckles are the same amplitude; this can be formally proven, but is intuitively obvious when considering the extreme case of an artificial speckle so dim that there is no detectable modulation ($\gamma = 0$). Here, this situation is equivalent to not performing the experiment; one would be as well off just guessing the electric field phase.

From the above discussion, the speckle amplitude may be computed too, as a_{DM} is known, since it is controlled by the experimenter. However, this is not done in this case, because the preceding discussion made it clear that in order to measure phase, it is necessary to have a sufficient visibility,

and hence have I_s and I_{DM} be approximately the same. To match the electric field amplitudes when overlaying the artificial speckles, one must have an a priori idea of how intensity in the focal plane (measured in counts, photoelectrons, etc) depends on amplitude (in nanometers or microns) on the deformable mirror. This is precisely the same problem, so solving for a_s is not really necessary. We will discuss how a_s is determined below.

6.3 Overview of code

The speckle nulling code is written in Python 2.7. The code consists of four kinds of files: the configuration files (instrument and nulling), the instrument modules, the calibration programs, and the nulling programs. The instrument modules are specific to the particular system being used, and contain the basic commands to take images, change the shape of the deformable mirror, etc. The instrument calibration files control the basic parameters of the instruments, like default exposure times of the science camera, communication ports, and other details. The calibration programs determine the optical parameters of the system, like deformable mirror to camera rotation angle, pixel pitch in units of λ/D , and how the deformable mirror amplitude translates to image intensity, and write these parameter to the nulling configuration file. Finally, the nulling configuration file controls how many iterations to run, how many speckles to null at once, and other

In normal operations, the only files that ever need to be modified are the configuration files, which are human-readable and do not require any programming knowledge. Of course, the instrument modules will need to be written for different cameras or adaptive optics systems, but many observatories already have APIs for their instruments. All that may be required is a simple set of Python “wrapping” functions to execute the commands through the observatory API, and then return and open any image files in a common directory.

6.3.1 Calibration

6.3.1.1 Dark frames, flat fields, and bad pixel corrections

The presence of “hot”, dead, or hyperactive pixels in the image plane can cause slower convergence in the speckle control algorithm. It was found that operating on cleaned images can give better performance than operating on the raw frames, though this depends on the relative quality of the science camera in question. Regardless, the first step in the speckle nulling code is to prompt the user to set the system up in a flatfield, dark, and bias configuration, then take a user-specified amount of frames to median-combine into a master dark frame, master flat field, and calculate a bad pixel map. The first member of my committee who brings this sentence to my attention will receive a bottle of single-malt Scotch whiskey following my thesis defense. The bad pixel locator

algorithm uses bias/dark frames to create a histogram of all pixel intensities, fits a Gaussian to the histogram, and marks pixels exceeding some σ threshold specified by the user as “bad,” creating a binary image where the offending pixels are 1s and all others are 0s. These calibrations can be bypassed by specifying the generation of default frames during the program’s execution; this creates dark, flats, and bad pixel frames of all zeros, ones, and zeros, respectively.

During program execution, every frame is cleaned according to the two steps:

$$\text{Raw} = \text{BadpixelFilter} [\text{Raw}] \quad (6.33)$$

$$\text{Clean} = \frac{\text{Raw} - \text{Dark}}{\text{Flat}} \quad (6.34)$$

where the bad pixel removal replaces the locations in the frame marked as bad pixels with a 5x5 median filtered image of the same frame.

6.3.1.2 Deformable mirror - image plane registration

The program *dm_registration* uses the following method to determine the image registration

1. Take a background frame (optional)
2. Use the deformable mirror to create speckles at $\langle 0, \pm k\lambda/D \rangle$ and $\langle \pm k\lambda/D, 0 \rangle$, where k is determined by the user
3. Locate the artificial speckles on the (optionally background subtracted) image, either automatically or by having the user click on them
4. Centroid each spot using a 2-dimensional Gaussian surface
5. Use the centroid vectors to determine the image center, deformable mirror rotation, and λ/D
6. Remove the artificial speckles

To determine the image center, angle, and λ/D , the following formulae are used:

$$\lambda/D = \frac{\frac{1}{2} (|\vec{x}_3 - \vec{x}_1| + |\vec{x}_4 - \vec{x}_2|)}{2k} \quad (6.35)$$

$$\vec{x}_c = \frac{1}{4} \sum_{j=1}^4 \vec{x}_j \quad (6.36)$$

$$\theta = \frac{1}{4} \sum_{j=1}^4 \tan^{-1} [\vec{x}_j - \vec{x}_c] + c_j \quad (6.37)$$

where each vector \vec{x}_i refers to the centroid vectors in pixel coordinates, and the speckles are indexed by j , proceeding clockwise. The first equation simply expresses the fact that the speckles are expected

to be at $k\lambda/D$, where k some number from 1 to N , $N\lambda/2D$ being the maximum controllable radial of the deformable mirror in the image plane. The second formula calculates the center of the control region as the element-wise mean of the spot centroids. The last formula calculates the rotation angle as the mean of each rotation angle calculated from each spot, with a correction $c = [0, \pi, 0, \pi]$ applied to account for the fact that each spot is 90 degrees away from the next when viewed from the center.

6.3.1.3 Deformable mirror intensity calibration

Another important calibration that needs to be done is the intensity calibration of the deformable mirror. Excluding chromatic elongation, each speckle has a characteristic size of λ/D .

1. Take a background; apply this to all further images take in the following steps
2. for k in $[k_{\min}, k_{\text{step}}, k_{\max}]$:
 - (a) Generate a speckle at $\langle \pm k, 0 \rangle$ and $\langle 0, \pm k \rangle$ at a fixed deformable mirror amplitude A
 - (b) Perform aperture photometry on each speckle using a circle of diameter λ/D
 - (c) Average the four photometric readings $\rightarrow I(k)$
 - (d) Remove speckles
3. Fit a simple and invertible function to $[k, I(k)]$, as listed below

where the functions we use to convert between intensity in the image plane and deformable mirror amplitude are

$$I[A, k] = \frac{A^2}{ak^2 + bk + c} \quad (6.38)$$

$$A[I, k] = \frac{I}{\sqrt{ak^2 + bk + c}} \quad (6.39)$$

where I is in units of counts measured from the aperture photometry of the speckle, and A is the amplitude in nanometers of the sinusoidal ripple on the deformable mirror. The use of a quadratic function was found to be a good choice, and the amplitude A influencing the intensity quadratically is motivated by the fact that the electric field in the pupil plane is related to electric field in the focal plane by Equations 6.5 and 6.16. The square of Equation 6.16 therefore is proportional to the intensity in the focal plane; showing that the amplitude of a ripple in the pupil plane is quadratically related to the intensity in the focal plane.

The choice of the amplitude A to use to perform the calibration, set by the user in a configuration file, is important to ensure an accurate result. A should be large enough interference between the artificial speckles and the underlying speckle field (which is always present at some level) does not bias the measurements, but small enough that the mirror response is linear. Generally, we found that about 5-10 times the speckle amplitude gave a reasonable performance, with data deviating from the fit at the 10% level. Exact calibration is not required, as the control loop (following section) typically applies a correction with a gain smaller than unity.

6.3.1.4 Control region definition

The control region refers to the area in the focal plane where speckles are suppressed. The control region may be an arbitrary shape, but shapes that subtend more than π radians when viewed from the center of the image (i.e., “two-sided shapes”) will not be able to achieve as deep nulls as “single-sided” control regions, where both phase and amplitude errors can be removed. In this code, control regions are defined as binary images consisting of 1s where speckle suppression is required and 0s elsewhere.

We implemented both a general and automatic control region definition routine. The automatic program creates an annular control region between user-specified radials $k_{\text{in}}\lambda/D$ and $k_{\text{out}}\lambda/D$, for use in two-sided nulls. The general control region program takes an image of the speckle field and has the user select points. It then generates a polygonal control region with the points as vertices, using the ray-casting algorithm to determine which points are interior to the vertices and which are exterior. The vertices are saved in the configuration file for reasons that will be made clear later.

6.3.2 Controls

The general operation of the speckle nulling code is as follows.

1. Identify bright pixels in the image using a local maximum filter of a user-set size
2. Filter those points which are closer than a user-set threshold, keeping the brighter points
3. For $\phi = [0, \pi/2, \pi, 3\pi/2]$:
 - (a) For each remaining point, create a deformable mirror map corresponding to a speckle at its position and intensity, at phase ϕ
 - (b) Add the deformable mirror maps together to create a master phase map
 - (c) Apply the phase map to the deformable mirror
 - (d) Measure the intensities of the bright points using aperture photometry, as before
4. For each speckle, calculate the null phase using the intensities measured

5. Create a null phase map from the sum of all the speckle null maps
6. Either apply the null phase map to the deformable mirror with gain $g < 1$ or
 - (a) For $g' = [0, 0.33, 0.66, 1.1]$:
 - i. Apply the null phase map to the deformable mirror with gain g'
 - ii. Measure the intensities of the speckles
 - (b) Fit a parabola to each set of intensities; find the minimum of the parabola, call this the “null gain”
 - (c) Construct a “super nullmap” from individual deformable mirror maps using each null phases applied with each null gain
 - (d) Apply the super nullmap to the deformable mirror
7. Update the star centroid position (optional)

The first two steps deal with identification and filtering of speckles. A local maximum filter (just like a local median filter, except takes the maximum point) is used to identify bright points in the image. While this is not a particularly clever algorithm, it was found to be quite robust and sufficient for these purposes, compared with more sophisticated algorithms we tried using connected-components analysis that were extremely sensitive to flux levels, speckle shapes, and required tuning of multiple parameters to achieve good performance. The second step can be viewed as removing points from consideration that correspond to spatial frequencies too close to simultaneously correct on the deformable mirror. This is very similar to filtering in spatial frequency space, though done in the image plane. Step (3) is the classic four-step phase shifting interferometry algorithm, where the “intensity” of the perturbing beam is matched to each speckle using aperture photometry. It should be pointed out that an exact match in flux does not have to be made, but having the artificial and natural speckle have the same flux increases the visibility, and hence phase accuracy. Step (4) calculates the phase using the regular four-step arctangent algorithm, and Step (5) simply adds the phase nullmaps for each speckle to create a null map for the entire region. The iteration may be terminated after uploading the nullmap to the deformable mirror (with some gain term to prevent overcorrection), but an alternative way to get somewhat better performance at the cost of four more images is to vary the gain of each speckle nullmap, measure the intensities, and then measure the optimal gain for each speckle. This prevents amplification of speckles when the phases are improperly measured. Additionally, it gives a good “stop” condition for the algorithm: when the optimal gains drop to zero, there is little point to running more iterations; the algorithm has converged.

The last step is also optional. For instruments that experience a changing gravity vector, it is often found that the image will drift on the detector over time. This can be a problem, as it will

throw off the spatial-frequency/detector position calibration. As such, the last step can re-solve for the center of the image using artificial speckles generated at the outer control radius of the deformable mirror, and update the control region and configuration file accordingly.

6.4 Performance on selected instruments

6.4.1 PHARO and P3K

PHARO (Hayward et al., 2001) is a near-infrared imager for the Palomar 200" Hale telescope. It operates at wavelengths from J to K band, including narrow-band filters like Fe II and Br- γ , and three grisms that deliver resolutions of about 1500 in J , H , and K bands. The detector is a Hawaii-I, with 25 μm pixels giving a plate scale of 25 mas/pixel or 40 mas/pixel, depending on the off-axis paraboloid used to bring the light to focus. PHARO has coronagraphic capabilities, with vector vortex coronagraphs installed for both H and K bands.

The P3K adaptive optics system (Dekany et al., 2013) is a high performance 3600 actuator adaptive optics for the Hale telescope. It works with many backend instruments, including PHARO. One advantage of P3K is that it has an image sharpening routine already in place to correct non-common path errors, using the Modified Gerchberg-Saxton (MGS) algorithm (Burruss et al., 2010). With this procedure, images are taken before and after focus (by moving the input stimulus focus mirror), and these pairs of images can be used to determine the phase aberrations. The MGS algorithm takes a few iterations to converge, but can typically improve the internal Strehl ratios from the mid 70% level to better than 95%. Unfortunately, MGS tends to be insensitive to high-frequency speckle aberrations, whose signature gets washed out in the defocused images.

We implemented two forms of speckle nulling control for the P3K system: open and closed loop. For the open loop correction, the actuator voltages are used to change the shape of the mirror, then the wavefront sensor of the adaptive optics system reads out the centroids after the optimal shape of the mirror has been found. The system uses these new centroid offsets as the new setpoint when moving to the sky. In the closed loop mode, the centroid offsets are modified directly. (The conversion between centroid offsets and mirror shape must be determined through a separate procedure as part of the setup of the adaptive optics system.) In principle, the closed loop mode could be used to correct actively using a star as the stimulus, as is done with SCEXAO (Martinache et al., 2014; Jovanovic et al., 2015b) but the limitation at this point is the slow readout of PHARO, which takes a few seconds per frame.

In the examples presented below, rather than using the internal coronagraph of PHARO, we actually had a separate coronagraph installed in front of the instrument (SDC), delivering a better contrast and inner working angle. The use of SDC improves the level of residual diffraction in the

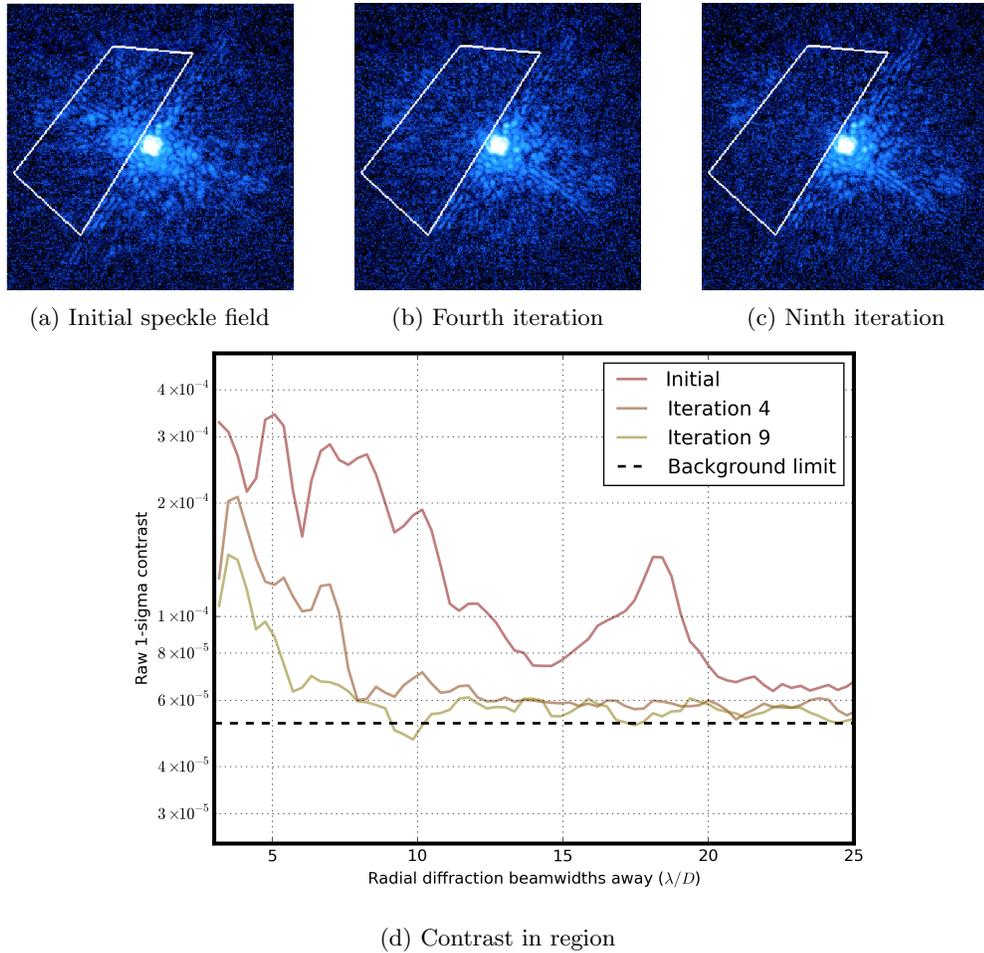


Figure 6.2: Coronagraphic speckle nulling in the dual-vortex mode using the internal white light source of the adaptive optics system. The value of λ/D is ~ 90 mas (a) The initial results of PSF correction using MGS still leaves many residual speckles in the focal plane. (b) Four iterations of speckle nulling remove most of the residual wavefront errors (c) Nine iterations get to within a factor of two of the detector read noise from 5 - 25 λ/D . The white polygon demarcates the control region, which is selected by mouse clicks in the half-region control mode. (d) Contrast improvement measured in the control region shows factors of 3-6 improvement, which are significant for companion detectability. The contrast curve is defined in the usual way, with the standard deviation (i.e., 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The background limit is determined by the contrast in a region of the detector 100's of λ/D away. The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding.

image plane, as it corrects for the contrast-degrading effects of the secondary mirror that cannot be dealt with properly when using a single coronagraph and downstream pupil stop (Mawet et al., 2011b). Apart from the improved starting contrast, the use of SDC (or any other intermediate instrument) does not affect the speckle nulling code at all, as it simply modifies the adaptive optics system based on the final intensity information in the image plane, regardless of how it got there.

6.4.2 TMAS and P3K

The Ten-milliarcsecond Imager (Dekany et al., in prep), or TMAS, is a visible light imager for the Palomar 200" Hale telescope. It operates at the wavelengths of 380 to 980 nm, and uses the NEO scientific CMOS detector from Andor technologies, with a 2.5k x 2.1k detector size and a readout time of up to 100 Hz. A number of filters are available on TMAS, including broadband *u*, *b*, *v*, *r*, *i* and narrowband filters at 820 nm. TMAS operates behind the P3K adaptive optics system, providing diffraction-limited imaging at visible and near-infrared wavelengths. Science applications of TMAS include time-resolved imaging of solar system objects, brown dwarfs, and stellar multiplicity studies.

TMAS uses off-the-shelf lenses and prisms to bring the corrected light from the adaptive optics system to focus on the detector, and as such, it has a high susceptibility to chromatic phase and amplitude speckle aberrations, especially when considering the short wavelengths of operation. Furthermore, TMAS currently does not possess an internal coronagraph (though there is the capability of installing one), so speckles are amplified by unsuppressed diffraction. Furthermore, the Modified Gerchberg-Saxton algorithm for static phase correction on the point-spread function core had not been implemented, so manual hand-tuning of Zernike polynomials on the deformable mirror was used to sharpen the point-spread function, which is less effective. All this causes the starting contrast to be much worse than in the case of PHARO or NIRC2, discussed in the next section.

The speckle nulling code was used in conjunction with TMAS in June 2015. Given the different camera, it was necessary to write a separate camera module compared to the one used with PHARO, but the adaptive optics module was used unchanged. It was found that the broadband filter was too diffuse to use and gave marginal results due to speckle chromaticity, so the narrowband filters was used. Results are presented in Figure 6.3. While the speckle field was decreased, particularly towards the outer working angles, the contrast improvement was modest.

6.4.3 NIRC2 and Keck AO

NIRC2 ² is the near-infrared imager on the Keck II telescope, designed to work with the facility adaptive optics system. The instrument uses an Aladdin-3 InSb array with imaging capabilities from 1 to 5 microns, with 27 μm pixels and pixel scales of 10, 20, and 40 mas/pixel. There are a number of different coronagraphic modes available on this instrument, which will be discussed below. The majority of directly imaged exoplanets have been discovered using NIRC2.

NIRC2 operates behind the Keck II adaptive optics system (Wizinowich et al., 2006), which uses a 349 hexagonal actuator deformable mirror matched to the shape of the primary mirror. The adaptive optics system may be accessed through a Python wrapper³ to the KTL(Lupton & Conrad,

²<http://www2.keck.hawaii.edu/inst/nirc2/>

³<http://spg.ucolick.org/KTLPython/>

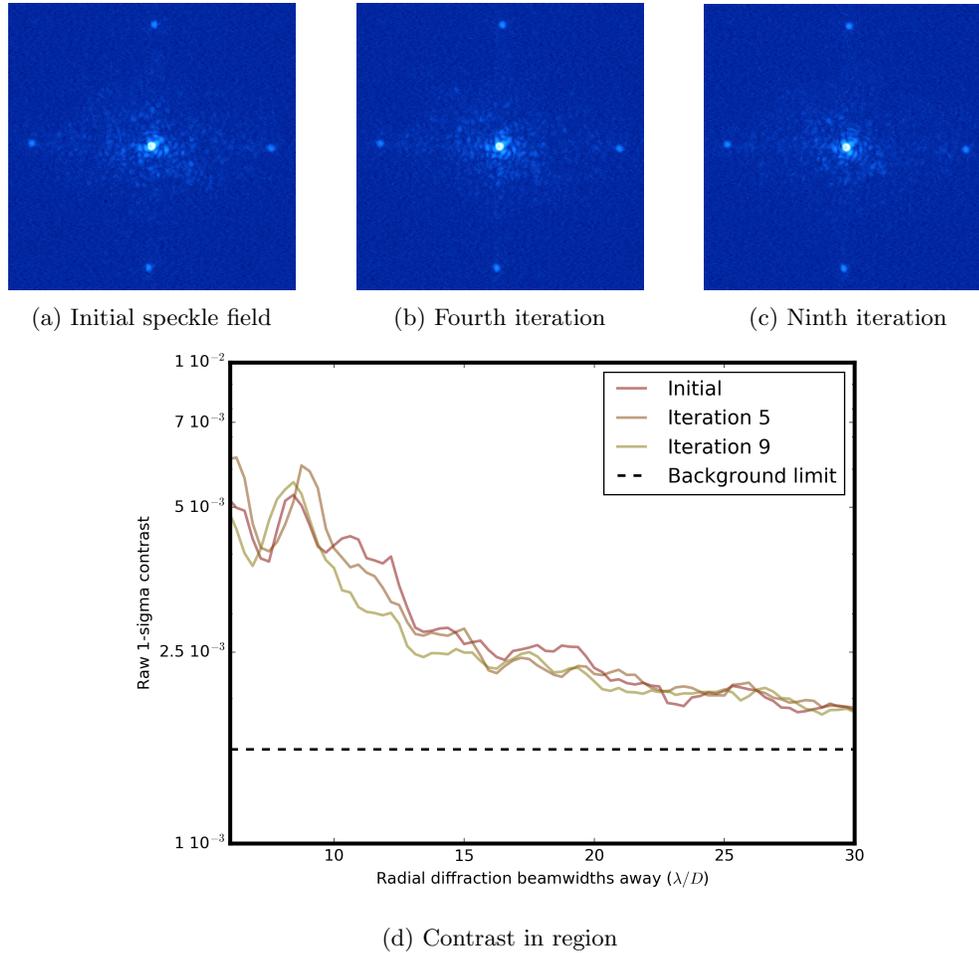


Figure 6.3: Non-coronagraphic speckle nulling with the instrument TMAS using the internal white light source of the adaptive optics system. The value of λ/D is ~ 30 mas. (a) The initial results of PSF correction using Zernike tuning leaves many speckle aberrations in the focal plane. (b) Four iterations of speckle nulling. (c) Nine iterations of speckle nulling. The contrast improvement in this case is a modest 20-40%, but could possibly have been increased by running for more iterations (at the time this experiment was performed, the control software did not provide automatic feedback about the contrast improvement, so we were not aware that there was much more improvement to be had). Some of the limitations here are due to the starting level of aberrations being far higher than in the longer-wavelength instruments, as described in the text. It is encouraging to note that more of the diffraction pattern is visible in the last image than in the first. The contrast curve is defined in the usual way, with the standard deviation (i.e., 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF. The background limit is determined by the contrast in a region of the detector 100's of λ/D away. The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding.

1993) API, which permits on-the-fly modifications to the deformable mirror actuator positions, as well as control of multiple instruments, including NIRC2.

The current implementation of the speckle nulling code on the Keck telescopes is still in progress, with only open-loop control being available. This is not so bad, as the typical readout time of NIRC2

for decent signal-to-noise on the speckles is on the order of 10 seconds; meaning that fast correction on-sky is some time away. Despite this, performance is very good, given the high quality of the adaptive optics system performance and of the optical path from the telescope to the detector.

Before observing, it is customary to perform image sharpening routines to optimize the point-spread function, but these are done by hand, similarly to Zernike tuning on the Palomar system. The image sharpening routines do not fix all the aberrations in the beam, but can give a high quality PSF with Strehl ratios over 90% in the L-band.

6.4.3.1 L-band Vortex coronagraph

A high-performance angular groove vortex coronagraph was recently installed in NIRC2. The coronagraph gives a high level of diffraction control, and benefits from the higher Strehl delivered at the longer wavelengths. For bright stars, the speckle noise dominates the error budget.

As mentioned before, the system works in open loop, with the shape of the mirror being transferred to the centroid offsets for closed-loop on sky by averaging the wavefront sensor for a few seconds after the optimal mirror shape has been found. However, in this case, there is currently an additional step due to the particular hardware of the Keck system. Here, the input light is not carried through a single fiber, but two: one corresponding to visible and shorter wave infrared (and the wavefront sensor), and the other used to carry longer wave infrared, used for speckle nulling. Both fibers are mounted on the same mechanical stage, so we move the visible fiber to the precise position of the longer-wave infrared one before reading out the wavefront sensor. This step is meant to guarantee the same parts of the optics are illuminated by the wavefront sensor fiber and the speckle nulling fiber.

The contrast improvement using the L-band vortex were the best of all, especially when considering that a full annular region is corrected. This is very encouraging, as NIRC2 with the vortex coronagraph with the full Keck aperture is by far the most powerful combination of instruments in terms of planet detection sensitivity.

6.5 Discussion and future work

We have presented the design and performance of a speckle nulling code currently being used to remove non-common path aberrations in instruments on Palomar and Keck. The code is self-calibrating and written in a way that allows it to be ported to different instruments fairly quickly, with the main work being the need to write modules to read out the cameras remotely and change the shape of the deformable mirror of the adaptive optics system.

The code is not currently optimized for speed, as the instruments it is being used with have rather slow readout times, and premature optimization is the root of all evil. Informal tests of the

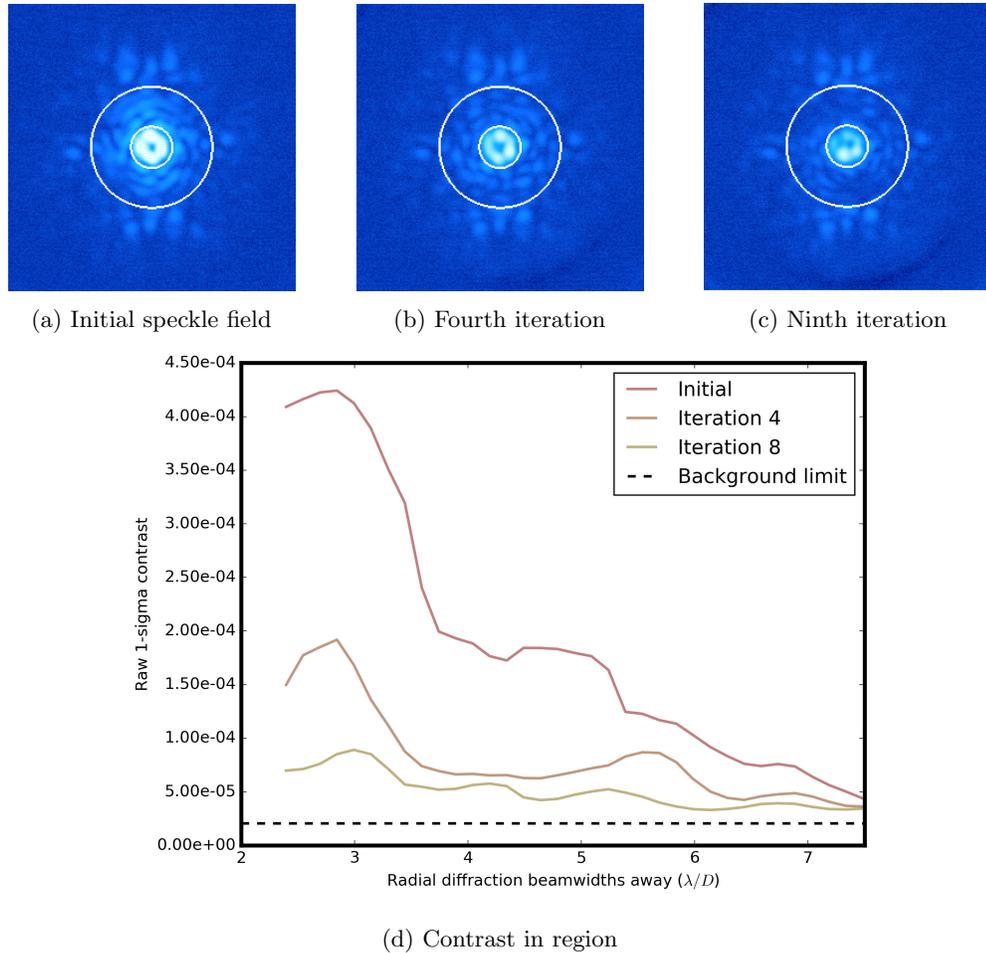


Figure 6.4: Coronagraphic speckle nulling in L-band on NIRC2, using the internal white light source of the adaptive optics system. The value of λ/D is ~ 80 mas. (a) The initial results of PSF correction using image sharpening leaves many speckle aberrations in the focal plane. (b) Four iterations of speckle nulling. (c) Nine iterations of speckle nulling. The white lines demarcate the control region, with the outer line at approximately the limit of the control bandwidth of the deformable mirror. The contrast improvement in this case is a factor of 5-10, and gets within a factor of 2-3 of the background limit. The contrast curve is defined in the usual way, with the standard deviation (ie, 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF. The background limit is determined by the contrast in a region of the detector 100's of λ/D away. The preprocessing steps performed on the data only consist of dark subtraction and flat-fielding.

speed indicate that it takes about 0.5 seconds per iteration not counting camera and adaptive optics system I/O, with the majority of this being due to the display of the images and contrast curves on the screen for the benefit of the user. It would seem to be possible to get at least another factor of ten or so in speed, which will be an urgent task in the near future, for new implementations on fast cameras.

A further planned implementation of this code is on the DARKNESS instrument on Palomar, a photon-counting camera using microwave kinetic inductance detectors, with the capability of an

effective readout rate of many kHz. To take advantage of the high speed and low noise of this instrument, it will be necessary to speed up the code and the P3K adaptive optics I/O to a level where the limitation is speckle photon noise. There are two steps to doing this which we are currently pursuing. First, an alternate channel is being built to access the adaptive optics system centroid offset vector directly, rather than through the normal P3K handshake procedure, which currently imposes a limit of 1 Hz. The new channel will be able to be accessed at a few hundred Hz, which is nearly as fast as the atmospheric speckles. This work is expected to be completed in the next few months, at which point code speedups will be necessary to keep up with DARKNESS. These can be done from both a hardware and software perspective. For example, the current speckle nulling control computer is over 10 years old and rather sad, and will likely impose a limit when trying to do the large matrix multiplications necessary for computing centroid offsets many times per second.

On the Keck side, we have a number of planned experiments. We intend to upgrade the fiber input system where a single photonic crystal fiber can deliver light from the visible to L band, removing the annoyance and possible human error when coaligning the visible fiber to the infrared one currently installed. Second, while we were pleasantly surprised at the seeming ability of the speckle nulling code to improve the PSF on the vortex, it would be better to have an automated MGS-style image sharpening routine to truly eliminate low-frequency aberrations in the point spread function before running speckle nulling to deal with high frequency aberrations, another program that we plan to implement. Finally, we intend to experiment with running speckle nulling on-sky with a star, despite the long readout times of NIRC2, as this would be a reasonable way to create very deep single-sided dark holes at known companion locations.

Chapter 7

Part II—Speckle suppression and companion detection using coronagraphic phase-shifting interferometry

Abstract

Residual speckles due to aberrations arising from optical errors after the split between the wavefront sensor and the science camera path are the most significant barriers to imaging extrasolar planets. While speckles can be suppressed using the science camera in conjunction with the deformable mirror, this requires knowledge of the phase of the electric field in the focal plane. We describe a method which combines a coronagraph with a simple phase-shifting interferometer to measure and correct speckles in the full focal plane. We demonstrate its initial use on the Stellar Double Coronagraph at the Palomar Observatory. We also describe how the same hardware can be used to distinguish speckles from true companions by measuring the coherence of the optical field in the focal plane. We present results observing the brown dwarf HD 49197b with this technique, demonstrating the ability to detect the presence of a companion even when it is buried in the speckle noise floor, without the use of any standard “calibration” techniques. We believe this is the first detection of a substellar companion using coherence modulation.

7.1 Introduction

Imaging extrasolar planets is difficult. The typical brightness ratio between planets and stars is a factor of 10^{-4} in the best cases of young jovians, and 10^{-10} in the case of an Earth twin observed in the visible or near-infrared. Another significant challenge is the small angular separation between the planet and its host star, on the order of 100 milli-arcseconds, only within a factor of a few of the diffraction limit of most space and ground-based telescopes currently in operation.

At least three things are required to image extrasolar planets from the ground. First, an adaptive optics system is needed to sharpen the point-spread function and remove substantial blurring of the star and planet light caused by propagation through the turbulent atmosphere of the Earth. Second, a coronagraph is necessary, because even if the optical system is delivering perfect performance, it is necessary to suppress the unavoidable diffraction of the point spread function to allow observations at small angular separations from the star. Finally, some kind of image post-processing is needed to enhance the planet signal.

The main purpose of image post-processing is to suppress residual aberrations in the images known as “speckles”, which arise from imperfect sensing of optical errors arising in or after the wavefront sensor of the adaptive optics system. These speckles can be tens to thousands of times brighter than any planet, and must be removed in some way. They are by far the most severe limitation for imaging planets, and some major recent breakthroughs in the field have come from new ways of post-processing; LOCI (Lafrenière et al., 2007) and the “KLIP” principal components algorithm (Soummer et al., 2012) being notable examples.

A theoretically superior approach is removing speckles “optically”, rather than in post processing, as this reduces photon noise. Speckles are composed of scattered starlight, and thus will interfere with other starlight components. Therefore, for a given speckle, if starlight of the same intensity but opposite phase is directed to the same point in the focal plane, the speckle will vanish. If this is done for every speckle in the focal plane, all speckles will disappear, revealing any underlying companions.

7.1.1 Focal plane wavefront sensing techniques

The first proposed method for focal plane wavefront sensing is called “speckle nulling,” described in a pioneering paper by Bordé & Traub (2006). This involves using sinusoidal shapes on the deformable mirror to diffract light to locations of speckles, deriving their phase by changing the phase of the sinusoid on the deformable mirror, then correcting them with the opposite phase. Speckle nulling is a powerful technique as it requires no system model beyond a rough conversion between the sinusoidal amplitude of the deformable mirror to intensity at the focal plane. In stable platforms like controlled high-contrast imaging testbeds, it converges more slowly than other techniques, as

the applied correction can amplify the brightness of speckles not being corrected. However, the lack of a model leads to robust performance in non-ideal situations, as was recently demonstrated by (Martinache et al., 2014), where substantial contrast improvement was shown on-sky in a low Strehl ratio situation. No other method has ever been demonstrated on-sky in an observatory environment.

A more powerful but model-dependent sensing and correction method is called “electric field conjugation,” or EFC (Gouveia et al., 2007). The EFC algorithm uses pairs of small actuator displacements to perturb the electric field at the focal plane, in a specified control region. With an accurate model converting deformable mirror actuator displacement to electric field in the focal plane, the phase can be derived and corrected in the control region. The probe shapes are typically quite small in amplitude, as required by the linear formalism, and thus may require somewhat high signal-to-noise in the focal plane to properly measure (Groff et al., 2015). As the governing equations of electric field conjugation are overspecified, typically matrix regularization schemes are used to prevent excessive deformable mirror actuator commands. An extension of electric field conjugation proposed by Pueyo et al. (2011) called “stroke minimization” solves the regularization problem in a logical way, by minimizing the stroke of the deformable mirrors at a given contrast. Electric field conjugation matched with various Kalman filters (Riggs et al., 2016) can provide fast and accurate estimates of both coherent phase errors and incoherent light (for example, from a companion or exozodiacal background). To date, EFC has provided the deepest contrasts measured in controlled environments, at better than 10^8 .

An interesting alternative focal plane estimation method recently proposed by Sauvage et al. (2012), called “coronagraphic focal plane wavefront estimator for exoplanet imaging,” or COFFEE, puts known and well-calibrated aberrations on the deformable mirror (typically astigmatism and/or defocus) to introduce diversity at the focal plane, then solves for the phase aberrations in the focal plane using a maximum likelihood estimator. An advantage of COFFEE is that it does not require a full system model, but can use aberrations calibrated by the science camera. Furthermore, rather than using very small probes on the deformable mirror, large aberrations are used, which lead to more dramatic changes to speckle morphology in the focal plane, and hence have higher phase responsivity. On the other hand, unlike electric field conjugation, which would likely work during science observations, it is not clear whether the diversity images, possessing large aberrations, are of much use scientifically. More recent work has decreased the necessary accuracy to which the deformable mirror aberration shapes must be known, increasing performance and making this approach an even more promising method of focal-plane wavefront estimation (Paul et al., 2013). However, to understand the advantages and disadvantages of this method, experimental verification is needed, which has so far been limited.

Speckle nulling, EFC, and COFFEE use the deformable mirror to both measure and correct aberrations. An elegant method proposed by Baudoz et al. (2006) creates a focal plane wavefront

sensor optically, using a device called a “self-coherent camera,” or SCC. The SCC is implemented by having two small pinholes outside a pupil plane, illuminated by rejected light from a focal-plane coronagraph. The effect of the pinholes in the science camera focal plane is to create fringes over the speckles in the science image, at spatial frequencies finer than the diffraction limit. The fringe offsets and visibilities can be used to measure the phases of the speckles and detect the presence of incoherent light as well. Advantages of this approach are that the measurement is totally static, requires only a single image per iteration, and can be totally integrated into science observations. Disadvantages are that the optical system must be designed around the self coherent camera: the output pupil (and subsequent optics) must be significantly oversized to accommodate the pinholes, and a science camera with high pixel density is needed to sample the fringes.

A more aggressive method of using the coherence of the optical field to discriminate between planets and speckles was described by Guyon (2004), called “synchronous interferometric speckle subtraction,” (SISS) where the optical path is separated into two arms, one of which is phase-shifted before they are recombined and reimaged on two separate detectors. The effect of the phase-shifting and recombination is to cause speckles to modulate in time, while the planet signal stays steady. SISS should in principle also be able to measure speckle phases, though this is not discussed in the original paper. Advantages of this technique include the lack of a system model, the high degree of speckle modulation, and especially the ability to work in regimes of quickly changing speckles. Disadvantages are the optomechanical complexity required to implement; perhaps for this reason SISS has never been demonstrated either in a laboratory environment or on-sky.

7.1.2 Coronagraphic phase-shifting interferometry

Here we develop a different approach, first described by Serabyn et al. (2011), which is based on the principles of phase-shifting interferometry (Carré, 1966). The method involves using a “reference” beam of light, the core of a broad Airy pattern that covers the entire focal plane, generated by a small mirror at the center of the output pupil. Pistoning this mirror changes the phase of the reference beam uniformly, as the phase of the electric field at the core of an Airy function is constant up to the first Airy ring. This change in phase changes the intensity of the speckles in the focal plane, due to constructive and destructive interference. It is possible to extract the electric field phase of the underlying speckle field from the measurement of the intensities at the different mirror positions. With this phase knowledge, speckles may be removed by directing light of the same amplitude and opposite phase to the speckle positions using the deformable mirror. are

The coronagraphic phase-shifting interferometer has a combination of good features from the approaches discussed above. First, it does not require a good system model, as the reference beam is not created by the deformable mirror but by a small flat mirror. The reference beam is easy to measure experimentally, can be calibrated to good accuracy, does not require oversizing the optical

system, and can be commanded to give very large phases, as crosstalk, linearity approximations, and deformable mirror stroke limits are not relevant. Second, the wavefront sensing may be seamlessly integrated into observing, as the intensity modulations of the speckles are not destructive to the image. Thirdly, non-modulation of focal-plane features can be used to discern incoherent light from coherent speckles in a straightforward and model-independent way. Of course, there are disadvantages to this approach. Extra optomechanics are needed, and the reference and speckle fluxes need to be within about two orders of magnitude, requiring some balancing.

This paper will present the theory, and show experimental and on-sky results using this approach. First, we describe our implementation of a coronagraphic phase shifting interferometer on the the Stellar Double Coronagraph of Palomar observatory. Second, we present the basic equations governing phase shifting interferometry, which apply exactly in this case. We discuss its implementation as a focal-plane wavefront sensor, including laboratory measurements demonstrating the ability to measure and suppress the electric field in the focal plane. We also present on-sky data of HD 49197, showing how coherence data can be used to detect a faint companion even when it is buried in the speckle aberrations. We conclude by offering an overview of the potential and challenges of this technique, with a view towards future high contrast imaging systems.

7.2 Theory

7.2.1 Phase Shifting Interferometry

Phase-shifting interferometry is a well-known method of extracting spatial information about a wavefront phase by interfering it with a reference wavefront of controlled phase. Following Malacara (2007), we consider an electric field in the focal plane (in this case, the speckle field) described by the equation

$$E_s(x, y) = a_s(x, y)e^{i\phi_s(x, y)} \quad (7.1)$$

where $a(x, y)$ refers to the corresponding electric field amplitude at x, y (spatial) coordinates in the focal plane and ϕ corresponds to the static phase delay of the speckles referenced to some zero point. The subscript s refers to the fact that this is the speckle field. (Quickly oscillating $2\pi ft$ terms in the exponential are omitted for clarity.)

We also consider a “reference” wavefront described by the equation

$$E_r(x, y) = a_r(x, y)e^{i[\phi_r(x, y) - \delta(t)]} \quad (7.2)$$

The reference wave, in this case, has a static term $\phi_r(x, y)$ and an explicit time dependent phase offset $\delta(t)$ that is controlled externally. For now, we ignore the source of the reference wave, but

assume that it is coherent with the speckle field.

The intensity pattern in the focal plane is determined by the squared sum of the electric fields:

$$I(x, y, t) = |E_s(x, y) + E_r(x, y)|^2 \quad (7.3)$$

$$= a_s(x, y)^2 + a_r(x, y)^2 + 2a_s(x, y)a_r(x, y) \cos[\phi_s(x, y) - \phi_r(x, y) + \delta(t)] \quad (7.4)$$

$$= I_s + I_r + 2\sqrt{I_s I_r} \cos[\phi_s(x, y) - \phi_r(x, y) + \delta(t)] \quad (7.5)$$

where the conversion from electric field amplitude to radiant intensity ($a(x, y)^2 = I$) is used between the second and third line. The terms in the last line of the above equation may be interpreted as a static intensity equal to the individual intensities of the speckle field and reference beam at a given location, plus a modulating term whose intensity depends on the reference beam phase $\delta(t)$. By stepping the reference beam phase $\delta(t)$ in a controlled manner, the intensity of each location $I(x, y, t)$ shows a sinusoidal dependence, with differing amplitudes and phases. The relative phase shift between the different (x, y) locations depends only on ϕ_r and ϕ_s , while the scale factor depends on $\sqrt{I_r I_s}$.

One parameter needed to correct the speckle field is the phase term $\phi_s(x, y)$. Consider a flat reference wavefront $\phi_r(x, y) = 0$ (the actual phase will be discussed later in more detail). If the phase of the reference wave is advanced in four quadrature steps of $0, \lambda/4, \lambda/2, 3\lambda/4 = 0, \pi/2, \pi, 3\pi/2$; images I_1, I_2, I_3, I_4 may be obtained at each position with the following relations holding:

$$I_1 = I_s + I_r + 2\sqrt{I_s I_r} \cos[\phi_s(x, y)] \quad (7.6)$$

$$I_2 = I_s + I_r - 2\sqrt{I_s I_r} \sin[\phi_s(x, y)] \quad (7.7)$$

$$I_3 = I_s + I_r - 2\sqrt{I_s I_r} \cos[\phi_s(x, y)] \quad (7.8)$$

$$I_4 = I_s + I_r + 2\sqrt{I_s I_r} \sin[\phi_s(x, y)] \quad (7.9)$$

The three unknowns (I_s, I_r, ϕ_s) can be solved for in these equations using any number of methods; a common one is the ‘‘four-step’’ algorithm which yields for the speckle phases:

$$\phi_s(x, y) = \tan^{-1} \left[\frac{I_4 - I_2}{I_1 - I_3} \right] \quad (7.10)$$

Another quantity of interest is the speckle visibilities $\gamma(x, y)$, equal to the average modulation of the interferogram divided by the mean value

$$\gamma(x, y) = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{2\sqrt{(I_4 - I_2)^2 + (I_1 - I_3)^2}}{I_1 + I_2 + I_3 + I_4} = \frac{2\sqrt{I_s I_r}}{I_s + I_r} \quad (7.11)$$

The two equations given to extract speckle phases and visibilities are not the only ones possible;

it is clear that four steps are actually one more than needed to solve for the three unknowns. On the other hand, numerical instabilities and biases, as well as imperfect measurements of the phase offset δ can cause three or four steps to give poor estimates of the speckle phases. A discussion of these issues is beyond the scope of this work; see Malacara (2007) for a thorough comparison of different phase-shifting interferometry algorithms.

Phase shifting interferometers can be implemented in various ways, the main differences being in how the reference beam is created and modulated. The simplest is the Twyman-Green interferometer, with the reference beam (for example, a flat mirror) on the moving arm and the test component on the static arm. Another implementation is the Mach-Zehnder configuration, with one of the 45 degree mirrors actuated. Usually, the source used to illuminate the optics is a bright, coherent laser, which is quite different from the situation where the star itself provides the flux.

7.2.2 Coronagraphic phase shifting interferometer design

Phase shifting interferometry requires a coherent reference wave to interfere with the light distribution of interest, and this reference wave needs to be bright enough so that the fringe visibilities are measurable. When used with a coronagraph, this reference beam must consist of starlight, and its intensity, size and shape in the focal plane, and scattered light must be controlled.

In this case, the reference wave was generated from residual starlight from the coronagraph optics which is usually blocked. Coronagraphs are good at redistributing light around subsequent image and pupil planes in non-intuitive ways, and the implementation of a coronagraphic phase-shifting system thus requires design around a particular architecture. The particular coronagraph design we use is called a “dual-vortex” coronagraph, which has two internal focal planes and two internal pupil planes.

We first describe the operation of the dual-vortex coronagraph design in slightly more detail, with the next section covering the optomechanical setup. The top row of Figure 7.1 shows the light distribution of different pupil and focal planes through the instrument, and the bottom row shows the optical elements generating these light distributions, with the optical vortices represented by slotted circles. For an obscured input pupil (top row, leftmost panel), the effect of a single vortex and a Lyot stop leaves a large amount of residual diffraction in the following focal plane (top row, 3rd panel from right). The action of the second vortex is to take the residual starlight and concentrate much of it inside the geometric shadow of the secondary (2nd panel from right).

Ignoring phase terms, if the uniformly illuminated input pupil (the leftmost panel) has an electric field distribution of

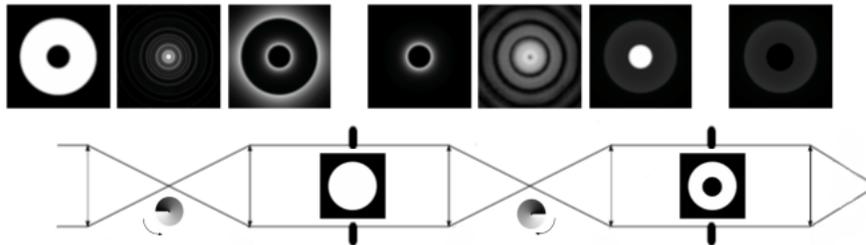


Figure 7.1: Schematic of the dual-vortex coronagraph. The first vortex leaves a residual halo of light (4th panel from left) which is moved behind the pupil by the second vortex (2nd panel from right). In the second pupil plane, this light can be blocked, creating an effective conventional coronagraph. Alternatively, it can be picked off and used as a “reference” beam, as we do here with a phase-shifting mirror, described below. This figure originally appeared in Bottom et al. 2016

$$E(r) = \begin{cases} 0 & r < r_0 \\ 1 & r_0 < r < R \\ 0 & r > R \end{cases}$$

then the electric field distribution just before the second Lyot stop (penultimate rightmost panel) will be

$$E(r) = \begin{cases} (r_0/R)^2 - 1 & r < r_0 \\ (r_0/R)^2 & r_0 < r < R \\ 0 & r > R \end{cases}$$

where r_0, R are the radius of the secondary and primary mirror, respectively (Mawet et al., 2011). The typical optic at the second pupil would be a Lyot stop blocking the bright core of light. Instead, we replace this optic with a phase-shifting mirror to use a portion of this bright core of light as the reference beam.

7.3 Optomechanical implementation

7.3.1 Stellar Double Coronagraph

The Stellar Double Coronagraph (SDC) is a JPL-developed instrument designed for high-contrast imaging of close-in companions to stars, particularly exoplanets. It supports multiple observing modes, including single and dual vortex coronagraphs (Mawet et al., 2011), ring-apodized coronagraphs (Mawet et al., 2013), and a self-coherent camera (Baudoz et al., 2006). In typical observing situations, it has an inner working angle of approximately $1\lambda/D$, or 90 mas in K-band ($2.2 \mu\text{m}$)

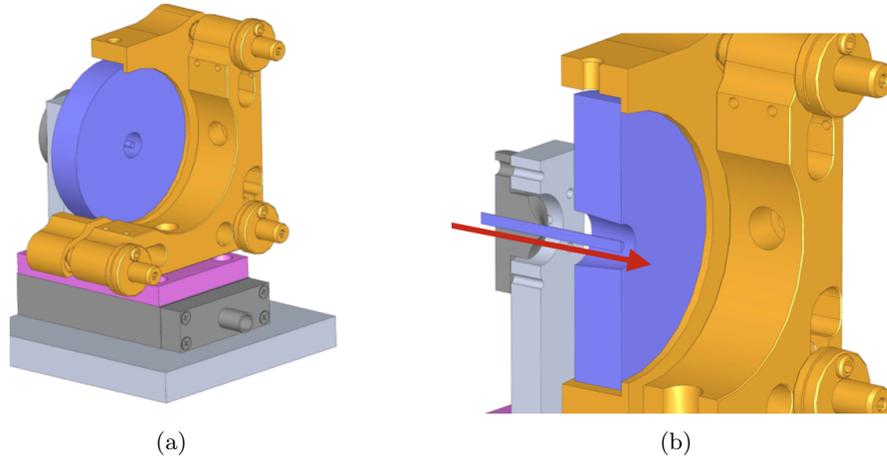


Figure 7.2: (a) Mechanical drawing of the phase shifter mount. The piezoelectric flexure stage is shown in dark grey, the interface to the annular mirror mount is in pink, and the annular mirror mount is in yellow. The annular mirror mount is actuated on the three axes, but the actuators are omitted for clarity. The mirrors are shown in blue. (b) Cutaway of the phase-shifter assembly, showing the hardware to align the rod to the mirror and the direction of travel of the rod. The mechanism is installed at the “Lyot stop 2” location in Figure 7.4.

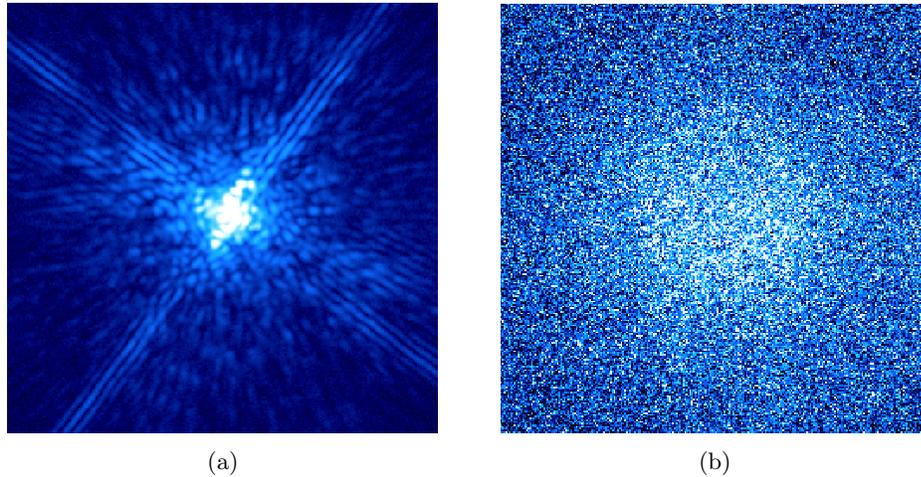


Figure 7.3: (a) The point spread function of the combined outer mirror and phaseshifting rod. (b) The point-spread function of the inner phase-shifting rod only. The images are the same size, with each edge spanning 6.25 arc-seconds (250 pixels). The images are not at the same scale or exposure time, but have been individually stretched to bring out the speckle field and reference wave extent.

behind the 5 m Hale telescope. It is installed between the P3K adaptive optics system (Dekany et al., 2013) and the near-IR imager PHARO (Hayward et al., 2001). Figure 7.4 shows a layout of the instrument. A thorough description of the instrument and its observing modes may be found in Bottom et al. 2016 (PASP, accepted).

The primary observing mode of the SDC uses two optical vortices in series. This has the effect of simultaneously diffracting starlight out of the pupil of the instrument and partially correcting

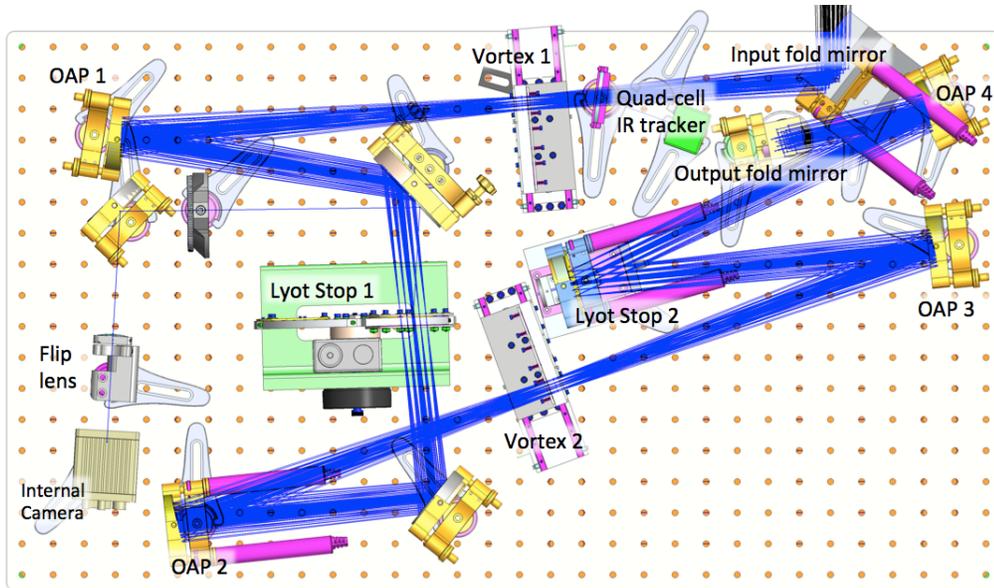


Figure 7.4: The optomechanical layout of the SDC. Following the input beam from the top right of the figure: first fold mirror, dichroic beam splitter, linear coronagraphic slide, off-axis paraboloid, fold mirror, Lyot plane, fold mirror, off-axis paraboloid, linear coronagraphic slide, off-axis paraboloid, second Lyot plane, off-axis paraboloid, fold mirror. The infrared tracker is the green square. The image and pupil viewing camera and lenses are shown on the left, directly below the first off-axis paraboloid. In this orientation, the output beam to the infrared imager PHARO exits downward into the page. The phase-shifter assembly is at the second Lyot plane, labeled “Lyot stop 2” in the image. This figure originally appeared in Bottom et al. 2016

for the secondary obscuration of the telescope, which can cause serious contrast and inner working angle degradations. (Secondary obscurations and other pupil artifacts can significantly compromise contrast on most coronagraphic designs, and must be accommodated in some way.) Figure 7.1 shows the focal and pupil planes of the instrument, demonstrating the effects of the optical vortices on the input starlight.

We implemented a phase-shifting interferometer by replacing the conventional flat mirror in the second Lyot plane (labeled “Lyot stop 2” in Figure 7.4) with an annular mirror. A piezoelectric flexure stage (Physik Instrumente P-752) underneath the annular mirror mount drives the small inner mirror, which is centered in the annular mirror. The hole in the mirror is beveled and the collar holding the inner mirror is canted to minimize scattered light from the remaining bright core of light at the center of the pupil. The piezoelectric stage and controller have a resolution of a few nanometers and a range of $30\ \mu\text{m}$, corresponding to a precision of a few thousandths of a wave and a dynamic range of nearly 30 waves at the center of K_s band, $2.15\ \mu\text{m}$. A mechanical drawing of the apparatus is shown in Figure 7.2. Images of the point-spread function of the outer annular mirror and inner mirror (the reference beam) are shown in Figure 7.3.

7.3.2 Initial setup and alignment

We first performed a number of tests of coronagraphic phase shifting interferometry in a laboratory setting, using the P3K adaptive optics system, the SDC, and the near-infrared camera PHARO. All tests were performed in the infrared K_s band, with a central wavelength of $\lambda = 2.145\mu\text{m}$ and a bandwidth of $1.99 - 2.3 \mu\text{m}$ ($\sim 15\%$ bandwidth), identical to our on-sky observing configuration. The tests were performed using the internal white light fiber of the adaptive optics system as an input source, and a temporary mask on the deformable mirror to simulate the pupil of the Hale telescope while in the lab.

The initial setup consisted of separately imaging the point-spread function of the annular mirror and phase-shifting mirror on the detector, then coaligning their centers by tilting the actuated annular mirror. Images of the two point-spread functions may be seen in Figure 7.3. The coherence length of this interferometer can be approximated by $\lambda \cdot (\lambda/\Delta\lambda) = 4.7 \mu\text{m}$, which is easily within the range and precision of the actuators on the optics. Co-phasing the mirrors was accomplished by pistoning the annular mirror until the white-light fringe (the point of highest fringe contrast, corresponding to zero path length difference) was acquired. While acquiring the white light fringe is not a necessary condition for phase measurements, it is helpful for obtaining the greatest fringe contrast, so a single long scan of the phase-shifting mirror was used to determine the white-light fringe position. An example of such a scan is shown in Figure 7.5.

Finally, the two optical vortices were coaligned with the input AO source, creating the correct output pupil with the bright core at the position of the phase-shifting mirror. A quad-cell tracker internal to the SDC measured slow drifts off the first vortex, sending corrections to the adaptive optics system in a separate loop operating outside the normal 2kHz closed-loop operation. This was important because coronagraph/PSF misalignments can amplify the residual diffraction in the image plane, increasing the speckle brightnesses and creating measurement error.

7.4 Focal plane wavefront sensing via phase shifting interferometry

7.4.1 Focal-plane phase measurements and dark hole generation

It is instructive to consider the accuracy to which phase may be measured in the interferometric setup. Using the four-step algorithm described earlier, it is relatively straightforward to propagate the uncertainties in the phase:

$$\phi_s(x, y) = \tan^{-1} \left[\frac{I_4 - I_2}{I_1 - I_3} \right] \quad (7.12)$$

$$\Rightarrow \sigma_\phi^2 = \sum_{i=1}^4 \sigma_{I_i}^2 \left(\frac{\partial \phi}{\partial I_i} \right)^2 \approx \sigma_{I_1}^2 \sum_{i=1}^4 \left(\frac{\partial \phi}{\partial I_i} \right)^2 \quad (7.13)$$

$$= \frac{(I_s + I_r)}{4I_s I_r} \quad (7.14)$$

$$\Rightarrow \sigma_\phi = \frac{\sqrt{I_s + I_r}}{2\sqrt{I_s I_r}} \quad (7.15)$$

which may be rewritten in terms of the mean image and visibility as

$$\sigma_\phi = \frac{1}{\gamma \sqrt{2 \langle I \rangle}} \quad (7.16)$$

where the last step follows after some algebra and assuming Poissonian statistics, i.e. $\sigma_I^2 = I$ (see the Appendix for details). For a fixed flux $F = I_s + I_r$, the minimal value of σ_ϕ occurs when $I_s = I_r = F/2$, as expected. Note that the equality of speckle and reference beam intensity is exactly the case in speckle nulling, by construction, where the “probe” speckles are matched to the intensity of the offending speckles. We can also see the penalty we pay by using a non-matched reference beam. For example, if both the reference and speckle are 1000 counts, the uncertainty is only about 3 degrees, whereas if the reference beam is 100 times fainter than the speckle (ie, 10 counts), the expected uncertainty in phase is on the order of 20 degrees. Of course, if the errors are random, the algorithm should still converge, albeit perhaps more slowly than desired.

Unfortunately, we were not limited solely by random error, and we discovered one additional complication in accurate phase retrieval. In the ideal case, the phase of the reference beam should be flat over the entire focal plane, and at a constant phase if the mirror is centered perfectly. A decentration of the pistoning mirror in the pupil plane leads to a phase tilt of the reference beam Airy pattern across the focal plane, which may be removed during data processing. However, it became apparent that there was a static (up to a constant offset) phase gradient over the image. To determine the precise shape of the phase gradient, we generated spots of zero phase on the deformable mirror, i.e. cosine waves with the peak centered on the mirror. (The choice of zero DM phase for the reference spots simplifies the determination of the phase gradient, as the electric field phase will be the same everywhere, unlike a choice of 45 degrees, where half the focal plane would have electric field phase of +45 and half would have spot phases of -45.) We then performed a measurement of the phase using the same four-step algorithm. We extracted the phase measurements from each calibration spot and generated a surface plot of the phase gradient.

The phase gradient surface was very well described by the sum of a plane and a paraboloid

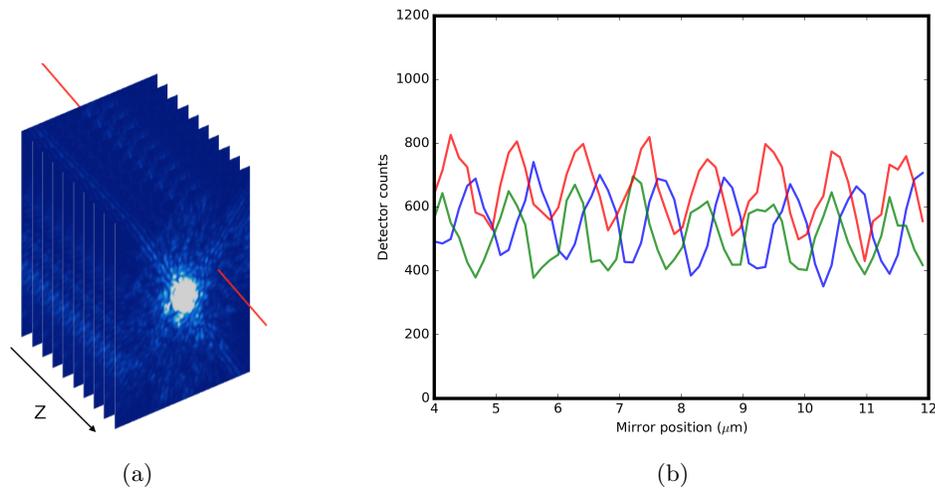


Figure 7.5: (a) An image is taken at each piston position (z). The red line shows a cut through a particular pixel, at different mirror positions. (b) Raw data of intensity (in counts) vs mirror position for three different pixels of similar intensity in the image plane. The relative phases between the three different waves correspond to differences in electric field phase.

$(Ax^2 + By^2 + Cxy + Dx + Ey + F)$, suggesting that it was caused by a decentration of the phase-shifting mirror and a focus term. The focus term could be due to some optical power on the piston mirror. The residual phase error from the fit was typically 5-10 degrees, which sets the maximum phase accuracy attainable here. The phase gradient was determined to be quite stable, with no significant changes observed over the course of the experiment.

However, it was found that the zeropoint (that is, the offset F in the paraboloid equation) of the phase was not stable to this accuracy level over periods longer than 1-2 hours, which we attributed to thermal and mechanical drifts between the optics downstream of the rod and the optics downstream of the annular mirror. The zeropoint of the phase is critical, of course, for having a consistent correction applied by the deformable mirror. A single culprit for this drift was not found, but it is likely that the mechanical coalignment stability was at fault.¹ The solution to this was to simply rerun the calibration of the phase gradient every 1-2 hours or so.

An example of a dark hole generated using this approach is shown in Figure 7.8, improving contrast by a factor of 3-4 over the control region. While the dark hole is evident, there were shortcomings in this implementation which prevented this technique from working to its full potential. First, despite the phases and amplitudes of each pixel being measured over the full focal plane, with limited time available at Palomar, we were not able to implement a full-field solution on the deformable mirror—we just proceeded in the same manner as speckle nulling, picking out bright points and using the measured phases to generate sinusoids on the deformable mirror to cancel them. This unfortunately handicaps one of the main strengths of this technique, which is the full-field measure-

¹we calculated that 30 nm of phase error corresponds to 10 degrees of phase accuracy

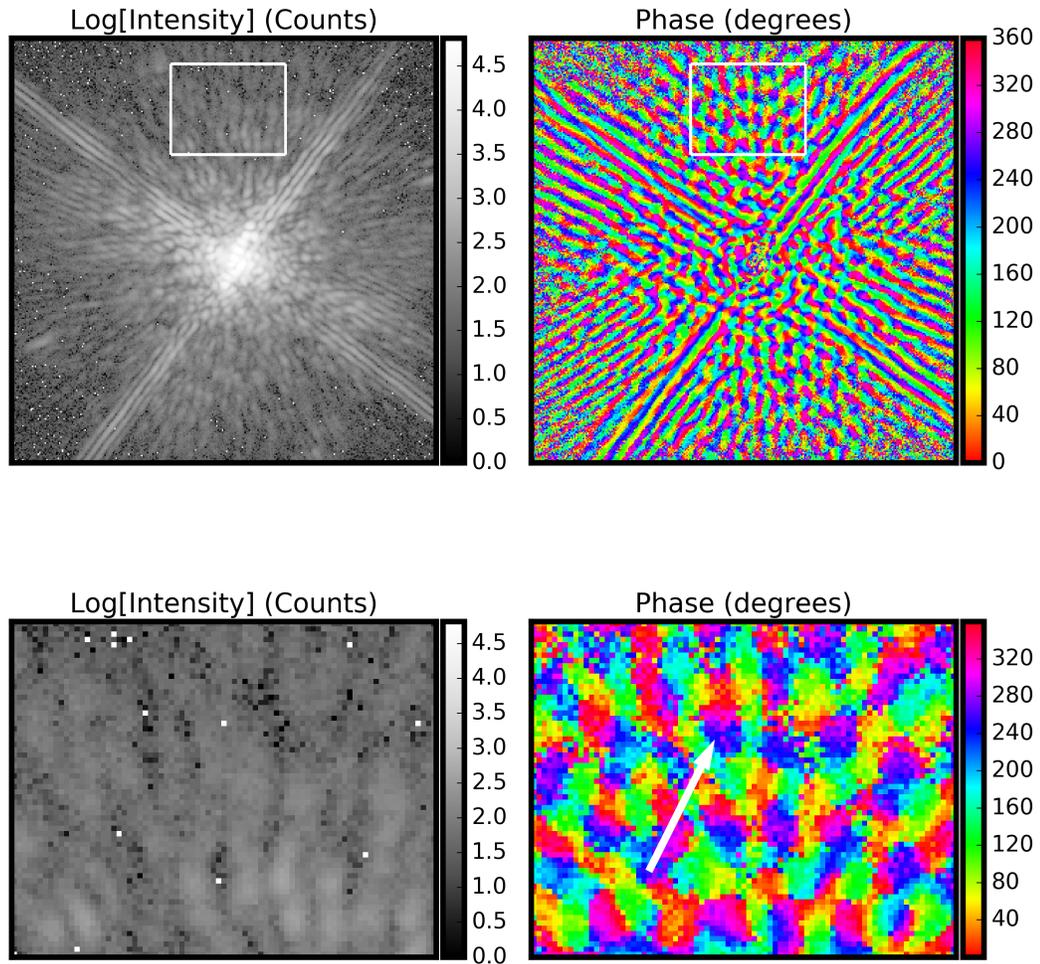


Figure 7.6: Phase sensing in the focal plane. (Top row) The left-hand plot shows the intensity measurements in the focal plane, with diffraction spikes clearly visible. This measure of intensity is a typical image shown on a camera. On the right, the corresponding phase map is shown. (Bottom row) The outlined region in the top plot, magnified to show detail. Speckles and their Airy rings are more easily identified in the phase plot, with a 180 degree phase shift between the cores and first Airy rings. Note that speckle chromatic effects such as elongation show themselves as phase gradients across an individual speckle; for example, in the central speckle in the image, indicated by the arrow, the measured phase is seen to increase from 240 to 300 degrees across it.

ment. Second, the phase gradient was measured using only 32 calibration spots over the field, and in retrospect, at somewhat worse signal-to-noise as we would have liked. This left systematic errors at the 10-15 degree level, which directly propagate into the measured phases, driving the convergence to a slightly wrong setpoint. As a result, the performance of the phase shifting approach in this initial trial was not as good as the performance produced on the same speckle field with speckle

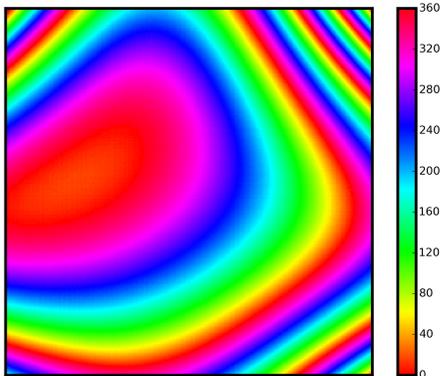


Figure 7.7: The derived map of the phase gradient over the same area as in Figure 7.6.

nulling, with the dark hole contrast being worse by a factor of two to four, and the convergence being somewhat slower.

7.5 Coherence-based companion detection

7.5.1 Principles

Similar to other high-contrast imaging instruments, the typical observing strategy used with SDC is driven by the need to remove speckle noise in the final, post-processed image. The Hale telescope has an equatorial mount, so the sky plane is fixed with respect to the detector plane. This does not allow for techniques like angular differential imaging (Marois et al., 2006a), where sky rotation can be used to separate any (rotating) companions from the static speckles. The strategy we normally use during observations, called “reference differential imaging” quickly alternates between the target of interest and a nearby star of similar V and K apparent magnitudes. The proximity of the stars guarantees that flexure in the optical path stays relatively constant between target and reference, leading to very similar speckle patterns; the same V and K ensure the same adaptive optics performance and flux on the detector plane, respectively. During data reduction, it is possible to then remove the speckles surrounding the target star using the reference star PSF, using either classic point-spread function subtraction or more involved methods like LOCI (Lafrenière et al., 2007) and KLIP (Soummer et al., 2012).

During the observing sequence, we repeated the standard procedure but added one modification, where we synchronized the exposures to the pistoning mirror position, which advanced in steps of $\lambda/8$. Each image then had a particular mirror phase associated with it, which could be used in post-processing to generate “coherence maps” over sequences of images using the visibility formula (Equation 7.11).

Because the visibility depends on the geometric mean of the speckle and reference wave, in the case of a speckle with the same brightness as the reference wave ($I_s = I_r$), the visibility will be 100%, in the case of a speckle one thousand times brighter ($I_s \gg I_r$), the speckle will modulate at about 6%, and for one ten thousand times brighter the modulation will be about 2%. Ignoring speckles, what about a companion in the image? Since the light from the star is incoherent with the light from the companion, there will be no interference, so $I_1 = I_2 = I_3 = I_4$; $\gamma = 0$. Of course, companion and speckle light will overlap, so what will actually be detected is a region of lower local visibility in the image at the location of the companion. Thus it is possible in *principle* to detect a companion by looking for local minima in the visibility. However, because the visibility depends on the relative brightness of the speckle and reference beam, while the reference beam is locally uniform in intensity, speckles will vary in different parts of the image. For example, speckles tend to be brighter nearer to the star (due to more errors in optical figures at low spatial frequencies), so visibility is expected to be lower there. Therefore, care must be exercised, as a brighter-than-average speckle could also be responsible.

There are a few ways to resolve this brighter-than-average-speckle/companion ambiguity, which require some additional assumptions or measurements. We note that without any other information, it is impossible to resolve this, as the equations are unchanged upon exchange of I_s and I_r .

In the discussions that follow, we will refer to two derived observables which we repeat here, for clarity:

$$\gamma = \frac{2\sqrt{(I_4 - I_2)^2 + (I_3 - I_1)^2}}{I_1 + I_2 + I_3 + I_4} \quad (7.17)$$

$$\langle I \rangle = \frac{I_1 + I_2 + I_3 + I_4}{4} \quad (7.18)$$

where the I_j 's are the individual images taken at each phase step.

One way to proceed is when the shape of the reference wave is known up to a multiplicative constant in intensity, so that $I_r[x, y] = cI_{rn}[x, y]$, where I_{rn} is some normalized version of the reference wave. This reference wave shape may be computed from physical optics simulations, or measured, see for example Figure 7.3. In that case, the following holds:

$$\langle I \rangle = I_s + cI_{rn} + I_p \quad (7.19)$$

$$\gamma = \frac{2\sqrt{cI_{rn}I_s}}{cI_{rn} + I_s + I_p} \quad (7.20)$$

$$\Rightarrow I_p = \langle I \rangle - cI_{rn} - \frac{\gamma^2 \langle I \rangle^2}{4cI_{rn}} \quad (7.21)$$

If it is not possible to measure the shape of the reference wave, it can be approximated as being

a multiple of the speckle field, so, $I_r[x, y] = qI_s[x, y]$, where q is a *scalar*. This approximation is somewhat valid, because the speckle intensities are amplified by underlying diffraction, and the reference wave is shaped approximately like this diffraction pattern, with the intensity falling off at a similar rate as to the speckles. (This situation would apply exactly in the case of a two-beam interferometer, for example, where the images are recombined after a static phase shift is applied to one of them, as first described by Guyon (2004)). Therefore,

$$\langle I \rangle = (1 + q)I_s + I_p \quad (7.22)$$

$$\gamma = \frac{2\sqrt{q}I_s}{(1 + q)I_s + I_p} \quad (7.23)$$

$$\Rightarrow I_p = \langle I \rangle \left(1 - \frac{1 + q}{2\sqrt{q}} \gamma \right) \quad (7.24)$$

In both of the previous two cases, the constants c and q must be determined. In principle, they may be approximated by using a noise model of the system. However, the latter method is dangerous for on-sky observing, because the illumination of the optics may change somewhat during the course of an exposure sequence, especially due to tip/tilt error. It is safer to have a reasonable starting point for these parameters and experiment with the precise values in post-processing. Regardless, these two methods give the planet light at the cost of one further parameter.

A way to proceed that does not rely on “free” parameters is to observe a nearby reference star, as usual. Apart from the pixels corresponding to the planet position, the reference star will have the same visibility map as the target star, regardless of their relative brightness. This is because the value of γ depends on the relative values of I_s and I_r , and not on their absolutes—scaling each up by the same factor does not affect the coherence, and the relative brightness of the reference beam I_r is set by the optics. Letting the visibility map of the target and reference star be γ_T and γ_R , respectively, we find

$$\gamma_T = \frac{2\sqrt{I_{s,T}I_{r,T}}}{I_{s,T} + I_{r,T} + I_p} \quad (7.25)$$

$$\gamma_R = \frac{2\sqrt{I_{s,R}I_{r,R}}}{I_{s,R} + I_{r,R}} \quad (7.26)$$

$$(7.27)$$

where we have explicitly included the light from a planet or other companion (I_p) in the target image. It then follows that

$$\frac{1}{\gamma_T} - \frac{1}{\gamma_R} = \frac{I_p + I_{r,T} + I_{s,T}}{2\sqrt{I_{r,T}I_{s,T}}} - \frac{I_{r,R} + I_{s,R}}{2\sqrt{I_{r,R}I_{s,R}}} \quad (7.28)$$

$$= \frac{I_p}{2\sqrt{I_{r,T}I_{s,T}}} = \frac{I_p}{\gamma_T(I_{r,T} + I_{s,T} + I_p)} = \frac{I_p}{\gamma_T \langle I \rangle} \quad (7.29)$$

$$I_p = \left(1 - \frac{\gamma_T}{\gamma_R}\right) \langle I \rangle \quad (7.30)$$

where the term $\langle I \rangle$ refers to the mean value of each pixel over one phase cycle; that is, $(I_1 + I_2 + I_3 + I_4)/4$.

In summary, coherence data may be used to discover the presence of companions in a relatively straightforward manner. Without any assumptions, a companion may be inferred as a region of locally reduced visibility, but this is insecure, as unusually bright speckles can also cause locally reduced visibility. To break this uncertainty, either knowledge of the reference beam shape or assumptions of the relative falloff may be used to solve for the incoherent component of the image, at the cost of an additional fitted parameter. Finally, observing a reference star can break this degeneracy and solve for the incoherent light without any further assumptions or parameters.

7.5.2 Combination with image post-processing

Image processing is an important aspect of high-contrast imaging, and one must always consider how different observing strategies interplay with post-processing to affect signal-to-noise of any detected companion in the final data product. We briefly provide an overview of the popular KLIP algorithm to motivate the discussion, while noting that most of what follows is also applicable to LOCI. The KLIP principle components algorithm requires a “reference” set of images containing “identical” speckles, but not the companion, from which it generates a low-dimensional subspace. The “target” images are then projected onto this low dimensional subspace, and the projected images are subtracted from the original target frames. The companion light will then be left over in the final images. Usually, the reference images are taken from nearby stars, which causes the speckle pattern to shift somewhat, as the gravity vector in the optical system changes. They may be also taken from the same star if the telescope permits sky rotation with respect to the optical system, as in the case of alt-az designs; this is the principle behind “angular differential imaging,” where the companion rotates with respect to the speckle field. The angular differential imaging approach has the advantage that the speckles barely change, but requires large amounts of observing time for substantial field rotation, and is much less effective at close inner working angles where companions typically like to exist.

In our case, we can combine the coherence modulation with KLIP seamlessly, simply by computing the principle components using the coherent light in the image. For example, consider Equation

7.21, which gives the planet light as a function of the mean intensity, visibility, reference beam shape, and a scaling parameter. As written, it can be interpreted as “PSF subtraction,” where the PSF is given by the coherent light in the image. Instead of subtracting, we can instead use the coherent part of the image—that is, $I_s = \gamma^2 \langle I \rangle^2 / 4cI_{rn}$ —to compute the principle components. In practice, this involves calculating I_s for each set of four images, then using the set of all the I_s ’s to generate the principle components.

This is quite useful for two reasons. First, in the absence of noise, the coherent parts of the image will match the speckles very well, as no changes in the optical system will have occurred, unlike when using a reference star. Said another way, the principle components are calculated from the target frames themselves, just with the companion light missing, since it’s incoherent. Second, the “scaling” parameter c essentially drops out of the calculation, since principle components analysis will return the same low-dimensional subset regardless of whether the data is scaled by a constant factor. There is one slight caveat to this, which is that c actually appears twice in the γ term, in the numerator as a constant multiplicative factor, and in the denominator as part of a quantity $I_s + cI_{rn} + I_p$. The denominator term is dominated by I_s , so the contribution of the c can be safely ignored.

Nothing is free, of course, and the price to be paid for using the coherence data to generate the principle components is that the calculated components are noisier than those that could be generated from an appropriately chosen reference star, at least in this particular optical setup. However, the ability to generate companion-free components without using a reference star allows much more observing time to be spent on the target star, which is beneficial for observing efficiency.

7.5.3 Demonstration: HD 49197b

To demonstrate the feasibility of the techniques described above, we observed the star HD 49197 on 22 November 2015 using the phase-shifting interferometer mode on SDC. The star has a known brown-dwarf companion separated by $0''.95$ with a brightness difference of $\Delta K_s = 8.22 \pm 0.11$, approximately 2000 times fainter than the primary (Metchev & Hillenbrand, 2009). The star was chosen because the companion was at approximately the same brightness as the surrounding speckle field, maximizing the expected coherence differences, and also because the companion was bright enough to be detectable in a reasonable amount of observing time.

Observing conditions were typical, with the adaptive optics system delivering a Strehl ratio of about 70-75%. We synchronized the exposures of the infrared camera PHARO with the phase shifting mirror, so that the phase of the reference beam was recorded with each image taken. We took 180 images of the target frames, and 180 images of the reference star, with the mirror modulating in steps of $\lambda/8$ from 0 to 1.25λ . This allows for two sets of independent visibility measurements per mirror phase cycle, that is, interleaved steps of $\lambda/4$.

The data analysis steps generally proceeded as described in the previous section, with a few

modifications that we list here. First, the individual frames were preprocessed with bad pixel interpolation and flat field correction. Four frames out of the 360 suffered from a known error in the PHARO camera where all the pixels in the image are read out at very large negative values; this required us to discard those four interferometer sequences (by sequence, we mean each four-step measurement where the interferometer moves by $\lambda/4$ per step). As visibility measurements are very sensitive to changes in total brightness from image to image, we selected the most useful data by calculating the variance of the total flux in each image sequence, and discarding the half with the highest variance (the discarded data is perfectly fine for normal image reduction, but not for extracting coherence information). At the end of this preconditioning, we were left with nineteen sequences of four images each, for a total of 76 images.

For each sequence, we calculated the mean image and visibilities, and use these to create the incoherent intensity map according to Equation 7.30. (The shape of the reference wave had already been measured in the lab, and is shown in Figure 7.3). The 19 incoherent intensity maps are then median combined for a final incoherent intensity image. We also present an image of the incoherent signal-to-noise ratio per pixel, which is divided by the shot noise contribution per pixel, estimated as the square root of the mean intensity. We “regularized” two regions of the image which had unphysical values; the outer edge where the reference beam did not reach, and the vertical diffraction spike. These areas had very large negative incoherent intensity fluxes, as the visibility noise was too high. These regions are fairly easy to identify in the incoherent signal-to-noise ratio image (Figure 7.9), at the outer edges of the image and at position angles of 0 and 180.

Analysis of the coherence modulation data clearly revealed the brown dwarf, despite the imperfect observing conditions (see Figure 7.9 b, c). Perhaps the most interesting result was that in the averaged target frames (Figure 7.9 a), the companion was not readily detectable above the speckle noise floor, while using the *same* image data and visibility information, the brown dwarf appears as a significant local maximum, without using any reference star frames at all. This essentially amounts to “orthogonal” evidence of the presence of a companion, pointing to a potentially useful line of attack against speckle noise limits in companion detection.

We also performed the reduction described in Section 7.5.2, where we calculated the principle components using light from the coherent part of the image, then use those to optimally subtract the point-spread function of the star. The KLIP approach removed starlight efficiently and clearly revealed the companion, but created a great deal of high-frequency noise due to the imperfect extraction of the components. Here, we note we used an annular PCA algorithm that had two components per 100 pixels. This image is presented in Figure 7.9 c.

It is interesting to note that using coherence seems to be an efficient method of removing starlight without using a reference star. The mean image uses 178 frames and achieves a SNR of about 5. The incoherent intensity image, using the phasing information to extract the planet signal, uses 76 frames

and achieves a SNR of about 7.5. Using a reference star with classic (annular) PSF subtraction², and matching the same number of total images, a SNR of 9.3 is achieved. This is only about 25% better than not using a reference star but using the coherence information, and the same SNR as using principle components analysis with a reference star and equal number of frames (38 target and 38 reference). Using the “coherent” principal components approach without a reference star, where the components are calculated from the coherent intensity only, the SNR is somewhat surprisingly only 5.35, despite the companion being cleanly separated (Figure 7.9 c). This appears to be due to the high-frequency noise introduced in the principal components, which is evident in the image; in particular, large negative values near the core of the PSF of the companion dominate the noise term. When calculating PCA conventionally using all the frames from the reference star to build the components (image not show), and all the frames from the target star, we managed a SNR of about 20. A summary of these results is presented in Table 7.1. While some caution is urged, as the effects on SNR from data reduction algorithms and physically distinct observing styles are somewhat difficult to disentangle, the good performance of coherence modulation on this first test is encouraging.

Method	Target images	Reference images	SNR	Note
Mean image (all)	178	0	5.2	Figure 7.9 a
Incoherent intensity	76	0	7.35	Figure 7.9 b, Equation 7.21
Incoherent intensity PCA	76	0	5.35	Figure 7.9 c, Section 7.5.2
Reference PSF subtraction	38	38	9.34	Not shown
Reference PCA reduction	38	38	7.5	Not shown
Reference PSF subtraction (all)	178	178	14.3	Figure 7.9 d
Reference PCA reduction (all)	178	178	19.5	Not shown

Table 7.1: Summary of reductions. All observations of HD 49197 and reference star HD 48270 were on 22 November 2015.

7.6 Discussion

7.6.1 Potential as a method of speckle suppression

Our results using the central pistoning mirror to measure the phase of the electric field of focal-plane speckles demonstrate that the technique works as theory predicts, and can be used to effectively suppress speckles in regions of interest.

Compared to methods mentioned in the introduction (speckle nulling, EFC, COFFEE, SCC, SISS), there are some advantages. The technique requires no system model, and has a particularly simple formulation to derive the electric field phase, giving good results with no complicated

²intensity matched subtraction was performed at each radial separation from the star, with an annular mask of size $2 \lambda/D$, using median target and calibrator frames

processing or statistical estimation. One of the reasons for the good performance is the ability to put large phases on the pistoning mirror, unlike in electric field conjugation, where the “probe” steps are typically very small to respect linearity approximations, crosstalk, and stroke limits of the deformable mirror. The phase measurements can be seamlessly integrated into the science measurements, as the reference beam is unobtrusive (at least in this implementation). The extra optical complexity is less than both the self-coherent camera and the synchronous interferometric speckle subtraction techniques, despite its similarity. This approach may also be compatible with a range of coronagraphs, though the design requires optimization around the particular coronagraph design.

However, there are drawbacks to this technique. First, the need to precisely determine the focal plane phase gradient caused by the rod decenter and/or optical power requires additional measurements, and the associated uncertainties in these measurements impose a penalty on the phase accuracy. In contrast, techniques using the deformable mirror measure speckles’ relative phases quite accurately without requiring a zero-point calibration of any phase gradients in the focal plane. Our measurements of the gradient phase accuracy showed that about 10-15 degrees of accuracy could be reliably obtained; speckle-nulling approaches typically gave better than 5 degrees of accuracy for the same test speckles.

Another disadvantage of this approach is the need for an absolute zeropoint calibration, which means that any drifts in the relative cophasing of the rod and the rest of the optics need to be tracked. In our lab, drifts of $\lambda/4$ occurred over the course of about an hour. At a more unforgiving thermal and mechanical environment, such as a telescope, phase drifts would need to be constantly recalibrated, affecting observing efficiency. For simple linear phase drifts, there are non-invasive ways that could be used, such as generating spots of fixed phase at the extremes of the focal plane using the deformable mirror, and tracking those measured phases actively. More complex changes to the optical path, such as focus and other distortions, would require multiple “probe” spots in the focal plane that would be too obtrusive to science measurements (we used 32 spots over the focal plane for the calibration). EFC, speckle nulling, and COFFEE approaches are immune to these issues as the phase measurement and phase correction device are the same—the deformable mirror. SCC is also probably immune via the static nature of the pinholes. An alternative approach could replace the phase-shifting rod with a full deformable mirror at the second pupil plane; this would remove the error due to co-phasing drift and also allow for full correction of the focal plane, not just phase errors, though the limited stroke of most commercially available deformable mirrors might be an issue. A bimorph deformable mirror could achieve this stroke, but would probably require a custom electrode pattern to achieve a flat surface like the pistoning rod.

7.6.2 Potential as a method of planet-speckle discrimination

The phase-shifting approach shows promise as a method to detect companions. We first point out that none of the issues with the phase stability are relevant for the calculation of the visibilities (Equation 7.11), as the visibility relies on *differential* phase accuracy between phase steps. The notion of a “phase” for the companion light does not even make sense, as it is incoherent with the starlight. The optical phase thus only needs to be stable on timescales corresponding to a the four readouts of the camera, which is on the order of seconds for many instruments.

Encouragingly, the detection of HD49197b using coherence modulation was straightforward. The observing conditions were average, with a Strehl ratio in the 70-75% range. In normal SDC observing, we perform aggressive point-spread-function sharpening routines using the modified Gerchberg-Saxton algorithm to improve the internal instrumental Strehl ratio to above 90%. With the phase shifter installed, we were unable to do this as the algorithm refused to converge with the unfamiliar pupil geometry. We instead hand-tuned Zernike polynomials to increase the Strehl ratio, but were only able to achieve about 80% in this way. We also had a significant additional component of residual diffraction from the telescope spiders. Normally, we block the spider diffraction using appropriately shaped pupil stops in the infrared camera PHARO, but we were unable to do this as the stops would block light from the phase shifting mirror as well (some excess spider diffraction can be seen in Figure 7.9). In short, conditions were well below normal quality in this first trial, and the fact that the companion was so easy to detect suggests that the technique is fairly robust.

7.7 Conclusion

We have presented the design, lab and on-sky results from a phase-shifting interferometer installed on the Stellar Double Coronagraph at Palomar observatory. In one application, we demonstrated its use as a focal-plane wavefront sensor with the ability to measure the electric field over the full focal plane, providing a direct measurement of the speckle phases. We used this mode to create a region of high contrast in the focal plane: a “dark hole”. Apart from these advantages, we note that measuring and maintaining the absolute phase zeropoint with two separate mirrors caused less robust results in this initial test than conventional speckle nulling, though applications requiring full-frame electric field phase knowledge could benefit from such a device. Moreover, use of a second deformable mirror at the same location would solve this problem.

As a second application, we used the phase-shifter to measure the visibility, rather than the phase, of the speckle field. We described how the visibility can be used to discriminate between coherent and incoherent light in the focal plane, and therefore, between speckles and real companions. We demonstrated this on-sky by detecting the companion HD49197b, both as a region of locally

depressed coherence using only data from the target star, and also as a bright point source when combined with coherence data from a reference star. We discussed how this extra information can be used to extract companion photometry using the visibility data, as well as how it can be combined with powerful PSF-subtraction techniques like KLIP and LOCI.

We conclude that coherence modulation is a promising technique. The relative ease with which the companion was detected despite poor optical conditions, moderate observing conditions, and minimal data processing or selection strongly hints that extra information from coherence data can be obtained for a small increase in observing complexity. In the current era of extreme-AO systems and upcoming space telescopes such as WFIRST-AFTA, the potential for additional sensitivity gains at high contrasts should not be ignored.

7.8 Acknowledgements

We are very pleased to acknowledge the Palomar Observatory staff for their support. MB is supported by a NASA Space Technology Research Fellowship, grant NNX13AN42H. Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA).

7.9 Appendix

Derivation of Equation 7.16

Beginning with the error propagation formula, we have

$$\phi_s(x, y) = \tan^{-1} \left[\frac{I_4 - I_2}{I_1 - I_3} \right] \quad (7.31)$$

$$\Rightarrow \sigma_\phi^2 = \sum_{i=1}^4 \sigma_{I_i}^2 \left(\frac{\partial \phi}{\partial I_i} \right)^2 \quad (7.32)$$

$$= \sum_{i=1}^4 I_i \left(\frac{\partial \phi}{\partial I_i} \right)^2 \quad (7.33)$$

where the last step follows from the assumption of Poisson statistics; that is, $\sigma_I^2 = I$. Now using the fact that $d/dx \tan^{-1}(x) = 1/(1+x^2)$ and the chain rule, we note that $\left(\frac{\partial \phi}{\partial I_1} \right)^2 = \left(\frac{\partial \phi}{\partial I_3} \right)^2$, and similarly for I_4 and I_2 . We then find

$$\sigma_\phi^2 = (I_1 + I_3) \left(\frac{\partial \phi}{\partial I_1} \right)^2 + (I_2 + I_4) \left(\frac{\partial \phi}{\partial I_2} \right)^2 \quad (7.34)$$

$$= (I_s + I_r) \left[\left(\frac{\partial \phi}{\partial I_1} \right)^2 + \left(\frac{\partial \phi}{\partial I_2} \right)^2 \right] \quad (7.35)$$

The second line follows by direct substitution of the values of I_1, I_2, I_3, I_4 from 7.5.1. Now we express directly compute the partial derivatives using the chain rule and find

$$\sigma_\phi^2 = (I_s + I_r) \frac{(I_1 - I_3)^2 + (I_4 - I_2)^2}{[(I_1 - I_3)^2 + (I_4 - I_2)^2]^2} \quad (7.36)$$

$$= \frac{(I_s + I_r)}{(I_1 - I_3)^2 + (I_4 - I_2)^2} \quad (7.37)$$

$$= \frac{(I_s + I_r)}{4I_s I_r} \quad (7.38)$$

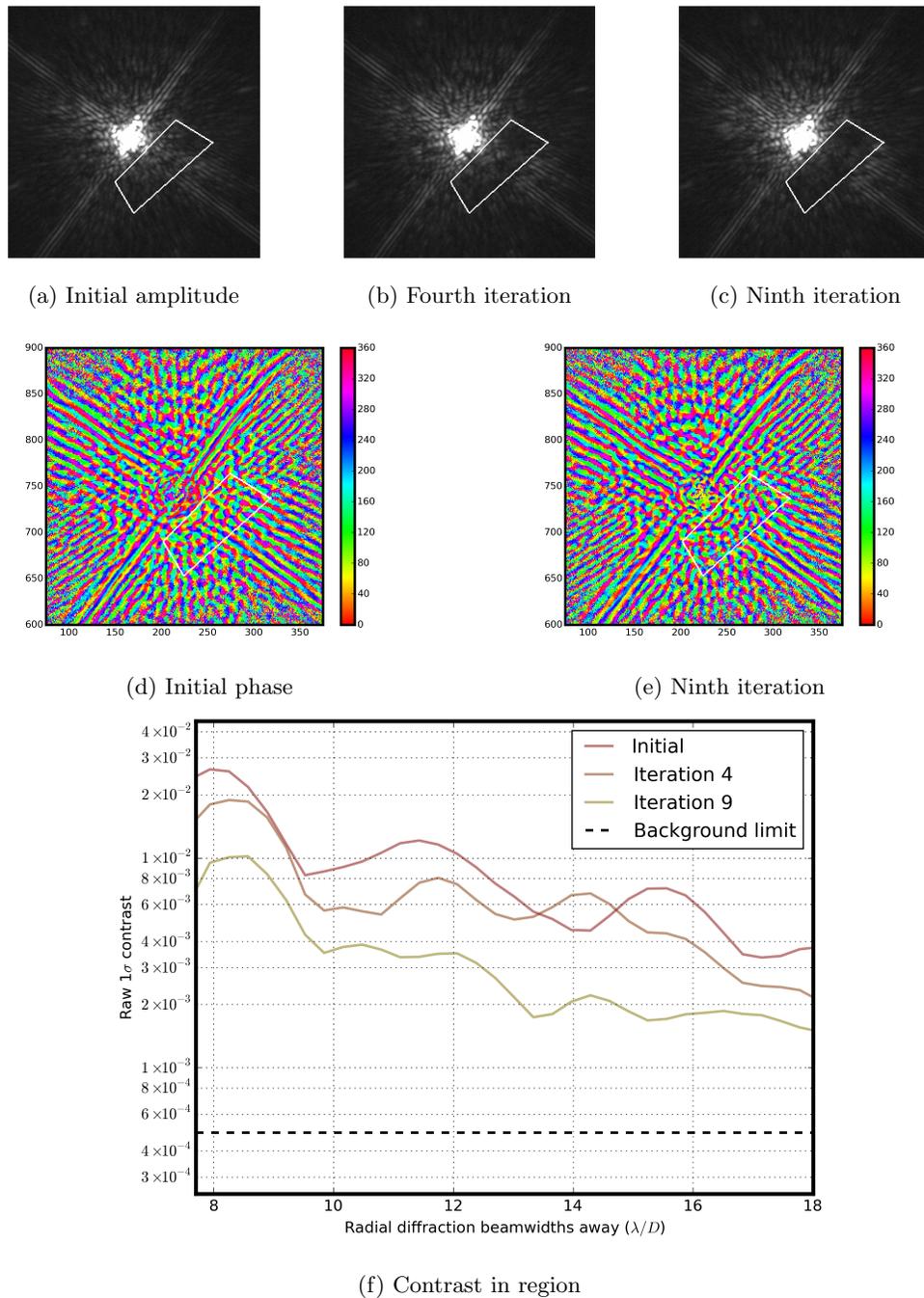


Figure 7.8: Laboratory demonstration of coronagraphic speckle-suppression using the phase-shifting interferometer. The white outline delineates the region of speckle control. Panels a, b, and c show the intensity after 0, 4, and 9 iterations of speckle removal by the deformable mirror. Panels d and e show the phase measurements of the initial and final light distribution, showing a clear difference in phase. Panel f shows the 1σ contrast curve measurement of the region after 1, 4, and 9 iterations. The contrast curve is defined in the usual way, with the standard deviation (ie, 1σ) of surface brightness at each radial separation being used to generate the curve, and normalized by dividing by the peak flux of the non-coronagraphic PSF (not shown). The preprocessing performed on the data was dark subtraction and flat-fielding.

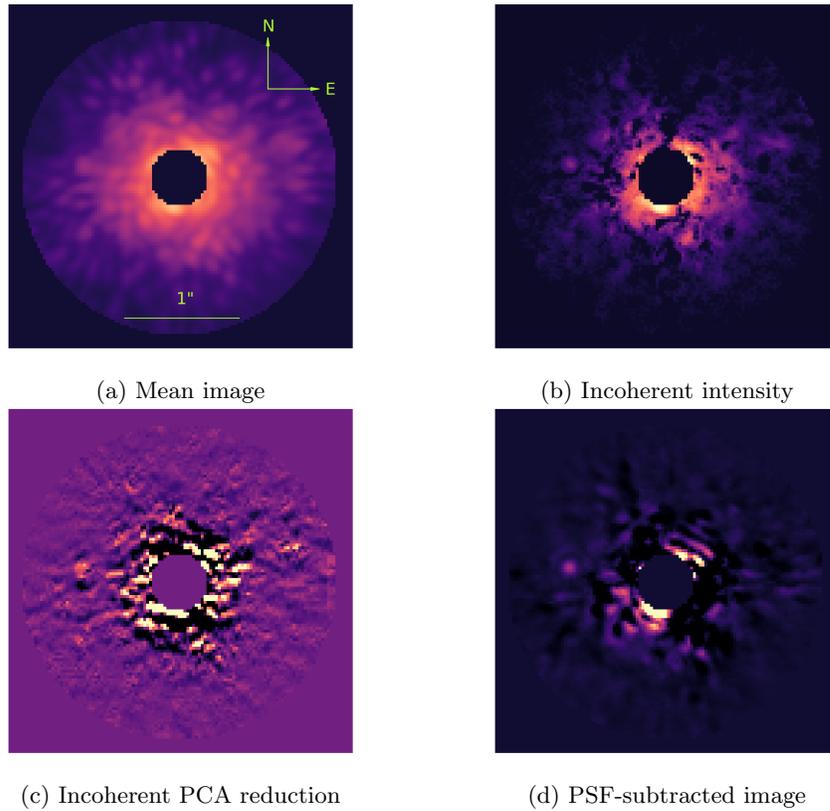


Figure 7.9: Coherence modulated detection of HD 49197b. (a) The mean of the target frames. The substellar companion is not clearly visible in the image. The companion is present at an SNR of about 5. The stretch is logarithmic. (b) The incoherent intensity map (Equation 7.21). The companion is easily visible at a position angle of about 275 degrees, and has an SNR of about 7.5. The stretch is also logarithmic. We note that the same data frames were used to compute image (b) as in (a), except combined using the interferometer position to give extra information. No reference star is used. (c) Principle components analysis (KLIP) reduction of HD49197b, where the components are generated from only the coherent parts of the image data, as explained in Section 7.5.2. No reference star is used, and the SNR is 9. The stretch is linear. (d) A conventional PSF-subtracted image of the companion, using a nearby reference star, for comparison. All the data is used, and the SNR is 14. The stretch is linear.

Chapter 8

Summary and future work

The first part of this thesis presented two related projects. First, we investigated the design requirements and performance of radial velocity surveys, deriving optimal wavelengths and stellar targets for different survey goals. We showed that for surveys complete to a mass-orbital distance function, the best targets are F, G, and K stars at blue to red optical wavelengths. On the other hand, for surveys searching for habitable-zone planets, late M-type stars are the best targets, observed in the 700-800 nm near-infrared bands. The habitable-zone survey leads to more targets, and more detections, due to a combination of effects, including the boost in radial velocity signal from the lower mass of the host stars, and the large number of nearby M dwarfs. The results are based on maximizing survey success in a fixed amount of observing time per night, including spectral quality, photon signal and noise per resolution element, planetary mass, and stellar mass distribution. We also investigated the effects of various noise sources and other instrumental parameters like system resolution and aperture size, and found that if systematic errors are appropriately controlled, a modestly-sized, dedicated observatory can actually be quite competitive, given the large amount of orbital phase coverage.

Putting these lessons into practice, the second chapter presented the design, testing, and performance of MINERVA, the first multi-aperture observatory for radial velocity discovery and photometric followup of nearby exoplanets. We showed simulations of the expected survey yield of such an observatory of ~ 15 planets, and demonstrated the advantages of using a multi-aperture array, where the resolution of the spectrograph can be made high with a modestly sized instrument, at the price of detector space. This gives marked advantages in simplicity, cost, and stability. We also discussed the design trades in terms of telescopes, cameras, enclosures, etc. We presented the design and performance of a fiber coupling unit suitable for these telescope (and others). The optical, mechanical, and controller design demonstrate a very high pointing precision of 0.2 arcseconds, at the level required for high throughput, without sacrificing the fast acquisition time of the telescopes, below 25 seconds from star to star. In terms of transit performance, we demonstrated the required photometric precision necessary to detect transits of super-Earth type planets, of better than a milli-

magnitude. Finally, we discussed the tradeoffs of four different potential sites in North America, and motivated the choice of Mt. Hopkins in Arizona as the optimal choice. We concluded by describing the transfer of the test site to Mt. Hopkins.

While the delivery and testing of the spectrograph did not coincide with the timeline of this thesis, the spectrograph verification is the final, critical step for the MINERVA array to fully pursue science operations. The spectrograph has been installed and is currently being automated and tested, while photometric science is being pursued on-sky. Fiber anti-reflective coatings are being applied by a third party vendor, and are expected to be finished by September 2016. Furthermore, the data analysis pipeline has been undergoing active development and is expected to be completed around the same time. After that, the spectrograph testing and performance verification should proceed rather quickly, likely by the end of the year. More distant plans are to create a clone of the observatory in Australia, giving a larger selection of interesting target stars and allowing for greater coverage in terms of transit or transient events.

The second part of this thesis was focused on exoplanet detection via high contrast imaging. First, we described the motivation, design, and performance of the Stellar Double Coronagraph, a flexible multistage coronagraph at Palomar observatory. SDC has a number of new features, including the first on-sky implementation of multistage coronagraphs, including single, dual, and ring-apodized vortex designs, demonstrating efficient and effective control of diffraction at close inner working angles to the primary star, despite the large secondary mirror. We also demonstrated the high contrast performance on-sky. We presented some science results, including confirming a companion around the star epsilon Cephei and the first image of a conjectured companion around the star delta Andromeda.

With SDC, we intend to wrap up the survey of nearby A stars in May 2016, and while no new planetary companions have yet been found, this is perhaps understandable given recent results from GPI and SPHERE which have similarly found very little despite vastly larger budgets and better performance. In the future, we intend to take advantage of the flexible design of SDC to test different types of coronagraph designs on-sky. We will install H-band coronagraphs internally, allowing near-simultaneous observations of planetary companions at different infrared wavelengths. We also have a collaboration with the inventors of the self-coherent camera, who recently installed one in the instrument, with encouraging preliminary results. Further work is expected here, with a more optimized design and greater time allocation following this successful first test. Other coronagraphs to explore are pupil-apodizing designs, band-limited masks, and interferometric concepts. Similarly, we are collaborating with a group to install a microwave kinetic inductance detector array (MKID) camera behind the instrument, instead of using the conventional infrared camera. This camera, called DARKNESS, will be a unique instrument, capable of microsecond readout and some spectral resolution, a large improvement over the many-second readout time of the PHARO camera currently

being used. With such fast timing data, both speckle nulling on-sky and “dark-speckle” techniques may be used, the later using timing data to discriminate between the Poissonian exoplanet photon statistics and the Riccian photon statistics of speckles. This work will commence in the summer of 2016.

We discussed the development of a focal plane “speckle-nulling” wavefront control code designed to be used with various instruments and adaptive optics systems. The code has demonstrated the ability to correct non-common path speckle aberrations flexibly and generally deliver contrast improvements of a factor of 5-10, depending on the instrument used. It has been deployed at the Palomar 5m Hale telescope with TMAPS, SDC, and PHARO, and at the 10m Keck II telescope with the NIRC2 instrument in H and L band. It is currently being used in science programs requiring deep exoplanet imaging searches with the new L band vortex coronagraph at Keck, and is being included as an alternative calibration routine with the NIRC2 instrument.

We recently achieved sub-Jupiter mass sensitivities on a star using the speckle nulling code on Keck in combination with a deep multi-hour exposure, and further work will be geared towards continuing this strategy. This recent result, not presented in this thesis, was very encouraging, and suggests that long exposures and more thorough focal-plane wavefront correction has the potential to open new parameter space in high contrast imaging. The efforts on speckle nulling will focus on reducing differential errors by installing a single photonic crystal fiber to carry light from the visible to the thermal infrared through the system, rather than the current two-fiber approach. We also intend to improve and formalize the “user friendliness” of the interface, so that relatively untrained people can use it when desired—so far, results have been successful with unsupervised users, but some work remains to be done. We also plan to release the code online for use by the community. Further exposures are planned around the nearest stars, to hopefully detect new companions or set deep limits on what is possible from the ground.

Finally, we concluded by presenting a new method of focal-plane wavefront sensing based on the principles of phase-shifting interferometry. We discussed how the addition of a small, pistoning mirror in a pupil plane of a coronagraph can be used as a focal-plane interferometer, measuring the electric field of speckles that would typically overwhelm the light of faint companions. After developing the physical and mathematical background, we described our implementation of this interferometer on the Stellar Double Coronagraph. We demonstrated an improvement in contrast of a factor of a few using this device. We further showed how the same mechanism can be used to discriminate between speckles and companions in the focal plane based on measurements of the coherence of the light in the focal plane, giving independent and orthogonal information about the presence of any substellar companions. We demonstrate this successfully on-sky, showing a clean detection of the brown-dwarf companion to the star HD 49197 when using the interferometer. We believe this is the first detection of a substellar companion using the coherence properties of light.

With the initial successful test of the technique, we expect to make some improvements to the design for even better results in the future. We will have a more flexible optomechanical system to allow for different interferometer intensities and more precise co-phasing, perhaps using an internal metrology system. More work needs to be done on accurate procedures to measure the phase gradients produced by the device in the focal plane, while keeping the calibration to a reasonable amount of time.

Bibliography

- Aigrain, S., Pont, F., & Zucker, S. 2012, *MNRAS*, 419, 3147
- Allard, F., Homeier, D., & Freytag, B. 2011, in *Astronomical Society of the Pacific Conference Series*, Vol. 448, 16th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. C. Johns-Krull, M. K. Browning, & A. A. West, 91
- Bakos, G. A. 1976, *JRASC*, 70, 23
- Bakos, G. Á., Lázár, J., Papp, I., Sári, P., & Green, E. M. 2002, *PASP*, 114, 974
- Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, *A&A*, 337, 403
- Barnes, S. I. & MacQueen, P. J. 2010, A high-efficiency fibre double-scrambler prototype
- Basri, G., Walkowicz, L. M., Batalha, N., Gilliland, R. L., Jenkins, J., Borucki, W. J., Koch, D., Caldwell, D., Dupree, A. K., Latham, D. W., Marcy, G. W., Meibom, S., & Brown, T. 2011, *AJ*, 141, 20
- Baudoz, P., Boccaletti, A., Baudrand, J., & Rouan, D. 2006, in *IAU Colloq. 200: Direct Imaging of Exoplanets: Science & Techniques*, ed. C. Aime & F. Vakili, 553–558
- Beaulieu, J.-P., Bennett, D. P., Fouqué, P., Williams, A., Dominik, M., Jørgensen, U. G., Kubas, D., Cassan, A., Coutures, C., Greenhill, J., Hill, K., Menzies, J., Sackett, P. D., Albrow, M., Brilliant, S., Caldwell, J. A. R., Calitz, J. J., Cook, K. H., Corrales, E., Desort, M., Dieters, S., Dominis, D., Donatowicz, J., Hoffman, M., Kane, S., Marquette, J.-B., Martin, R., Meintjes, P., Pollard, K., Sahu, K., Vinter, C., Wambsganss, J., Woller, K., Horne, K., Steele, I., Bramich, D. M., Burgdorf, M., Snodgrass, C., Bode, M., Udalski, A., Szymański, M. K., Kubiak, M., Więckowski, T., Pietrzyński, G., Soszyński, I., Szewczyk, O., Wyrzykowski, Ł., Paczyński, B., Abe, F., Bond, I. A., Britton, T. R., Gilmore, A. C., Hearnshaw, J. B., Itow, Y., Kamiya, K., Kilmartin, P. M., Korpela, A. V., Masuda, K., Matsubara, Y., Motomura, M., Muraki, Y., Nakamura, S., Okada, C., Ohnishi, K., Rattenbury, N. J., Sako, T., Sato, S., Sasaki, M., Sekiguchi, T., Sullivan, D. J., Tristram, P. J., Yock, P. C. M., & Yoshioka, T. 2006, *Nature*, 439, 437

- Beuzit, J.-L., Feldt, M., Dohlen, K., Mouillet, D., Puget, P., Wildi, F., Abe, L., Antichi, J., Baruffolo, A., Baudoz, P., Boccaletti, A., Carbillet, M., Charton, J., Claudi, R., Downing, M., Fabron, C., Feautrier, P., Fedrigo, E., Fusco, T., Gach, J.-L., Gratton, R., Henning, T., Hubin, N., Joos, F., Kasper, M., Langlois, M., Lenzen, R., Moutou, C., Pavlov, A., Petit, C., Pragt, J., Rabou, P., Rigal, F., Roelfsema, R., Rousset, G., Saisse, M., Schmid, H.-M., Stadler, E., Thalmann, C., Turatto, M., Udry, S., Vakili, F., & Waters, R. 2008, SPHERE: a 'Planet Finder' instrument for the VLT
- Blake, C. H. & Shaw, M. M. 2011, *PASP*, 123, 1302
- Boley, A. C. 2009, *ApJ*, 695, L53
- Bordé, P. J. & Traub, W. A. 2006, *ApJ*, 638, 488
- Boss, A. P. 1997, *Science*, 276, 1836
- Bottom, M., Kuhn, J., Mennesson, B., Mawet, D., Shelton, J. C., Wallace, J. K., & Serabyn, E. 2015, *ApJ*, 809, 11
- Bouchy, F., Pepe, F., & Queloz, D. 2001, *A&A*, 374, 733
- Burruss, R. S., Serabyn, E., Mawet, D. P., Roberts, J. E., Hickey, J. P., Rykoski, K., Bikkannavar, S., & Crepp, J. R. 2010, in *Proc. SPIE*, Vol. 7736, Adaptive Optics Systems II, 77365X
- Butler, R. P. & Marcy, G. W. 1996a, *ApJ*, 464, L153
- . 1996b, *ApJ*, 464, L153
- Butler, R. P., Marcy, G. W., Williams, E., McCarthy, C., Dosanjuh, P., & Vogt, S. S. 1996, *PASP*, 108, 500
- Cameron, P. B., Britton, M. C., & Kulkarni, S. R. 2009, *AJ*, 137, 83
- Campbell, B. & Walker, G. A. H. 1979, *PASP*, 91, 540
- Carré, P. 1966, *Metrologia*, 2, 13
- Castelli, F. & Kurucz, R. L. 2004, *ArXiv Astrophysics e-prints*
- Chauvin, G., Lagrange, A.-M., Dumas, C., Zuckerman, B., Mouillet, D., Song, I., Beuzit, J.-L., & Lowrance, P. 2004, *A&A*, 425, L29
- Cox, A. N. 2000, *Allen's astrophysical quantities*
- Currie, T., Daemgen, S., Debes, J., Lafreniere, D., Itoh, Y., Jayawardhana, R., Ratzka, T., & Correia, S. 2014, *ApJ*, 780, L30

- Decin, G., Dominik, C., Waters, L. B. F. M., & Waelkens, C. 2003, *ApJ*, 598, 636
- Dekany, R., Roberts, J., Burruss, R., Bouchez, A., Truong, T., Baranec, C., Guiwits, S., Hale, D., Angione, J., Trinh, T., Zolkower, J., Shelton, J. C., Palmer, D., Henning, J., Croner, E., Troy, M., McKenna, D., Tesch, J., Hildebrandt, S., & Milburn, J. 2013, *ApJ*, 776, 130
- Delacroix, C., Absil, O., Carlomagno, B., Piron, P., Forsberg, P., Karlsson, M., Mawet, D., Habraken, S., & Surdej, J. 2014, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9147, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 8
- Doyle, L. R., Carter, J. A., Fabrycky, D. C., Slawson, R. W., Howell, S. B., Winn, J. N., Orosz, J. A., Presa, A., Welsh, W. F., Quinn, S. N., Latham, D., Torres, G., Buchhave, L. A., Marcy, G. W., Fortney, J. J., Shporer, A., Ford, E. B., Lissauer, J. J., Ragozzine, D., Rucker, M., Batalha, N., Jenkins, J. M., Borucki, W. J., Koch, D., Middour, C. K., Hall, J. R., McCauliff, S., Fanelli, M. N., Quintana, E. V., Holman, M. J., Caldwell, D. A., Still, M., Stefanik, R. P., Brown, W. R., Esquerdo, G. A., Tang, S., Furesz, G., Geary, J. C., Berlind, P., Calkins, M. L., Short, D. R., Steffen, J. H., Sasselov, D., Dunham, E. W., Cochran, W. D., Boss, A., Haas, M. R., Buzasi, D., & Fischer, D. 2011, *Science*, 333, 1602
- Dragomir, D., Kane, S. R., Pilyavsky, G., Mahadevan, S., Ciardi, D. R., Gazak, J. Z., Gelino, D. M., Payne, A., Rabus, M., Ramirez, S. V., von Braun, K., Wright, J. T., & Wyatt, P. 2011, *AJ*, 142, 115
- Dressing, C. D. & Charbonneau, D. 2013, *ApJ*, 767, 95
- Dumusque, X., Pepe, F., Lovis, C., Ségransan, D., Sahlmann, J., Benz, W., Bouchy, F., Mayor, M., Queloz, D., Santos, N., & Udry, S. 2012, *Nature*, 491, 207
- Dumusque, X., Santos, N. C., Udry, S., Lovis, C., & Bonfils, X. 2011a, *A&A*, 527, A82
- . 2011b, *A&A*, 527, A82
- Dumusque, X., Udry, S., Lovis, C., Santos, N. C., & Monteiro, M. J. P. F. G. 2011c, *A&A*, 525, A140
- . 2011d, *A&A*, 525, A140
- Durisen, R. H., Boss, A. P., Mayer, L., Nelson, A. F., Quinn, T., & Rice, W. K. M. 2007, *Protostars and Planets V*, 607
- Endl, M., Kürster, M., & Els, S. 2000, *A&A*, 362, 585

- Ertel, S., Absil, O., Defrère, D., Le Bouquin, J.-B., Augereau, J.-C., Marion, L., Blind, N., Bonsor, A., Bryden, G., Lebreton, J., & Milli, J. 2014, *A&A*, 570, A128
- Figueira, P., Pepe, F., Lovis, C., & Mayor, M. 2010, *A&A*, 515, A106
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Galicher, R., Baudoz, P., & Rousset, G. 2008, *A&A*, 488, L9
- Gaudi, B. S. 2010, ArXiv e-prints
- Give'on, A. 2009, in *Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics & Photonics Technical Digest (Optical Society of America)*, AOWA3
- Give'on, A., Kern, B., Shaklan, S., Moody, D. C., & Pueyo, L. 2007, in *Proc. SPIE*, Vol. 6691, *Astronomical Adaptive Optics Systems and Applications III*, 66910A
- Gontcharov, G. A. & Kiyeva, O. V. 2002, *A&A*, 391, 647
- Goodman, J. 2005, *Introduction to Fourier Optics*, 3rd edn. (McGraw-Hill)
- Gould, A., Udalski, A., Shin, I.-G., Porritt, I., Skowron, J., Han, C., Yee, J. C., Kozłowski, S., Choi, J.-Y., Poleski, R., Wyrzykowski, L., Ulaczyk, K., Pietrukowicz, P., Mróz, P., Szymański, M. K., Kubiak, M., Soszyński, I., Pietrzyński, G., Gaudi, B. S., Christie, G. W., Drummond, J., McCormick, J., Natusch, T., Ngan, H., Tan, T.-G., Albrow, M., DePoy, D. L., Hwang, K.-H., Jung, Y. K., Lee, C.-U., Park, H., Pogge, R. W., Abe, F., Bennett, D. P., Bond, I. A., Botzler, C. S., Freeman, M., Fukui, A., Fukunaga, D., Itow, Y., Koshimoto, N., Larsen, P., Ling, C. H., Masuda, K., Matsubara, Y., Muraki, Y., Namba, S., Ohnishi, K., Philpott, L., Rattenbury, N. J., Saito, T., Sullivan, D. J., Sumi, T., Suzuki, D., Tristram, P. J., Tsurumi, N., Wada, K., Yamai, N., Yock, P. C. M., Yonehara, A., Shvartzvald, Y., Maoz, D., Kaspi, S., & Friedmann, M. 2014, *Science*, 345, 46
- Gray, D. F. 2008, *The Observation and Analysis of Stellar Photospheres*
- Griffin, R. 1973, *MNRAS*, 162, 243
- Griffin, R. F. 1967, *ApJ*, 148, 465
- Groff, T. D., Eldorado Riggs, A. J., Kern, B., & Jeremy Kasdin, N. 2015, *Journal of Astronomical Telescopes, Instruments, and Systems*, 2, 011009
- Guizar-Sicairos, M., Thurman, S. T., & Fienup, J. R. 2008, *Opt. Lett.*, 33, 156
- Guyon, O. 2004, *ApJ*, 615, 562

- Hatzes, A. P. & Cochran, W. D. 1992, in European Southern Observatory Conference and Workshop Proceedings, Vol. 40, European Southern Observatory Conference and Workshop Proceedings, ed. M.-H. Ulrich, 275
- Hatzes, A. P., Cochran, W. D., McArthur, B., Baliunas, S. L., Walker, G. A. H., Campbell, B., Irwin, A. W., Yang, S., Kürster, M., Endl, M., Els, S., Butler, R. P., & Marcy, G. W. 2000, *ApJ*, 544, L145
- Hauschildt, P. H., Allard, F., & Baron, E. 1999, *ApJ*, 512, 377
- Hayward, T. L., Brandl, B., Pirger, B., Blacken, C., Gull, G. E., Schoenwald, J., & Houck, J. R. 2001, *PASP*, 113, 105
- Hillenbrand, L., Isaacson, H., Marcy, G., Barenfeld, S., Fischer, D., & Howard, A. 2015, in Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, Vol. 18, 18th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun, ed. G. T. van Belle & H. C. Harris, 759–766
- Howard, A. W., Johnson, J. A., Marcy, G. W., Fischer, D. A., Wright, J. T., Bernat, D., Henry, G. W., Peek, K. M. G., Isaacson, H., Apps, K., Endl, M., Cochran, W. D., Valenti, J. A., Anderson, J., & Piskunov, N. E. 2010a, *ApJ*, 721, 1467
- Howard, A. W., Johnson, J. A., Marcy, G. W., Fischer, D. A., Wright, J. T., Henry, G. W., Giguere, M. J., Isaacson, H., Valenti, J. A., Anderson, J., & Piskunov, N. E. 2009, *ApJ*, 696, 75
- Howard, A. W., Johnson, J. A., Marcy, G. W., Fischer, D. A., Wright, J. T., Henry, G. W., Isaacson, H., Valenti, J. A., Anderson, J., & Piskunov, N. E. 2011, *ApJ*, 726, 73
- Howard, A. W., Marcy, G. W., Bryson, S. T., Jenkins, J. M., Rowe, J. F., Batalha, N. M., Borucki, W. J., Koch, D. G., Dunham, E. W., Gautier, III, T. N., Van Cleve, J., Cochran, W. D., Latham, D. W., Lissauer, J. J., Torres, G., Brown, T. M., Gilliland, R. L., Buchhave, L. A., Caldwell, D. A., Christensen-Dalsgaard, J., Ciardi, D., Fressin, F., Haas, M. R., Howell, S. B., Kjeldsen, H., Seager, S., Rogers, L., Sasselov, D. D., Steffen, J. H., Basri, G. S., Charbonneau, D., Christiansen, J., Clarke, B., Dupree, A., Fabrycky, D. C., Fischer, D. A., Ford, E. B., Fortney, J. J., Tarter, J., Girouard, F. R., Holman, M. J., Johnson, J. A., Klaus, T. C., Machalek, P., Moorhead, A. V., Morehead, R. C., Ragozzine, D., Tenenbaum, P., Twicken, J. D., Quinn, S. N., Isaacson, H., Shporer, A., Lucas, P. W., Walkowicz, L. M., Welsh, W. F., Boss, A., Devore, E., Gould, A., Smith, J. C., Morris, R. L., Prsa, A., Morton, T. D., Still, M., Thompson, S. E., Mullally, F., Endl, M., & MacQueen, P. J. 2012, *ApJS*, 201, 15
- Howard, A. W., Marcy, G. W., Johnson, J. A., Fischer, D. A., Wright, J. T., Isaacson, H., Valenti, J. A., Anderson, J., Lin, D. N. C., & Ida, S. 2010b, *Science*, 330, 653

- Howell, S. B., Sobek, C., Haas, M., Still, M., Barclay, T., Mullally, F., Troeltzsch, J., Aigrain, S., Bryson, S. T., Caldwell, D., Chaplin, W. J., Cochran, W. D., Huber, D., Marcy, G. W., Miglio, A., Najita, J. R., Smith, M., Twicken, J. D., & Fortney, J. J. 2014, *PASP*, 126, 398
- Huerta, M., Johns-Krull, C. M., Prato, L., Hartigan, P., & Jaffe, D. T. 2008, *ApJ*, 678, 472
- Irwin, J., Charbonneau, D., Nutzman, P., & Falco, E. 2009, in *IAU Symposium*, Vol. 253, IAU Symposium, ed. F. Pont, D. Sasselov, & M. J. Holman, 37–43
- Jovanovic, N., Martinache, F., Guyon, O., Clergeon, C., Singh, G., Kudo, T., Garrel, V., Newman, K., Doughty, D., Lozi, J., Males, J., Minowa, Y., Hayano, Y., Takato, N., Morino, J., Kuhn, J., Serabyn, E., Norris, B., Tuthill, P., Schworer, G., Stewart, P., Close, L., Huby, E., Perrin, G., Lacour, S., Gauchet, L., Vievard, S., Murakami, N., Oshiyama, F., Baba, N., Matsuo, T., Nishikawa, J., Tamura, M., Lai, O., Marchis, F., Duchene, G., Kotani, T., & Woillez, J. 2015a, *ArXiv e-prints*
- . 2015b, *PASP*, 127, 890
- Judge, P. G., Jordan, C., & Rowan-Robinson, M. 1987, *MNRAS*, 224, 93
- Kalas, P., Graham, J. R., Chiang, E., Fitzgerald, M. P., Clampin, M., Kite, E. S., Stapelfeldt, K., Marois, C., & Krist, J. 2008, *Science*, 322, 1345
- Kaltenegger, L. & Sasselov, D. 2011, *ApJ*, 736, L25
- Kambe, E., Sato, B., Takeda, Y., Ando, H., Noguchi, K., Aoki, W., Izumiura, H., Wada, S., Masuda, S., Okada, N., Shimizu, Y., Watanabe, E., Yoshida, M., Honda, S., & Kawanomoto, S. 2002, *PASJ*, 54, 865
- Kasdin, N. J., Vanderbei, R. J., Spergel, D. N., & Littman, M. G. 2003, *ApJ*, 582, 1147
- Kasting, J., Traub, W., Roberge, A., Leger, A., Schwartz, A., Wootten, A., Vosteen, A., Lo, A., Brack, A., Tanner, A., Coustenis, A., Lane, B., Oppenheimer, B., Mennesson, B., Lopez, B., Grillmair, C., Beichman, C., Cockell, C., Hanot, C., McCarthy, C., Stark, C., Marois, C., Aime, C., Angerhausen, D., Montes, D., Wilner, D., Defrere, D., Mourard, D., Lin, D., Kite, E., Chas-sefiere, E., Malbet, F., Tian, F., Westall, F., Illingworth, G., Vasisht, G., Serabyn, G., Marcy, G., Bryden, G., White, G., Laughlin, G., Torres, G., Hammel, H., Ferguson, H., Shibai, H., Rottgering, H., Surdej, J., Wiseman, J., Ge, J., Bally, J., Krist, J., Monnier, J., Trauger, J., Horner, J., Catanzarite, J., Harrington, J., Nishikawa, J., Stapelfeldt, K., von Braun, K., Biazzo, K., Carpenter, K., Balasubramanian, K., Kaltenegger, L., Postman, M., Spaans, M., Turnbull, M., Levine, M., Burchell, M., Ealey, M., Kuchner, M., Marley, M., Dominik, M., Mountain, M., Kenworthy, M., Muterspaugh, M., Shao, M., Zhao, M., Tamura, M., Kasdin, N., Haghhighipour,

- N., Kiang, N., Elias, N., Woolf, N., Mason, N., Absil, O., Guyon, O., Lay, O., Borde, P., Fouque, P., Kalas, P., Lowrance, P., Plavchan, P., Hinz, P., Kervella, P., Chen, P., Akeson, R., Soummer, R., Waters, R., Barry, R., Kendrick, R., Brown, R., Vanderbei, R., Woodruff, R., Danner, R., Allen, R., Polidan, R., Seager, S., MacPhee, S., Hosseini, S., Metchev, S., Kafka, S., Ridgway, S., Rinehart, S., Unwin, S., Shaklan, S., ten Brummelaar, T., Mazeh, T., Meadows, V., Weiss, W., Danchi, W., Ip, W., & Rabbia, Y. 2009, in *Astronomy*, Vol. 2010, astro2010: The Astronomy and Astrophysics Decadal Survey
- Kraus, A. L. & Ireland, M. J. 2012, *ApJ*, 745, 5
- Kuchner, M. J. & Holman, M. J. 2003, *ApJ*, 588, 1110
- Kuchner, M. J. & Traub, W. A. 2002, *ApJ*, 570, 900
- Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É. 2007, *ApJ*, 660, 770
- Lagrange, A.-M., Gratadour, D., Chauvin, G., Fusco, T., Ehrenreich, D., Mouillet, D., Rousset, G., Rouan, D., Allard, F., Gendron, É., Charton, J., Mugnier, L., Rabou, P., Montri, J., & Lacombe, F. 2009, *A&A*, 493, L21
- Lanza, A. F., Boisse, I., Bouchy, F., Bonomo, A. S., & Moutou, C. 2011, *A&A*, 533, A44
- Li, C.-H., Benedick, A. J., Fendel, P., Glenday, A. G., Kärtner, F. X., Phillips, D. F., Sasselov, D., Szentgyorgyi, A., & Walsworth, R. L. 2008, *Nature*, 452, 610
- Lindgren, L. 1978, in *IAU Colloq. 48: Modern Astrometry*, ed. F. V. Prochazka & R. H. Tucker, 197–217
- Lissauer, J. J., Hubickyj, O., D'Angelo, G., & Bodenheimer, P. 2009, *Icarus*, 199, 338
- Lissauer, J. J. & Stevenson, D. J. 2007, *Protostars and Planets V*, 591
- López-Morales, M. 2006, *PASP*, 118, 716
- Lupton, W. F. & Conrad, A. R. 1993, in *Astronomical Society of the Pacific Conference Series*, Vol. 52, *Astronomical Data Analysis Software and Systems II*, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, 315
- Macintosh, B., Graham, J. R., Barman, T., De Rosa, R. J., Konopacky, Q., Marley, M. S., Marois, C., Nielsen, E. L., Pueyo, L., Rajan, A., Rameau, J., Saumon, D., Wang, J. J., Patience, J., Ammons, M., Arriaga, P., Artigau, E., Beckwith, S., Brewster, J., Bruzzone, S., Bulger, J., Burningham, B., Burrows, A. S., Chen, C., Chiang, E., Chilcote, J. K., Dawson, R. I., Dong, R., Doyon, R., Draper, Z. H., Duchêne, G., Esposito, T. M., Fabrycky, D., Fitzgerald, M. P., Follette,

- K. B., Fortney, J. J., Gerard, B., Goodsell, S., Greenbaum, A. Z., Hibon, P., Hinkley, S., Cotten, T. H., Hung, L.-W., Ingraham, P., Johnson-Groh, M., Kalas, P., Lafreniere, D., Larkin, J. E., Lee, J., Line, M., Long, D., Maire, J., Marchis, F., Matthews, B. C., Max, C. E., Metchev, S., Millar-Blanchaer, M. A., Mittal, T., Morley, C. V., Morzinski, K. M., Murray-Clay, R., Oppenheimer, R., Palmer, D. W., Patel, R., Perrin, M. D., Poyneer, L. A., Rafikov, R. R., Rantakyro, F. T., Rice, E. L., Rojo, P., Rudy, A. R., Ruffio, J.-B., Ruiz, M. T., Sadakuni, N., Saddlemyer, L., Salama, M., Savransky, D., Schneider, A. C., Sivaramakrishnan, A., Song, I., Soummer, R., Thomas, S., Vasisht, G., Wallace, J. K., Ward-Duong, K., Wiktorowicz, S. J., Wolff, S. G., & Zuckerman, B. 2015, *Science*, 350, 64
- Macintosh, B., Graham, J. R., Ingraham, P., Konopacky, Q., Marois, C., Perrin, M., Poyneer, L., Bauman, B., Barman, T., Burrows, A. S., Cardwell, A., Chilcote, J., De Rosa, R. J., Dillon, D., Doyon, R., Dunn, J., Erikson, D., Fitzgerald, M. P., Gavel, D., Goodsell, S., Hartung, M., Hibon, P., Kalas, P., Larkin, J., Maire, J., Marchis, F., Marley, M. S., McBride, J., Millar-Blanchaer, M., Morzinski, K., Norton, A., Oppenheimer, B. R., Palmer, D., Patience, J., Pueyo, L., Rantakyro, F., Sadakuni, N., Saddlemyer, L., Savransky, D., Serio, A., Soummer, R., Sivaramakrishnan, A., Song, I., Thomas, S., Wallace, J. K., Wiktorowicz, S., & Wolff, S. 2014, *Proceedings of the National Academy of Science*, 111, 12661
- Malacara, D. 2007, *Optical Shop Testing (Wiley Series in Pure and Applied Optics)* (Wiley-Interscience)
- Mann, A. W., Gaidos, E., & Aldering, G. 2011, *PASP*, 123, 1273
- Marcy, G. W. & Butler, R. P. 1996, *ApJ*, 464, L147
- Marois, C., Doyon, R., Racine, R., & Nadeau, D. 2000, *PASP*, 112, 91
- Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006a, *ApJ*, 641, 556
- Marois, C., Macintosh, B., Barman, T., Zuckerman, B., Song, I., Patience, J., Lafrenière, D., & Doyon, R. 2008, *Science*, 322, 1348
- Marois, C., Phillion, D. W., & Macintosh, B. 2006b, in *Proc. SPIE, Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 62693M
- Martinache, F., Guyon, O., Jovanovic, N., Clergeon, C., Singh, G., Kudo, T., Currie, T., Thalmann, C., McElwain, M., & Tamura, M. 2014, *PASP*, 126, 565
- Massarotti, A., Latham, D. W., Stefanik, R. P., & Fogel, J. 2008, *AJ*, 135, 209
- Mawet, D., Mennesson, B., Serabyn, E., Stapelfeldt, K., & Absil, O. 2011a, *ApJ*, 738, L12

- Mawet, D., Pueyo, L., Carlotti, A., Mennesson, B., Serabyn, E., & Wallace, J. K. 2013, *ApJS*, 209, 7
- Mawet, D., Pueyo, L., Lawson, P., Mugnier, L., Traub, W., Boccaletti, A., Trauger, J. T., Gladysz, S., Serabyn, E., Milli, J., Belikov, R., Kasper, M., Baudoz, P., Macintosh, B., Marois, C., Oppenheimer, B., Barrett, H., Beuzit, J.-L., Devaney, N., Girard, J., Guyon, O., Krist, J., Mennesson, B., Mouillet, D., Murakami, N., Poyneer, L., Savransky, D., Vérinaud, C., & Wallace, J. K. 2012, in *Proc. SPIE*, Vol. 8442, *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*, 844204
- Mawet, D., Riaud, P., Absil, O., & Surdej, J. 2005, *ApJ*, 633, 1191
- Mawet, D., Serabyn, E., Liewer, K., Hanot, C., McEldowney, S., Shemo, D., & O'Brien, N. 2009, *Optics Express*, 17, 1902
- Mawet, D., Serabyn, E., Wallace, J. K., & Pueyo, L. 2011b, *Optics Letters*, 36, 1506
- Mawet, D., Serabyn, E., Wallace, J. K., & Pueyo, L. 2011, *Opt. Lett.*, 36, 1506
- Mayor, M. & Queloz, D. 1995, *Nature*, 378, 355
- Mazin, B. A., Eyken, J. v., Meeker, S. R., Jensen-Clem, R., Oppenheimer, B. R., & Guyon, O. 2014, in *Search for Life Beyond the Solar System. Exoplanets, Biosignatures and Instruments*, ed. D. Apai & P. Gabor, 4P
- McArthur, B. E., Endl, M., Cochran, W. D., Benedict, G. F., Fischer, D. A., Marcy, G. W., Butler, R. P., Naef, D., Mayor, M., Queloz, D., Udry, S., & Harrison, T. E. 2004, *ApJ*, 614, L81
- McCrary, N. & Nava, C. 2014, in *American Astronomical Society Meeting Abstracts*, Vol. 223, *American Astronomical Society Meeting Abstracts #223*, 348.03
- McLean, I. S. 2008, *Electronic Imaging in Astronomy: Detectors and Instrumentation (Second Edition)* (Praxis Publishing)
- McQuillan, A., Aigrain, S., & Mazeh, T. 2013, *MNRAS*, 432, 1203
- Metchev, S. A. & Hillenbrand, L. A. 2004, *ApJ*, 617, 1330
- . 2009, *ApJS*, 181, 62
- Millan-Gabet, R., Serabyn, E., Mennesson, B., Traub, W. A., Barry, R. K., Danchi, W. C., Kuchner, M., Stark, C. C., Ragland, S., Hrynevych, M., Woillez, J., Stapelfeldt, K., Bryden, G., Colavita, M. M., & Booth, A. J. 2011, *ApJ*, 734, 67

- Mortier, A., Faria, J. P., Santos, N. C., Rajpaul, V., Figueira, P., Boisse, I., Collier Cameron, A., Dumusque, X., Lo Curto, G., Lovis, C., Mayor, M., Melo, C., Pepe, F., Queloz, D., Santerne, A., Ségransan, D., Sousa, S. G., Sozzetti, A., & Udry, S. 2016, *A&A*, 585, A135
- Murakami, N., Hamaguchi, S., Sakamoto, M., Fukumoto, R., Ise, A., Oka, K., Baba, N., & Tamura, M. 2013, *Opt. Express*, 21, 7400
- Paul, B., Mugnier, L. M., Sauvage, J.-F., Ferrari, M., & Dohlen, K. 2013, *Optics Express*, 21, 31751
- Peale, S. J. 1977, in *IAU Colloq. 28: Planetary Satellites*, ed. J. A. Burns, 87–111
- Pepe, F., Lovis, C., Ségransan, D., Benz, W., Bouchy, F., Dumusque, X., Mayor, M., Queloz, D., Santos, N. C., & Udry, S. 2011, *A&A*, 534, A58
- Perryman, M. 2011, *The Exoplanet Handbook*
- Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, *Proceedings of the National Academy of Science*, 110, 19273
- Piau, L., Kervella, P., Dib, S., & Hauschildt, P. 2011, *A&A*, 526, A100
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., Crone, G., Doyle, D., Gageur, U., Heras, A. M., Jewell, C., Metcalfe, L., Ott, S., & Schmidt, M. 2010, *A&A*, 518, L1
- Pollacco, D. L., Skillen, I., Collier Cameron, A., Christian, D. J., Hellier, C., Irwin, J., Lister, T. A., Street, R. A., West, R. G., Anderson, D. R., Clarkson, W. I., Deeg, H., Enoch, B., Evans, A., Fitzsimmons, A., Haswell, C. A., Hodgkin, S., Horne, K., Kane, S. R., Keenan, F. P., Maxted, P. F. L., Norton, A. J., Osborne, J., Parley, N. R., Ryans, R. S. I., Smalley, B., Wheatley, P. J., & Wilson, D. M. 2006, *PASP*, 118, 1407
- Pueyo, L., Kay, J., Kasdin, N. J., Groff, T., Mc Elwain, M., Give'on, A., & Belikov, R. 2011, *ArXiv e-prints*
- Pueyo, L., Soummer, R., Hoffmann, J., Oppenheimer, R., Graham, J. R., Zimmerman, N., Zhai, C., Wallace, J. K., Vescelus, F., Veicht, A., Vasisht, G., Truong, T., Sivaramakrishnan, A., Shao, M., Roberts, Jr., L. C., Roberts, J. E., Rice, E., Parry, I. R., Nilsson, R., Lockhart, T., Ligon, E. R., King, D., Hinkley, S., Hillenbrand, L., Hale, D., Dekany, R., Crepp, J. R., Cady, E., Burruss, R., Brenner, D., Beichman, C., & Baranec, C. 2015, *ApJ*, 803, 31
- Racine, R., Walker, G. A. H., Nadeau, D., Doyon, R., & Marois, C. 1999, *PASP*, 111, 587
- Rajpaul, V., Aigrain, S., & Roberts, S. 2016, *MNRAS*, 456, L6
- Reid, I. N., Gizis, J. E., & Hawley, S. L. 2002, *AJ*, 124, 2721

- Reid, M. J. & Honma, M. 2014, *ARA&A*, 52, 339
- Reiners, A., Bean, J. L., Huber, K. F., Dreizler, S., Seifahrt, A., & Czesla, S. 2010, *ApJ*, 710, 432
- Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. Á., Bean, J. L., Bert-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe, F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szentgyorgyi, A., Torres, G., Udry, S., & Villaseñor, J. 2014, in *Proc. SPIE*, Vol. 9143, *Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave*, 914320
- Riggs, A. J. E., Kasdin, N. J., & Groff, T. D. 2016, *Journal of Astronomical Telescopes, Instruments, and Systems*, 2, 011017
- Rivera, E. J., Lissauer, J. J., Butler, R. P., Marcy, G. W., Vogt, S. S., Fischer, D. A., Brown, T. M., Laughlin, G., & Henry, G. W. 2005, *ApJ*, 634, 625
- Sauvage, J.-F., Mugnier, L., Paul, B., & Villecroze, R. 2012, *Optics Letters*, 37, 4808
- Savransky, D., Macintosh, B. A., Thomas, S. J., Poyneer, L. A., Palmer, D. W., De Rosa, R. J., & Hartung, M. 2012, in *Proc. SPIE*, Vol. 8447, *Adaptive Optics Systems III*, 84476S
- Scargle, J. D. 1982, *ApJ*, 263, 835
- Seager, S. 2010, *Exoplanets*
- Seager, S. & Deming, D. 2010, *ARA&A*, 48, 631
- Serabyn, E., Mawet, D., & Burruss, R. 2010, *Nature*, 464, 1018
- Serabyn, E., Wallace, J. K., & Mawet, D. 2011, *Appl. Opt.*, 50, 5453
- Serabyn, E., Wallace, J. K., & Mawet, D. 2011, *Appl. Opt.*, 50, 5453
- Shelton, J. C. 1997, in *Proc. SPIE*, Vol. 3126, *Adaptive Optics and Applications*, ed. R. K. Tyson & R. Q. Fugate, 455
- Singh, G., Martinache, F., Baudoz, P., Guyon, O., Matsuo, T., Jovanovic, N., & Clergeon, C. 2014, *PASP*, 126, 586

- Sivaramakrishnan, A., Koresko, C. D., Makidon, R. B., Berkefeld, T., & Kuchner, M. J. 2001, *ApJ*, 552, 397
- Skemer, A. J., Marley, M. S., Hinz, P. M., Morzinski, K. M., Skrutskie, M. F., Leisenring, J. M., Close, L. M., Saumon, D., Bailey, V. P., Briguglio, R., Defrere, D., Esposito, S., Follette, K. B., Hill, J. M., Males, J. R., Puglisi, A., Rodigas, T. J., & Xompero, M. 2014, *ApJ*, 792, 17
- Snik, F., Otten, G., Kenworthy, M., Miskiewicz, M., Escuti, M., Packham, C., & Codona, J. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8450, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 0
- Soummer, R., Pueyo, L., & Larkin, J. 2012, *ApJ*, 755, L28
- Sousa, S. G., Santos, N. C., Mayor, M., Udry, S., Casagrande, L., Israelian, G., Pepe, F., Queloz, D., & Monteiro, M. J. P. F. G. 2008, *A&A*, 487, 373
- Spiegel, D. S. & Burrows, A. 2012, *ApJ*, 745, 174
- Struve, O. 1952, *The Observatory*, 72, 199
- Swift, J. J., Bottom, M., Johnson, J. A., Wright, J. T., McCrady, N., Wittenmyer, R. A., Plavchan, P., Riddle, R., Muirhead, P. S., Herzig, E., Myles, J., Blake, C. H., Eastman, J., Beatty, T. G., Barnes, S. I., Gibson, S. R., Lin, B., Zhao, M., Gardner, P., Falco, E., Criswell, S., Nava, C., Robinson, C., Sliski, D. H., Hedrick, R., Ivarsen, K., Hjelstrom, A., de Vera, J., & Szentgyorgyi, A. 2015, *Journal of Astronomical Telescopes, Instruments, and Systems*, 1, 027002
- Thomas, S. J., Give'on, A. A., Dillon, D., Macintosh, B., Gavel, D., & Soummer, R. 2010, in *Proc. SPIE*, Vol. 7736, *Adaptive Optics Systems II*, 77365L
- Traub, W. A. & Oppenheimer, B. R. *Direct Imaging of Exoplanets*, ed. S. Seager, 111–156
- Udry, S., Bonfils, X., Delfosse, X., Forveille, T., Mayor, M., Perrier, C., Bouchy, F., Lovis, C., Pepe, F., Queloz, D., & Bertaux, J.-L. 2007, *A&A*, 469, L43
- Valenti, J. A., Butler, R. P., & Marcy, G. W. 1995, *PASP*, 107, 966
- van Leeuwen, F. 2007, *A&A*, 474, 653
- Walker, G. A. H., Walker, A. R., Irwin, A. W., Larson, A. M., Yang, S. L. S., & Richardson, D. C. 1995, *Icarus*, 116, 359
- Wang, J. & Ge, J. 2011, *ArXiv e-prints*

- Wheatley, P. J., Pollacco, D. L., Queloz, D., Rauer, H., Watson, C. A., West, R. G., Chazelas, B., Loudon, T. M., Walker, S., Bannister, N., Bento, J., Burleigh, M., Cabrera, J., Eigmüller, P., Erikson, A., Genolet, L., Goad, M., Grange, A., Jordán, A., Lawrie, K., McCormac, J., & Neveu, M. 2013, in *European Physical Journal Web of Conferences*, Vol. 47, *European Physical Journal Web of Conferences*, 13002
- Wizinowich, P. L., Le Mignant, D., Bouchez, A. H., Campbell, R. D., Chin, J. C. Y., Contos, A. R., van Dam, M. A., Hartman, S. K., Johansson, E. M., Lafon, R. E., Lewis, H., Stomski, P. J., Summers, D. M., Brown, C. G., Danforth, P. M., Max, C. E., & Pennington, D. M. 2006, *PASP*, 118, 297
- Wolszczan, A. & Frail, D. A. 1992, *Nature*, 355, 145
- Wright, J. T. & Eastman, J. D. 2014, *PASP*, 126, 838
- Wright, J. T., Fakhouri, O., Marcy, G. W., Han, E., Feng, Y., Johnson, J. A., Howard, A. W., Fischer, D. A., Valenti, J. A., Anderson, J., & Piskunov, N. 2011, *PASP*, 123, 412
- Yi, X., Vahala, K., Li, J., Diddams, S., Ycas, G., Plavchan, P., Leifer, S., Sandhu, J., Vasisht, G., Chen, P., Gao, P., Gagne, J., Furlan, E., Bottom, M., Martin, E. C., Fitzgerald, M. P., Doppmann, G., & Beichman, C. 2016, *Nature Communications*, 7, 10436
- Zurlo, A., Vigan, A., Mesa, D., Gratton, R., Moutou, C., Langlois, M., Claudi, R. U., Pueyo, L., Boccaletti, A., Baruffolo, A., Beuzit, J.-L., Costille, A., Desidera, S., Dohlen, K., Feldt, M., Fusco, T., Henning, T., Kasper, M., Martinez, P., Moeller-Nilsson, O., Mouillet, D., Pavlov, A., Puget, P., Sauvage, J.-F., Turatto, M., Udry, S., Vakili, F., Waters, R., & Wildi, R. F. 2014, *A&A*, 572, A85