Observational, Numerical, and Laboratory Methods in High Contrast Imaging

by

Benjamin Lionel Gerard
B.A., University of Colorado, 2014

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Physics and Astronomy

Observational, Numerical, and Laboratory Methods in High Contrast Imaging

by

Benjamin Lionel Gerard
B.A., University of Colorado, 2014

Supervisory Committee

Dr. C. Marois, Co-Supervisor
(Department of Physics and Astronomy)

Dr. K. Venn, Co-Supervisor
(Department of Physics and Astronomy)

Supervisory Committee

Dr. C. Marois, Co-Supervisor
(Department of Physics and Astronomy)

Dr. K. Venn, Co-Supervisor
(Department of Physics and Astronomy)

ABSTRACT

The search to directly image and characterize exoplanets that are initially hidden below the stellar and instrumental noise relies on the use of both extreme adaptive optics (AO) and a subsequent point spread function (PSF) subtraction pipeline. In this thesis I present my research on both real-time AO techniques and post-processing PSF subtraction techniques. First, I present a new PSF subtraction algorithm designed to image the HR 8799 debris disk using the *Hubble Space Telescope*. I find an over-luminosity after PSF subtraction that may be from the inner disk and/or planetesimal belt components of this system, but ultimately conclude that this is likely a non-detection as a result of telescope stability and broadband chromatic effects. Thus, assuming a non-detection, I derive upper limits on the HR 8799 dust belt mass in small grains, consistent with measurements of other debris disk halos. Next, I present a new PSF subtraction algorithm applied to current campaign data from the Gemini Planet Imager (GPI), designed to optimize the GPI planet detection sensitivity of narrow orbit planets. My results, while still being investigated, seem to show that current algorithms are already optimized, and that limited gains can be achieved with my new algorithm. Finally, I apply a new real-time AO nulling technique, called super-Nyquist wavefront control (SNWFC), to be used on future 30 m class telescopes to image wide-orbit exoplanets. I demonstrate application of SNWFC in both a deterministic laboratory experiment and coronagraphic simulations using an interferometric nulling technique, suggesting that this technique would allow higher SNR characterization of wide-orbit exoplanets on future telescopes.

Contents

List of Tables

List of Figures

CO-AUTHORSHIP

The following co-authors contributed to the content of this thesis:

- whole thesis - **Dr. Christian Marois**, for the advising, input, and editing throughout this thesis, which draws from my three publications: Gerard et al. (2016a), Gerard et al. (2016b), and Gerard & Marois (2016).
- Chapter 2 - **Dr. Samantha Lawler**, for editing and co-authorship of this chapter, originally published in Gerard et al. (2016a), particularly in an analysis of blowout grain ejection timescale and a comparison to other systems.
- Chapter 2 - **Megan Tannock**, for the initial work and direction on the HR 8799 debris-disk project as part of a summer co-op.
- Chapter 3 - **GPIES collaboration**. For the design, testing, construction, commissioning, data reduction, and references therein of the Gemini Planet Imager Exoplanet Survey (GPIES), from which data in this chapter is used for my analysis and pipeline development.

ACKNOWLEDGEMENTS

I would like to thank:

my Supervisory Committee for your comments and suggestions in editing this thesis.

Masen Lamb and Maaike Van Kooten for your help in learning the ways of the bench.

AO coffee for your help, instruction, and discussion on adaptive optics.

SAS Journal Club for your help, instruction, and discussion on debris disks and Kuiper belt objects.

Jared Keown for listening and engaging in discussions of dark holes.

Jared Keown, Mara Johnson-Groh, Austin Davis, and Connor Bottrell for your support as friends and classmates to make it though a year of coursework.

the UVic astrograds for your support as astronomers and friends to look at and/or think about cool things in the sky.

Charli Sakari for your comments and suggestions in editing the introduction, conclusion, background sections of this thesis.

my parents for your love and support, without which I would not be where I am today.

my grandfather for your persistence in scientific thinking as a role model in physics and engineering.

Also, I gratefully acknowledge research support of the Natural Sciences and Engineering Council (NSERC) of Canada. Samantha M Lawler gratefully acknowledges support from the NRC Canada Plaskett Fellowship. This research used the facilities of the Canadian Astronomy Data Centre operated by the National Research Council of Canada with the support of the Canadian Space Agency. STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for MAST for non-*HST* data is provided by the NASA Office of Space Science via grant NNX09AF08G and by other grants and contracts.

I thank the anonymous referee for his or her comments and suggestions that have improved Chapter 2, originally published in Gerard et al. (2016a). The GPI project has been supported by Gemini Observatory, which is operated by AURA, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the NSF (USA), the National Research Council (Canada), CONICYT (Chile), the Australian Research Council (Australia), MCTI (Brazil) and MINCYT (Argentina).

DEDICATION

To my family and friends that helped me along the way

Table 1: Table of Acronyms

Acronym	**Description**
ADI	angular differential imaging
AO	adaptive optics
DM	deformable mirror
ELT	extremely large telescope
ExAO	extreme adaptive optics
FM	forward model
FOV	field of view
FPM	focal plane mask
FWHM	full width at half maximum
GPI	Gemini Planet Imager
GPIES	GPI Exoplanet Survey
HST	Hubble Space Telescope
IFS	integral field spectrograph
IRAS	Infrared Astronomical Telescope
IWA	inner working angle
JWST	James Webb Space Telescope
KLIP	Karhunen-Loève Image Projection
LOCI	locally optimized combination of images
MAST	Mukilski Archive for Space Telescopes
MTF	modulation transfer function
NCPA	non common path aberration
NIR	near infrared
NNLS	non-negative least-squares
OSFi	Optimized Spatially Filtered
OTF	optical transfer function
PSF	point spread function
RSDI	reference star differential imaging
RTC	real time control system (algorithm for AO)
SC	science camera
SCAO	single conjugate AO
SCC	self-coherent camera
SED	spectral energy distribution

SHWFS	Shack Hartmann wavefront sensor
SNR	signal to noise ratio
SOSIE	speckle-optimized subtraction for imaging exolanets
SR	Strehl ratio
SSDI	simultaneous spectral differential imaging
STIS	Space Telescope Imaging Spectrograph
SVD	singular value decomposition
TLOCI	template locally optimized combination of images
WFIRST	Wide Field Infrared Survey Telescope
WFS	wavefront sensor

Chapter 1

Introduction

The Solar System was formed in a cloud of collapsing gas and dust. Although this process of stellar and planetary formation is undoubtedly common, we have only recently, in the past ∼30 years, begun to accumulate enough data and formulate consistent working theories to understand our place in the Galaxy and Universe, such as the statistical prevalence of Jovian planets, terrestrial planets, habitable planets, and the formation and evolution mechanisms for all of the above. These observational and theoretical fields of exoplanet astronomy have now branched into many "flavours." The focus of this thesis will be on the direct imaging exoplanet detection technique, and so it is noted that all the introductory and contextual information in this section and at the beginning of each Chapter is biased to this method.

1.1 What is an Exoplanet?

Answer: A planet in another stellar system, which brings to question *what is a planet*? But there are a number of subtleties to this simple question. In the past 25 years, the detection and characterization of Kuiper Belt Objects (KBOs) in our own Solar System and brown dwarfs/exoplanets in other solar systems required a more unambiguous definition, and thus the infamous IAU resolution declared that (IAU, 2006)

> "A planet is a celestial body that:
>
> (a) is in orbit around the Sun,
>
> (b) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and

(c) has cleared the neighbourhood around its orbit."

Thus, by item (c) Pluto and other round KBOs are not planets. However, the information in items (b) - (c) is often unobtainable with any current exoplanet detection method (§1.3), none of which are sensitive to planet masses (M_P) below $M_P \lesssim 1\ M_\oplus$, and thus there is no current working IAU definition to distinguish low mass exoplanets from "dwarf exoplanets" or "small solar system bodies" (Boss et al., 2012).

However, on the high mass end, more directly applicable to exoplanets, a planet can be considered as "a round object, not capable of core fusion, that orbits an object capable of core fusion" (Basri & Brown, 2006). At Solar metallicity, the lower mass limit for hydrogen fusion is $M \sim 0.07 - 0.074\ M_\odot \equiv 73 - 78\ M_J$, the lithium burning limit (via the reactions $^7\mathrm{Li} + p \rightarrow 2\alpha$ and $^6\mathrm{Li} \rightarrow \alpha +^3 \mathrm{He}$) is $M \sim 63\ M_J$, and the deuterium burning limit (via the reaction $^2\mathrm{H} + p \rightarrow \gamma +^3 \mathrm{He}$) is $M \sim 13\ M_J$ (Burrows et al., 2001). Thus, throughout this thesis I will use this $M \sim 13\ M_J$ upper limit to distinguish "planets"[1] from "brown dwarfs," which is also the same working definition agreed upon by the 2003 IAU Working Group on Extrasolar Planets (Boss et al., 2003):

> "Objects with true masses below the limiting mass for thermonuclear fusion of deuterium (currently calculated to be 13 Jupiter masses for objects of solar metallicity) that orbit stars or stellar remnants are 'planets' (no matter how they formed). "

It is noted that the above limits are metalicity-dependent (with no metals, the hydrogen limit is $M \sim 96\ M_J$; Burrows et al. 2001), and that mass is not a directly observable quantity in direct imaging (§1.2.2). Thus, even today, there is still ambiguity in distinguishing a planet from a brown dwarf in this $\sim 13\ M_J$ regime (see Bowler (2016), Table 1, "Candidate Planets and Companions Near the Deuterium-Burning Limit," and references therein). Also based on the definitions in Boss et al. (2003), the claimed detections via deep optical and near IR imaging of "free floating planets" in star forming regions (e.g., Quanz et al., 2010), although below 13 M_J, should instead be referred to as "sub-brown dwarfs," ultimately forming from the low mass end of molecular cloud mass fragmentation as opposed to gravitational instability and/or grain growth within a circumstellar disk (§1.2.1), as no such detections show evidence of being ejected from a previous planetary system (Perryman, 2011).

[1] Also throughout this thesis, I will use the words "planet" and "exoplanet" interchangeably.

1.2 Theoretical Methods

In this section I will discuss the current theories of high mass planet formation (§1.2.1) and brown dwarf/exoplanet atmospheres (§1.2.2) relevant to the direct imaging exoplanet detection method.

1.2.1 Planet Formation

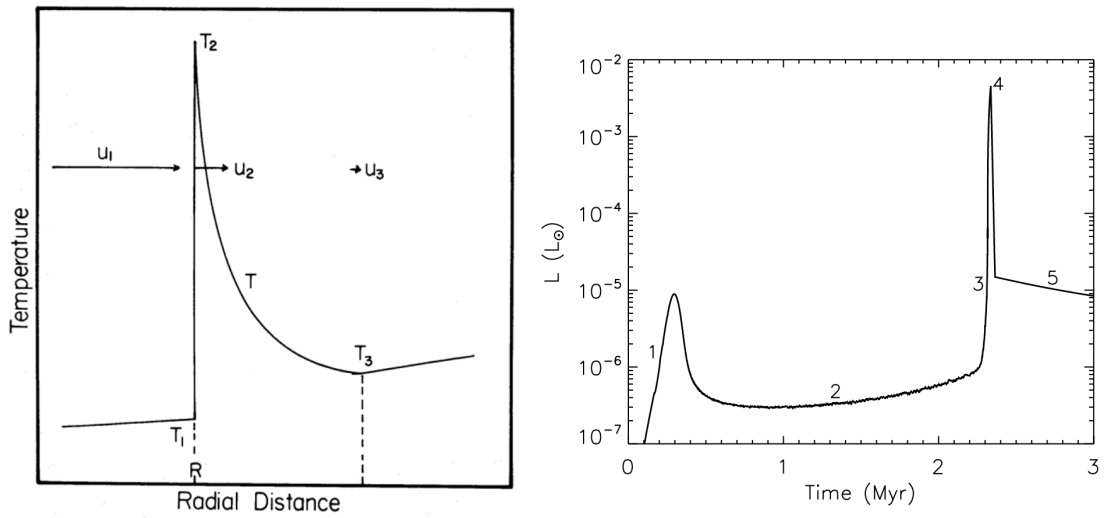
There are two main theories of massive planet formation: disk instability and core accretion:

1. In the disk instability scenario, gas undergoes hydrodynamical collapse to form planets within a protostellar accretion disk (an accretion disk of gas and dust surrounding a newly formed star; see §2.1) on ∼100 year orbital timescales (Boss, 2000).

2. The alternative core accretion scenario involves the initial creation of a solid core followed by gas accretion onto the core, which instead occurs on timescales of a few Myr (Marley et al., 2007).

Core accretion is considered a "cold start" relative to the "hot start" disk instability because in this scenario the gas loses energy while accreting onto the planet core (Marley et al., 2007). The mechanism for energy loss is called an "accretion shock." An illustration of this concept, initially proposed to explain protostar collapse by Stahler et al. (1980), and the expected luminosity signature for core collapse planet formation by Marley et al. (2007) is shown in Figure 1.1. The previously radiation-dominated accreting gas is shocked to such a high temperature because it (1) has a high incoming gas velocity and (2) encounters the much heavier mean molecular weight at the core boundary, both of which significantly increase the kinetic energy. By the Virial Theorem, this enormous amount of kinetic energy is approximately the same as the core's gravitational energy, which sets the temperature at the centre of the core. Accordingly, the gas must then cool to the temperature at the core's surface, and so it radiates away this excess energy generated from the shock.

In comparison, a "hot start" disk instability is thought to retain a much higher gas temperature during formation (Boss, 2000), although after ∼100 Myr a planet formed this way will eventually cool to the same temperature as a planet formed through core accretion (Spiegel & Burrows 2012; Figure 1.2). Thus, in young planetary systems on



(a) (b)

Figure 1.1 (a) An illustration of the gas cooling process during an accretion shock from Stahler et al. (1980). Accreting gas moves towards the planet core from left to right. The accretion shock occurs at point 2. The accreting gas at temperature T_1 and velocity u_1 reaches the dense hydrostatic core (which is at temperature T_3), slowing the gas velocity and shocking the gas temperature to $T_2 >> T_3$. Between point 2 and 3 the gas cools until it reaches T_3. (b) The stages of cold start planet formation over time for a 1 M_J planet, from Marley et al. (2007). The planet core grows in step 1, followed by gas accretion in step 2 until the core and envelope masses become equal, followed by runaway gas accretion in step 3, ultimately leading to the accretion shock in step 4, and in step 5 the planet slowly cools to its final, stable state.

the order of tens of Myr, planets formed through disk instability can be about 4.5 to 9 magnitudes brighter than planets formed through core accretion. Accordingly, the youngest planetary systems are the best laboratories for understanding the context and diversity of giant planet formation scenarios, a bias in favour of direct imaging (§1.4).

Gravitational disk instability is accepted as the formation mechanism for brown dwarfs, and most directly imaged planets thus far fall in this regime (Bowler, 2016). However, core accretion is still believed to be the main mechanism of planet formation within ∼50 AU (Perryman, 2011), which includes Jupiter (Hubickyj et al., 2005). The sensitivity level reached with current and future direct imaging surveys should be sufficient to detect a larger sample of young planets in order to better understand the different regimes for each formation scenario. Currently, 51 Eridani b (Macintosh

et al., 2015, see §1.2.2, §1.4.3) is the only directly imaged exoplanet consistent with cold start models, illustrated in Figure 1.2.

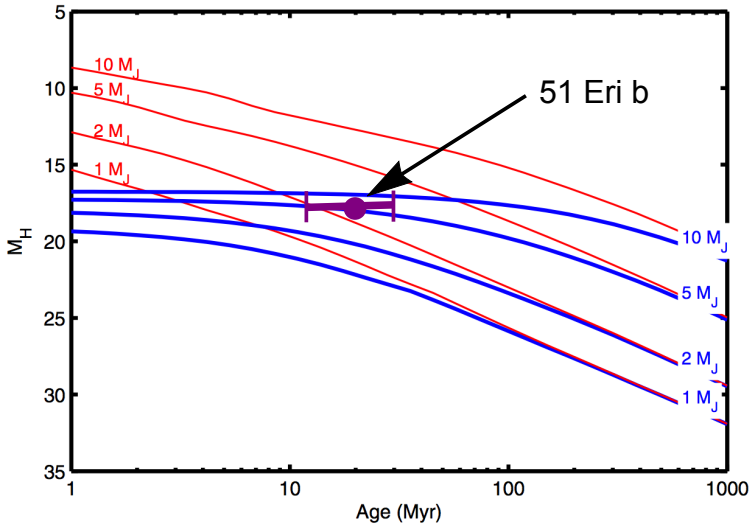


Figure 1.2 A comparison of the evolutionary cooling curves, showing absolute H-band magnitude vs. age, for both hot start models (red) and cold start models (blue) from Spiegel & Burrows (2012). The exoplanet 51 Eridani b from Macintosh et al. (2015) is also shown in purple (H-band magnitude error bars are within the size of the data point), indicating that its formation scenario is consistent with both models.

1.2.2 Exoplanet Atmospheres

The three main classes of brown dwarfs are the L, T, and Y spectral types, which we will see below informs the study of giant exoplanet atmospheres. As summarized in Burrows et al. (2001), L-type dwarfs are known to be dominated above ∼1500 K by H_2O, CO, and silicate grains, and also by TiO and VO above ∼2000 K, although the presence of clouds (see below) produces a relatively flat near infrared (NIR) spectrum. Transitioning to L-type dwarfs below ∼1500 K, clouds become less effective, and the dominant opacity sources are then H_2O, CH_4, NH_3, H_2, and alkali metals. NIR low resolution L-type spectra are typically characterized by strong methane absorption in the J and H bands as well as an overall blue colour that is enhanced relative to the blackbody value (see below). Finally, below ∼500 K, water clouds begin to form in the atmosphere, signalling the start of the Y dwarf spectral class.

Spectral observations thus far of giant exoplanets are empirically linked to the known L, T, and Y brown dwarf spectral classes (Figure 1.3a; Bowler, 2016). Thus,

analogous to brown dwarf evolution, giant exoplanets are also understood to transition in spectral type from L to T to Y as they cool and age. There are, however, a number of caveats to this assumption, such as the dependence of evolution timescale on formation pathway (disk instability vs. core accretion). In brown dwarfs, deuterium burning (which is still not sufficient in balancing radiative losses) may play a role in prolonging the L-T transition (Bowler, 2016). Figure 1.3b shows the brown

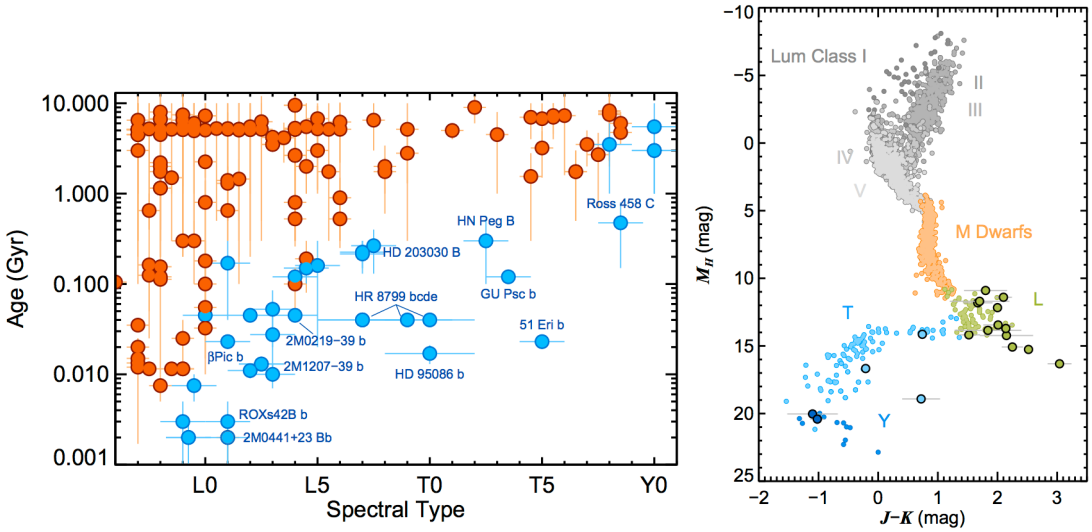


Figure 1.3 (a) A catalog of spectral type vs. age for directly imaged planets (blue) and brown dwarfs (red), in both cases illustrating spectral type evolution as a function of age (i.e., there are more T and Y types at older vs. younger ages). (b) The colour magnitude diagram for brown dwarfs, including M dwarfs, main sequence, and post-main sequence stars for reference. L, T, and Y brown dwarfs are colour coded to match with directly imaged planets of the same spectral type, shown with bolded circles. Both figures are from Bowler (2016).

dwarf colour magnitude diagram, including main sequence and post main sequence stars for reference, overlaid with the current giant exoplanet detections as bold points.

Because of the few number of Y dwarf detections as of yet (e.g., Kirkpatrick et al., 2012), most of the focus in the exoplanet atmospheres modelling community has been on the L-T transition in brown dwarfs and giant exoplanets, which I will discuss below. The L-T transition is believed to be caused by an evolution of cloud opacity as a function of temperature (Saumon & Marley, 2008). At higher temperatures, the presence of certain gaseous molecular species in clouds causes greater uniform absorption (i.e., a larger opacity) as a function of wavelength compared to lower

temperatures. For example, at $T_{\rm eff} \gtrsim 2000$ K, L-type dwarfs have clouds composed of TiO and VO, but by $T_{\rm eff} \lesssim 1500$ K these molecules undergo sedimentation, or "rain out," where they condense into solids and no longer act as an opacity source (Burrows et al., 2001; Saumon & Marley, 2008). Saumon & Marley (2008) show that in a simple model, varying this sedimentation effect with temperature can reproduce good fits to the L-T transition, although they note that as of yet there is no cloud model that includes more realistic physics. After this transition, T dwarfs, which have lower opacity clouds than L dwarfs, appear bluer in the NIR, as in Figures 1.3b and 1.4. This effect is due to

- the decreased absorption cross section of water (H_2O) and molecular hydrogen (H_2) at bluer wavelengths, ultimately deviating above the Planck function in the NIR by as much as two to five orders of magnitude (Burrows et al., 2001), and

- increased methane (CH_4) absorption in the H and K bands compared to J (Burrows et al., 2001; Saumon & Marley, 2008, and see Figure 1.4).

One important result that has emerged from direct imaging data is that massive exoplanets appear to transition from L to T spectral type at a redder, cooler, and later stage in evolution in comparison to the brown dwarf field population, as shown in Figure 1.3b (Bowler, 2016, i.e., along the L sequence, the bolded points generally appear redder and dimmer than the unbolded points). The mechanisms causing this later, cooler L-T transition, are currently understood to be a result of additional changes in cloud properties (Marley et al., 2012), although the specific temperature range where this cooler transition occurs is not yet well understood, mainly due to the lack of T-type detections of giant exoplanets thus far (Macintosh et al., 2015).

A detailed review of the current input physics involved in modelling the properties of giant exoplanet atmospheres (including radiative transfer, dynamics and mixing, chemistry, gas opacities, and clouds and condensates) can be found in the review paper by Marley & Robinson (2015) and references therein, and is beyond the scope of this section. A review on the numerous techniques in direct imaging used specifically to obtain M_p can be found in Bowler (2016). Below, we outline an example case of one such technique: spectral template fitting.

The recent detection and characterization of 51 Eridani b by Macintosh et al. (2015) represents the first and only directly imaged T-type giant exoplanet thus far, an important step in understanding the L-T transition in this regime. Best fit models are

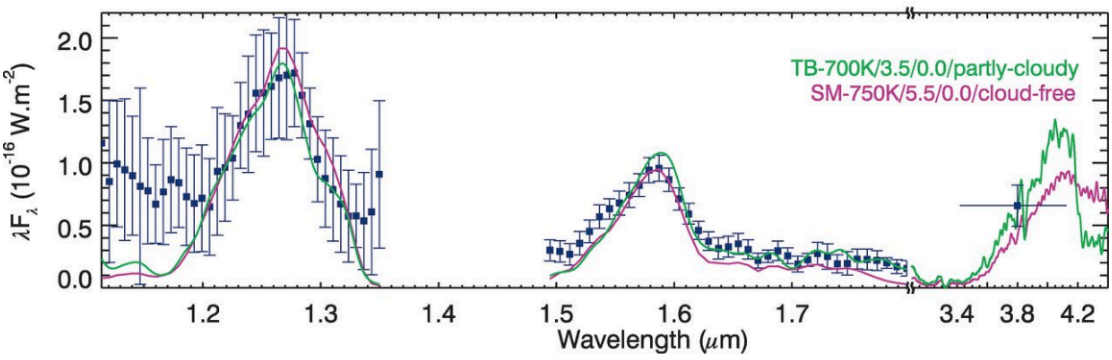


Figure 1.4 Best fit spectral templates of the directly imaged exoplanet 51 Eridani b, from Macintosh et al. (2015), showing greater methane absorption in the red (i.e., more absorption in H compared to J), characteristic of a T-type spectral class. The J and H spectroscopy ($R \approx 45$) and Lp photometry (all blue points) are fit using both a cloudless model (purple) and hybrid partial cloud model (green).

computed using a grid-based approach as in Madhusudhan et al. (2011), minimizing χ^2 on a $\log(g)$ vs. $T_{\rm eff}$ grid, for which evolutionary tracks can also constrain age, luminosity, radius (R_p), and M_p. Figure 1.4 shows the best two model fits from Macintosh et al. (2015) to the J and H spectrum + Lp photometry, using:

1. cloud-free evolutionary tracks similar to those in Saumon & Marley (2008, Figure 4). By fitting the first two parameters (e.g., $\log(g)$ and $T_{\rm eff}$), the four others can then be determined through evolutionary track constraints, which, e.g., sets a fixed relationship between M_p and R_p.

2. partly cloudy evolutionary tracks similar to the "hybrid" model in Saumon & Marley (2008, §4.2), which changes the amount of cloud sedimentation as a function of $T_{\rm eff}$. First, M_p is determined to be $\sim$2 M_J using independent luminosity vs. age evolutionary tracks by measuring the planet luminosity and assuming that the planet and star are the same age, as in Figure 1.2. In this case they assume a hot start model as in Saumon & Marley 2008. Cold start models are less precise in this context, constraining the mass only between 2 and 12 M_J. The age for 51 Eridani is determined by radial velocity, proper motion, and distance association with the β Pictoris moving group (Zuckerman et al., 2001), a system of $\sim$20 stars co-moving with the star β Pictoris, known through stellar evolution modelling to be $\sim$20 Myr old. Knowing M_p and age a priori, a similar fitting procedure as in model 1 is carried out using hybrid model evolutionary tracks, but instead without fixing a M_p-R_p relationship, to

obtain T_{eff}, R_p, and $\log(g)$.

Parameter results for each model are shown in Macintosh et al. (2015, Table 2). Model 2 assumes the age is the same as the β Pictoris moving group, resulting in more realistic values for age and M_p than model 1, although both models are consistent with results from similar models in other systems (e.g., Madhusudhan et al., 2011). Results such as these are the first step towards a better understanding of planet formation and wide orbit massive exoplanets in a statistical context, a goal which first requires that we simply detect and characterize exoplanets.

1.3 Observational Methods

The many different exoplanet detection methods encompass a vast range of physical parameter space, as illustrated in Figure 1.5. Data from these techniques are indeed beginning to fill in the statistical distributions of exoplanets, including those like Earth, although every observing technique is biased. Below, we briefly outline the

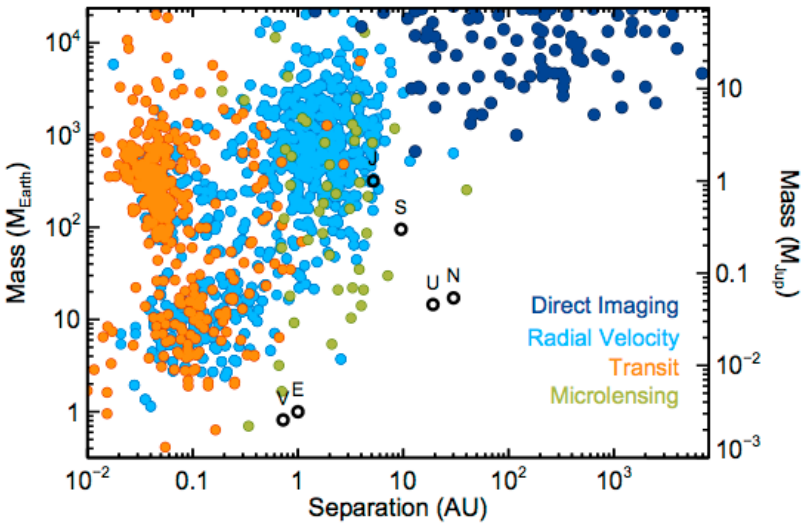


Figure 1.5 The distribution of exoplanet mass vs. separation for different observational techniques, as of April 2016, obtained from Bowler (2016) and references therein, including the exoplanet.eu database (Schneider et al., 2011). Some Solar System planets are shown as black open circles for reference.

basics of each exoplanet detection method listed in Figure 1.5.

1.3.1 Radial Velocity

Because a planet-star system orbits about its barycentre, stellar reflex motion causes the star to exhibit periodic radial velocity variations as a function of time. The shape and primary frequency of the radial velocity curve, measured spectroscopically, allows for a measurement of $M_p \sin(i)$, where i is the planet orbital inclination. However, i cannot be disentangled from M_p, and this technique thus only provides an upper limit on M_p (Perryman, 2011).

Radial velocity detections are biased towards massive planets with short orbital periods closer to edge-on inclinations. A more massive planet produces a more detectable radial velocity amplitude, K ($K \propto M_p$), and the duration of observations should be at least on the order of the detectable orbital period. Radial velocity observations are also biased against planets around younger, hotter, more massive A-type stars due to the comparative lack of spectral features; older, cooler, less-massive F and G type stars have more absorption lines, enabling a more precise radial velocity determination.

1.3.2 Transit

When a planetary system is aligned nearly edge-on with respect to the Earth, a planet eclipsing its host star can produce a periodic dimming effect, similar to the periodic radial velocity signal. If the stellar mass, $M_\star$, and radius, $R_\star$, are known a priori (e.g., from stellar modelling), the period, transit depth, and transit duration can be used to measure M_p and R_p, thus providing a direct estimate of bulk planet density (Perryman, 2011).

A confirmed transit detection typically requires follow-up observations and detection through radial velocities in order to rule out eclipsing binaries. As with radial velocity detections, transit detections are similarly biased to short period, large R_p planets near edge-on inclination; the transit depth scales as $(R_p/R_\star)^2$, and for a circular orbit the probability of a transit alignment scales as $0.005\,(R_\star/R_\odot)\,(a/(1\text{ AU}))^{-1}$ (Perryman, 2011).

1.3.3 Gravitational Microlensing

A background star aligned with a foreground planetary system can produce a temporal gravitational lensing effect. This lensing geometry creates multiple images of the

background source star if aligned with the planetary system to within the Einstein radius ($\sim$1 mas for a 1 $M_\odot$ lens near the Galactic centre; Perryman 2011), which remains unresolvable with any current or future planned telescopes. Thus, as the background star moves in and out of alignment relative to the foreground planetary system, these lensed images contribute to an unresolved magnification effect over time, ultimately allowing the determination of M_p, $M_\star$, and a.

Although these microlensing events last on a timescale of a few days to weeks, depending on the relative source-lens proper motion, they can only ever be detected once. Thus, planet detections and parameters are instead determined from unambiguous model fits to extremely high SNR data; e.g., some events can produce magnification as large as $\sim$3000 (Dong et al., 2006). However, the probability of detecting such a microlensing event around a given background field star is $\lesssim 10^{-8}$, requiring simultaneous monitoring of $\gtrsim 10^8$ stars in order to detect any one given event (Perryman, 2011). Additionally, there may be further uncertainty in measuring $M_\star$ and M_p if light from the planetary system's host star cannot be detected after the microlensing event (Perryman, 2011), also presenting a barrier to confirming a detection with other techniques.

1.3.4 Direct Imaging

Topics within the direct imaging technique will compose the remainder of this thesis. In §1.4 I outline the challenges, tools, and history of the field, including adaptive optics and coronagraphy (§1.4.1), point spread function (PSF) subtraction (§1.4.2), and the past, current, and future instruments and campaigns (§1.4.3).

Direct imaging is biased to young, self-luminous exoplanets (§1.2.1, §1.4) on wide orbits ($\gtrsim$10 AU, §1.4), but is mostly independent of i and has the advantage of providing detection and spectroscopy at a single epoch for the uniquely-occupied parameter space in Figure 1.5.

1.4 Observational Methods: Direct Imaging of Exoplanets

Another name for this observational technique is "high contrast imaging." As the name suggests, direct imaging of exoplanets requires obtaining a contrast sensitivity below the planet-to-host-star flux ratio. In reflected light, a Jovian mass planet to star

contrast scales with a as $\sim 2\times10^{-9}(a/5\text{ AU})^{-2}$ (Graham et al., 2007), and detecting a terrestrial planet in a Sun-Earth analogue of the Solar System would require a contrast of $\sim 10^{-10}$ (Perryman, 2011). In contrast (pun intended), detecting a planet's own thermal emission is independent of separation but dependent on planet age; younger, hotter planets are more self-luminous (§1.2) and therefore require a smaller contrast.

Ground-based high contrast imaging instruments can reach the best contrast using adaptive optics (AO) and coronagraphy in the NIR (§1.4.1). Imaging self-luminous, wide-orbit, young Jovian planets represents the "tip of the iceberg" for detections with direct imaging; observing at 1.6 μm, a 3 M_J planet with age 10 Myr, or a 7 M_J planet with age 100 Myr, around a G2V star requires a contrast of 4×10^{-6} (Graham et al., 2007).

1.4.1 Extreme Adaptive Optics

In a vacuum, an infinitely small point source imaged with a telescope has a finite angular size. This effect is called "diffraction," and occurs when light incoming at the telescope's primary mirror (called the telescope pupil, or "pupil plane") is focused onto a science camera (which lies in the "focal plane"). More specifically, in Fourier optics, this Fraunhofer far field approximation states that at an infinite distance (a valid approximation for astronomical objects beyond the Solar System), the wavefront (a complex-valued quantity representing amplitude, A, and phase, ϕ, of the electric field from electromagnetic radiation) in the focal plane is the Fourier transform of the wavefront in the pupil plane (Steck, 2015). Because the electric field from a point source at the telescope pupil is $Ae^{i\phi}$ (a scalar plane wave solution to Maxwell's wave equation), and on a camera we observe the energy of the electric field (i.e., $|\text{wavefront}|^2$), the image on a science camera for an infinite point source, called the point spread function (PSF), is related to the pupil plane wavefront via

$$\text{PSF} = |FT\{Ae^{i\phi}\}|^2, \tag{1.1}$$

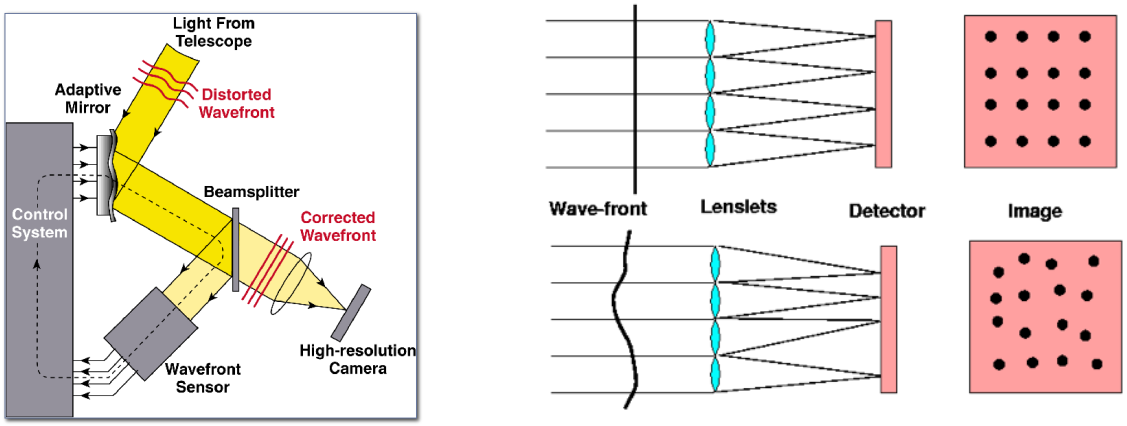
where $FT\{\}$ is the Fourier transform operator. For a telescope with a circular primary mirror of diameter D, equation 1.1 produces a specific Bessel function called an Airy disk. The Airy disk contains a central core with a finite angular size, for which the full width at half maximum (FWHM) is $\sim\lambda/D$ in radians, surrounded by concentric rings, known as Airy rings. This effect, known as "diffraction," causes:

1. high spatial frequency ringing in the PSF, removing light away from the central $\sim\lambda/D$ core, and

2. an infinite point source to appear in images with a finite size, often referred to as the "convolution of a point source with a circular aperture," ultimately limiting exoplanet imaging detections to planet-star separations that are $\gtrsim \lambda/D$. For example, in the NIR ($\lambda = 1.65$ μm, H band) with a 10 metre telescope ($\lambda/D = 34$ mas), a planet-star separation of 1 AU can only be resolved for stars closer than ~ 30 pc.

Without adaptive optics (AO, below) on the ground or going to space, the central core is much greater than λ/D, making exoplanet imaging increasingly difficult. Then, with a diffraction limited image, a coronagraph and/or apodization (both further below) can further improve the contrast below the Airy rings.

With ground-based telescopes, however, atmospheric turbulence (e.g., seeing$\sim$500 mas, Andersen 2014) limits a 10 metre class telescope to $\sim$15 times worse than the diffraction limit, "blurring" our view of the sky. AO is used to correct for this effect, as schematically outlined in Figure 1.6a. A simple AO system contains the following:



(a) (b)

Figure 1.6 (a) A schematic outline of an AO system, from Max (2016), including a DM, WFS, RTC, and SC. (b) An illustration of how a SHWFS can measure atmospheric phase aberration, from Tokovinin (2005).

wavefront sensor (WFS) - an optical element used to measure incoming wavefront aberrations, the simplest of which is a Shack-Hartmann WFS (SHWFS), illustrated in Figure 1.6b. In this case, an array of lenslets samples the pupil plane

wavefront such that the SHWFS detector shows a grid of micro-PSFs that lie at the centre of their subapertures for a flat, unaberrated wavefront. For an aberrated wavefront, the x and y offsets of each micro-PSF from its subaperture centre, called the "slopes," provide the amount of phase aberration[2], or "slope of the wavefront", that can be corrected to obtain a flat wavefront using a deformable mirror (next point).

deformable mirror (DM) - a mirror that can change shape on the kHz frequencies needed to correct for atmospheric turbulence, the simplest of which is controlled by piezoelectric actuators that can typically push or pull as much as $\sim$10 μm (Max, 2016). AO systems typically work in the NIR compared to the optical. At a set DM actuator density, the fit to the aberrated wavefront relative to the wavelength of light is better at longer wavelengths, thus allowing for a lower residual noise floor in the NIR vs. optical after each AO correction. Thus, an optical AO system needs to be run at faster kHz speeds than in the NIR in order to compensate for this effect with more iterations per coherent timescale of atmospheric turbulence, typically changing every few tens of milliseconds (Andersen, 2014).

real time control system (RTC) - a control algorithm that translates slope information from the WFS into actuator commands on the DM at kHz speeds. The iterative, communicative process between the WFS and DM via the RTC is referred to as "closing the loop" because effects from the DM correction are immediately seen by the WFS, thus confirming if the DM correction creates a flat wavefront.

science camera (SC) - a science imaging camera, which without AO would be limited in resolution by seeing, but instead with AO is limited by diffraction.

The performance of an AO system is usually characterized by a number between 0 and 1 called the Strehl ratio (SR), which is the peak value of the obtained PSF core divided by the peak value of a flux-normalized, purely diffraction-limited core. In H band, facility-class AO systems usually perform around $0.3 \lesssim \mathrm{SR} \lesssim 0.5$ on dimmer stars, whereas on brighter stars exoplanet imaging systems typically produce $\mathrm{SR} \gtrsim 0.8$ (Andersen, 2014).

[2]The SHWFS is typically not used to correct amplitude aberrations, although in principle this can also be measured from the micro-PSF intensities.

The contrast from a close-to-diffraction-limited PSF is still not sufficient to image exoplanets. Stellar noise from the PSF core and Airy rings limit contrast at a few λ/D to $\sim 10^{-3}$. In order to reach deeper contrasts in a close-to-diffraction-limited image and prevent the bright central star from quickly saturating, we must block out the central stellar core and suppress the Airy rings, both of which can be done with a coronagraph. A simple Lyot coronagraph, illustrated in Figure 1.7, can improve contrast to $\sim 10^{-6}$. The two main components of a Lyot coronagraph are:

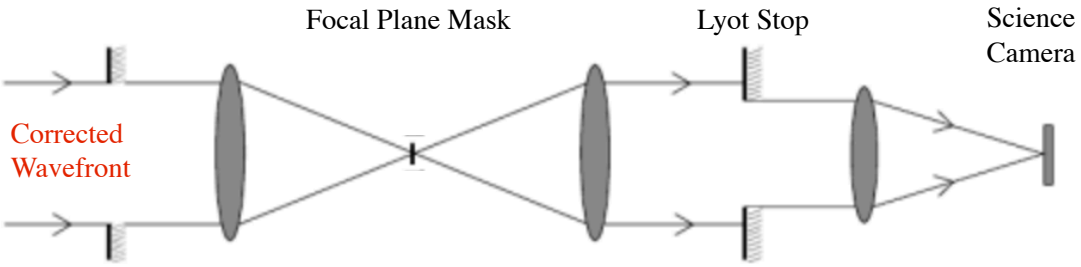


Figure 1.7 An Illustration of the Lyot coronagraph optical design, adapted from Perryman (2011).

focal plane mask (FPM) - an amplitude mask placed in the focal plane to block light from the ∼few λ/D PSF core. However, this alone will not suppress the Airy rings, which thus requires a Lyot stop (next point).

Lyot stop - a washer-shaped amplitude mask placed in the pupil plane. Because of Fourier optics, the pupil plane can be thought of as a map of spatial frequencies of the focal plane, with the origin at the image centre (i.e., on-axis), lower spatial frequencies near the centre, and higher spatial frequencies away from the centre. Thus, a washer-shaped amplitude mask in the pupil plane acts as a low pass filter of the focal plane, greatly suppressing the high spatial frequency Airy rings in the final SC image.

Finally, light from an off-axis planet will miss the FPM and Lyot stop, greatly increasing the achievable contrast.

Either in addition to or instead of using a coronagraph, Airy rings can also be suppressed with apodization—a remapping of the entrance pupil through relay optics in order to create a smooth, symmetric amplitude profile between the edge and centre of the re-imaged telescope aperture. This smooth gradient effect greatly removes the ringing from diffraction (i.e., Airy rings), normally due to the hard-edged binary pupil

mask. Instead, e.g., a Gaussian apodization in the pupil plane Fourier transforms to a Gaussian PSF in the image plane, significantly removing ringing effects compared to an Airy disk. Apodization can be implemented in either amplitude, by smoothly changing the transmissivity (between 0% and 100%) in a re-imaged telescope pupil as a function of separation from the central optical axis (e.g., Soummer et al., 2006), or in phase, by using aspherical optics to redistribute the light more towards the centre of the pupil (e.g., Guyon, 2003). Although the former is more commonly in use on most current exoplanet imaging instruments due to convenience of manufacturing and robustness against pointing errors, the latter is an interesting new technique, mostly still in laboratory testing and development, that limits the throughput effects of amplitude apodization.

The combination of AO, coronagraphy, and apodization is known as extreme adaptive optics (ExAO) and in addition to the final SC or integral field spectrograph (IFS; e.g., Maire et al. 2014) comprises the main design requirements for a high contrast imaging instrument. In space, there is also a need for ExAO, but AO is beyond the capability of any current space telescope imaging systems (but see §1.4.3). Although there is no longer a turbulent atmosphere, and thus no longer a need for ∼kHz AO, the main limiting source of reaching a deeper contrast on either ground-based or space-based telescopes is from optical diffraction effects of instrumental aberrations, called "quasi-static speckles" (see 1.4.2), which change on timescales of minutes to hours (Marois et al., 2008a). For space telescopes, one such example is telescope breathing (Chapter 2), which causes increased defocus throughout an observing sequence during a single orbit, limiting achievable contrast to $\sim 10^{-6}$ (Schneider et al., 2014; Gerard et al., 2016a). This limitation could be corrected for using a DM to reach deeper contrasts.

Another advantage of a space telescope with ExAO capabilities, which is also applicable to future ground-based high contrast imagers, is the use of focal plane wavefront sensing (e.g. Bordé & Traub, 2006; Baudoz et al., 2006; Give'On et al., 2007). This is designed specifically for exoplanet imaging as opposed to general AO, where the goal is to cause the stellar light to destructively interfere at a specific location in the focal plane where a planet could be found; the planet light is not removed because it is incoherent with the stellar light. This method improves contrast compared to conventional ExAO by using the SC as a WFS (called "focal plane wavefront sensing"), which removes differential aberrations between SC and WFS path, called non common path aberration (NCPA). Further chromatic NCPA effects

arise when using a dichroic (instead of a beam splitter as in Figure 1.6a) to separate the WFS and SC path. This alternative is commonly used for NIR AO systems to optimize throughput on the SC, whereas focal plane wavefront sensing is not affected by chromaticity or throughput loss. Focal plane wavefront sensing also allows for simultaneous correction of phase and amplitude aberrations, an improvement, e.g., over the standard SHWFS, which typically only corrects for phase aberration (see the "WFS" bullet point and footnote 2 above). In a vacuum, the best laboratory demonstration of wavefront control thus far has reached a contrast of 5×10^{-10} at $2 - 4\ \lambda/D$ separations (Kern et al., 2013). Wavefront control algorithms are still in a development phase, taking too long to converge for practical use on ground-based telescopes and is ultimately geared towards future space-based ExAO. See Chapter 4 for a specific application of wavefront control to the self coherent camera (SCC) technique (Baudoz et al., 2006).

1.4.2 Point Spread Function Subtraction

After obtaining a diffraction-limited, coronagraphic image, additional post-processing can improve the contrast even further. The main limitations to improving contrast in this regime are "quasi-static speckles"—sources of instrumental noise that vary on the timescale of minutes to hours (Soummer & Aime, 2004; Marois et al., 2008a). These quasi-static aberrations can originate from polishing errors in the instrument optics, although a comprehensive list of such sources for a given instrument is usually unknown. One example could be polishing errors in the SC path that can contribute static NCPA, not seen by the WFS. However, thermal fluctuations of the instrument on ∼minutes to hours timescales can cause quasi-static NCPA. Such thermal variations are apparent, e.g., on the Hubble Space Telescope (*HST*) throughout a full orbit, where the temperature difference between the day side and night side creates images that become increasingly out of focus with time as the optics expand and contract, respectively, also known as "breathing" (see Chapter 2). Because these quasi-static speckles may not change over an ∼hour observing sequence, averaging the images in an observing sequence and/or integrating longer than ∼30 seconds (after which a coronagraphic image becomes quasi-static speckle noise dominated) will not improve contrast. The best approaches used thus far to further suppress the quasi-static speckle noise of a post-coronagraphic image are least-squares-based PSF subtraction (Lafrenière et al., 2007a) or principal component analysis-based PSF

subtraction (Soummer et al., 2012) combined with angular differential imaging (ADI; Marois et al., 2006b) and/or simultaneous spectral differential imaging (SSDI; Racine et al., 1999; Marois et al., 2000; Sparks & Ford, 2002), all of which are discussed below. The goal in these techniques is to minimize the residual noise in a post-coronagraphic image with a least-squares fit and then decorrelate the quasi-static speckles such that they can be further attenuated over a large number of images across an observing sequence.

The main two PSF subtraction algorithms are

- locally optimized combination of images (LOCI; Lafrenière et al., 2007a) algorithm, a least-squares-based approach, and
- Karhunen-Loève Image Projection (KLIP; Soummer et al., 2012), a principal component analysis-based approach.

Both are designed to reconstruct the target image (creating a "least-squares target image") using a set of reference images, or "references" for short, selected from the target sequence[3] such that the noise in the final PSF subtracted image (i.e., the square of the difference between the target image and the reconstructed target image) is minimized. The least-squares image is then subtracted from the target image to obtain a PSF subtracted image. In both LOCI and KLIP, the references are selected by a user-defined "aggressiveness," which requires that any reference image at a given radial separation (i.e., considering an annulus of the target image to subtract) must be rotated and magnified above a minimum ADI-mode field of view (FOV) rotation and SSDI-mode radial magnification (see below), typically $\sim 0.5\lambda/D$ (also see the introduction in Chapter 3 for a further explanation of how this algorithmically implemented). The purpose of this reference selection procedure, called the "selection criteria," is to remove any images where the planet signal significantly overlaps with the target image, thus minimizing its self-subtraction. However, LOCI and KLIP also differ in how they reconstruct the target image and correct for algorithm throughput losses. In KLIP, K selected references creates a $K \times K$ covariance matrix, for which the K eigenvalues and eigenvectors create a basis onto which the target image is projected to create a least-squares target image. The number of eigenvectors and eigenvalues

[3]The "target sequence" refers to every available image during a sequence in ADI observing mode (see below), in time for a single broadband SC or in both time and wavelength for an IFS or multi-band SC, whereas a "target image" is a single image in the target sequence whose on-axis PSF, but not potential planet, we want to subtract (ultimately iterating over all images in the target sequence).

used from the available basis, called "KL modes," sets the planet SNR in the final PSF subtracted image (i.e., using more KL modes lowers the noise but also lowers the planet signal; Soummer et al. 2012). In LOCI, the same covariance matrix is used to generate least-squares subtraction coefficients that are then multiplied by the references and summed to create the least-squares target image. These subtraction coefficients are obtained from a separate region of the image (called the "optimization region") than the desired region to subtract (called the "subtraction region") in order to subtract the noise but not the planet signal (Lafrenière et al., 2007a; Marois et al., 2010a). This LOCI formalism using optimization and subtraction region geometry is illustrated in Figure 1.8, which also illustrates the same procedure applied to KLIP but using only a subtraction region.

A simple schematic of ADI is outlined in Figure 1.9. In practice there are additional more robust steps not included in Figure 1.9b. In the explanation below, I use the terminology of the image vectors A and C and least-squares target image B from Figure 1.9b. If the images A_i represent the target sequence for a set of images obtained using ADI, and the the desired target image to subtract the PSF is, e.g., A_0, the least squares-target image, B, is constructed using LOCI or KLIP (i.e., instead of a median) with a selection criteria that does not include A_0 in the set of references. This process is then repeated to subtract each target image, A_i, with a least-squares, ultimately creating the image vector C with each corresponding PSF-subtracted target image, C_i, after which the de-rotation and median combine procedure is the same as in Figure 1.9b.

Diffraction is chromatic, and so the radial separation of a speckle will change with λ/D. In SSDI, before running a least-squares, compensating for this effect by magnifying images to align speckles allows for additional noise suppression, complimentary to ADI (although before the most recent generation of ExAO instruments that include an IFS, discussed below in §1.4.3, SSDI was typically not as effective and still non-Gaussian compared to ADI; Marois et al. 2008a, 2014). Using an IFS, a reduced datacube (multiple images as a function of wavelength across the IFS broadband spectral window) shows that a speckle radial separation will linearly increase with wavelength, scaled by $\sim \lambda/D$. Thus, the general SSDI procedure is as follows: images similar to A in Figure 1.9b, but across a range of wavelengths instead of times, are radially magnified to align speckles, then subtracted using LOCI or KLIP (where here an assumption of the planet spectrum is necessary; Marois et al. 2014), then demagnified to the original wavelength, and then combined into a single image using a

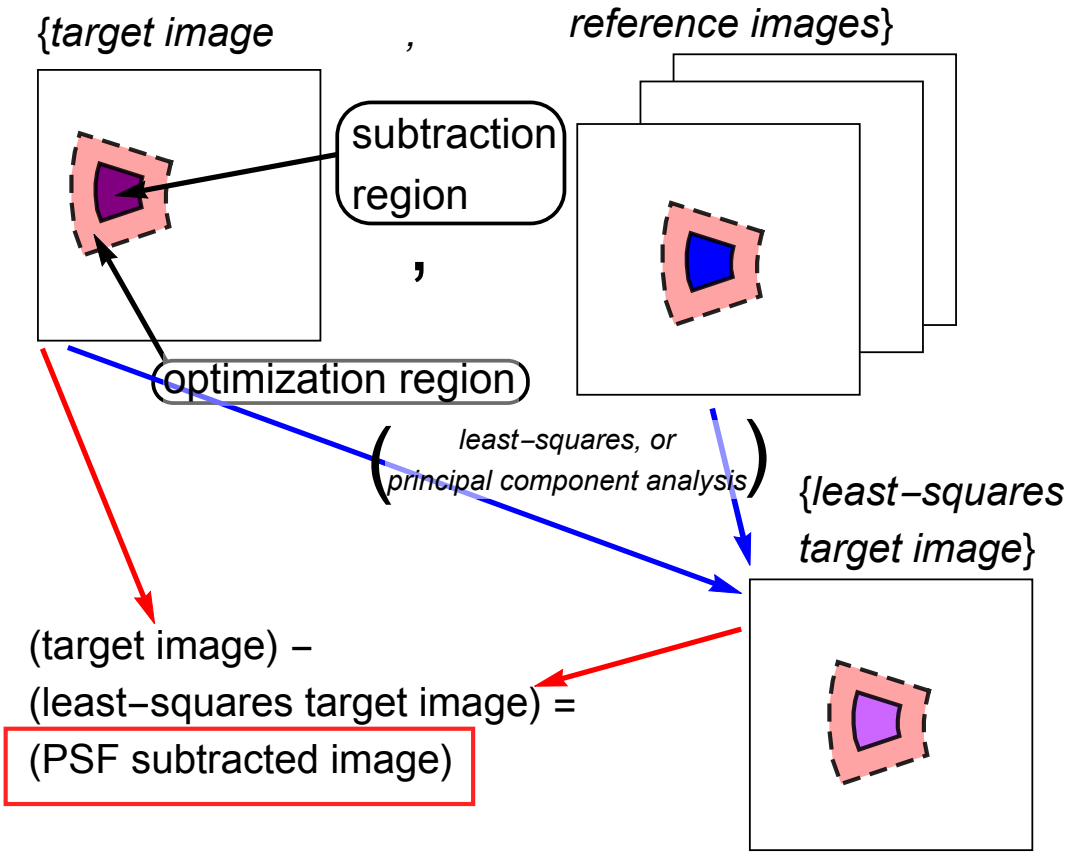
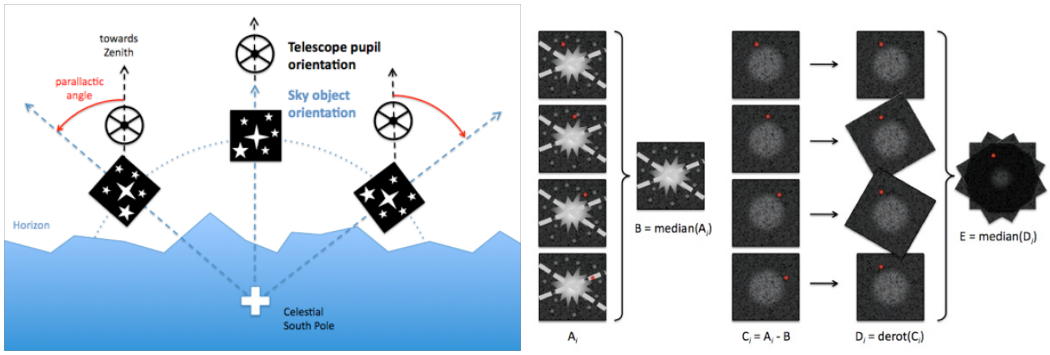


Figure 1.8 A schematic of the LOCI (Lafrenière et al., 2007a) and KLIP (Soummer et al., 2012) PSF subtraction algorithms, showing the subtraction region and, for LOCI, the surrounding optimization region. In LOCI, the subtraction coefficients are determined using the optimization regions in the target image and reference images and then applied to the subtraction region in the set of references to create a least-squares target image. In KLIP, the least-squares target image is instead created by selecting the number of KL modes to include in the projection of the target image onto the KL basis of the reference image covariance matrix. The least-squares target image is then subtracted from the target image to produce a final PSF-subtracted image.

weighted average of the assumed planet spectrum. Analogous to the final de-rotation step in ADI, the final demagnification step in SSDI is necessary to re-align the potential planet in the PSF subtracted target image—whose position does not change with λ/D (i.e., in comparison to speckles, the radial separation of an astrophysical source is not affected by diffraction)—before collapsing the full target sequence into



(a) (b)

Figure 1.9 (a) A schematic diagram of the observing mode with ADI (Marois et al., 2006a), where the azimuthal rotator is turned off while tracking an object, such that the pupil position remains fixed during an observing sequence, but any astrophysical object will rotate around the on-axis PSF. For example, on an altitude-azimuth telescope at the Cassegrain focus, this is done by tracking an object as the Earth rotates, but without rotating the instrument in order to keep the instrument and telescope optics aligned. (b) A simple illustration of ADI PSF subtraction, allowing quasi-static speckle suppression after the sequence of images are rotated north up (decorrelating the quasi-static speckles in each image) and then median combined (Soummer & Aime, 2004; Marois et al., 2008a). Both images are obtained from Vyacheslav (2014).

a single image.

1.4.3 Past, Present, and Future Instruments and Surveys

In this section I will give a brief chronological overview of exoplanet imaging instruments and surveys, including the past (early and first generation surveys), present (second generation surveys), and future. A thorough review of early, first, second generation surveys can be found in Bowler (2016), and so unless explicitly stated, all information below on this topic is from Bowler (2016) and references therein. I will also present highlights of narrow-orbit ($\lesssim$ 70 AU, of interest to Solar System and radial velocity studies) exoplanet detections by direct imaging thus far (Bowler, 2016, Table 1, "Directly Imaged Planets and Planet Candidates with Inferred Masses $\lesssim 13\ M_{\rm Jup}$").

The first early surveys before ~2005 were not designed specifically for high contrast imaging. On *HST*, a coronagraph allowed for starlight suppression, but no AO correction system was/is available onboard, thus observations are limited by un-

correctable quasi-static speckle noise (e.g., see Chapter 2). PSF subtraction was performed using roll subtraction, where in a second image the telescope is physically rotated to use as a reference, ultimately influencing the invention of ground-based ADI. Ground-based facility class AO imaging systems that included coronagraphy were still unoptimized in design for the high SR, high contrast regime. No exoplanets were directly imaged from these surveys.

From ∼2005-2012, the first generation of surveys with dedicated high contrast imaging instruments and/or advanced speckle suppression techniques began, including optimized Lyot coronagraphy, adaptive secondary mirrors, multi-band SSDI observing/PSF subtraction, ADI observing/PSF subtraction, and LOCI PSF subtraction. These era produced the first wave of directly imaged exoplanets, including:

HR 8799 b,c,d, and e (Marois et al., 2008b, 2010b),

This was the first directly imaged planetary system, and is the only imaged multi-planetary system thus far, with four ∼5-7 M_J planets at separations between ∼14 and 68 AU. This detection came from the International Deep Planet Survey (IDPS; PI: C. Marois; Galachier et al., submitted), a survey of ∼300 young stars, using multi-band SDI and ADI observing with optimized instruments at the VLT, Keck, Gemini-South, and Gemini-North between 2009 and 2012.

β Pictoris b (Lagrange et al., 2009, 2010), and

Imaging of a warped debris disk (see §2.1) around the star β Pictoris by Smith & Terrile (1984) suggested that this system may contain an unresolved inner planet (Lecavelier Des Etangs et al., 1995). This young (∼23 Myr) massive (∼13 M_J) planet was first detected with the VLT in 2003 on the northeast side of the star (Lagrange et al., 2009), and then detected again in 2009 on the southwest side of the star (Lagrange et al., 2010), confirming detection via orbital motion. In both instances Lagrange et al. use a reference star PSF subtraction scheme that produced similar results compared to ADI.

HD 95086 b (Rameau et al., 2013c,b).

This detection produced the lowest mass (∼4-5 M_J) exoplanet from first generation surveys. Observations were made using the VLT Survey of Young Nearby Dusty Stars (Rameau et al., 2013a), observing with ADI, and targeting stars with ages $\lesssim$200 Myr and distances $\lesssim$65 pc.

A number of additional first generation surveys used the VLT, MMT, Gemini-South, Gemini-North, Subaru, and Keck Observatories, none of which produced detections of additional near orbit exoplanets (see Bowler 2016 and references therein), suggesting a low frequency of massive planets beyond ∼10 AU (see below).

With the second generation high contrast imaging surveys (2012-present) came the first generation of ExAO instruments, which include:

Project 1640 (P1640) (PI: R. Oppenheimer) Using the 3000 actuator AO system on the Polomar Observatory 5.1 meter Hale telescope, combined with an optimized coronagraph and first ever IFS in a high contrast imager, P1640 began operations in 2012 and since 2013 has been undergoing a three year survey of nearby massive stars.

LBTI (PI: A. Skemer) Using the twin 8.4m LBT mirrors, high order WFSing, deformable secondary mirrors, ADI, and particularly L′ band (3-4 μm) imaging allow for an ongoing ∼70 night survey of stars $\lesssim$1 Gyr-old. The IFS was recently installed and will soon be available for use.

Gemini Planet Imager (GPI) (PI: B. Macintosh) With polarimetry capabilities and a high-order, multiple DM ExAO system on the 8.2 m Gemini-South telescope, the ongoing 890 hour, ∼600 star GPI Exoplanet Survey (GPIES) detected 51 Eridani b (Macintosh et al., 2015, see §1.2.2), the lowest mass (∼2 M_J) directly imaged exoplanet to date.

Spectro-Polarimetric High-contrast Exoplanet Research (SPHERE) (PI: J.-L. Beuzit) includes an ongoing 200 night survey of young stars, similar to GPIES but on an enormous ExAO instrument with dual band imaging, long-slit spectroscopy, and optical imaging and polarimetry.

Additional instruments not yet operating but currently upgrading to ExAO capabilities include Magellan AO (MagAO; PI: L. Close) on the 6.5 m Clay telescope and Subaru Coronagraphic ExAO (SCExAO; PI: O. Guyon) on the 8.2 m Subaru telescope.

Other than 51 Eridani b, no new narrow-orbit (< 100 AU) exoplanet detections have thus far come from second generations surveys. In general, the theme from early, first, and second generation surveys thus far is that Jovian exoplanets on wide orbits

are rare; the meta-analysis of direct imaging surveys thus far from Bowler (2016, §4.5, Table 3) suggests that $\lesssim$1 % of all stars harbour a 5-13 M_J planet beyond 10 AU.[4]

In the future, the field of exoplanet imaging will split into space-based and ground-based ExAO. The 6.5m James Webb Space Telescope *JWST* will carry four different coronagraphs on board, adapted to use across a broad $10 - 20$ μm wavelength range (Boccaletti et al., 2015). One coronagraph will be a Lyot coronagraph, while the other three are designed using the four quadrant phase mask (FQPM; Rouan et al., 2000), which replaces the FPM from the classical Lyot design with a π phase shift in two of the four quadrants of the focal plane, causing on-axis destructive interference with an inner working angle (IWA) down to 1 λ/D and significantly less throughput loss. Although *JWST* will not have a DM, increased stability compared to *HST* in design and being in a trailing orbit should remove breathing effects and significantly improve achievable space-based contrast, despite a factor of $\sim$10 larger IWA than current ground based ExAO instruments. Additionally, the 2.4 m Wide Field Infrared Survey Telescope (WFIRST) was recently approved as a NASA mission (launch date: mid 2020s) and will be the first space telescope with a DM, Lyot coronagraph, and WFS, aiming for contrasts greater than 10^{-9} (after post-processing) in the optical at separations greater than 200 mas (Spergel et al., 2015). Seventy-six target stars that are already known from radial velocity studies could be accessible through direct imaging and spectroscopic charactization at $R = 70$ with WFIRST, although it is still unlikely that an Earth-sized planet at 1 AU could be imaged (Spergel et al., 2015).

On the ground, first generation instruments for 30 m-class telescopes are not yet funded to include any ExAO capabilities, and so exoplanet imaging in this era may instead require use of the facility class AO instruments (e.g., Herriot et al., 2014) to do high contrast imaging. Performance in this regime may be comparable to the performance of the current second generation ExAO instruments (e.g., Marois et al., 2012). Planning and design for dedicated second generation 30 m-class ExAO instruments has only recently begun, although early science studies suggest that more than 20 known radial velocity planets, and potentially $\sim$10 rocky planets within 15 pc, could be imaged at $3 - 10$ μm (Quanz et al., 2015).

[4]However, this frequency is notably higher for B and A stars: $\sim$8 % for 5 $M_J < M <$ 13 M_J and 10 AU $< a <$ 100 AU (Bowler, 2016).

1.5 Agenda

In Chapter 2, I present my work to directly image the HR 8799 debris disk, a circumstellar disk in this multi-planetary system. I develop a new PSF subtraction algorithm to recover this diffuse residual emission.

In Chapter 3, I develop a new PSF subtraction pipeline for the Gemini Planet Imager, designed to obtain higher sensitivity for narrow-orbit planets in the inner 100 to 300 mas ($\sim 2-7\ \lambda/D$). This technique is based on the use of an archival PSF library using campaign data from the Gemini Planet Imager Exoplanet Survey.

In Chapter 4, I present simulations and a laboratory experiment for a new high contrast imaging technique to directly image wide orbit exoplanets, called Super Nyquist Wavefront Control.

In Chapter 5, I summarize my work in this thesis and discuss future work.

Chapter 2

Searching for the HR 8799 Debris Disk with the *Hubble Space Telescope*

The following text was adapted from Gerard et al. (2016a).

2.1 Background

Over a timescale of ∼10 Myrs, gas and dust from as far out as ∼1 pc are first accreted onto a forming star from the surrounding environment to form a "protostellar accretion disk." The dust grains then begin to coagulate to form planetesimals, rocky planets, and Jovian planet cores (i.e., if formed through the core accretion scenario) in a "protoplanetary accretion disk." Finally, radiation and stellar wind from the newborn star causes the gas and dust to either be "blown out" of the system or fall into the star in a "transition disk," ultimately revealing a residual system of planets and planetesimals (Broekhoven-Fiene, 2012; Draper, 2014).

Then, on a time scale of $\lesssim$ a few Gyrs, the remaining large rocky planetesimals (with diameters $\gtrsim$1000 km) form a "debris disk"—the result of collisions between these residual bodies that are then ground down into smaller and smaller grains, ultimately creating a grain size distribution ranging from submm dust particles to 1000 km bodies (Broekhoven-Fiene, 2012). This size distribution is typically modelled as a power law: $n(D_{\rm par}) \propto D_{\rm par}^{-q}$, where $D_{\rm par}$ is the grain size and q is a variable

positive constant. In an ideal case where all collisions are self-similar[1] and there is no maximum and minimum grain size, $q = 3.5$ (Dohnanyi, 1969). This case is referred to as a steady state "collisional cascade" and is often assumed for grain sizes that are above the blow out limit (e.g., Su et al., 2009). The blow out limit is the particle size at which stellar radiation overcomes the force of gravity, pushing smaller grains onto more eccentric orbits until they become unbound. In an idealized case for large, spherical particles, the ratio of radiative force to gravitational force is $\beta \equiv F_{\rm rad}/F_{\rm grav} \propto L_\star/(D_{\rm par}\ M_\star)$, where $L_\star$ and $M_\star$ are the star mass and luminosity, respectively (Wyatt, 2009).

Debris disk detections can be spatially resolved from reflected stellar light (e.g., Smith & Terrile, 1984; Draper et al., 2016) or thermal emission (e.g., Matthews et al., 2014), or instead unresolved but detected through an excess submm spectral energy distribution (SED; e.g., Sadakane & Nishida 1986). Optical and NIR scattered light observations (i.e., $\lesssim$ a few microns) of debris disks are usually sensitive to grains below the blow out limit (which is typically $\gtrsim$ a few microns). These small unbound grains are ejected on $\sim$orbital timescales, which means that they are constantly being replenished by collisions from larger bodies (Broekhoven-Fiene, 2012). The (bound) debris disk dust grains, typically below $\sim$1 mm in diameter, will absorb stellar radiation at $\lambda \lesssim 1$ mm and then reemit thermal radiation at the temperature of the disk. These temperatures, typically less than a few hundred Kelvin, are well below the stellar effective temperature, and therefore an unresolved debris disk can still be detected as a submm "bump" on the Raleigh Jean's tail of the stellar SED, showing a second unique temperature component from the debris disk (and in some cases a third component when there are two "bumps;" see below).

Of particular interest to high contrast imaging is the possible exoplanet–debris disk connection; the planetary systems of the four know directly imaged narrow orbit exoplanets—HR 8799, β Pictoris, HD 95086, and 51 Eridani—all have debris disks. Resolved debris disk morphology studies, which often observe warps, clumps, and asymmetries, could predict the presence of an unseen planet through dynamical planet/disk interactions (e.g., Lecavelier Des Etangs et al., 1995). Observing gaps in a debris disk at longer wavelengths could also infer the presence of planets clearing their orbits (e.g., ALMA Partnership et al., 2015). Furthermore, HR 8799, HD 95086,

[1]That is, any collision between one impactor planetesimal with diameter $D_{\rm im}$ and a target planetesimal with diameter $D_{\rm tar}$ is the same for a given ratio of $D_{\rm im}/D_{\rm tar}$, independent of any individual particle size.

and 51 Eridani all have debris disks with multiple SED temperature components, one hotter component interior to the planet(s) and one colder component exterior to the planet(s), perhaps analogous to the Solar System's asteroid belt, Jovian planet, and Kuiper belt architecture (Bowler, 2016). For this reason, and since debris disks are typically found in young systems, many recent direct imaging surveys have targeted already known debris disks in search of planets (e.g., Rameau et al., 2013a; Macintosh et al., 2014). However, the fraction of planet detections in these surveys is still considerably low, and as of yet no study has shown a statistical correlation between the two. It may be interesting in the future for high contrast imaging surveys to specifically target known two temperature component debris disks, since the recent work by Kennedy & Wyatt (2014) suggests that these systems generally also possess two spatial components, perhaps analogous to planetary systems such as the Solar System.

2.2 Introduction

The era of direct imaging of extrasolar planets is upon us. High contrast images of HR 8799 have revealed the presence of four planets orbiting their host star (Marois et al., 2008b, 2010b), and more recent integral field spectrographs have provided spectra of their atmospheres (Bowler et al., 2010; Barman et al., 2011; Ingraham et al., 2014). In contrast to radial velocity or transit exoplanet detection methods, these giant planets at large separations trace a range of parameter space often closer to our own Solar System, and so understanding the physical and chemical properties of this and other similar systems is crucial to better understanding the process of Solar System and planet formation.

One key step in better understanding the planet formation process is the formation and stability of protoplanetary disks and their remnant debris disks. The HR 8799 debris disk was first measured by Sadakane & Nishida (1986) from an unresolved SED infrared excess at 60 μm with the Infrared Astronomical Telescope (*IRAS*) point source catalogue. The debris disk was later slightly spatially resolved by Su et al. (2009) (hereafter S09) also with an SED infrared excess. S09 used the *Spitzer Space Telescope* at 24, 70, and 160 μm to measure one component inside the known orbiting planets (the planets are between 15 and 68 AU; Marois et al., 2008b, 2010b) at $\sim 6-15$ AU—the inner disk—and another two components outside the planets at ∼90-300 AU—the planetesimal belt—and ∼300-1000 AU—the halo. Matthews et al.

(2014) have also spatially resolved the HR 8799 planetesimal belt and halo at 70, 100, and 160 μm and marginally resolved at 250, 350, and 500 μm with the *Herschel Space Telescope*, also measuring a similar three component debris disk from image and surface brightness profile modelling. Matthews et al. (2014) also measure a disk inclination of $26 \pm 3^{\circ}$, close to face-on orientation.

However, none of the HR 8799 debris disk components have been imaged in optical scattered light due to the close-to-face-on nature of the disk. ADI PSF subtraction (Marois et al., 2006a), a standard optical/near infrared high contrast imaging technique used to construct a reference stellar/instrumental PSF from a sequence of science images, is not optimized for close-to-face-on planetary disks and causes self-subtraction effects. Additionally, a low spatial frequency close-to-face-on disk is hard to distinguish from focus and pointing errors, which at minimum need to be kept stable at the contrast needed for a detection. High contrast close-to-face-on disk imaging with ground based telescopes is generally avoided due to both limited stability from atmospheric turbulence, even with adaptive optics, and self-subtraction effects. Observing a separate reference star of matching spectral type in the same sequence as the target star using ground based telescopes is also generally avoided for close-to-face-on disks due to limited stability. Imaging close-to-face-on disks from space is still difficult but more feasible in both highly correlated stability and by observing a separate star as the reference PSF to prevent self-subtraction effects.More recent robust least-squares-based PSF subtraction analysis using *HST* data (e.g., Soummer et al., 2012; Galicher et al., 2013; Currie et al., 2012; Choquet et al., 2014; Soummer et al., 2014; Choquet et al., 2016) have made significant improvements over classical PSF subtraction techniques (e.g., see §2.6.1).

The optical scattering properties of small grains in the HR 8799 debris disk (i.e., the contrast needed for a *HST* detection) is unknown. The *HST* Space Telescope Imaging Spectrograph (STIS) imaging CCD can be used with a Lyot coronagraph to reach contrasts similar to ground based extreme adaptive optics instruments ($\sim 10^{-6}$, e.g., Schneider et al., 2014). The relatively wide 52×52 arcsecond STIS field of view provides an optimal opportunity to image the HR 8799 planetesimal belt in optical wavelengths.

In this paper we present a new PSF subtraction algorithm, called Optimized Spatially Filtered (OSFi) normalization, that was developed to image the HR 8799 planetesimal belt with STIS. In §2.3 we describe the STIS data. In §2.4 we describe our new OSFi normalization algorithm. In §2.5 we present the results of OSFi normaliza-

tion applied to the HR 8799 STIS data. In §2.6 we analyze a number of false-positive detection heuristics based on our OSFi algorithm results, ultimately indicating that we are seeing a non-detection. In §2.7 we present an upper limit analysis to constrain the mass contained in the sub-micron grain planetesimal belt, and in §4.5 we conclude and summarize our results. Throughout this paper we use the following measurements for HR 8799: distance $d_\star = 39.9$ pc, radius $R_\star = 1.34 R_\odot$, effective temperature $T_\star = 7250$ K, stellar luminosity $L_\star = 4.9\ L_\odot$, and stellar mass $M_\star = 1.5\ M_\odot$ (Sadakane, 2006; Gray et al., 2006; Su et al., 2009; Matthews et al., 2014). The HR 8799 spectral type varies in the literature between A5V and F0V with characteristics of an anomalous λ Boo star (Gray et al., 2006). The STIS plate scale is 0.05078 arcseconds per pixel (Hernandez & et al., 2012).

2.3 STIS HR 8799 Data

We obtained public archival STIS data of HR 8799 (Clampin, 2010) from the Mukilski Archive for Space Telescopes (MAST). STIS coronographic observations of HR 8799 are present from three epochs: 2011 November 2 (hereafter epoch 1), 2011 November 12, and 2012 October 4 (hereafter epoch 2). Epochs 1 and 2 each contain three 2300s exposures of HR 8799, one exposure per *HST* orbit. The three successive images in epoch 1 and three successive images in epoch 2 are hereafter referred to as science images 0, 1, 2, 3, 4, and 5, respectively. One 2300s exposure of HIP 117990, a separate reference star of similar color to HR 8799, is present in the final orbit of epoch 1 and 2 (hereafter referred to as reference images 1 and 2, respectively). Data from the 2011 November 12 epoch contained nine 120 s and three 240 s science exposures and one 2226 s reference exposure and is therefore unusable due to the different noise floors between reference and science images. All of the HR 8799 images within a given epoch are taken at different telescope roll angles. We use the dark-subtracted, flat-fielded, cosmic ray-rejected `.crj` images (Hernandez & et al., 2012).

2.4 OSFi normalization: A New PSF subtraction algorithm

2.4.1 Image Registration

Before we can implement PSF subtraction between a science and reference image (hereafter called reference-subtraction) or two different science images (hereafter called self-subtraction), all images must be registered to a common center with sub-pixel accuracy. To do this, after normalizing the flux in each image to 1/(exposure time), we place an aperture mask around the four diffraction "spiders" in the regions unaffected by the coronagraph FPM and set the remainder of the image to zero. We then run upsampled cross-correlation on the masked spider image with the same image rotated 180 degrees about its center using an image registration Python package[2] to determine the image center to sub-pixel accuracy and then shift the unmasked images to a common center. We used cubic spline interpolation for sub-pixel shifting to avoid ringing effects from a Fourier-based shifting algorithm.

2.4.2 OSFi PSF subtraction

A classical LOCI PSF subtraction (Lafrenière et al., 2007a) is not optimal to minimize the noise by constructing a reference image from the science images for a dataset with only two images per epoch[3]. Additionally, as mentioned in §2.2, constructing the reference image from the science images with ADI will cause self-subtraction of the close-to-face-on HR 8799 debris disk.

We also considered the option of using reference star differential imaging (RSDI) through the Archival Legacy Investigation of Circumstellar Environments (ALICE) pipeline (Choquet et al., 2014), where a large archive of HST reference images is used in a principal component analysis-based least-squares calculation (Soummer et al., 2012). Although RSDI removes the risk of self-subtraction for disks, we concluded that this approach is not optimal for HR 8799. Because we are looking for a close-to-face-on disk with a smooth radial profile (Matthews et al., 2014), using additional reference images from different epochs may have (1) been observed at a slightly different relative focus from the HR 8799 observations, and (2) a different stellar spectral

[2]https://github.com/keflavich/image_registration

[3]Science images 0 and 3 are unusable due to decreased pointing stability (§2.5).

type compared to HR 8779, which we will see in §2.6 can cause a similar radial profile that can be confused with a disk.

Instead, normalizing the reference PSF subtraction (i.e., using the single separately acquired reference image in the same sequence as HR 8799) to the high spatial frequency noise can be thought of as "minimizing" the noise contaminating the expected low spatial frequency, close-to-face-on disk. This is, in essence, a least-squares using a locally optimized region for face-on disks that is designed to work on small datasets. This is exactly the rationale behind an Optimized Spatially Filtered (OSFi) PSF subtraction: we subtract the high spatial frequency noise from the science image, leaving the residual low spatial frequency astrophysical disk. This is done by normalizing the subtraction of a separately acquired reference image ($\mathrm{im_{ref}}$) from the science image ($\mathrm{im_{sci}}$) to the high spatial frequency noise in each PSF:

$$\mathrm{im_{OSFi}} = \mathrm{im_{sci}} - \mathrm{im_{ref}} \left(\frac{\sigma \left[\mathcal{F}_H \left(\mathrm{im_{sci}} \right) \right]}{\sigma \left[\mathcal{F}_H \left(\mathrm{im_{ref}} \right) \right]} \right), \tag{2.1}$$

where σ is the robust standard deviation operator (Dowell et al., 2012a) in a user-defined optimization region sampling the PSF, $\mathcal{F}_H$ is the image high-pass filter operator, and $\mathrm{im_{OSFi}}$ is the residual PSF-subtracted image. Although there are many methods of high-pass filtering in image processing (e.g., Fourier filtering, convolution kernels, etc.), our initial approach was to first create a low-pass filtered image, $\mathcal{F}_L(\mathrm{im})$, using a basic 11 pixel-wide median boxcar filter and then create the high-pass image, $\mathcal{F}_H(\mathrm{im})$, via:

$$\mathcal{F}_H(\mathrm{im}) = \mathrm{im} - \mathcal{F}_L(\mathrm{im}) \tag{2.2}$$

In §2.6 we further discuss the dependency of different high-pass filtering techniques on our OSFi algorithm throughput. We found that an 11 pixel standard deviation Gaussian convolution kernel used to construct $\mathcal{F}_L(\mathrm{im})$, and by equation 2.2 $\mathcal{F}_H(\mathrm{im})$, has the best throughput performance in equation 2.1, and so we use this technique for the rest of our analysis (see §2.6.1.1). We chose our optimization region as an annulus between 3.5 arcseconds (140 AU; the STIS region outside which there are no saturated pixels leaking from behind the FPM, the effective inner working angle) and 7.5 arcseconds (300 AU) in agreement with the measured ∼90-300 AU planetesimal belt in S09. To create the optimization region, the image is first median filtered using a 3 by 3 pixel boxcar to remove individual bad pixels. Masks are then also placed over the spiders, coronagraph wedge, and any additional remaining bad pixels/regions

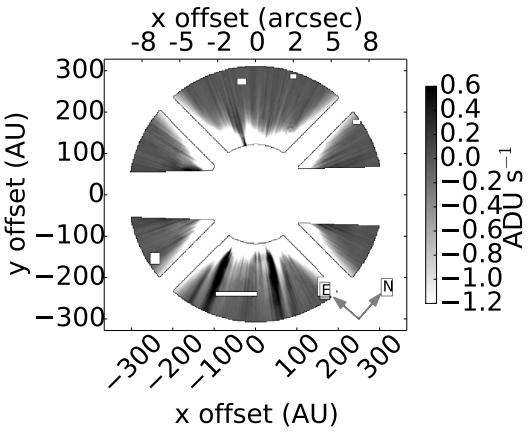


Figure 2.1 An example of a typical OSFi optimization region (here showing science image 1) after both masking the coronagraph, diffraction spiders, and bad pixels and high-pass filtering the image. The coordinate grid shows x and y offset from the registered image center in both arcseconds and AU.

before high-pass filtering with an 11 pixel standard deviation Gaussian kernel. An example high spatial frequency optimization region is shown in Figure 2.1.

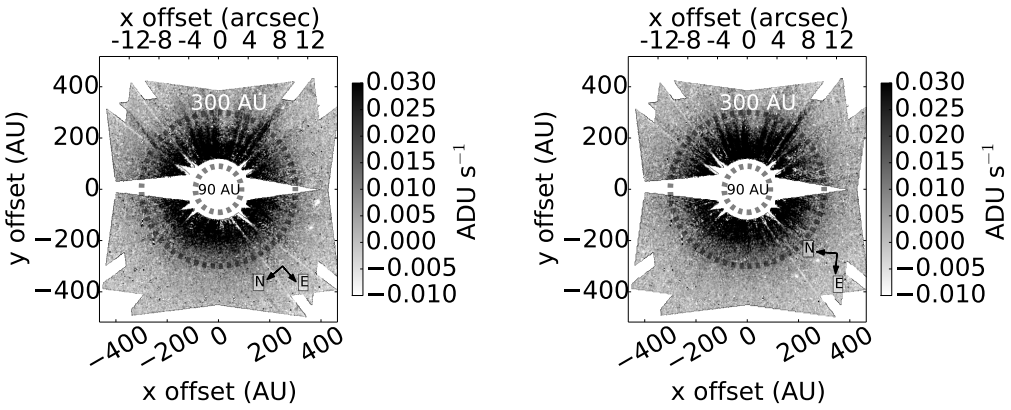
Finally, multiple PSF-subtracted images taken at different telescope roll angles within a single epoch can be combined to additionally reduce the residual quasi-static high spatial frequency noise by rotating each PSF-subtracted image to north up and median combining all de-rotated images. This does not cause ADI self-subtraction as discussed in §2.2 because the reference image is a separate, stellar PSF with no disk.

Self-subtraction using the OSFi algorithm should yield a featureless noise map because there is a close-to-face-on disk in both images. This is essentially a measure of contrast if both science images have the same diffuse close-to-face-on halo features that are independent of telescope roll angle and the low spatial frequencies of the two PSFs are normalized. As we will see in §2.5, the latter is not the case if the observing sequence experiences a focus evolution due to telescope breathing.

On a similar note, we further emphasize that our OSFi algorithm is designed for small, correlated datasets. OSFi RSDI would not work well to image a diffuse face-on halo because we are only subtracting the high spatial frequencies of the PSF, assuming that the low spatial frequencies are normalized to the same coefficient. This is not a fair assumption if the observing sequence experiences a focus evolution (§2.5,2.6) and/or a difference between the reference and target star spectral type (§2.6). These are low spatial frequency effects that are not seen by the OSFi algorithm due to high

pass filtering, and so including a larger diversity of datasets with a different spectral type and focus evolution effects would ultimately limit OSFi sensitivity to detect a face-on, diffuse disk rather than improve it.

2.5 Results



(a) (b)

Figure 2.2 Reference-subtracted OSFi-normalized images from 2011, November 2 (left; epoch 1), and 2012, October 4 (right; epoch 2), using the separately acquired reference images of HIP 117990 during the same observing sequence to subtract from the HR 8799 images. The observed residual over-luminosity is consistent between both epochs. All following PSF-subtracted images are derotated to the same matching pupil orientation for comparison and show the measured S09 90–300 AU planetesimal belt region outlined by dashed circles.

OSFi reference-subtraction and self-subtraction for both epochs are shown in Figures 2.2 and 2.3, respectively. The images have been cropped to the central ∼30 arcsecond region of interest, derotated to the match pupil orientation in each respective epoch, and show the measured S09 90–300 AU planetesimal belt region. Reference-subtraction in shows a diffuse, close-to-face-on halo, whereas self-subtraction shows hardly any residual flux. The radial profiles for both epochs of self- and reference-subtraction are shown in Figure 2.4 as the solid lines, illustrating that reference-subtraction is consistent between both epochs and confirming that self-subtraction yields no residual.

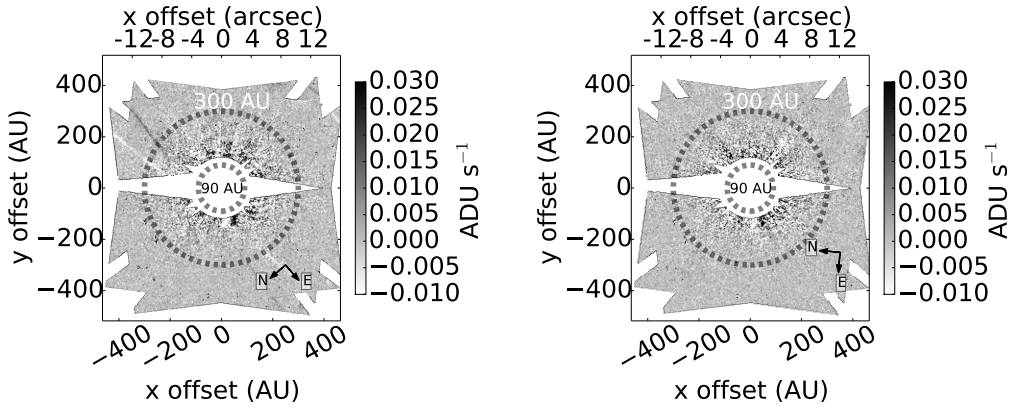
These results use only science images 1, 2, 4, and 5 from epochs 1 and 2. We found

in self-subtraction that use of science image 0 or 3, corresponding to the first orbit in each sequence of images, shows a noisier residual than self-subtraction without either of these images, illustrated in Figure 2.5. We may be seeing here that the first orbit in an *HST* observing sequence is has lower pointing stability compared to subsequent orbits.

One potential problem with OSFi PSF subtraction is that it is optimized to find *any* low spatial frequency residual. High spatial frequency normalization means that this residual could be entirely from any unnormalized low spatial frequencies in the science and/or reference PSF (i.e., from a difference in spectral type and/or focus between the two science images; see §2.6.3). The absence of any low spatial frequency residual in Figure 2.3 shows that there is no detected instability or focus evolution effects between the two consecutive science images in either epoch.

2.6 Detection Heuristics

We discuss a number of possible false-positive origins for the over-luminosity from OSFi reference-subtraction.



(a) (b)

Figure 2.3 The epoch 1 (left) and epoch 2 (right) self-subtracted, OSFi-normalized images. Here, the reference image is a science image at the same epoch but different roll angle, and so the displayed featureless residuals from OSFi PSF subtraction show that the same low spatial frequencies in both science images are independent of telescope roll angle (i.e., there is no detected focus evolution).

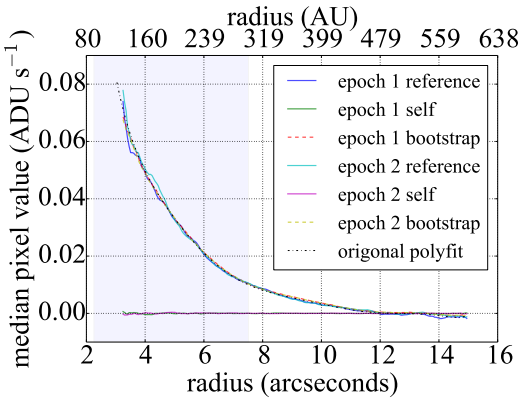


Figure 2.4 Radial profiles for OSFi reference- and self-subtracted images (§2.5), bootstrapped images (§2.6.1), and a fitted polynomial profile to reference-subtracted images (§2.6.1). The reference subtracted images use HIP 117990 as the reference star, whereas the self-subtracted images use HR 8799 at a different telescope roll angle as the reference star. The bootstrapped images use HR 8799 at a different telescope roll angle as a reference image, but the raw HR 8799 target image includes an added polynomial radial profile to measure our OSFi algorithm throughput. The shaded 90–300 AU region represents the measured S09 planetesimal belt.

2.6.1 Algorithm Effects

Bootstrapping

We define bootstrapping as a form of PSF subtraction to measure algorithm throughput. This is done by adding a power law to a registered science image, $\mathrm{im}_{\mathrm{sci}}$, and then running OSFi PSF subtraction on that "bootstrapped" image using the other regular science image in the same epoch as the reference image, $\mathrm{im}_{\mathrm{ref}}$. The residual should return the same PSF-subtracted radial profile as the original input power law as long as the added polynomial profile does not change the normalization coefficient, $\left(\frac{\sigma[\mathcal{F}_H(\mathrm{im}_{\mathrm{sci}})]}{\sigma[\mathcal{F}_H(\mathrm{im}_{\mathrm{ref}})]}\right)$. If the residual is not the same as the input, this would indicate a non-unity algorithm throughput, resulting in either over- or under-subtraction. If this effect is severe, it could amplify the noise in a non-detection to look like a detection.

We chose to input the fitted polynomial profile from mean radial profile of the reference-subtracted images, but any other profile would serve the same purpose. We found that using a Gaussian 11 pixel standard deviation convolution kernel for $\mathcal{F}_L$ in equation 2.2 best recovers the added input radial profile after self-subtraction. No other filters we tried returned a similar unity throughput, including 5, 11, 21, and 31 pixel median kernels and 5, 21, and 31 pixel standard deviation Gaussian kernels.

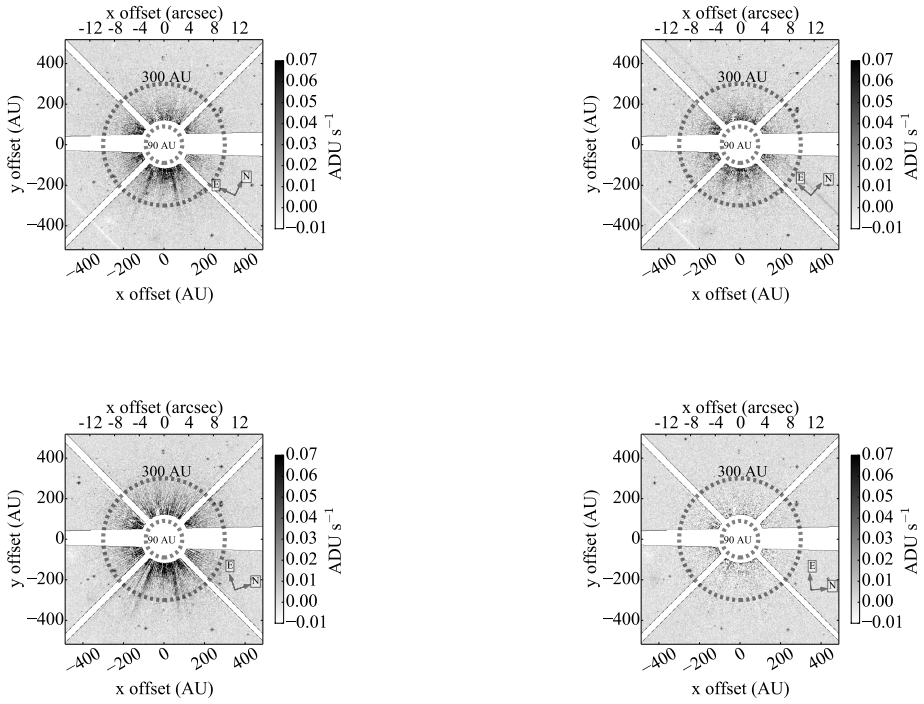
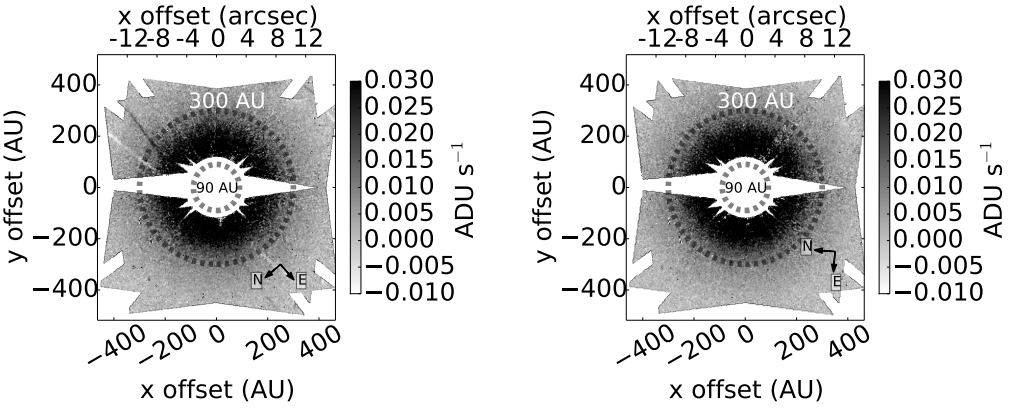


Figure 2.5 OSFi self-subtraction with (left column) and without (right column) the first image of each observing sequence. The first row is from epoch 1, the second is from epoch 2. Comparing the left column to the right column shows that the first image of each observing sequence contains additional low spatial frequencies, and so we exclude science 0 and science 3 from our set of reference images.

The two PSF-subtracted, bootstrapped images, are shown in Figure 2.6,

showing a similar residual halo component as with OSFI reference-subtraction (Figure 2.2). Figure 2.4 shows the OSFi bootstrapped radial profiles as the dashed lines, output from the input polynomial dashed-dotted line, illustrating that a Gaussian 11 pixel standard deviation convolution kernel yields close to unity OSFi algorithm throughput. We also confirmed that bootstrapping (1/3)×(original polynomial profile) similarly recovers the expected output radial profile (not shown in Figure 2.4).

The agreement between bootstrapping and reference-subtraction indicates that the recovered over-luminosity is not amplified or over-subtracted by our OSFi algorithm.



(a) (b)

Figure 2.6 Epoch 1 (left) and epoch 2 (right) bootstrapped OSFi PSF-subtracted images.

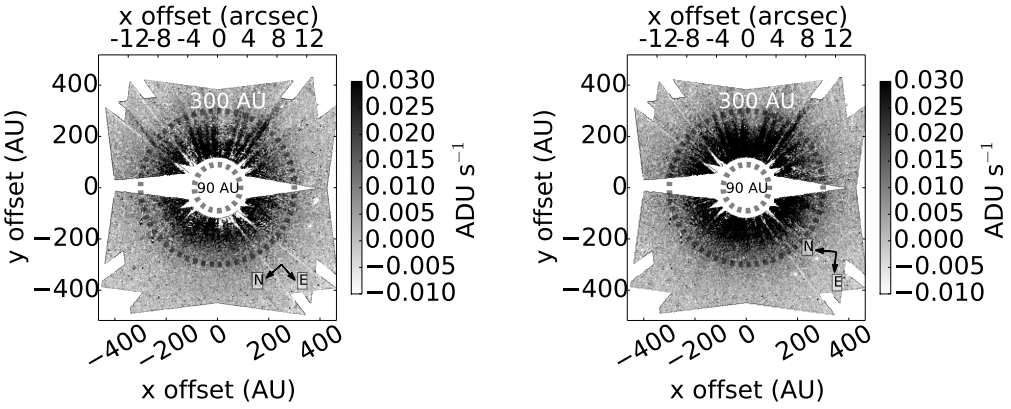
OSFi vs. spider-normalized PSF Subtraction

The classical iterative *HST* PSF subtraction technique has been to remove the diffraction spiders by eye (e.g., Schneider et al., 2014). This is similar to normalizing reference-subtraction to the cumulative flux ratio between the science and reference image in an aperture mask around the spiders, similar to the spider aperture mask used in our image registration pipeline (§2.4.1). In this case, the coefficient in equation 2.1 changes from $\left(\frac{\sigma[\mathcal{F}_H(\mathrm{im_{sci}})]}{\sigma[\mathcal{F}_H(\mathrm{im_{ref}})]}\right)$ to $\left(\frac{\Sigma(\mathrm{im_{sci}})}{\Sigma(\mathrm{im_{ref}})}\right)$, where Σ refers to the cumulative flux within the spider aperture mask.

Reference-subtracted images using spider-normalized PSF subtraction are shown Figure 2.7. Their radial profiles are shown in Figure 2.8 along with spider-normalized self-subtraction and bootstrapping results.

First, we see a general inconsistency between OSFi and spider-normalized reference-subtraction. These differences are most likely due to how the PSF subtraction algorithms determine the subtraction coefficient (e.g., different amounts of high spatial frequency leakage into the optimization region or spider apertures).

Second, in contrast to OSFi reference-subtraction, spider-normalized reference-subtraction shows an inconsistency between epochs 1 and 2. This discrepancy may be from the fact that the spiders are close to Nyquist sampled ($\sim$3 pixels across the FWHM), occasionally causing the brightest photon to fall undetected in between two pixels which may add a bias in calculating the normalization coefficient, $\left(\frac{\Sigma(\mathrm{im_{sci}})}{\Sigma(\mathrm{im_{ref}})}\right)$. Furthermore, radial profiles of spider-normalized bootstrapping (Figure 2.8) show a slight amount of over-subtraction. This is likely due to flux from the low spatial frequency disk (and/or instrumental halo) contaminating the spider flux in the bootstrapped image, thus causing a larger subtraction coefficient that results in over-subtraction. Although this bootstrapping over-subtraction is expected for spider normalization of close-to-face-on disks, an over-luminosity due to scattered light off of halo dust grains should in principle be consistent between different epochs and PSF subtraction algorithms after accounting for differences in throughput loss. This is not the case for spider normalization, suggesting that these inconsistencies may simply be from the non-optimal use of spider-normalized PSF subtraction for close-to-face-on disks.



(a) (b)

Figure 2.7 Epoch 1 (left) and epoch 2 (right) reference-subtracted, spider-normalized images, where the reference star is HIP 117990 and the subtraction coefficient is now determined by normalizing reference and target images to the cumulative flux in four apertures around the spiders. This is a separate PSF subtraction method from OSFi normalization.

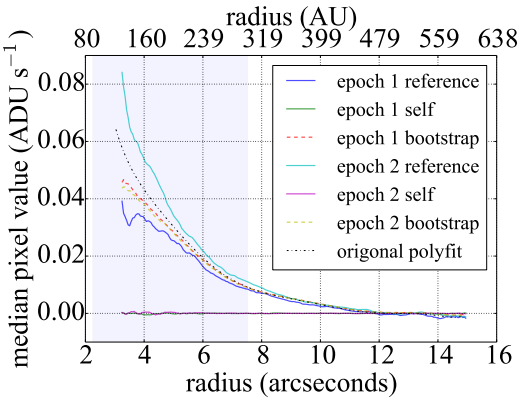


Figure 2.8 Radial profiles of PSF-subtracted HR 8799 images using spider normalization, showing reference-subtraction, self-subtraction, and bootstrapping. The shaded 90–300 AU region represents the measured S09 planetesimal belt.

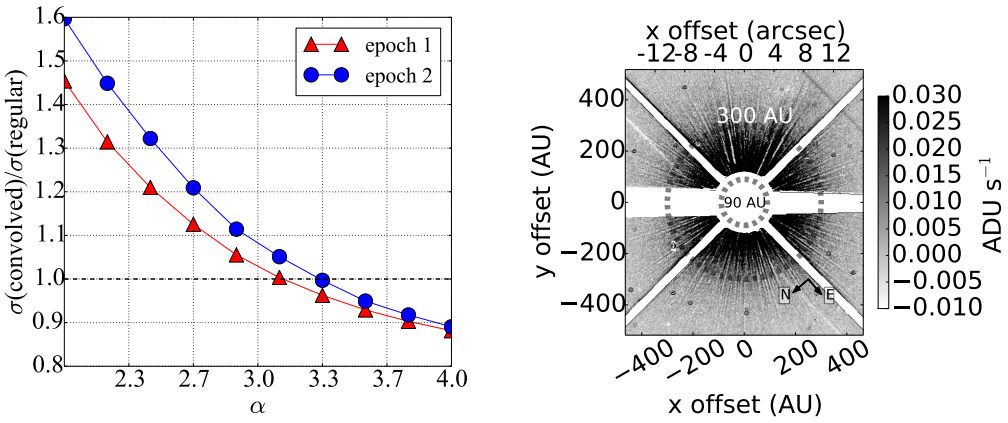
2.6.2 PSF Effects

Inner Disk Convolution

In Figures 2.2 and 2.7, shown at the same pupil orientation, high spatial frequency residual noise in the upper half of each image appears to be matching across both epochs, suggesting that these features are from the PSF. Thus, we consider the possible explanation that we are seeing the effect of the measured ∼6-15 AU inner disk (S09) on the STIS PSF. This spatially unresolved inner disk should affect the PSF via convolution of a ∼6-15 AU annular kernel with the stellar PSF, ultimately causing some amount of high spatial frequency noise after subtracting the "sharper," un-convolved reference PSF from the "smoother," inner disk-convolved science PSF. However, the relative optical brightness of the inner disk to the stellar PSF component is unknown due to the unknown HR 8799 debris disk optical grain scattering properties, and so we set this as a free parameter in the equation below:

$$\begin{aligned} \mathrm{PSF}_{\mathrm{conv}} &= \mathrm{im}_{\mathrm{sci}}(1 - 1/\alpha) + [\mathrm{im}_{\mathrm{sci}} * d_{\mathrm{inner}}]\,(1/\alpha), \\ \mathrm{PSF}_{\alpha} &= \mathrm{PSF}_{\mathrm{conv}} - \mathrm{im}_{\mathrm{sci}'}\left(\frac{\sigma\left[\mathcal{F}_H\left(\mathrm{PSF}_{\mathrm{conv}}\right)\right]}{\sigma\left[\mathcal{F}_H\left(\mathrm{im}_{\mathrm{sci}'}\right)\right]}\right), \end{aligned} \tag{2.3}$$

where α represents the relative star-to-inner disk brightness and is an unknown scalar value to be optimized in the corresponding reference-subtraction, d_{inner} is a 6-15 AU annular convolution kernel (3 to 7 pixels from the image center) whose sum is normalized to one, $\mathrm{im}_{\mathrm{sci}}$ is the input science image to be convolved with the d_{inner} kernel,



(a) (b)

Figure 2.9 Left: The robust standard deviation in $\mathcal{F}_H\,(\mathrm{PSF}_\alpha)$—indicated as σ(convolved)—relative to regular OFSi reference-subtraction—$\mathcal{F}_H\,(\mathrm{im}_{\mathrm{OSFi}})$, indicated as σ(regular)—vs. α. α is the relative star-to-inner disk brightness. We use $\mathrm{im}_{\mathrm{sci}}$ = science image 2 and $\mathrm{im}_{\mathrm{sci}'}$ = science image 1. The black dotted line shows the desired amount of high spatial frequency noise in $\mathcal{F}_H\,(\mathrm{PSF}_\alpha)$ needed to match the noise in $\mathcal{F}_H\,(\mathrm{im}_{\mathrm{OSFi}})$. Right: PSF_α for $\alpha = 3.2$, shown at the same pupil orientation as in Figure 2.2.

$\text{im}_{\text{sci}'}$ is the reference image to be used in self-subtraction (i.e., another science image in the same epoch as im_{sci}), and $*$ is the convolution operator. The two factors of $(1/\alpha)$ and $(1-1/\alpha)$ in equation 2.3 along with the cumulative normalization of d_{inner} are set so that flux is conserved in transforming im_{sci} into PSF_{conv}. We are running self-subtraction to observe only the PSF effects of inner disk convolution, whereas reference-subtraction would produce inner disk convolution effects in addition to the residual already seen in Figure 2.2. To find the optimal value of α, we first measure the standard deviation of PSF_{α} in the same optimization region as in regular reference-subtraction after 11 pixel Gaussian high-pass filtering. We then compare this standard deviation to the standard deviation of the corresponding regular im_{sci} OSFi reference-subtraction in the same high-pass filtered optimization region for a given value of α. Standard deviation of the high-pass filtered PSF-subtracted image serves as a useful metric to measure the high spatial frequency noise coming from inner disk convolution.

Figure 2.9 shows the relative standard deviation between PSF_{α} and regular OSFi reference PSF subtraction plotted vs. α, indicating that the optimal α value between both epochs is $\alpha \sim 3.2$. The right hand panel of Figure 2.9 shows $\text{PSF}_{\alpha=3.2}$. Upon initial inspection, Figure 2.9 shows that high spatial frequency noise is present in both positive and negative values. However, Figure 2.2 only shows the equivalent positive "spikes" and is thus most likely not from the effect of imaging the inner disk.

Pointing Accuracy

It could also be possible that the residual we are seeing is a result of a difference in alignment of the STIS coronagraph wedge (i.e., FPM) with respect to the star center between the science and reference images. The instrumental PSF component from the FPM is at the same position with respect to the image center for every image, but pointing accuracy can cause the star position to move with respect to the FPM. We can test this accuracy by measuring the difference in offsets of the reference and science images from image registration. The width of the FPM at the position covering the star is $\sim$45 pixels, and we find a mean radial offset difference between science and reference images of $\sim$0.07 pixels, or $\sim$3 miliarseconds, and so this small relative pointing difference difference with respect to the FPM size is an unlikely cause of the observed over-luminosity.

2.6.3 Breathing, Spectral Difference

The origin of over-luminosity may be from defocus between the reference and science images. As mentioned in §2.3, the reference star was observed after the science images in epochs 1 and 2. Accordingly, the same pointing difference to slew from the science to the reference star can cause the same amount breathing defocus in both epochs, due to the same difference in incident sunlight angle with respect to *HST* (J. Krist, private communication), which could explain the matching radial profiles between the two epochs.

In addition to focus, a difference in spectral type between the science and reference star can also cause a residual after PSF subtraction that looks like a close-to-face-on disk (Grady et al., 2003). HR 8799 spectral type characterization varies between A5V and F0V with characteristics of an anomalous λ Boo star (Gray et al., 2006), and the reference star, HIP 117990, is an F2IV (Grenier et al., 1999). Despite this difference, the broadband optical colors are closely matching: $(\mathrm{B}-\mathrm{V})_{\mathrm{HIP\ 117990}} = 0.26$, $(\mathrm{B}-\mathrm{V})_{\mathrm{HR\ 8799}} = 0.32$ (Høg et al., 2000), and differing broadband spectral shape would be more deterministic of a residual halo feature than differing anomalous spectral lines.

Tiny Tim Simulations

To further investigate the possibility of a difference in focus and/or spectral type being the cause of the observed over-luminosity after PSF subtraction, we initially ran simulations with Tiny Tim, an *HST* PSF simulator (Krist et al., 2011b). Using the closest matching Pickles (1998) synthetic spectra to the HR 8799 and reference star spectral type and the typical 5.8 μm defocus caused from breathing Grady et al. (2003), our initial results showed that neither spectral difference nor defocus are consistent with the observed over-luminosity. However, further evidence is needed to fully rule out instrumental PSF effects. Ultimately, to rule out the origin from spectral difference and/or coronagraph effects we would need real spectra for both the reference and science star over the full $0.2-1\mu$m STIS imaging bandpass used in a STIS PSF simulator that includes effects from the coronagraph. The STIS Tiny Tim PSF simulator does not include a coronagraph FPM or Lyot stop, which can cause low spatial frequency residuals from misalignment, and for this reason we instead ran OSFi PSF subtraction on another similar dataset, described below.

HD 10647

If breathing defocus and/or spectral difference is the source of over-luminosity, we should see a similar radial profile after OSFi PSF subtraction of a different STIS target that is similar to HR 8799 in color difference and angular separation between the corresponding target and reference star. A similar color difference should create a similar effect over the STIS bandpass, and a similar angular separation should create a similar breathing defocus effect[4].

Thus, we used the Canadian Astronomical Data Center (CADC) *HST* Cache (Stoehr et al., 2009), astroquery (Ginsburg et al., 2013), and MAST to obtain STIS observations of HD 10647, a known edge-on disk (Krist et al., 2011a). The HD 10647 and HR 8799 datasets are closely matching in color difference and angular separation between the target and reference star used in each observing sequence: $\Delta(B-V)_{\text{(HR 8799 - reference)}} = 0.06$ (Høg et al., 2000), $\Delta(B-V)_{\text{(HD 10647 - reference)}} = 0.05$ (Ducati, 2002), $\Delta(\rho)_{\text{(HR 8799 - reference)}} = 11.3°$, and $\Delta(\rho)_{\text{(HD 10647 - reference)}} = 9.1°$, where ρ is the angular separation between the given target and reference star (van Leeuwen, 2007). Since HD 10647 is an edge-on disk, our rationale here is that if we see a radial profile similar to HR 8799 (i.e., resembling a close-to-face-on disk), we know in this case we are seeing instrumental and/or bandpass effects, which would suggest that this is also what we are seeing with HR 8799.

Initially, we chose to use the images with exposure times scaled to match the target (HD 10647) to reference (HD 7570) flux ratio, so that all images had the same cumulative detector flux, as with HR 8799[5]. We then applied the same OSFi pipeline as described above to this sequence, optimizing throughput via bootstrapping with a 21 pixel (standard deviation) Gaussian kernel (see Figure 2.11 below). Each HD 10647 PSF-subtracted image was also flux-normalized based on the HD 10647 to HR 8799 flux ratio, which we determined using the spiders for each target.

An initial run through of the OSFi pipeline on HD 10647 showed that the science images experience a focus evolution throughout the observing sequence. OSFi self-subtracted images on this sequence, using science 3a as a reference, are shown in

[4]The absolute angle of sunlight (i.e., position on the sky) will also matter in determining the amount of breathing defocus in addition to the relative target-reference separation, and so the separation alone should only give us a rough order of magnitude estimate. However, as we will see below, there are likely additional factors that cause breathing effects.

[5]We name the consecutive HD 10647 observations, identified in the MAST dataset as OB1J09010, OB1J09020, OB1J10010, and OB1J10020 as science 0a, 1a, 2a, and 3a, respectively, and use the only scaled reference star observation, OB1J11030.

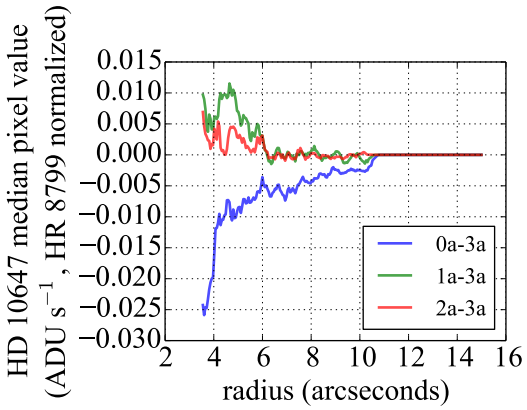
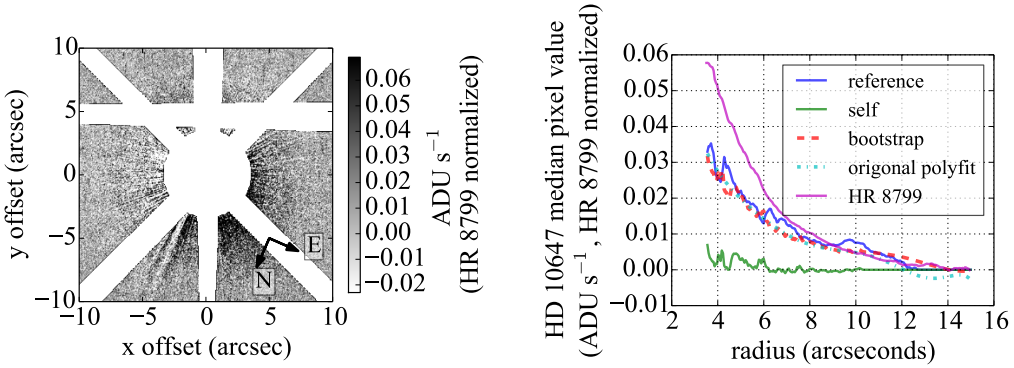


Figure 2.10 OSFi self-subtracted radial profiles of HD 10647 (Krist et al., 2011a), using science 3a as the reference image and science 0a (blue), 1a (green), and 2a (red) as the science images. We can see that 2a-3a has the most stable self-subtraction, indicating that there is a focus evolution throughout the HD 10647 observing sequence.



(a) (b)

Figure 2.11 Our OSFi reference-subtracted image of HD 10647 (left), a known edge-on disk (Krist et al., 2011a), and corresponding radial profiles (right). We do not detect the disk. We found the best match in bootstrapping (input: light blue, output: red) when using a 21 pixel (standard deviation) Gaussian kernel. To prevent effects from focus evolution, HD 10647 reference-subtraction is computed using 3a - reference and self-subtraction using 2a-3a. The similar HD 10647 and HR 8799 radial profiles (dark blue and pink, respectively) suggest that we are not seeing an astrophysical disk in either dataset.

Figure 2.10. We can see that the most stable self subtraction in this sequence is science 2a-3a, indicating that focus is changing throughout the sequence. We also see the same pattern using other images as the reference, further supporting the notion

that two images are more stable with respect to one another if taken consecutively in the observing sequence.

Using this information, we only selected science 3a as the science image (taken just before the reference image) in our OSFi reference-subtraction, discarding science 0a, 1a, and 2a. Accordingly, our OSFi reference-subtracted image and various radial profiles of HD 10647 are shown in Figure 2.11. We do not detect the known disk, instead seeing a similar halo profile as with HR 8799. We note that both self- and reference-subtraction are considerably noisier here than with HR 8799, as expected due to (1) a factor of $\sim$10 smaller in exposure times, (2) using only one target image and therefore removing the speckle suppression effects of ADI, and (3) the possible presence of additional focus evolution between science 3a and the HD 10647 reference star (we saw no such evidence for a similar focus evolution in HR 8799 self-subtraction).

Figure 2.11 shows that the reference-subtracted HD 10647 radial profile is within a factor of $\sim$2 of the HR 8799 reference-subtracted radial profile, suggesting that we are seeing a similar non-astrophysical source of over-luminosity in both targets. Thus, we proceed in modelling the HR 8799 debris disk from optical photometry assuming a non-detection, using the detected over-luminosity as the noise floor of our OSFi post-processing technique.

2.7 Dust Disk Model, Upper Limits

Assuming the observed over-luminosity is a non-detection, in the following argument we present an upper limit analysis on the mass in small (optical) grains, $M_{\rm dust}$, in the HR 8799 debris disk system. We assume the disk's reflecting dust particles are spherical with radius a and cross section πa^2. If so, an infinitesimal shell at distance $D_{\rm shell}$ from the host star with grain density ρ has a mass of

$$M_{\rm shell} \sim \frac{16\pi}{3} \rho \left(D_{\rm shell}\right)^2 a \frac{L_{\rm shell}}{L_\star}, \tag{2.4}$$

where $L_{\rm shell}$ and $L_\star$ are the luminosity of the shell and star, respectively (Jura et al., 1995). We assume $a = 0.6$ μm optical scattering grain radius (central wavelength in the 0.2-1 μm STIS bandpass) and small grain density of $\rho = 2$ g cm^{-3} as in Galicher et al. (2013). We determine $L_\star$ by integrating the Plank blackbody function over the STIS bandpass with $R_\star = 1.34 R_\odot$ (Gray et al., 2006) and $T_\star = 7250$ K (Sadakane, 2006). $L_{\rm shell}$ is the luminosity from a single unresolved shell of width one pixel within

the measured ∼90-300 AU region.

To determine $L_{\rm shell}$ we first recompute the reference-subtracted images in Figures 2.2 and 2.7 using the "PHOTFLAM" header parameter in the initial pipeline-reduced `.crj` files to obtain OSFi and spider reference-subtracted images in physical units of f_λ = erg cm^{-2} sec^{-1} Å^{-1} (Hernandez & et al., 2012). This radial profile is then extrapolated via fifth order polynomial fitting, similar to §2.6.1.1, beyond the 3.5 arcsecond saturated region to 90 AU (2.3 arcseconds) in order to cover the entire measured region of the planetesimal belt. The entire image is finally multiplied by 5 to obtain a "5 σ" upper limit[6].

We then compute the cumulative flux, $f_{\rm shell}$, in a one pixel wide annulus within the dust belt 90-300 AU region by taking the median value in each annulus times the number of pixels in the annulus, removing any additional bad pixels and instrumental high spatial frequencies. The corresponding shell luminosity at the distance of HR 8799, $d_{\rm HR\ 8799}$, is

$$L_{\rm shell} = f_{\rm shell}\ 4\pi\ (d_{\rm HR\ 8799})^2 \tag{2.5}$$

Combining equations 2.4 - 2.5, the dust belt mass is then computed as a discrete sum over one pixel-wide annuli within the S09 planetesimal belt region:

$$M_{\rm dust} \leq \frac{16\pi}{3}\ \frac{\rho}{L_\star}\ a \sum_{\rm shell\ =\ 90\ AU}^{300\ \rm AU} \left[5 L_{\rm shell}\ (D_{\rm shell})^2\right] \tag{2.6}$$

The resulting mass upper limits for OSFi normalization epoch 1, OSFi normalization epoch 2, spider normalization epoch 1, and spider normalization epoch 2 are $1.4 \times 10^{-4}\ M_\oplus$, $1.2 \times 10^{-4}\ M_\oplus$, $1.1 \times 10^{-4}\ M_\oplus$, and $1.7 \times 10^{-4}\ M_\oplus$, respectively, yielding a final upper limit on the mass in small dust grains of

$$M_{\rm dust} \leq 1.7 \times 10^{-4}\ M_\oplus. \tag{2.7}$$

S09 (Table 3 and §4.2) also determine the dust mass within HR 8799's 90-300 AU belt region, but from entirely different methods, using SED fitting from thermal emission at 24, 70, and 160 μm. They find a mass of $M_{\rm dust} \sim 1.9 \times 10^{-2}\ M_\oplus$ in grains between 1 and 10 μm in the 90 – 300 AU planetesimal belt. They assume a standard collisional cascade grain size distribution power law of $n(a) \propto a^{-q}$ with $q = 3.5$ (Dohnanyi,

[6]If the residual in the PSF-subtracted image is a non-detection, this represents a "1 σ" threshold, assuming Gaussian statistics, for which an astrophysical detection in that image should be at least 5 times brighter.

1969). Integrating over this size distribution from the minimum to maximum grain sizes for the infrared (1-10 μm) and optical (0.2-1 μm) observations[7] gives the mass in each size bin

$$m(a) \propto \int_{a_{\min}}^{a_{\max}} n(a)\rho a^3 da \propto a^{0.5}\Big|_{a_{\min}}^{a_{\max}}. \tag{2.8}$$

S09 finds an optical grain dust mass consistent with their observations of 5×10^{-3} $M_\oplus$, almost an order of magnitude larger in mass than the upper limit found above. In other words, these *HST* observations imply far less mass in small grains than one would expect from observations of dust at longer wavelengths and a collisional power-law size distribution.

The blowout limit for HR 8799, using $L_\star = 4.9$ $L_\odot$, $M_\star = 1.5$ $M_\odot$ (Gray et al., 2006) and $\rho = 2$ g cm^{-3} is $a_{bl} \sim 2$ μm. As our observation wavelength of 0.2-1 μm is smaller than a_{bl}, a shallower-than-collisional-cascade power-law for optical grains makes sense (i.e., there should be less mass in grains smaller than a_{bl} compared to what is expected from steady state collisions). Thus, as expected, the optical grains we are sensitive to must be ejected from the HR 8799 system, removed faster than they can be produced in collisions.

In order to estimate the mass-loss rate from the system by ejected small grains, we must estimate a timescale over which these grains are visible. As these grains are on hyperbolic orbits, we will use the orbital timescale of barely bound grains to conservatively estimate the ejection timescale. Barely bound grains will have highly eccentric orbits, with pericenters within the "birth ring" of parent bodies.

Assuming a pericenter at the inner edge of the main planetesimal ring ($\sim$100 AU; S09), and an apocenter at the outermost observed extent of the halo ($\sim$2000 AU; Matthews et al., 2014), the orbit of a barely bound grain will have a period on the order of 10^4 years. Assuming that the observed mass is ejected on this timescale, using equation 2.7 we find that $\lesssim 1.7\times10^{-8}$ $M_\oplus$/year will be ejected from the system.

If this mass ejection rate is held constant over the age of HR 8799 ($\sim$30 Myr; Marois et al., 2010b), only $\lesssim$0.5 $M_\oplus$ should have been ground into dust and blown out of the system. This order-of-magnitude upper limit on the mass is not worryingly large; it is smaller than estimates of parent body masses in most debris disk systems (e.g., Krivov, 2010), including our own solar system (e.g., Tsiganis et al., 2005). Thus, this mass ejection rate alone does not imply a low-probability event such as a recent

[7]Here we are assuming the light that is scattered or re-emitted by dust grains is the same wavelength as the grain size.

catastrophic collision to explain.

A few other stars with debris disks and extended dust haloes have been studied. Due to its favorable edge-on viewing angle, β Pic has been known for many years to host an extremely large dust halo visible in scattered light, extending at least $\sim$1000 AU from the star (e.g. Kalas & Jewitt, 1995). Detailed collisional and radiative simulations have shown that the Vega system's halo is consistent with collisional production (Müller et al., 2010; Sibthorpe et al., 2010), and Fomalhaut is known to have an extended halo of very small grains (Espinoza et al., 2011). Su et al. (2015) suggest that the HD 95086 system's halo could be produced by collisions within a wide planetesimal belt, similar to the scattering disk component of the Kuiper Belt.

Interestingly, *Herschel* and *Spitzer* observations of HR 8799 find a large mass of dust grains in the outer halo (Matthews et al., 2014). However, because this study is based on mid- and far-IR observations, it would not be sensitive to the very small dust grains seen in optical scattered light. Grains observed by *Herschel* and *Spitzer* will be well above the blowout limit and thus bound to the star. These larger grains would be ejected onto wide halo orbits by a different mechanism than the small grains discussed above.

The reader is reminded that in §2.6.3 we concluded that our observations show a deep non-detection of the HR 8799 debris disk. Thus, we conclude here that our analysis of a non-detection produces an upper limit on the mass loss rate of small dust grains in HR 8799 that is consistent with observations of other systems.

2.8 Summary & Conclusions

We have developed a new PSF subtraction algorithm, called Optimized Spatially Filtered (OSFi) normalization, for direct imaging of close-to-face-on disks using space telescope reference PSF subtraction. The algorithm is optimized to normalize PSF subtraction to the high spatial frequency noise in each reference and science image, preventing contamination from a residual low spatial frequency close-to-face-on disk. Unlike the classical spider normalization reference PSF subtraction technique, OSFi normalization is not affected by throughput contamination from a close-to-face-on disk.

After applying this algorithm to *HST*/STIS HR 8799 data, our main findings are as follows:

- A low spatial frequency residual is present after reference-subtraction and is consistent between the two epochs separated by a year.
- In comparison, self-subtraction between two separate roll angles within a single epoch yields no residual, showing that there is no detected focus evolution between science images.
- Residual high spatial frequency noise is also present after reference PSF subtraction. We evaluate the possibility that this may be from imaging the PSF of the known inner ∼6-15 AU disk (Su et al., 2009; Matthews et al., 2014), ultimately showing that this is unlikely and that we are likely seeing an instrumental effect.
- Spider normalization reference-subtraction is inconsistent between the two epochs and also inconsistent with OSFi normalization, illustrating that classical spider normalization is unoptimized for face-on disks.
- By running OSFi PSF subtraction on HD 10647, a similar dataset to the HR 8799 sequence, we demonstrate that the observed over-luminosity is likely from a combination of defocus and difference in spectral type over the STIS broadband between the reference and science images in both epochs.
- Using the photometry of this result, we determine an upper limit on the planetesimal belt mass in optical grains, which is smaller than expected from the Su et al. (2009) results, which assume a collisional cascade power law size distribution. This suggests that these optical grains are not steady state, as is expected for dust grains smaller than the blowout limit.
- For these sub-blowout limit dust grains, we estimate an ejection timescale that, if constant over the lifetime of HR 8799, would grind down $\leq 0.5 M_{\oplus}$ in parent bodies, consistent with estimates for other known debris disk systems.

Chapter 3

Planet detection down to a few λ/D: an RSDI/TLOCI approach to PSF subtraction

The following text was adapted from Gerard et al. (2016b).

3.1 Background

A typical approach to determine the exoplanet parameter space accessible by direct imaging is to extrapolate from known radial velocity distributions (e.g., Lafrenière et al., 2007b; Graham et al., 2007; Nielsen et al., 2008). Lafrenière et al. (2007b) use the radial velocity distributions of planet mass, M_p, and semi major axis, a, from Cumming et al. (2008)[1] in order to estimate survey completeness for the Gemini Deep Planet Survey (GDPS), a survey of 85 nearby young stars using the Gemini North Altair AO system that ultimately found no new exoplanet detections. Lafrenière et al. (2007b) determine that less than 28% of stars harbour at least one planet with $0.5M_J < M_p < 13M_J$ and 10 AU $< a <$ 25 AU, and that this fraction is only weakly dependent on the assumed radial velocity power law distributions.

Graham et al. (2007) fit the M_p and a distributions ($\frac{\partial n}{\partial M_p} \propto M_p^{-1}$ and $\frac{\partial n}{\partial a} \propto a^{-0.5}$, respectively) using radial velocity data compiled from eight different surveys, in total

[1]Cumming et al. (2008) find a mass distribution of $\frac{\partial n}{\partial M_p} \propto M_p^{-1.3}$ and semi major axis distribution of $\frac{\partial n}{\partial a} \propto a^{-0.7}$ using data from a 10 year exoplanet radial velocity survey with Keck high resolution spectroscopy, accounting for the bias of a better radial velocity sensitivity to detect a lower M_p at a shorter a (see §1.3.1).

detecting ∼200 exoplanets from the monitoring of ∼2000 stars. Extrapolating $\frac{\partial n}{\partial a}$ to wider orbits accessible by direct imaging, these distributions are then used in a Monte Carlo simulation to randomly populate field stars within 50 pc, including a range of orbital parameters[2]. Then, using the distance and stellar age and assuming a hot start formation scenario, Graham et al. (2007) populate a scatter plot of contrast vs. the apparent angular separation needed for a planet detection by direct imaging. Assumptions about the planet age and formation scenario are combined with the planet mass to infer a model atmosphere and then obtain a NIR flux (e.g. Spiegel & Burrows, 2012), which can then be converted to contrast using the distance. When compared to the expected performance for GPI (i.e., the sensitivity curve of contrast vs. angular separation, known as a "contrast curve"), this simulation suggests that GPIES should detect ∼100 exoplanets. Currently, about half way through the survey, GPIES has detected one planet, 51 Eridani b (Macintosh et al., 2015). This apparent discrepancy suggests that either the expected GPI contrast curve was initially overestimated or the Graham et al. (2007) Monte Carlo model included unrealistic physical assumptions. If the latter is true, this could, e.g., suggest a different population of cold start vs. hot start planets, and/or (given the weak dependence on radial velocity power law distributions in Lafrenière et al. 2007b) that there is a much steeper change in $\frac{\partial n}{\partial M_p}$ and $\frac{\partial n}{\partial a}$ towards higher mass and wider separations than predicted from radial velocity studies. Alternatively, the assumed distributions may be entirely different around young B and A stars, which are typically avoided in radial velocity surveys (§1.3.1). Research is currently ongoing for both the former and latter options, such as the work presented in the remainder of this chapter.

With this information in mind, GPI and other similar second generation ExAO instruments were designed to work at separations near the diffraction limit at $\sim 2\lambda/D$, compared to previous first generation surveys which only accessed separations greater than $\sim 7\ \lambda/D$ (e.g., Galicher et al. 2016, submitted). The $\frac{\partial n}{\partial a}$ distributions from Graham et al. (2007) and Cumming et al. (2008) suggest that it is more likely to find planets at these smaller separations, of particular interest to future extremely large telescope (ELT) ExAO instruments that will be able to access planet separations ∼3-4 times smaller than any current 10 m-class telescope. For WFIRST, even with

[2]Although not explicitly stated in Graham et al. (2007), most orbital parameters can be sampled randomly between their minimum and maximum values, as in Nielsen et al. (2008). However, also as in Nielsen et al. (2008), the distribution of orbital eccentricities is instead obtained from radial velocity data and is assumed to be identical for wider orbits. Alternatively, the similar Monte Carlo simulations in Bowler (2016) assume entirely circular orbits.

$\sim 10^{-9}$ contrasts that could be deep enough to image a habitable planet, at $\lambda = 550$ nm and $D = 2.4$ m, a planet with 1 AU separation is only resolved within 21 pc, and so working at or near the diffraction limit is also critical to maximizing the number of accessible targets. Thus, my work in the remainder of this chapter on improving GPI PSF subtraction at $2-7$ λ/D represents an initial study on achievable post-processing sensitivity in this regime what to expect with future ELTs and space telescopes.

3.2 Introduction

In the search to detect and characterize exoplanets by direct imaging, the best achievable contrast requires suppression effects from both a realtime AO and coronagraphic system as well post-processing of these images to remove the residual PSF. Current state of the art PSF subtraction algorithms are limited in sensitivity at small angular separations from the on-axis PSF, close to the coronagraph's FPM. Exoplanet population predictions suggest that with the current generation of high contrast imaging instruments, more detections are possible at smaller separations (Graham et al., 2007), or alternatively there could be additional planets in existing data that could be seen with better post-processing sensitivity. It is suggested that the distribution of radial-velocity detected planets as a function of separation follows an inverse power law (Cumming et al., 2008), continuing to wider separations accessible by direct imaging (Lafrenière et al., 2007b), and thus improved post-processing performance at smaller angular separations is of great interest to the exoplanet community.

The main factor limiting PSF subtraction performance near the IWA is the selection criteria for angular differential imaging (ADI; Marois et al., 2006a) and simultaneous spectral differential imaging (SSDI; Racine et al., 1999; Marois et al., 2000; Sparks & Ford, 2002). This selection criteria requires that reference images, used in some form of a least-squares-based (Lafrenière et al., 2007a; Soummer et al., 2012) PSF subtraction algorithm, have a certain planet signal threshold (known as "aggressiveness") to limit planet self-subtraction in the target image. The amount of field of view (FOV) rotation or spectral magnification required with ADI or SSDI, respectively, usually between $\sim$1.5 and 3 λ/D, limits fewer available references from the target sequence at smaller separations (i.e., with SSDI, for a flat spectrum, compared to a larger separation, diffraction at a smaller separation moves speckles a smaller absolute radial distance as a function of wavelength[3], and with ADI, at a set FOV ro-

[3]For a spectrum with strong spectral features like methane absorption, the selection criteria will

tation, arc length decreases with decreasing radial separation). Thus, this ADI+SSDI selection effect limits the optimal PSF subtraction sensitivity to planet detection at small angular separations near the instrument IWA.

A solution to this problem is RSDI, where PSF subtraction on a target image that may contain a hidden planet can access a large archive of "planet-less" references images. The key to increasing sensitivity at small IWA using RSDI, as with any PSF subtraction technique, is to use a set of reference images that are highly correlated to the target image. With this in mind, the archival legacy investigations of circumstellar environments (ALICE) pipeline (Choquet et al., 2014) was recently developed for *HST* PSF subtraction, mostly to recover disks, but no ground-based first generation AO surveys, generally accessing a higher sensitivity and smaller IWA than *HST*, have yet acquired enough data for RSDI in the $\sim$2-7λ/D regime[4].

With the next generation of high contrast imaging instrument surveys such as GPIES (Macintosh et al., 2014), we can test the performance of RSDI down to $\sim$2 λ/D. Our initial work on this topic acts as demonstrator for RSDI performance gain in this regime with future high contrast imaging survey instruments. In this paper, we present an analysis using GPIES to increase planet sensitivity near the IWA with a least-squares-based RSDI algorithm. In §3.3 we describe our procedure used to create PSF library reference images, in §3.4 we outline the specifics of our algorithm, in §3.5 we present the results of our algorithm applied to December 2014 GPI 51 Eri data (Macintosh et al., 2015), and in §3.6 we summarize our work and consider possible future improvements.

This analysis is based entirely on GPI H band 51 Eri target sequence data from December 2014 (Macintosh et al., 2015) and an additional PSF library archive from the GPIES campaign through December 2015. We only consider performance close to the edge of the FPM ($\sim$125 mas; Soummer et al., 2009) in the inner 100 mas to 300 mas annulus. Matching the 51 Eri b detection, all of the following analysis is carried out using a methane (T8) dwarf spectrum. We use zero indexing to refer to frames and slices in the target sequence. We use a GPI pixel scale of 14.166 mas/spaxel, determined from all GPIES and lab astrometric data (Konopacky et al., 2014). We note that a similar RSDI procedure with GPIES was recently developed

allow many more references to include in the least-squares at the peak vs. trough of the spectrum, thus "counteracting" the small angular separation selection bias.

[4]Ground-based RSDI was initially attempted with pervious-generation generation high contrast imaging surveys beyond $\sim$7 λ/D, but with no performance gain, likely due to stability issues (e.g., Galicher et al. 2016, submitted)

using the KLIP algorithm (Soummer et al., 2012), mostly designed for broadband disk detection (M. Millar-Blanchaer et al., in prep) and broadband planet detection at wider separations (D. Vega et al, in prep), and so the work presented in this paper is complimentary, instead using a least-squares designed for small IWA planet detection and spectral extraction.

3.3 PSF Library

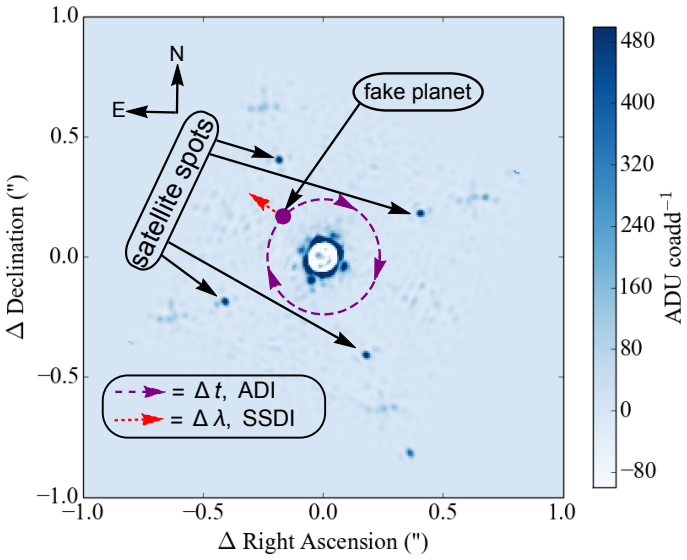


Figure 3.1 A schematic diagram, from a GPIES target sequence at a single time and wavelength after steps 1 - 2 above, illustrating how creating the PSF archive images in step 3 removes any planet signal by medianing across time and wavelength, where the planet position changes with respect to the telescope pupil due to ADI and SSDI.

Figure 3.1 shows a schematic of how we create the GPIES PSF library such that any potential planet is removed from the data. In each sequence of images for a given target, we use the datacubes produced from the GPI data reduction pipeline (Perrin et al., 2014) so that all images at a given time and wavelength are

1. registered to a common center using cubic spline interpolation, magnified as a function of wavelength to align speckles (the first step in SSDI), and flux normalized to flatten the stellar spectrum, all using the GPI satellite spot (Wang et al., 2014) positions,

2. high-pass filtered using a 11 by 11 pixel median boxcar filter to remove low spatial frequency noise, and

3. median combined in both time and wavelength, removing any possible planet signal as illustrated in Figure 3.1.

Any exoplanet will be removed after medianing all images in step 3 because (1) the planet position in ADI observing mode changes azimuthally with time (i.e., without de-rotating the images to a common position angle), and (2) after step 1 the planet position changes radially with wavelength. Thus, for every target sequence we obtain one PSF library reference image, yielding a total of 207 available archive references.

3.4 PSF Subtraction Algorithm

In this section we present our adaptation of the Speckle-Optimized Subtraction for Imaging Exolanets (SOSIE; Marois et al., 2010a) and Template Locally Optimized Combination of Images (TLOCI; Marois et al., 2014) least-squares pipelines to use with optimized RSDI. In §3.4.1 we review the principles of SOSIE and TLOCI in application to our pipeline, in §3.4.2 we explain our reference selection criteria based on image correlation, in §3.4.3 we outline our algorithm to optimize planet signal to noise ratio (SNR), and in §3.4.4 we discuss additional optimization parameters. For all images in the 51 Eri target sequence, unless explicitly stated, we apply all the steps in the TLOCI algorithm (Marois et al., 2014), using an input T8 spectrum.

3.4.1 TLOCI, SOSIE Architecture

As in the original least-squares-based LOCI algorithm (Lafrenière et al., 2007a), we define the region of interest to subtract the PSF as the "subtraction region" as well as a larger "optimization region." The optimization region is used to obtain the least-squares subtraction coefficients from a region that does not contain any exoplanet light, thus fitting for the speckle noise while preventing the algorithm from fitting the planet signal. The subtraction coefficients are then applied to the subtraction region to minimize the noise but also preserve the planet signal. In order to obtain these subtraction coefficients, a least-squares algorithm is run using a correlated set of reference images (with the same geometry of optimization and subtraction regions as in the target image) in order to minimize the noise in the target image optimization

region. In the LOCI algorithm, the subtraction region lies within the optimization region, causing the least-squares to fit any planet signal in the subtraction region, thus significantly affecting the algorithm throughput and planet SNR. In contrast, the SOSIE algorithm is designed to allow better planet throughput by

1. masking the subtraction region from the optimization region, and
2. correcting the remaining throughput reduction from self-subtraction by using a forward model (FM).

Our choice of masking geometry in step 1 is illustrated in Figure 3.2. Any SOSIE

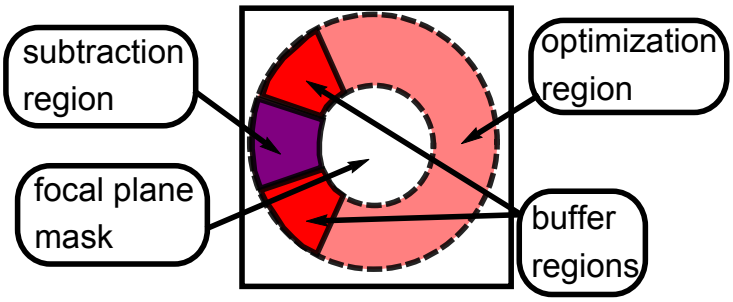


Figure 3.2 A schematic of our chosen optimization and subtraction region geometry, adding buffer regions between the two. A least-squares algorithm (Lafrenière et al., 2007a) is run on the optimization region in the target image and set of references to generate subtraction coefficients that are then applied to the target image subtraction region.

implementation relies on the assumption that noise in the optimization region is spatially correlated with noise in the subtraction region, and thus works best with highly correlated data obtained in stable conditions. We chose the optimization region geometry in Figure 3.2. We additionally chose to add a "buffer region" on either side of the optimization region, each extending azimuthally by the amount of target sequence FOV rotation (usually less than $\sim 30°$). The rationale for adding this buffer region is that if a planet lies at the edge of the subtraction region, the buffer region prevents the least-squares from fitting the planet light in an adjacent optimization region, potentially preventing throughput loss, although further testing of different SOSIE geometries, which are beyond the scope of this initial paper, are needed to better understand the sensitivity gain by adding a buffer region, as well as the validity of assuming the noise is azimuthally symmetric (see §3.6).

In step 2, we use a TLOCI approach (Marois et al., 2014) to generate spectrum-weighted template PSFs as a function of time and wavelength, applying a T8 template spectrum to select reference images at a given aggressiveness (see §3.4.2). We then apply the least-squares subtraction coefficients to the target and reference template PSFs, just as is done to subtract the real target images with a set of references, creating a residual noiseless FM to estimate the amount of self-subtraction and apply the appropriate throughput correction. Instead of using the FM flux value at the target PSF location to calculate a throughput correction, we use the peak FM flux value in an aperture centered on the target PSF location of width $\sim\lambda/D$ because the former can be biased by aggressive self-subtraction effects.

3.4.2 Reference Image Selection

Normalizing reference images to a given target image using the robust standard deviation (Dowell et al., 2012b), we then used the robust standard deviation of the target-reference difference (Figure 4.6a) to quantify target image "correlation," which converges to zero at the target image and is lower for more correlated reference images. Figures 4.6b, 3.3c, and 3.3d show the same robust correlation metric after applying the TLOCI selection criteria with aggressiveness 0.9, 0.5, and 0.1, respecively. An aggressiveness of 0.9 requires that for a given reference image, the cumulative FM flux in a λ/D diameter aperture, centered on the target image FM location, be less than 90% of the target image cumulative FM flux in that aperture (and for an aggressiveness of 0.5, less than 50%, etc.). In addition to aggressiveness, we require that any reference for a given target image must be at either the same time or wavelength. This allows for a decreased overall computation time in order to determine the correlation of a given target image to all the available references. Although this limits the total available number images from which to select references, we know that images at the same time and wavelength are already $\sim$90% correlated in wavelength and $\sim$50% correlated in time (Marois et al., 2014). Thus, images at a different time and wavelength from the target image should accordingly decrease correlation, and so with this simplification we are still keeping the most correlated images.

Assuming use of the 25 most correlated reference images (after the TLOCI selection criteria to minimize self-subtraction effects; but see §3.4.3) to subtract the target image, Figures 4.6b, 3.3c, and 3.3d show that at high and medium aggressiveness, most references are from the target sequence (24 of 25), whereas at low aggressive-

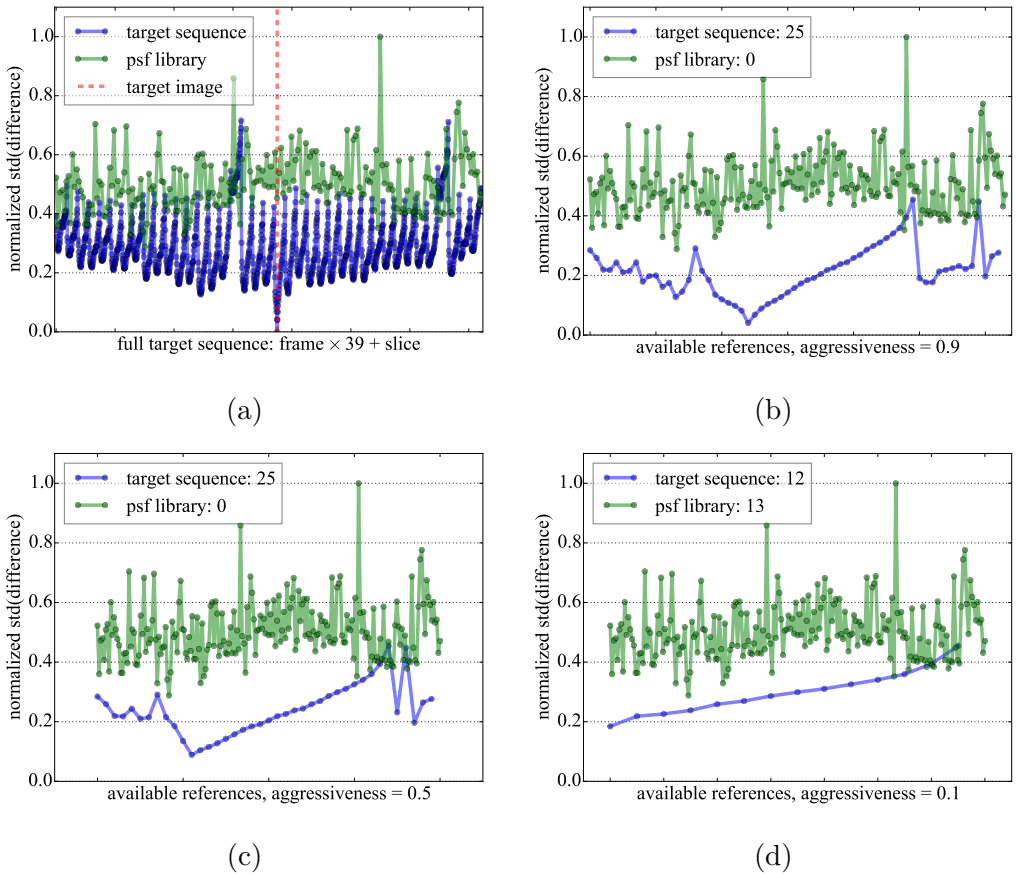


Figure 3.3 An example of how the correlation for a single target image (frame 20, slice 10) varies with aggressiveness. (a) Our correlation metric for an example target image no selection criteria, using the robust standard deviation of the difference between the target image and a given reference image, normalized to the highest value. As expected, the robust correlation converges to 0 at the sample target image. Using the reference selection criteria to chose the 25 most correlated images, an aggressiveness of 0.9 (b) and 0.5 (c) yields only one PSF library image to include in the set of references, whereas an aggressiveness of 0.1 (d) yields references mostly from the PSF library, suggesting that a gain in contrast by adding the PSF library is dependent on aggressiveness.

ness, a greater number of the most correlated references come from the PSF library (11 of 25). For low aggressiveness, we can see that all the target sequence images are more correlated than any PSF library image, but this aggressiveness limits the total number of available references from the target sequence to 14, thus requiring use of the PSF library if we want to use 25 total references. This discrepancy as a function of aggressiveness, in both available number of correlated images from the

target sequence and number of correlated images from the PSF library, suggests that adding the PSF library could improve contrast in at least some cases that require low aggressiveness, such as the spectral extraction of a companion.

3.4.3 Optimization Algorithm

Although the SOSIE and TLOCI algorithms have made improvements over the original LOCI algorithm to correct for throughput using masking and forward modelling, thereby effectively optimizing the planet SNR vs only minimizing the noise in the subtraction region, there are still (unaccounted for) free parameters which can change the planet SNR, including

1. number of references,

   Too few references limits the diversity of the least-squares to optimally minimize the noise in the subtraction region, whereas too many reference images will overfit the noise in the optimization region, also causing a higher noise in the subtraction region.

2. aggressiveness,

   A lower aggressiveness inherently requires selecting less correlated reference images from the target sequence but also causes less self-subtraction, whereas a higher aggressiveness will use more correlated images from the target sequence but cause more self-subtraction.

3. singular value decomposition (SVD) cutoff,

   Less correlated reference images can cause noise to propagate into the covariance matrix inversion step in the least-squares algorithm, effectively overfitting the noise in the optimization region (Marois et al., 2010a), thus causing a higher noise in the subtraction region, similar to step 1. This problem can be addressed by changing the SVD cutoff, which truncates the matrix inversion below a cutoff (singular) value.

4. different optimization and subtraction region geometries.

The core methodology of our algorithm is to address items 1 and 2 by optimizing what we call the "FM SNR." We also briefly address items 3 and 4 in §3.4.4 and also item 4 in §3.6.

To optimize the SNR as a function of number of reference images and aggressiveness in each subtraction region, we compute a grid search with our aforementioned SOSIE/TLOCI algorithm, varying the number of the most correlated references between 3 and 25 with a step size of 1 and varying the aggressiveness between 0.1 and 0.9 with a step size of 0.1. For each iteration, we compute the FM SNR by dividing the peak FM flux within a λ/D diameter aperture centered on the target PSF location (as discussed in §3.4.1) by the robust standard deviation in the subtraction region of PSF-subtracted image. Before calculating the noise in the subtraction region, we first convolve the original PSF-subtracted image with a λ/D diameter Airy disk kernel (Robitaille et al., 2013), whose cumulative flux is normalized to one so that flux is conserved in the convolved image. This convolution step filters out pixel-to-pixel noise in order to distinguish λ/D-sized speckles as the noise floor. Finally, we adopt the parameters of number of references and aggressiveness for a given subtraction region that maximize the FM SNR.

Throughout this process, in each least squares image, we found cases where negative least-squares subtraction coefficients generated an artificial increase in forward model flux, which bias the FM SNR optimiztion even though the planet is not actually brighter. The solution is to restrict the subtraction coefficients to be positive in the least-squares algorithm, also known as a non-negative least-squares (NNLS; Lawson & Hanson, 1987), and so we use this algorithm instead of a regular least-squares for the rest of our analysis. We also chose to optimize the FM SNR as opposed to bootstrapping a signal into the real image because the former is a noiseless image that can isolate throughput effects from noise effects, unlike the latter.

To test our algorithm performance in the rest of this paper, we will consider optimized PSF subtraction using only the target sequence (hereafter "tar opt"), unoptimized PSF subtraction using only the target sequence (hereafter "un-opt"), and optimized PSF subtraction using both the target sequence and PSF library (hereafter "tar+lib opt"). We do not consider optimized PSF subtraction using only the PSF library, without the target sequence. A full reduction with only the PSF library would not do better than just the target sequence or target sequence + PSF library because Figure 3.3 shows that the most correlated images are from the target sequence, independent of aggressiveness, and that these also outnumber the most correlated images to include from from the PSF library, so that on average for a given target image, the given aggressiveness is such that the majority of most references will be from the target sequence.

Two examples comparing our various optimization schemes are shown in Figure 3.4, along with the mean behavior across the full target sequence. Figure 3.4 illus-

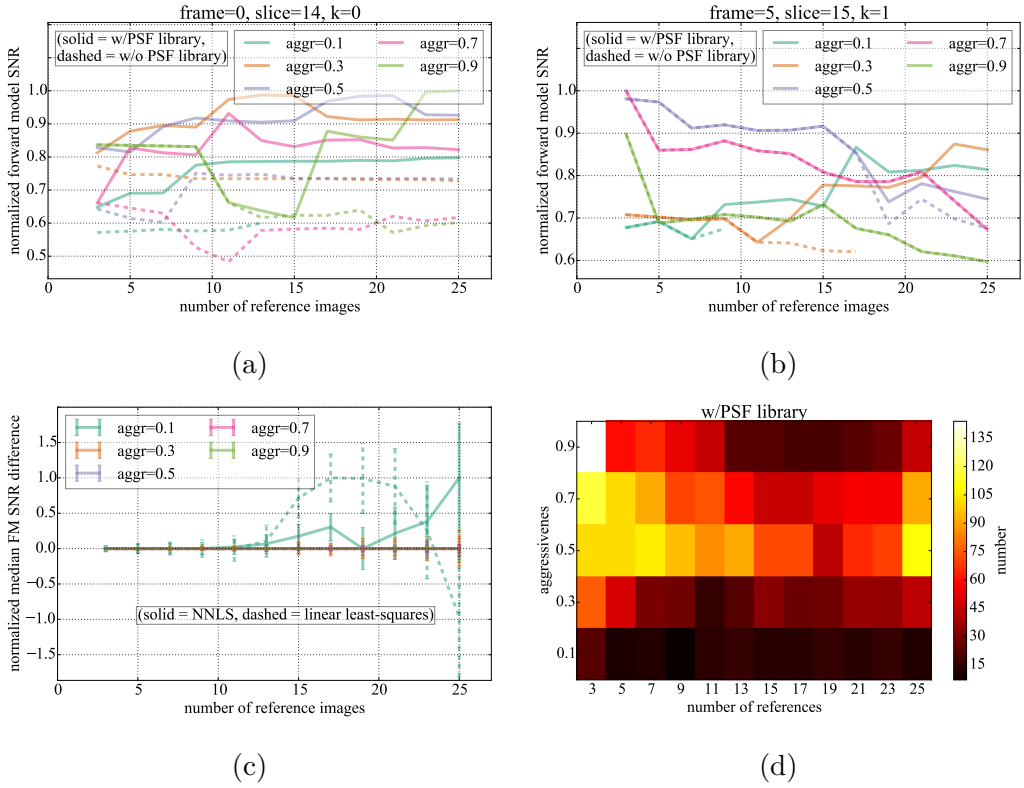


Figure 3.4 (a)-(b) FM SNR as a function of number of reference images and aggressiveness, with and without the PSF library ("tar+lib opt" and "tar opt," respectively), for two test cases. The variable "k" refers to the specific subtraction region considered, for which, using two subtraction regions across the full annulus, k=0 represents the lower half of the annulus and k=1 represents the upper half. (c) the difference in FM SNR values between the tar+lib opt and tar opt codes for each individual least-squares image, averaged across the full sequence, as a function of number of reference images and aggressiveness. Error bars are determined from the standard deviation of each averaged FM SNR difference divided by the square root of the number of available values at that given number of references and aggressiveness. The non-negative least-squares (NNLS) algorithm is shown in solid for comparison with the standard linear lease-squares algorithm, shown as dashed lines. A positive value on the y-axis indicates that adding the PSF library improves the FM SNR. (d) A two-dimensional histogram of the optimal values for aggressiveness and number of references as determined from FM SNR optimization across the full target sequence using the tar+lib opt code.

trates a number of important concepts that we discuss below:

- In Figure 3.4a, the FM SNR is consistently higher in the tar+lib opt code than in the tar opt code (i.e., for a given color, the solid lines generally lie above the dotted lines) by up to a factor of ∼2, suggesting that adding the PSF library should show significant improvement in SNR, relatively independent of aggressiveness and number of references. However, this was a specially picked case where we found an unusually high number of correlated PSF library images to the target image.

- Figure 3.4b is a much more typical example FM SNR optimization. In this case, we can see that:

    1. There is a ∼10-20 % improvement in FM SNR when using the PSF library, but only at medium to low aggressiveness.
    2. without using the PSF library, a low aggressiveness selection criteria may provide less than 25 available references from the target sequence, meaning that the number of references can only be optimized out to less than this value. This effect agrees with Figure 3.3, showing that at a lower aggressiveness there are less available references from the target sequence.

    The two points above suggest that adding the PSF library is most effective at low aggressiveness when there are less available references from the target sequence. High aggressiveness cases do not show any improvement because there are always enough available references from the target sequence that are all more correlated with the target image than any PSF library image.

- Figure 3.4c shows that

    1. adding the PSF library does on average improve FM SNR throughout the full sequence, but only at low aggressiveness and more than ∼10 references, consistent with Figure 3.4b.
    2. at low aggressiveness, the negative coefficients in a linear least-squares algorithm are biasing the tar+lib opt FM to performance that is both better and worse than tar opt code, illustrating that using a NNLS algorithm should yield a more consistent performance improvement in this regime.

- Figure 3.4d shows the diversity of optimal parameters for aggressiveness and number of references across the full sequence. This suggests that the standard

approach to PSF subtraction, which uses a set number of references and aggressiveness across the full sequence (i.e. the un-opt code), should not perform as well as our optimized approach.

3.4.4 Additional Parameters

In order to choose an optimal least-squares SVD cutoff for matrix inversion ($\mathrm{SVD}_{\mathrm{cutoff}}$)[5] and number of subtraction regions around the full annulus (n), we ran a grid search with $10^{-1} < \mathrm{SVD}_{\mathrm{cutoff}} < 10^{-7}$ (with step size of $\Delta\log_{10}(\mathrm{SVD}_{\mathrm{cutoff}}) = -1$) and $2 < n < 5$ (using a step size of $\Delta n = 1$) for a number of test cases at a fixed number of references and aggressiveness, using the same subtraction region geometry as in Figure 3.4. In a number of test cases for single target images, we generally found a maximum FM SNR at $\mathrm{SVD}_{\mathrm{cutoff}} = 10^{-3}$ and $n = 2$, and so we adopt these parameters for the rest of this paper (but see §3.6).

Relatedly, in order to run this optimization routine on the full target sequence in a reasonable amount of time (∼24 hours on our 2.7 GHz processor, running in serial), we simplified the above optimization routine by binning the step size in number of reference images to two instead of one and in binning in aggressiveness to 0.2 instead of 0.1, finding no evidence of a significant decrease in forward model SNR from these effects for multiple test cases.

3.5 Results

Results after running our tar+lib opt, tar opt, and un-opt codes on the full 51 Eri sequence, each of which include a run with and without a bootstrapped methane planet, are shown in Figure 3.5. Each image is rotated to north up (ADI) and demagnified (SSDI) before median combining in time to create a final PSF-subtracted datacube. The datacube is then collapsed across wavelength with a weighted mean using the input spectrum (Marois et al., 2014). When running the un-opt code, we chose to use the "typical" parameters applied throughout the full sequence (independent of the results in Figure 3.4d): 20 reference images and aggressiveness = 0.5. For the bootstrap reductions, we add the satellite spot PSF into the raw images, normalized to the robust standard deviation in slice 9 (the peak of a T8 spectrum) of the non-bootstrap

[5]The SVD cutoff was only investigated for matrix inversion in our testing with a linear least-squares algorithm (but see §3.6). We did not consider optimizing the matrix inversion in the NNLS algorithm.

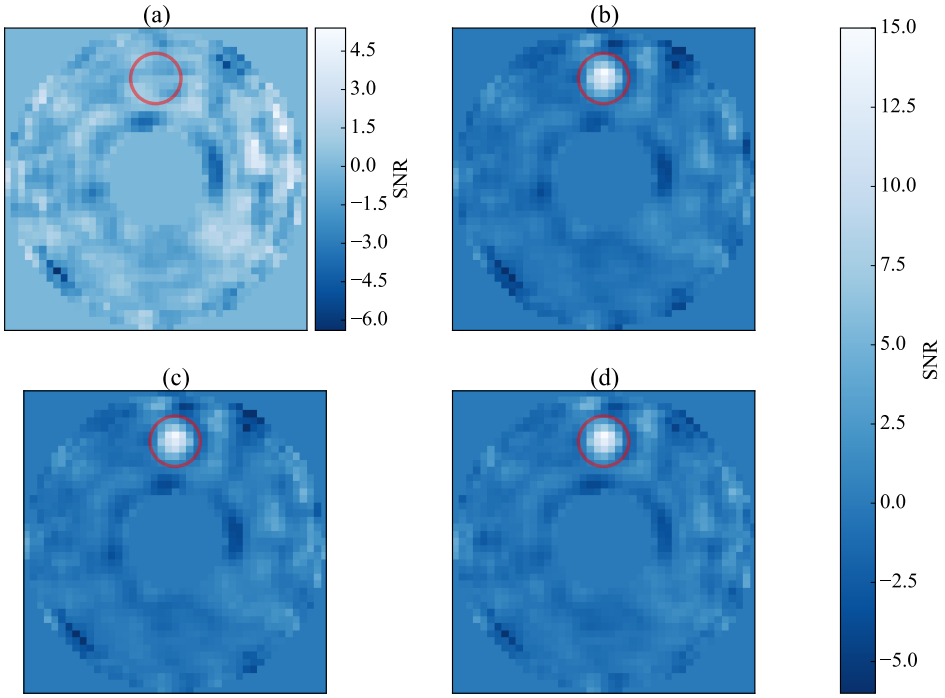


Figure 3.5 SNR maps for (a) the optimized code using the PSF library, (b) bootstrapping with the optimized code using the PSF library, (c) bootstrapping with the optimized code using only the target sequence, and (d) bootstrapping with the un-optimized code (20 reference images, aggressiveness = 0.5 throughout the full sequence) using only the target sequence. The red circle shows the bootstrap location. The bootstrap SNR in (b), (c), and (d) is 12.55, 12.45, and 12.95, respectively.

tar+lib opt collapsed cube within a 2 pixel-wide annulus around the bootstrap radial separation (200 mas), multiplied by 8, which represents an "$\sim 8\sigma$" detection. For a methane planet, the bootstrap signal is also a function of wavelength, and so we additionally normalize the bootstrap signal to the mean of the normalized input spectrum (accordingly, in the final time-collapsed cube the T8 peak should be above 8σ and the methane absorption band should be below 8σ, and thus the bootstrap signal in the wavelength-collapsed image should be detected at $\sim 8\sigma$).

To compute the SNR maps in Figure 3.5, we create a "back-rotated" time-collapsed datacube, where each frame is rotated by its amount of FOV rotation, but in the opposite direction of north up, thus medianing out any planet signal that is no longer

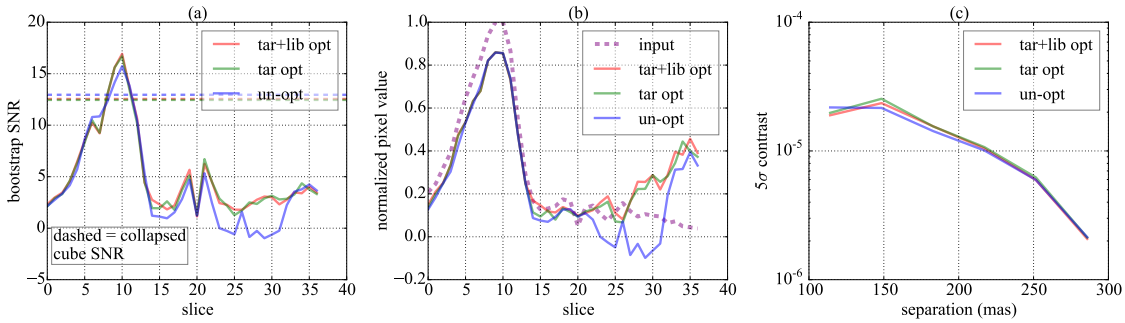


Figure 3.6 Performance after bootstrapping a simulated methane planet using our three different codes: optimized using the target sequence and PSF library (tar+lib opt), optimized using only the target sequence (tar opt), and un-optimized (20 reference images, aggressiveness = 0.5 throughout the full sequence) using only the target sequence (un-opt). (a) Bootstrap SNR as a function of wavelength, where the dashed lines of the same color show the corresponding collapsed cube bootstrap SNR as in Figure 3.5 (b)-(d), indicating that neither our FM SNR optimization scheme nor using the PSF library improves performance. (b) Output bootstrap signal normalized to the peak value of the input bootstrap signal, indicating that all algorithms more or less recover the input signal, although the methane absorption band is bootstrapped well below the PSF-subtracted noise floor and thus biased by any brighter speckles, and the ∼20% discrepancy at the T8 peak suggests that further work is needed to understand why our FM throughput correction is not performing as well in this regime. (c) 5σ contrast as a function of separation on the collapsed, back-rotated, bootstrapped images, again showing that all codes have a similar performance.

spatially aligned, but preserving the radial noise characteristics. Within a given 2 pixel-wide annulus, the signal values in the correctly rotated north-up, time-collapsed image are then divided by the robust standard deviation in the same annulus of the corresponding back-rotated image, repeated over all annuli of the 100 to 300 mas region to produce the SNR maps in Figure 3.5.

None of the non-bootstrapped codes show an obvious point source detection above 5σ in the inner annulus. We find a bootstrap SNR for the tar+lib opt, tar opt, and un-opt codes of 12.55, 12.45, and 12.95, respectively. These results suggest that neither our FM optimization scheme nor adding the PSF library actually improves planet SNR. The latter is consistent with the average results in Figures 3.4c, which shows that on average the PSF library shows no improvement except in cases of low aggressiveness, for which Figure 3.4d shows that most of the sequence is optimized at higher aggressiveness. However, the lack of improvement from our FM SNR optimization

scheme is puzzling, and discussed further below as well as in §3.6.

Figure 3.6 shows the bootstrap spectral performance and contrast curves for our three different PSF subtraction codes. The left panel shows the output bootstrap SNR spectrum (wavelength increases with increasing slices, from 0 to 36 across H band (Maire et al., 2014)), confirming the collapsed cube results in Figure 3.5 as a function of wavelength: neither our FM SNR optimization scheme nor adding the PSF library shows any improvement in bootstrap SNR over existing methods. The middle panel shows the input and output bootstrap signals as a function of wavelength, normalized to the peak input signal, indicating that in all codes our noiseless FM throughput correction mostly recover the input signal. In the methane absorption band there is some discrepancy between the input and recovered signal, but in this low SNR regime the bootstrap signal is well below the noise and may be biased by brighter speckles. However, around the T8 peak, all codes are ∼20% dimmer than the input signal. This throughput loss suggests that our FM throughput correction is not performing well in this regime, and further work is needed to understand why (§3.6). The right panel shows the 5σ contrast as a function of position. In order to remove bias from the bootstrap signal, we calculate contrast on the bootstrapped, collapsed, back-rotated images. I compute 5σ contrast at a given radial separation within the inner annulus, $c(r)$, via

$$c(r) = \frac{5\left(\sigma_{\text{robust}}\{\text{im}(r)\}\right)}{\langle f_{\text{sat}}\rangle \left(f_{\text{star/sat}}\right)}, \tag{3.1}$$

where $\text{im}(r)$ are the image pixel values within a two pixel-wide annulus centered at separation r, $\sigma_{\text{robust}}\{\}$ is the robust standard deviation operator, $\langle f_{\text{sat}}\rangle$ is the average satellite spot flux of the full sequence, and $f_{\text{star/sat}}$ is the GPI star to satellite spot flux ratio. I compute $\langle f_{\text{sat}}\rangle$ by median collapsing the full 51 Eri sequence (after registration, magnification, stellar spectrum normalization, and high pass filtering) in time and wavelength and then averaging the four peak satellite spot flux values. In H band $f_{\text{star/sat}}$ is 4726, or $\Delta m_{\text{star/sat}} = 9.23$ mags (Wang et al., 2014, P. Ingraham, private communication). As suggested by Figures 3.6 (a) and (b), Figure 3.6 (c) shows a lower contrast noise floor for tar+lib opt compared to the other two codes.

3.6 Conclusion

ADI and SSDI limit PSF subtraction sensitivity to detect and characterize planets at small angular separations. This problem can be addressed using RSDI. RSDI on ground-based high contrast imaging instruments at ∼2-7 λ/D has only recently become possible with sufficient campaign data from GPIES (Macintosh et al., 2014). Our main conclusions are as follows:

- We have developed a new method of optimized RSDI PSF subtraction using the SOSIE (Marois et al., 2010a) and TLOCI (Marois et al., 2014) least-squares formalism, which includes
    1. reference image selection based on an input spectral template and robust correlation to the target image,
    2. an ensemble of PSF library reference images that removes any planet signal through medianing, many of which are correlated enough to images in the target sequence to allow using RSDI,
    3. an improved planet throughput compared to the LOCI (Lafrenière et al., 2007a) formalism, based on an optimization and subtraction region masking scheme and a subsequent FM throughput correction, and
    4. an optimization routine designed to maximize the planet SNR as a function of the number of reference images and aggressiveness.
- When running this PSF subtraction routine on the inner 100 to 300 mas annulus of the GPI December 2014 51 Eri dataset (Macintosh et al., 2015), we find
    1. no obvious planet detection,
    2. when bootstrapping a fake methane planet into the raw datacubes, there is no apparent gain in planet SNR when adding the PSF library and/or using our FM optimization scheme compared to current non-RSDI-based PSF subtraction methods.

These results present the first attempt to improve planet SNR using RDSI in the ∼2-7 λ/D regime. This method should be further explored in the context of current and future high contrast imaging survey instruments working near the diffraction limit. Future work on this initial study will proceed for a number of different topics. However, for any further adjustments to the optimized code, the first step is to parallelize

the now serial Python-based optimization code so that it can reduce the full sequence in only a few hours on our 16 core machine rather than ∼24 hours. Afterwards, there are a number of different possible avenues to explore:

- Understand the discrepancy between input and output bootstrap signal near the T8 peak, perhaps originating from problems with applying our FM throughput correction in this regime. This throughput loss may also be affecting our FM SNR optimization scheme.

- Test performance of the three codes using additional spectral templates, such as a DUSTY (Baraffe et al., 2002) spectrum, which should be less effective than a methane spectrum when using SSDI due to the lack of spectral features, thus more sensitive to increased performance with RSDI.

- Test performance of different optimization and subtraction region geometries, e.g., using all or portions of the adjacent, more outer annulus and/or portions of the inner annulus to define the optimization region geometry, still not overlapping with the subtraction region. The rational here is that there could be a more optimal geometry that samples the noise of the subtraction region rather than assuming azimuthal symmetry in the same annulus (e.g., assuming radial symmetry).

- To further investigate the performance tradeoff between a linear least-squares and a NNLS, include a loop in the linear least-squares to optimize the SVD cutoff. Although we found that using two subtraction regions consistently gave the highest FM SNR, we found that the optimal SVD cutoff varied between 10^{-3} and 10^{-4}. We did not include a loop to optimize this parameter due to computational limits, which would have at minimum doubled the overall computation time in serial. However, there may be little or no gain from adding such a loop, since optimizing the SVD cutoff is similar to optimizing the number of references.

- Create PSF library images by medianing only across wavelength. This could allow for a greater ensemble of references from which to choose. However, medianing only in wavelength is also less effective at medianing out the planet flux in the inner annulus, and so there may be additional throughput effects from this. The rational here is that by median collapsing in time and wavelength, we

could be missing a potentially more correlated PSF library image due to time instability.

- Run the tar+lib opt code on the all of the GPIES campaign data acquired thus far to search for any undetected planets already in the existing data, using multiple spectral templates. Most datasets are far less ideal than our 51 Eri data set, particularly in stability and FOV rotation, for which the former and the latter can be improved by using an SNR optimization scheme and a PSF library, respectively. Thus, RSDI and/or FM SNR optimization may show more improvement on other targets.

Chapter 4

High contrast imaging of exoplanets on ELTs using a super-Nyquist wavefront control scheme

The following text was adapted from (Gerard & Marois, 2016).

4.1 Background

Recall from §1.4.1 that the wavefront in the pupil plane ($Ae^{i\phi}$) and focal plane are related by a Fourier transform, and that the DM in the pupil plane has a finite number of actuators, N_{act}. The result of this effect is that in the focal plane there is only a finite correctable region obtainable by an AO system, called the "AO control radius," or the "Nyquist limit." For a DM with N_{act} actuators superimposed across the telescope pupil with diameter D,[1] the highest possible spatial frequency along x or y achievable by the DM is a sine wave with $N_{\mathrm{act}}/2$ cycles across the effective pupil diameter. The Fourier transform of this pupil plane sine wave creates two delta functions in the focal plane, each of which have an angular separation of $(N_{\mathrm{act}}/2)(\lambda/D)$ on either side of the image centre. Because there is still a circular aperture, these delta functions are convolved with the original on-axis PSF, effectively "copying" the PSF

[1] Relay lenses can change the beam size to re-image the telescope pupil onto the DM, still placed in a conjugate pupil plane.

to the highest achievable focal plane separations. This defines the Nyquist limit: a box around the on-axis PSF centre with dimensions $N_{\rm act}(\lambda/D) \times N_{\rm act}(\lambda/D)$, outside of which no corrections to the wavefront can be made. A larger AO control radius is achieved by adding more actuators to a DM or by observing at a longer wavelength and decreased by using a telescope with a larger diameter.

In this chapter, I explore the concept of super-Nyquist wavefront control, or focal plane wavefront sensing beyond the Nyquist region. This technique requires (1) a super-Nyquist optical element, such as a grating or phase plate, and (2) a compatible wavefront control technique. I simulate the use of a specific focal plane wavefront sensing technique (§1.4.1) called the "self-coherent camera" (SCC), described below as in Baudoz et al. (2006) and subsequent papers. In this technique, the coronagraph design is modified to add a small, off-axis hole placed in the Lyot stop. Then, in the focal plane downstream of the Lyot stop, the on-axis beam recombines with the light from this off-axis hole, acting as an interferometer. This interference pattern creates fringes that can be seen on the science camera, shown in Figure 4.1. However, light from an off-axis planet will not be fringed because is it not coherent with the stellar light, and so isolating and minimizing the light from the fringed speckles in a specific region of the focal plane would create a "dark hole," significantly increasing the contrast in that region but without affecting the planet throughput. In order to isolate this fringed stellar light, we use the optical transfer function (OTF)—the Fourier transform of the PSF. Using the notation from equation 1.1, the OTF is given by

$$\begin{aligned} \text{OTF} &= FT\{\text{PSF}\} \\ &= FT\{|FT\{Ae^{i\phi}\}|^2\} \end{aligned} \tag{4.1}$$

Equation 4.1 shows that the OTF is a complex-valued image. When using the SCC design described above, taking the amplitude of the OTF in the focal plane where the two beams recombine shows an on-axis peak with two off-axis lobes, $F\{I_+\}$ and $F\{I_-\}$, on either side, shown in Figure 4.1. The two $F\{I_+\}$ and $F\{I_-\}$ lobes are only from fringed stellar light and contain no planet light. The amplitude of the OTF is called the modulation transfer function (MTF), which is an image of the spatial frequencies of the PSF [2]. Placing a binary mask on the MTF (i.e., an amplitude mask

[2]Accordingly, the relative amount of energy at the centre vs. edges of the MTF represents the relative power in low vs. high spatial frequencies of the PSF, respectively.

on the OTF) to isolate either lobe, e.g., $F\{I_-\}$, and then centring and inverse Fourier transforming the masked OTF yields an image similar to the PSF, called I_-, except that it (1) is complex-valued, and (2) only contains stellar light. Minimizing both the phase and amplitude of I_- in a specific region of the image corresponds to minimizing the energy of the stellar light in the real focal plane, ultimately creating a dark hole with improved contrast in that region. Figures of the SCC interferometric PSF and the MTF are shown in Figure 4.1.

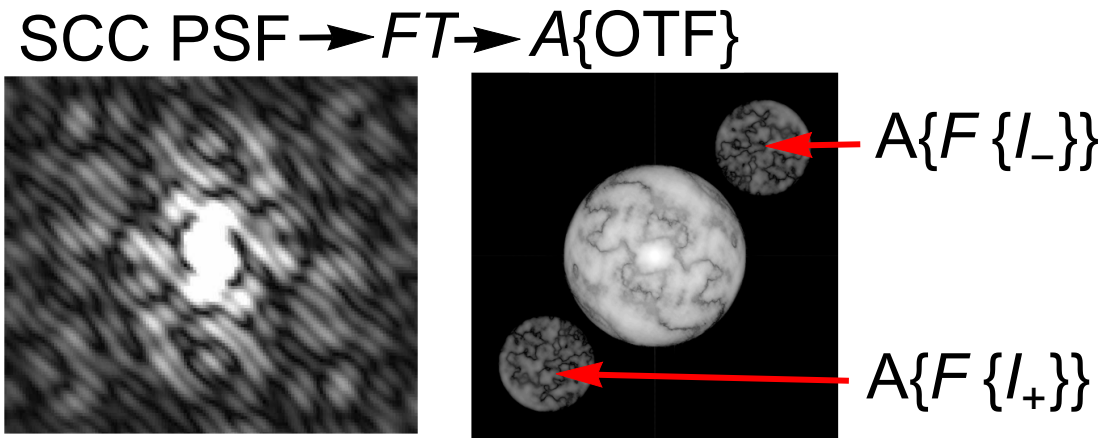


Figure 4.1 Left: an example image of the fringed PSF at the focal plane downstream of the SCC Lyot stop design, from Galicher et al. (2010). Right: the amplitude of the OTF of the fringed SCC PSF, showing two off-axis lobes, $F\{I_+\}$ and $F\{I_-\}$, from Mazoyer et al. (2014). $A\{\}$ denotes the amplitude operator, and FT denotes the Fourier transform operator. Placing an amplitude mask on either $F\{I_+\}$ or $F\{I_-\}$ and then centring and inverse Fourier transforming yields an image with isolated stellar light that can be minimized in order to maximize planet SNR.

4.2 Introduction

The direct imaging of exoplanets is more sensitive to planets beyond ∼5-10 AU. Although direct imaging has seen less planet detections than radial velocity or transit techniques, the past eight years have revealed a handful of directly imaged planets, including multiple planets around HR 8799 (Marois et al., 2008b, 2010b), HD 95086 b (Rameau et al., 2013c), Beta Pic b (Lagrange et al., 2010), and most recently 51 Eri b (Macintosh et al., 2015).

Detailed characterization of these existing planets with ELTs may be a difficult task, since these systems may lie either at the edge of or outside of the typical ∼0.4

arcsecond ELT AO Nyquist control region when observing in the near infrared (Herriot et al., 2014). This region is set by the DM actuator pitch projected onto the telescope pupil, and for a square grid DM is a $(N_{\rm act})(\lambda/D) \times (N_{\rm act})(\lambda/D)$ region around the on-axis PSF, where $N_{\rm act}$ is the number of actuators in width across the telescope pupil, λ is the wavelength of light, and D is the telescope diameter (Thomas et al., 2015). Thus, with classical, single conjugate AO (SCAO), uncorrected atmospheric turbulence and and quasi-static speckles will lower the planet SNR.

A recent technique has recently been proposed to allow wavefront control outside the Nyquist control region, called super-Nyquist wavefront control (SNWFC; Thomas et al., 2015). The main hardware component in this technique requires the use of a super-Nyquist element in an AO system, such as a mild pupil plane diffraction grating with a spacing between lines that is smaller than the DM actuator pitch relative to the pupil size. The pupil plane imprint creates a PSF copy in the focal plane that is outside the DM Nyquist region, allowing wavefront control to work in a similar $(N_{\rm act})(\lambda/D) \times (N_{\rm act})(\lambda/D)$ control region around this super-Nyquist PSF copy.

In this paper, we present the results of a laboratory experiment and simulations for a future experiment to show that it is possible to use SNWFC on an ELT AO system system to directly image already known and new exoplanets. In §4.3 we describe the laboratory experiment design (§4.3.1), simulations of expected lab performance (§4.3.2), and results in the lab (§4.3.3). In §4.4 we describe the setup and results of our simulation using the SCC (Baudoz et al., 2006; Mazoyer et al., 2014). In §4.5, we summarize our results and discuss future work.

4.3 Deterministic Laboratory Experiment

We first design, simulate, and test a simple deterministic laboratory experiment that demonstrates the possible performance improvement using SNWFC. The goal here is to show that performance improvement is possible in an idealized case to motivate more realistic SNWFC simulations (§4.4), the future testing of this technique in a laboratory, and the future ELT applications for high contrast imaging of exoplanets.

4.3.1 Experiment Design

In this section we describe the methodology and algorithm setup for our deterministic speckle nulling experiment. Figure 4.2 shows a schematic layout for our SNWFC

experiment.

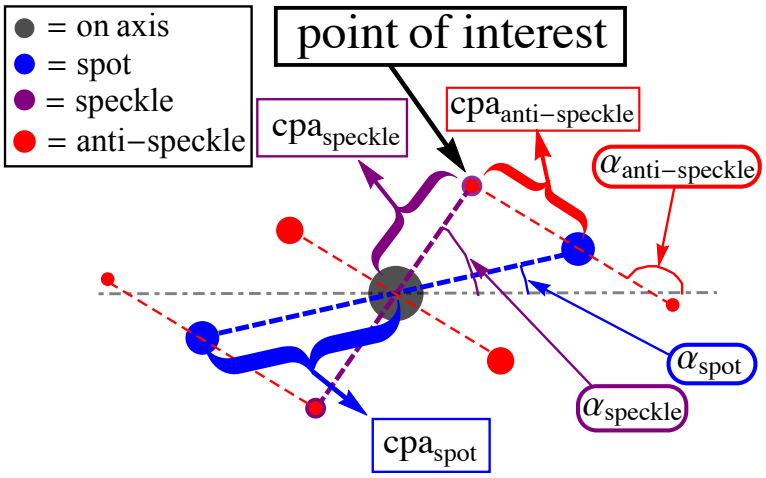


Figure 4.2 A schematic diagram of the focal plane for our deterministic SNWFC laboratory experiment. "α" represents the sine wave position angle, counter clockwise from the $+x$ axis. "cpa" is the number of cycles per aperture for a given DM sine wave, and it represents radial location of the sine wave PSF copy in the focal plane in units of λ/D (the DM sine wave angular frequency is $f = 2\pi(\text{cpa})/(D_{\text{DM}})$,where D_{DM} is the DM pupil plane diameter).

The basic structure involves the placement of three sine waves on the DM:

- **spot**: Shown in blue in Figure 4.2, the spot represents the effect of a PSF copy from a pupil plane sine wave phase plate at a higher frequency than the DM Nyquist frequency, thus creating a copy of the on-axis PSF outside the AO control region. In the absence of a phase plate for our experiment, we use the highest DM sine wave frequency to represent the spot. We note that with this design there is no actual super-Nyquist element, since the spot is, by definition, sub-Nyquist. However, one can instead imagine a system with a lower order DM where the DM Nyquist frequency creates the speckle, in which case a super-Nyquist phase plate creates the spot. In this case, SNWFC is needed to null the speckle using the spot. But, the super-Nyquist phase plate for a lower order DM and our real DM spot sine wave have the same optical effect, and so for the purposes of this experiment we are still demonstrating SNWFC, but instead using less than the full DM to represent the Nyquist region.

- **speckle**: Shown in purple in Figure 4.2, the speckle is at a slightly lower frequency than the spot, again sub-Nyquist in this experiment, and is meant to represent speckle noise hiding the signal from a planet at the **point of interest** (Figure 4.2).

- **anti-speckle**: Shown in red in Figure 4.2, the goal of this experiment is to copy the anti-speckle sine wave from the spot so that electric field in the focal plane at the point of interest is minimized, allowing the off axis planet light at that point to then be seen. Unlike the speckle intensity, the planet light is not removed because it is incoherent with the stellar light.

The frequency, position angle, amplitude, and phase of the spot and speckle are user-defined parameters. In order to cancel the speckle at the point of interest, the anti-speckle frequency and position angle are already determined geometrically based on the user-defined spot and speckle parameters. The phase and amplitude of the anti-speckle are then determined iteratively using the following methodology and procedure:

1. Using the image plane intensity, which is approximately a measure of amplitude squared, the anti-speckle amplitude is then approximately

$$a_{\text{anti-speckle}} = (f) a_{\text{speckle}} \sqrt{\frac{\text{im}_{\text{spot}}}{\text{im}_{\text{speckle}}}} \tag{4.2}$$

   in an image, im, with only two sine waves, the spot and speckle, where a is the sine wave amplitude. The fudge parameter, f is equal to 1 in this step and explained further below in step 3.

2. The anti-speckle phase is unknown, so we simply loop through -π to π in phase on the anti-speckle in order to find the phase that minimizes image plane intensity at the point of interest.

3. Using $f = 1$ in equation 4.2 is only a rough estimate for $a_{\text{anti-speckle}}$ because we are using a linear approximation of the wavefront (wavefront = $a\ e^{i\,\phi} \sim a(1+i\phi)$, so $|\text{wavefront}|^2 \sim O(a^2)$) to null the speckle, neglecting the higher order terms. So, to correct for these higher order terms we add a fudge factor as necessary in the anti-speckle amplitude, where $0.5 \lesssim f \lesssim 2$.

In simulations (§4.3.2) and in the lab (§4.3.3), finding a precise optimal anti-speckle phase is done iteratively, first finding a rough estimate, and then using a finer grid spacing in phase around that rough estimate to get a more precise value. The same iterative procedure is applied in finding the amplitude fudge factor. Throughout the remainder of §4.3.1, we calculate contrast in a given image normalized to the peak value in that same image, since we do not use a coronagraph in this laboratory experiment.

4.3.2 Simulations

We ran simulations of two different sine wave configurations: one where the point of interest lies on a bright Airy ring of the on-axis PSF, and one where the point of interest lies on a nearby dark ring. The initial and final parameters for both simulations are shown in Table 4.1. We use an image size of 2048×2048 pixels, beam ratio of 4, and a circular greyscale pupil (instead of a binary mask, the pupil mask is mean binned 10×10 from an original binary mask of image size of 20480×20480 to prevent numerical pixelation effects).

Table 4.1: Initial simulation and final parameters for our three sine wave deterministic speckle nulling procedure for both our dark ring and Airy ring simulations. The spot and speckle parameters are set initially, while the anti-speckle phase and amplitude are determined iteratively via the focal plane wavefront sensing scheme described in §4.3.1.

sine wave	α **(degrees)**	**cpa**	**phase (radians)**	**amplitude (radians)**
Airy ring simulation				
spot	1.0	4.0	0	0.13
speckle	28.7	2.95	0	0.064
anti-speckle	135.0	2.0	-2.57	0.94
dark ring simulation				
spot	1.0	4.0	0	0.13
speckle	35.0	2.57	0	0.064
anti-speckle	142.0	2.4	0.84	0.85

The results of our simulations after anti-speckle phase and amplitude iterations are shown in Figure 4.3. Figure 4.6a shows the results after only phase iterations, to

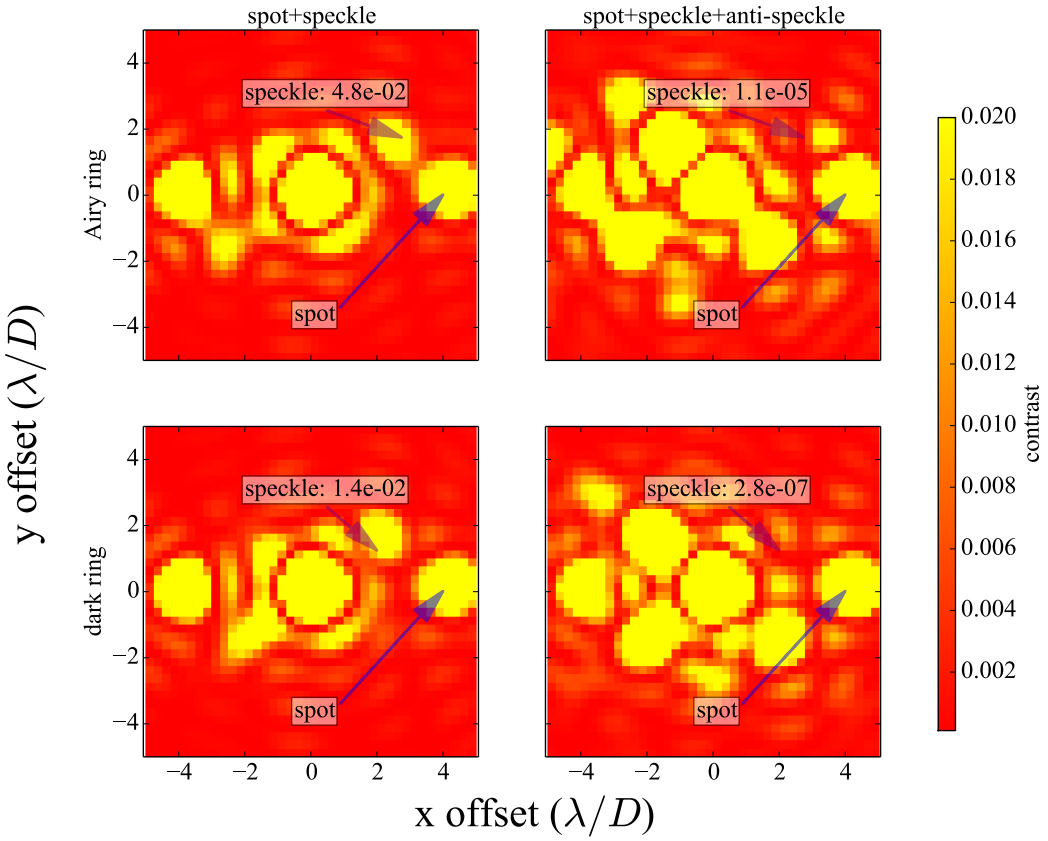


Figure 4.3 Images of our Airy ring (first row) and dark ring (second row) deterministic speckle nulling simulations. The left column shows the initial configuration with only two sine waves (spot+speckle), and the right column shows the configuration with all three sine waves after reaching the minimum anti-speckle phase and amplitude.

be compared later with our laboratory results in §4.3.3. Our two simulations show that changing the point of interest from a bright Airy ring to a nearby dark ring changes the achievable contrast as well as the absolute anti-speckle phase. The final contrasts reached after anti-speckle amplitude iterations (equation 4.2) for the Airy ring and dark ring simulations are 1.1×10^{-5} and 2.8×10^{-7}, respectively. After amplitude iterations for the Airy ring simulation we found $f = 1.786$, whereas for the dark ring simulation we found $f = 1.004$. The fudge factor results are expected, since higher order terms in the Taylor expansion of the PSF fall off more quickly in the dark ring compared to the Airy ring (Perrin et al., 2003).

4.3.3 Laboratory Results

We use a simple SCAO setup in the Adaptive Optics Laboratory at the National Research Council, Astronomy and Astrophysics (NRC), shown in Figure 4.4. We use a monochromatic 655 nm fiber-fed laser diode as a light source, an 11×11 actuator ALPAO DM (circular pupil, 97 total actuators), a 32×40 subaperture Imagine Optic HASO SHWFS, ultimately using a 29 subaperture diameter to remove edge effects.

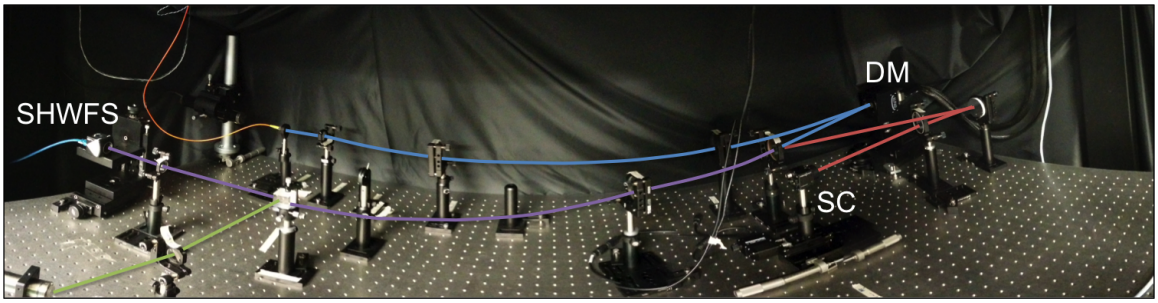


Figure 4.4 The NRC Adaptive Optics Laboratory wavefront sensor bench. The common path is in blue, the SHWFS path is in purple, the non-common path to the science camera (SC) is in red, and a separate pyramid WFS path (not fully shown) is in green.

We apply sine waves to the DM in a closed loop procedure, using the expected SHWFS slopes for the desired sine wave amplitude, frequency, and phase, described below. The x and y SHWFS slopes for a DM sine wave at a subaperture, i (using 1 indexing), are the mean values of the derivative of the wavefront phase along x and y, respectively:

$$\begin{aligned} x_{\text{slope}}(i) &= \left\langle \frac{\partial \Phi}{\partial x} \right\rangle_i \\ &= \frac{\int_{x_i-d/2}^{x_i+d/2} \int_{y_i-d/2}^{y_i+d/2} \frac{\partial \Phi}{\partial x} dxdy}{\int_{x_i-d/2}^{x_i+d/2} \int_{y_i-d/2}^{y_i+d/2} dxdy}, \\ y_{\text{slope}}(i) &= \left\langle \frac{\partial \Phi}{\partial y} \right\rangle_i \\ &= \frac{\int_{x_i-d/2}^{x_i+d/2} \int_{y_i-d/2}^{y_i+d/2} \frac{\partial \Phi}{\partial y} dxdy}{\int_{x_i-d/2}^{x_i+d/2} \int_{y_i-d/2}^{y_i+d/2} dxdy}, \end{aligned} \tag{4.3}$$

where Φ is the wavefront phase, d is the size of one subaperture ($d = D_{\text{pup}}/n_{\text{subapertures}}$, where D_{pup} and $n_{\text{subapertures}}$ are the diameter and number of subapertures across the

SHWFS pupil, respectively), and (x_i, y_i) is the center of subaperture i in units of pupil diameter[3]. After calibrating a flat wavefront with the reference slopes from closing the loop on zeros, a sine wave on the DM imparts a wavefront phase shift of

$$\Phi = a \sin\left(\frac{2\pi}{D_{\text{pup}}}\text{cpa}(x\cos(\alpha) - y\sin(\alpha)) - \phi\right), \tag{4.4}$$

where a, α, and ϕ are the sine wave the amplitude, position angle, and phase, respectively. Combining equations 4.3 and 4.4 allows us to apply a sine wave on the DM in closed loop at any amplitude, frequency (sub-Nyquist), position angle, and phase.

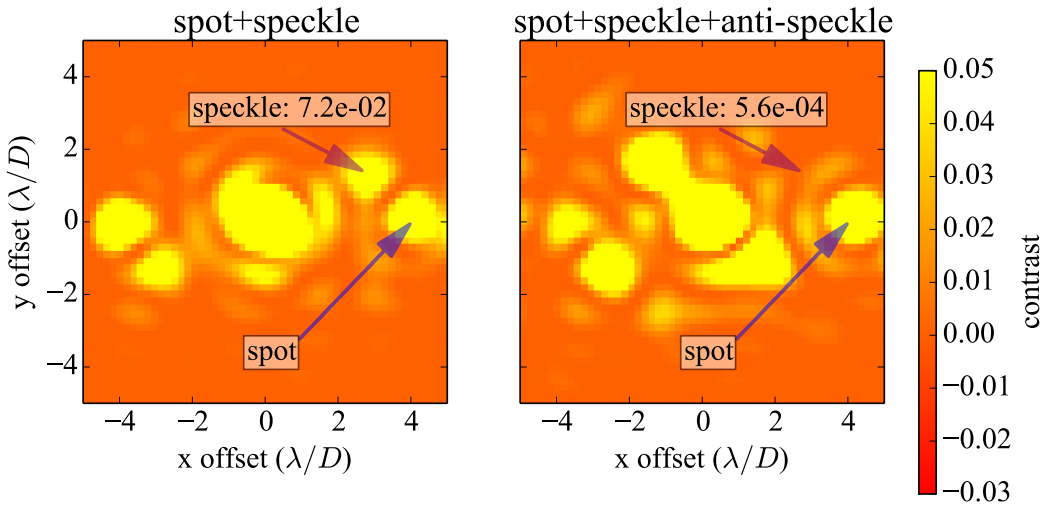
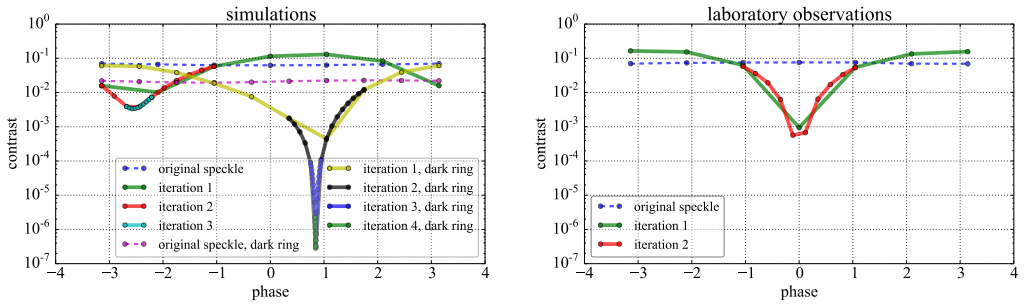


Figure 4.5 Laboratory science camera images from our deterministic SNWFC experiment, using the Airy ring simulation parameters from Table 4.1. The left image shows the initial speckle contrast for the spot+speckle DM sine wave combination, and the right image shows the spot+speckle+anti-speckle combination at the minimum anti-speckle phase of $\phi = -0.12$. The images are background/dark-subtracted, hence the negative contrast scale.

Using this methodology and the same parameters as our Airy ring simulation in Table 4.1 (but not the anti-speckle phase and amplitude, which we again determine

[3]Using this coordinate system, the lower left corner of the lower left subaperture in use is $(x_1, y_1) = (0, 0)$ and the upper right corner of the upper right subaperture is $(x_{n_{\text{subapertures}}}, y_{n_{\text{subapertures}}}) = (D_{\text{pup}}, D_{\text{pup}})$.



(a) (b)

Figure 4.6 Speckle contrast plotted vs. anti-speckle phase (radians) at the point of interest. (a) Noiseless simulations for our SNWFC laboratory experiment, showing both the replicated laboratory setup, where the speckle location lies on an Airy ring, and a second simulation showing the effect of moving the speckle location into a nearby dark ring. (b) Results from our laboratory data. The minimum contrast and phase are both in between our two simulations in (a), suggesting that noise on the bench is moving the speckle location closer to a dark ring from the original Airy ring location.

iteratively), we implemented the deterministic speckle nulling procedure described in 4.3.1, but after two phase iterations we reached the SC detector noise floor at the point of interest, and so we did not do any amplitude iterations. The before and after images comparing the spot+speckle with the minimum spot+speckle+anti-speckle are shown in Figure 4.5, and our laboratory phase iterations are shown in Figure 4.6b.

We found a final contrast at the point of interest of 5.6×10^{-4}, using $a_{\text{anti-speckle}} = 88$ nm and $\phi_{\text{anti-speckle}} = -0.12$ (from equations 4.2 and 4.4, respectively).

Comparing the final anti-speckle phase and contrast from our lab results in Figure 4.6b to our simulation results in Figure 4.6a suggests that in the lab, the speckle lies somewhere in between a bright and dark ring. Despite our laboratory setup identical to the Airy ring simulation, we suspect that noise on the bench distorts the on-axis Airy ring locations so that the speckle is located near a dark ring, suggested by the deeper contrast and closer absolute phase shift for the dark ring simulation in Figure 4.6a.

4.4 Simulations Using the Self-Coherent Camera

Although our deterministic speckle nulling procedure in §4.3 was designed to show that SNWFC is possible experimentally, we would like in the future to further test SNWFC by creating a more realistic dark hole as opposed to nulling a single speckle. In the absence of infrastructure for more stable lab conditions, a higher order DM, and a super-Nyquist phase plate, we instead run simulations to represent the possible performance on a future ELT high contrast imaging instrument.

The SCC, first developed by Baudoz et al. (2006) and evolving most recently to the design and algorithm in Mazoyer et al. (2014), is a wavefront control method to remove speckles in the focal plane (Baudoz et al., 2006; Mazoyer et al., 2014). Use of an off-axis hole placed in the Lyot stop causes fringes in the focal plane, but only from the stellar light. Isolating these fringed speckles is done using an amplitude mask on the complex-valued optical transfer function (OTF), yielding a complex-valued image, I^-, which contains phase and amplitude information of only the fringed stellar speckles. By constructing a set reference images in I^- (i.e., an interaction matrix) composed of sines and cosines (each with zero phase) at every λ/D interval within a symmetric half-Nyquist region (offset by $0.5\lambda/D$ to preserve symmetry), a least squares fit to an aberrated target image (e.g., with phase and amplitude errors in a sub-Nyquist half-DH region) yields the negative sine and cosine amplitude coefficients to apply to the DM such that the fringed speckles in the focal plane are minimized. See Mazoyer et al. (2014) and references therein for a more detailed description of the SCC wavefront control algorithm.

We choose to use the SCC wavefront control method instead of, e.g., electric field conjugation (EFC; Give'On et al., 2007), because of the algorithm timescale to converge. It is shown experimentally in in Mazoyer et al. (2014) that the SCC algorithm converges immediately after one iteration[4], and only relies on the time required to compute the target image correlation vector (i.e., after the initial daytime calibration has been done to compute and invert the interaction matrix). In contrast, EFC, even algorithmically, takes tens of iterations to converge, as in Give'On et al. (2007). With the already calibrated interaction matrix, the SCC timescale to

[4]We note that Mazoyer et al. (2014) do apply a second iteration after the first dark hole is achieved by recalibrating a new interaction matrix, offset with the coefficients already obtained from the initial least squares. We do not perform this second calibration in our simulations, as this is likely done to correct for laboratory limitations that we do not consider in our simulations, such as DM hysteresis, influence functions, fitting error, etc.

compute the target image correlation vector, sine and cosine amplitude coefficients, and resulting corrected image is ∼30 seconds on our 2.7 GHz Mac desktop using one core, but this could ultimately run much faster with a fully parallel code.

The main component of super-Nyquist wavefront control using the SCC, or with any other wavefront control method, requires an optical element that creates a PSF copy outside of the DM Nyquist region. In Thomas et al. (2015), they consider the use of (1) influence functions that are narrow relative to the DM actuator pitch, such as the known manufacturing residual print-through pattern on many DMs, and (2) a mild transmissive diffraction grating (i.e., at a super-Nyquist frequency), and perform SNWFC simulations using option (1). Here, we consider the use of a super-Nyquist sine wave phase plate, initially proposed in Marois et al. (2006) and Sivaramakrishnan et al. (2006) for astrometric and photometric calibration purposes (Marois et al., 2006b; Sivaramakrishnan & Oppenheimer, 2006).

We use a 60 nm amplitude sine wave phase plate to create the SCC reference array described above and in Mazoyer et al. (2014), thus copying every sine and cosine from the sub-Nyquist region to the analogous super-Nyquist region. The construction of these reference images and the corresponding least-squares covariance matrix, was done without noise, using 21 nm amplitude sine waves (Lafrenière et al., 2007a). This interaction matrix procedure could be implemented on an instrument during daytime calibrations, provided the conditions are stable between daytime calibration and night time observations.

Similar to Marois et al. (2012), our target image contained a 40 nm rms phase screen wavefront error and 1% amplitude errors, both with a -2 power law in spatial frequency (Marois et al., 2012). To simulate wavefront error, we first use Gaussian random numbers between 0 and 2π in phase and a -1 power law in amplitude (which makes a -2 power law in intensity) combined into a complex array; we then Fourier transform that array, then take the real part, and finally normalize the standard deviation to the desired wavefront error amplitude. To simulate the effect of an AO system, we removed the first 21 Zernike polynomials within the DM Nyquist region, using a least squares on the target image with a set of Zernike reference images. We used a 32 by 32 actuator DM, beam ratio of 4 pixels, image plane dimensions of 1012 by 1012, wavelength of 1.65 μm, and sine wave phase plate cpa of 1.1 (N_{act}), placing the super-Nyquist region just beyond the sub-Nyquist region. We used the GPI aperture and apodized Lyot coronagraph (APLC) design (Soummer et al., 2006). To determine the separation and diameter of the off-axis hole in the Lyot stop that

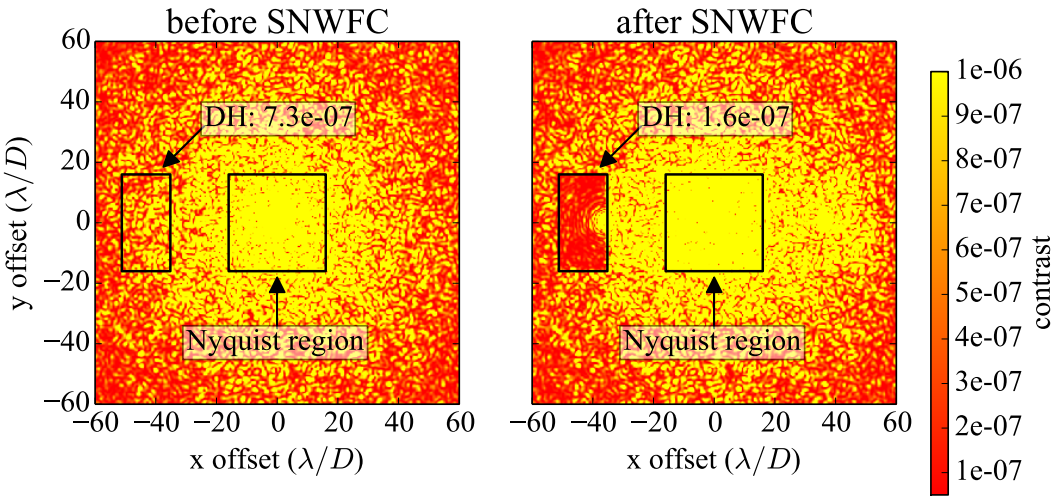


Figure 4.7 Before (left) and after (right) SCC SNWFC simulations, using 40 nm rms wavefront error, 1% amplitude error, 60 nm amplitude super-Nyquist phase plate (using cpa = $1.1 * N_{\mathrm{act}}$), a 32×32 DM, and an APLC. The contrast improves by a factor of ∼5 after dark hole correction.

makes the SCC, we used the equations from Mazoyer et al. (2014) ($\epsilon_0 > 1.5\ D_{\mathrm{pupil}}$, and $D_R < (1.22\sqrt{2}/N_{\mathrm{actuators}})D_{\mathrm{pupil}}$, respectively; Mazoyer et al. 2014).

Our simulation results are shown in Figure 4.7. The DH contrast in the raw image is 7.3×10^{-7}, and in the SCC image it goes down to 1.6×10^{-7}, a gain of ∼5 in contrast. We chose a half DH region left of the on-axis PSF, on the left side of the super-Nyquist region, but more generally the phase plate position angle and choice of which side of the spot to build a half DH can be set by the user, if, e.g., the planet location is already known. Images are converted to units of contrast by normalizing the values in the coronagraphic image to the peak value in the initial non-coronagraphic image. To calculate contrast in the super-Nyquist DH-region we take the robust standard deviation (Dowell et al., 2012a) in the contrast-normalized image within an aperture that removes 1 λ/D off each edge of the half-DH (to remove edge effects) and 5 λ/D radially around the super-Nyquist spot (to remove brighter diffraction effects from the spot that are not present in the raw image without a spot).

4.5 Conclusion

SNWFC allows already known and new exoplanets to be directly imaged at a higher SNR than is otherwise possible with the current AO design for ELTs. Our main conclusions in testing this technique are as follows:

- We have demonstrated in the lab that a deterministic SNWFC speckle nulling scheme increases performance, consistent with simulations.
- We demonstrated in simulation that SNWFC performance improvement is possible using the SCC after only one iteration, which is of particular interest to wavefront control on shorter timescales for ground-based telescopes.

There is still a lot of unexplored parameter space to get a better understanding of how well SCC-based SNWFC would perform on a telescope. Some additional factors in our simulations that are beyond the scope of this initial exploratory paper but worth further research are performance in polychromatic light, using a segmented pupil, using different coronagraphs and/or apodization, dependence on higher super-Nyquist phase plate frequencies, dependence on using an inner and outer scale in chosen wavefront error power law, dependence on a higher zero pad sampling for the off-axis SCC Lyot hole, and additional improvement based on diffraction suppression techniques (Pueyo et al., 2010). Ultimately, the next step before this technique can be tested on-sky is to demonstrate it in the lab.

Chapter 5

Conclusions and Future Work

In this thesis I have demonstrated a number of new techniques in both wavefront control and PSF subtraction in application to the direct imaging of exoplanets. These two approaches to improving contrast are complimentary to one another; the former is executed in real time during an observing run, while the latter can be used to further suppress the noise after the images are already taken. Although my work on wavefront control is based on a laboratory demonstration and simulations, wavefront control algorithms are constantly evolving towards on-sky use, especially for future ExAO space telescopes. In this scenario, real-time focal plane wavefront sensing and post-processing PSF subtraction techniques will both be necessary to reach the best achievable contrasts.

On the post-processing side, I have developed a new algorithm (1) for space telescope imaging of close-to-face-on disks and (2) in order to reach better contrasts near the inner working angle of an ExAO system. My main conclusions from each project are listed below:

1. My new OSFi algorithm is designed to recover the expected low spatial frequency (close-to-face-on) HR 8799 debris disk with *HST*/STIS. I found

   (a) a recovered over-luminosity, but after a number of false-positive heuristics I determined that this was from an instrumental focus evolution and/or spectral bandpass effect.

   (b) Using this non-detection as an upper limit, I determine an upper limit on the HR 8799 planetesimal belt mass in optical grains, consistent with other known debris disk systems.

2. I have developed a new PSF subtraction technique, optimized to improve contrast near the inner working angle of the Gemini Planet Imager (GPI). This technique shows no improvement compared to existing methods, suggesting that such methods are already optimized in this regime.

On the wavefront control side, I have tested the a new focal plane wavefront sensing technique called super-Nyquist wavefront control (SNWFC), designed to image wide-orbit exoplanets on future ELTs that would otherwise not be possible with a conventional AO system. I have demonstrated that SNWFC can improve contrast

1. in a deterministic optical laboratory experiment, using a DM, wavefront sensor, and a science camera, and

2. in simulation, using the self-coherent camera (SCC) focal plane wavefront sensing technique (Baudoz et al., 2006).

Laboratory

In the future, a number of further improvements are possible. For the GPI PSF subtraction algorithm, a number of additional free parameters have not yet been explored and/or optimized, including the forward model throughput discrepancy, sensitivity to the input spectral type, dependence on different optimization/subtraction region geometries, and the SVD cuttoff. Ultimately this algorithm should be run to process the entire GPIES campaign dataset, for which there are many less correlated datasets that may show much better performance improvement than on the GPI 51 Eri dataset. The next steps in testing SNWFC with the SCC would be to simulate broadband performance and ultimately demonstrate this technique with an ExAO laboratory setup. Despite the recent detection of the HR 8799 debris disk with ALMA (Booth et al., 2016), detection with additional *HST* observations is still unlikely due to focus stability and bandpass effects, although imaging this debris disk system in the optical/NIR should be possible on future, more stable space telescopes, such as JWST and WFIRST. Thus, since the stability of HST represents the best current instrument to image face-on disks on $\sim$arcminute scales, this non-detection is an important result, showing astronomers that this is the best we can do with current instrumentation.

Bibliography

ALMA Partnership, Brogan, C. L., Pérez, L. M., et al. 2015, The Astrophysical Journal Letters, 808, L3

Andersen, D. 2014, Lecture notes for directed studies in adaptive optics: `http://www.astro.uvic.ca/~sara/A511/`

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 2002, Astronomy and Astrophysics, 382, 563

Barman, T. S., Macintosh, B., Konopacky, Q. M., & Marois, C. 2011, The Astrophysical Journal, 733, 65

Basri, G., & Brown, M. E. 2006, Annual Review of Earth and Planetary Sciences, 34, 193

Baudoz, P., Boccaletti, A., Baudrand, J., & Rouan, D. 2006, in IAU Colloq. 200: Direct Imaging of Exoplanets: Science and Techniques, ed. C. Aime & F. Vakili, 553–558

Boccaletti, A., Lagage, P.-O., Baudoz, P., et al. 2015, Publications of the Astronomical Society of the Pacific, 127, 633

Booth, M., Jordán, A., Casassus, S., et al. 2016, Monthly Notices of the Royal Astronomical Society: Letters, Advance Access, arXiv:1603.04853

Bordé, P. J., & Traub, W. A. 2006, The Astrophysical Journal, 638, 488

Boss, A., Lecavelier des Etangs, A., Mayor, M., et al. 2012, Transactions of the International Astronomical Union, Series A, 28, 138, `http://home.dtm.ciw.edu/users/boss/C53TR.pdf`

Boss, A. P. 2000, The Astrophysical Journal Letters, 536, L101

Boss, A. P., Basri, G., Kumar, S. S., et al. 2003, in IAU Symposium, Vol. 211, Brown Dwarfs, ed. E. Martín, 529

Bowler, B. P. 2016, ArXiv e-prints, arXiv:1605.02731

Bowler, B. P., Liu, M. C., Dupuy, T. J., & Cushing, M. C. 2010, The Astrophysical Journal, 723, 850

Broekhoven-Fiene, H. 2012, Master's thesis, University of Victoria, Canada

Burrows, A., Hubbard, W. B., Lunine, J. I., & Liebert, J. 2001, Reviews of Modern Physics, 73, 719

Choquet, É., Pueyo, L., Hagan, J. B., et al. 2014, in Proceedings of the SPIE, Vol. 9143, Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave, 914357

Choquet, É., Perrin, M. D., Chen, C. H., et al. 2016, The Astrophysical Journal Letters, 817, L2

Clampin, M. 2010, STIS Coronagraphic Imaging of the Kuiper Belt Surrounding the HR 8799 Planetary System., HST Proposal, ,

Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, Publications of the Astronomical Society of the Pacific, 120, 531

Currie, T., Debes, J., Rodigas, T. J., et al. 2012, The Astrophysical Journal Letters, 760, L32

Dohnanyi, J. S. 1969, Journal of Geophysics Research, 74, 2531

Dong, S., DePoy, D. L., Gaudi, B. S., et al. 2006, The Astrophysical Journal, 642, 842

Dowell, J., Wood, D., Stovall, K., et al. 2012a, Journal of Astronomical Instrumentation, 1, http://fornax.phys.unm.edu/lwa/doc/lsl-0.4.x/robust.html

—. 2012b, Journal of Astronomical Instrumentation, 1, http://fornax.phys.unm.edu/lwa/doc/lsl-0.4.x/robust.html

Draper, Z. H. 2014, Master's thesis, University of Victoria, Canada

Draper, Z. H., Duchêne, G., Millar-Blanchaer, M. A., et al. 2016, ArXiv e-prints, arXiv:1605.02771

Ducati, J. R. 2002, VizieR Online Data Catalog, 2237

Espinoza, P., Su, K., Rieke, G., & Stapelfeldt, K. 2011, in Bulletin of the American Astronomical Society, Vol. 43, American Astronomical Society Meeting Abstracts #217, 339.01

Galicher, R., Baudoz, P., Rousset, G., Totems, J., & Mas, M. 2010, Astronomy and Astrophysics, 509, A31

Galicher, R., Marois, C., Zuckerman, B., & Macintosh, B. 2013, The Astrophysical Journal, 769, 42

Gerard, B., Lawler, S., Marois, C., et al. 2016a, The Astrophysical Journal, 823, 149

Gerard, B. L., & Marois, C. 2016, in Proceedings of the SPIE

Gerard, B. L., Marois, C., & GPIES team. 2016b, in Proceedings of the SPIE

Ginsburg, A., Robitaille, T., Parikh, M., et al. 2013, 10.6084/m9.figshare.805208.v2

Give'On, A., Belikov, R., Shaklan, S., & Kasdin, J. 2007, Optics Express, 15, 12338

Grady, C. A., Proffitt, C. R., Malumuth, E., et al. 2003, Publications of the Astronomical Society of the Pacific, 115, 1036

Graham, J. R., Macintosh, B., Doyon, R., et al. 2007, ArXiv e-prints, arXiv:0704.1454

Gray, R. O., Corbally, C. J., Garrison, R. F., et al. 2006, The Astronomical Journal, 132, 161

Grenier, S., Baylac, M.-O., Rolland, L., et al. 1999, Astronomy and Astrophysics, Supplement, 137, 451

Guyon, O. 2003, Astronomy and Astrophysics, 404, 379

Hernandez, S., & et al. 2012, Space Telescope Imaging Spectrograph Instrument Handbook for Cycle 21 v. 12.0

Herriot, G., Andersen, D., Atwood, J., et al. 2014, in Proceedings of the SPIE, Vol. 9148, Adaptive Optics Systems IV, 914810

Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, Astronomy and Astrophysics, 355, L27

Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2005, Icarus, 179, 415

IAU. 2006, https://www.iau.org/static/resolutions/Resolution_GA26-5-6.pdf

Ingraham, P., Marley, M. S., Saumon, D., et al. 2014, The Astrophysical Journal Letters, 794, L15

Jura, M., Ghez, A. M., White, R. J., et al. 1995, The Astrophysical Journal, 445, 451

Kalas, P., & Jewitt, D. 1995, The Astronomical Journal, 110, 794

Kennedy, G. M., & Wyatt, M. C. 2014, Monthly Notices of the Royal Astronomical Society, 444, 3164

Kern, B., Guyon, O., Kuhnert, A., et al. 2013, in Proceedings of the SPIE, Vol. 8864, Techniques and Instrumentation for Detection of Exoplanets VI, 88640R

Kirkpatrick, J. D., Gelino, C. R., Cushing, M. C., et al. 2012, The Astrophysical Journal, 753, 156

Konopacky, Q. M., Thomas, S. J., Macintosh, B. A., et al. 2014, in Proceedings of the SPIE, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 914784

Krist, J. E., Hook, R. N., & Stoehr, F. 2011a, in Proceedings of the SPIE, Vol. 8127, Optical Modeling and Performance Predictions V, 81270J

Krist, J. E., Hook, R. N., & Stoehr, F. 2011b, in Proceedings of the SPIE, Vol. 8127, Optical Modeling and Performance Predictions V, 81270J

Krivov, A. V. 2010, Research in Astronomy and Astrophysics, 10, 383

Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É. 2007a, The Astrophysical Journal, 660, 770

Lafrenière, D., Doyon, R., Marois, C., et al. 2007b, The Astrophysical Journal, 670, 1367

Lagrange, A.-M., Gratadour, D., Chauvin, G., et al. 2009, Astronomy and Astrophysics, 493, L21

Lagrange, A.-M., Bonnefoy, M., Chauvin, G., et al. 2010, Science, 329, 57

Lawson, C., & Hanson, R. J. 1987, Society for Industrial and Applied Mathematics, http://docs.scipy.org/doc/scipy-0.17.0/reference/generated/scipy.optimize.nnls.html

Lecavelier Des Etangs, A., Deleuil, M., Vidal-Madjar, A., et al. 1995, Astronomy and Astrophysics, 299, 557

Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, Proceedings of the National Academy of Science, 111, 12661

Macintosh, B., Graham, J. R., Barman, T., et al. 2015, Science, 350, 64

Madhusudhan, N., Burrows, A., & Currie, T. 2011, The Astrophysical Journal, 737, 34

Maire, J., Ingraham, P. J., De Rosa, R. J., et al. 2014, in Proceedings of the SPIE, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 914785

Marley, M. S., Fortney, J. J., Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2007, The Astrophysical Journal, 655, 541

Marley, M. S., & Robinson, T. D. 2015, Anual Review of Astronomy And Astrophysics, 53, 279

Marley, M. S., Saumon, D., Cushing, M., et al. 2012, The Astrophysical Journal, 754, 135

Marois, C., Correia, C., Galicher, R., et al. 2014, in Proceedings of the SPIE, Vol. 9148, Adaptive Optics Systems IV, 91480U

Marois, C., Doyon, R., Racine, R., & Nadeau, D. 2000, Publications of the Astronomical Society of the Pacific, 112, 91

Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006a, The Astrophysical Journal, 641, 556

Marois, C., Lafrenière, D., Macintosh, B., & Doyon, R. 2006b, The Astrophysical Journal, 647, 612

—. 2008a, The Astrophysical Journal, 673, 647

Marois, C., Macintosh, B., Barman, T., et al. 2008b, Science, 322, 1348

Marois, C., Macintosh, B., & Véran, J.-P. 2010a, in Proceedings of the SPIE, Vol. 7736, Adaptive Optics Systems II, 77361J

Marois, C., Véran, J.-P., & Correia, C. 2012, in Proceedings of the SPIE, Vol. 8447, Adaptive Optics Systems III, 844726

Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T. 2010b, Nature, 468, 1080

Matthews, B., Kennedy, G., Sibthorpe, B., et al. 2014, The Astrophysical Journal, 780, 97

Max, C. 2016, Lecture notes for Astr 289: Adaptive Optics: `http://www.ucolick.org/~max/289/`

Mazoyer, J., Baudoz, P., Galicher, R., & Rousset, G. 2014, Astronomy and Astrophysics, 564, L1

Müller, S., Löhne, T., & Krivov, A. V. 2010, The Astrophysical Journal, 708, 1728

Nielsen, E. L., Close, L. M., Biller, B. A., Masciadri, E., & Lenzen, R. 2008, The Astrophysical Journal, 674, 466

Perrin, M. D., Sivaramakrishnan, A., Makidon, R. B., Oppenheimer, B. R., & Graham, J. R. 2003, The Astrophysical Journal, 596, 702

Perrin, M. D., Maire, J., Ingraham, P., et al. 2014, in Proceedings of the SPIE, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 91473J

Perryman, M. 2011, The Exoplanet Handbook (Cambridge University Press)

Pickles, A. J. 1998, Publications of the Astronomical Society of the Pacific, 110, 863

Pueyo, L., Wallace, K., Troy, M., et al. 2010, in Proceedings of the SPIE, Vol. 7736, Adaptive Optics Systems II, 77362A

Quanz, S. P., Crossfield, I., Meyer, M. R., Schmalzl, E., & Held, J. 2015, International Journal of Astrobiology, 14, 279

Quanz, S. P., Goldman, B., Henning, T., et al. 2010, The Astrophysical Journal, 708, 770

Racine, R., Walker, G. A. H., Nadeau, D., Doyon, R., & Marois, C. 1999, Publications of the Astronomical Society of the Pacific, 111, 587

Rameau, J., Chauvin, G., Lagrange, A.-M., et al. 2013a, Astronomy and Astrophysics, 553, A60

—. 2013b, The Astrophysical Journal Letters, 779, L26

—. 2013c, The Astrophysical Journal Letters, 772, L15

Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, Astronomy and Astrophysics, 558, A33

Rouan, D., Riaud, P., Boccaletti, A., Clénet, Y., & Labeyrie, A. 2000, Publications of the Astronomical Society of the Pacific, 112, 1479

Sadakane, K. 2006, Publications of the Astronomical Society of Japan, 58, 1023

Sadakane, K., & Nishida, M. 1986, Publications of the Astronomical Society of the Pacific, 98, 685

Saumon, D., & Marley, M. S. 2008, The Astrophysical Journal, 689, 1327

Schneider, G., Grady, C. A., Hines, D. C., et al. 2014, The Astronomical Journal, 148, 59

Schneider, J., Dedieu, C., Le Sidaner, P., Savalle, R., & Zolotukhin, I. 2011, Astronomy and Astrophysics, 532, A79

Sibthorpe, B., Vandenbussche, B., Greaves, J. S., et al. 2010, Astronomy and Astrophysics, 518, L130

Sivaramakrishnan, A., & Oppenheimer, B. R. 2006, The Astrophysical Journal, 647, 620

Smith, B. A., & Terrile, R. J. 1984, Science, 226, 1421

Soummer, R., & Aime, C. 2004, in Proceedings of the SPIE, Vol. 5490, Advancements in Adaptive Optics, ed. D. Bonaccini Calia, B. L. Ellerbroek, & R. Ragazzoni, 495–503

Soummer, R., Aime, C., Ferrari, A., et al. 2006, in IAU Colloq. 200: Direct Imaging of Exoplanets: Science and Techniques, ed. C. Aime & F. Vakili, 367–372

Soummer, R., Pueyo, L., & Larkin, J. 2012, The Astrophysical Journal Letters, 755, L28

Soummer, R., Sivaramakrishnan, A., Oppenheimer, B. R., et al. 2009, in Proceedings of the SPIE, Vol. 7440, Techniques and Instrumentation for Detection of Exoplanets IV, 74400R

Soummer, R., Perrin, M. D., Pueyo, L., et al. 2014, The Astrophysical Journal Letters, 786, L23

Sparks, W. B., & Ford, H. C. 2002, The Astrophysical Journal, 578, 543

Spergel, D., Gehrels, N., Baltay, C., et al. 2015, ArXiv e-prints, arXiv:1503.03757

Spiegel, D. S., & Burrows, A. 2012, The Astrophysical Journal, 745, 174

Stahler, S. W., Shu, F. H., & Taam, R. E. 1980, The Astrophysical Journal, 241, 637

Steck, D. A. 2015, Classical and Modern Optics, available online at `http://steck.us/teaching` (revision 1.5.2, 28 June 2015)

Stoehr, F., Durand, D., Haase, J., & Micol, A. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 155

Su, K. Y. L., Morrison, S., Malhotra, R., et al. 2015, The Astrophysical Journal, 799, 146

Su, K. Y. L., Rieke, G. H., Stapelfeldt, K. R., et al. 2009, The Astrophysical Journal, 705, 314

Thomas, S., Belikov, R., & Bendek, E. 2015, The Astrophysical Journal, 810, 81

Tokovinin, A. 2005, Adaptive optics tutorial at CTIO, `http://www.ctio.noao.edu/~atokovin/tutorial/`, ,

Tsiganis, K., Gomes, R., Morbidelli, A., & Levison, H. F. 2005, Nature, 435, 459

van Leeuwen, F. 2007, Astronomy and Astrophysics, 474, 653

Vyacheslav. 2014, https://habrahabr.ru/post/205480/

Wang, J. J., Rajan, A., Graham, J. R., et al. 2014, in Proceedings of the SPIE, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 914755

Wyatt, M. 2009, Graduate lecture course on planetary systems: https://www.ast.cam.ac.uk/~wyatt/lecture7_debrisdiskmod.pdf

Zuckerman, B., Song, I., Bessell, M. S., & Webb, R. A. 2001, The Astrophysical Journal Letters, 562, L87