LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-220665

Planet Formation Instrument for the Thirty Meter Telescope

Bruce Macintosh, Mitchell Troy, James Graham, Rene Doyon

April 18, 2006

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Document No. TMT.IAO.CDD.006.005.REL01

Planet Formation Instrument for the Thirty-Meter Telescope

Feasibility Study Final Report

Lawrence Livermore National Laboratory
Jet Propulsion Laboratory
University of Montreal
University of California, Berkeley

February 15, 2006

Page Intentionally Left Blank

CONTENTS

Tables

Figures

Revision History

Date	Rev.	Revised By	Description	Approval
02-15-06	1.23	Bruce Macintosh	Release version	Bruce Macintosh

List of Acronyms

ADC	Analog-to-digital converter
ALMA	Atacama Large Millimeter Array, a planned synthesis radio telescope that will operate at millimeter and submillimeter wavelengths
AO	Adaptive Optics, the practice of detecting and correcting for dynamic wavefront errors.
APS	Alignment Phasing System, the system that will phase the segments of TMT
BMC	Boston Micro-machines Corporotation, a company which produces, among other products, MEMS DMs
CCD	Charge Coupled Device, an electronic unit used in astronomy for detecting photons across some area.
CFHT	Canada-France-Hawaii Telescope, a 3.6 meter telescope
CoDR	Conceptual Design Review
DM	Deformable mirror.
DRP	Data reduction pipeline
DSS	Diffraction Suppression System. For this report, the baseline DSS is a double-linear-shear nuller.
EUVL	Extreme Ultraviolet Lithography. A demonstration project to produce high-quality aspheric optics by a consortium including Tinsley Labs and LLNL
ExAO	Extreme Adaptive Optics -- another term for high-contrast imaging
ExAOC	Original name for an ExAO system for the Gemini Telescope
FFT	Fast Fourier Transform, an efficient algorithm for computing discrete Fourier transforms
FOV	Field of view
FPGA	Field-programmable gate array
FQPM	Four-quadrant phase mask, a method of suppressing the light of an on-axis source
FT	Fourier transform
FWHM	Full-width at half-maximum, a measure of the width of a function or distribution's peak
GPA	Gray Pixel Approximation. A method of constructing a grayscale telescope transmission via its analytical Fourier transform
GPI	Gemini Planet Imager: New name of the Gemini ExAO system
GSMT	Giant Segmented-Mirror Telescope. Typically, a large telescope whose primary mirror is composed of multiple mirrored segments.
HST	Hubble Space Telescope
IFS	Integral field spectrograph
IFU	Integral field unit
IRAS	Infrared Astronomical Satellite, a joint project of the US, UK and the Netherlands
IRIS	Infrared Imaging Spectrometer, an instrument for TMT

IWA	The smallest angle at which a given contrast is to be achieved. For PFI, the nominal IWA is 3 lambda/D, or about 34 mas
JPL	Jet Propulsion Laboratory
JWST	James Webb Space Telescope
LBT	Large Binocular Telescope, a binocular telescope under construction consisting of two 8.4-meter mirrors on a common mount
LLNL	Lawrence Livermore National Laboratory
LW	Frequency range from 3 to 5 μm
M1	Primary mirror.
M2	Secondary mirror.
M3	Tertiary mirror.
mas	Milliarcsecond.
MEMS	Micro-electrical mechanical systems
MIRI	Mid InfraRed Instrument, an infrared camera and spectrometer for the JWST
MWI	Mount Wilson Institute
Myr	Million years.
NACO	Adaptive optics instrument on the UT4 telescope of the VLT.
NCA	National Committee for Astronomy
NFIRAOS	Narrow Field Infrared Adaptive Optic System, an instrument for TMT
NICMOS	Near Infra-Red Camera and Multi-Object Spectrometer, an instrument on the HST
OAP	Off-axis paraboloid
OCDD	Operational Concept Definition Document
OVC	Optical vortex coronagraph, a relatively new proposed method of suppressing the light of an on-axis source
pc	Parsec. One parsec is ~3.26 light years
PCB	Printed Circuit Board
PCS	a Shack-Hartmann type wavefront sensor, which is permanently mounted at the left bent Cassegrain focal station of the Keck Telescope
PFI	Planet Formation Instrument. The proposed TMT instrument about which this report is written.
PM	Primary mirror -- see M1
PSD	Spectral power density - the portion of a signal's power (energy per unit time) falling within given frequency bins.
PSF	Point-spread function. The response of a telescope to an incoming wavefront.
PTI	Palomar Testbed Interferometer, originally built as a technology testbedfor advanced telescope technologies
P-V	Peak - Valley, a statistic describing the maximum range of a function or distribution
RFP	Request for proposal.
RMS	Root-mean-square, a statistic describing an expected deviation from zero of a function or

	distribution
S-H	Shack-Hartmann, a wavefront sensor acting as a phase gradient detector using a lenslet array
SIM	The Planetquest Space Interferometry Mission
SM	Secondary mirror -- see M2
SNR	Signal to noise ratio
SRD	Science requirements document
SW	Frequency range from 1.1 to 2.4 μm
TMT	Thirty Meter Telescope. A project to construct an extremely large telescope based on more than 700 hexagonal-shaped mirror segments that stretch a total of 30 meters in diameter.
TPF	Terrestrial Planet Finder.
TRIDENT	Near-infrared camera based on the SSDI concept at CFHT
VLT	An array of four 8-meter telescopes which can work independently or in combined mode
WCAM	Wavefront camera
WFE	Wavefront error
WFS	Wavefront sensor
WH	Warping harness.
ZPC	Zernike phase contrast, a wavefront sensor

1 INTRODUCTION AND DOCUMENT OVERVIEW

In the closing years of the 20th Century humankind began its exploration of the planetary systems in the solar neighborhood. Precision radial velocity measurements have now yielded the discovery of over 160 planets. Direct imaging of these planets, as opposed to detection of the effects of orbital motion on their parent star, is now feasible, and the first young planet in a wide orbit may have been detected using adaptive optics systems. Gemini and the VLT are building the first generation of high contrast adaptive optics systems, which deliver planet-imaging performance within few Airy rings of the host star. These systems will make the first surveys of the outer regions of solar systems by detecting the self-luminous radiation of young planets. These instruments will establish whether Jovian planets form predominantly through "top-down" (global gravitational instability) or "bottom-up" (core accretion) processes. The 8-m "extreme" AO systems cannot see close enough to the host stars to image Doppler planets, and they cannot reach the relatively distant, young clusters and associations where planets are forming. The Planet Formation Instrument will use the nearly four-fold improved angular resolution of TMT to peer into the inner solar systems of Doppler-planet bearing stars to yield a unified sample of planets with known Keplerian orbital elements and atmospheric properties. In star formation regions, where T Tauri stars (young solar type stars) are found in abundance, PFI can see into the snow line, where the icy cores of planets like Jupiter must have formed. Thus, TMT will be the first facility to witness the formation of new planets.

1.1 PFI Science Mission

Indirect planet searches are biased because they detect orbital motion caused by a companion planet. Only after the completion of one or more orbits can a secure detection be claimed. Thus, the median semimajor axis for the current sample of Doppler planets is < 1 AU. Kepler's third law enforces diminishing returns because the search domain grows only as $t^{2/3}$, where t is the duration of the survey. At small semimajor axis separations AO systems are blinded by the glare of the host star, and on 8-m telescopes only young planets (< 1 Gyr) beyond 5 AU are discovered. Moreover, chromospheric activity excludes young (< 2 Gyr) host stars from the Doppler surveys. Thus, Doppler and AO imaging techniques select disjoint samples of exoplanets. Higher angular resolution is required to unify these populations. The atmospheres of exoplanets, with effective temperature between that of Jupiter (126 K) and the T dwarfs ($T_{\rm eff}$ > 800 K) remain the last frontier of classical stellar atmospheres, and include such *terra incognita* as the H_2O and NH_3 cloud condensation. Low-resolution exoplanet spectra ($\lambda/\Delta\lambda$ < 50 & 500) will measure $T_{\rm eff}$ and composition as well as yield an independent measure of surface gravity. A unified sample is crucial because only planets with Doppler detections and dynamical mass measurements will have precisely known surface gravity. When log g and $T_{\rm eff}$ are known it is then possible to measure the age of the planet—it is likely that the most reliable ages for main sequence stars will be determined from the ages of their planets!

The core-accretion model is the dominant paradigm for planet formation—and at least one Jovian planet, Jupiter, formed by core accretion. Whereas, opacity-limited gravitational fragmentation will generally result in minimum masses of 3–9 M_J, the accumulation of icy cores consisting of 5–15 $M_\oplus$ is a cornerstone of the core-accretion scenario. Neptune-mass exoplanets have been discovered using the Doppler technique, offering what appears to be

a significant confirmation of the core-accretion hypothesis. Nonetheless, the core accretion still faces outstanding difficulties. For example, it is unknown how core accretion can make Jovian-mass planets so fast. Reconstructing how planets form from Doppler surveys, which show us the circumstances after billions of years of dynamical evolution, presents a monumental challenge. We propose to address planet formation directly by observing systems that are forming planets now. There are a handful of young systems in the solar neighborhood, e.g., the 20 Myr-old TW Hya association stars that might show recently formed planets. However, to catch run away gas accretion in action we must find targets that are an order of magnitude younger: class II and class III young stellar objects, which are only found in abundance in dark clouds such as Taurus and Ophiucius. Thus, PFI has four core science missions:

1. Broad and systematic studies of the extrasolar planet population in the solar neighborhood, spanning semimajor axis separations $0.5 < a/\mathrm{AU} < 50$ and exoplanet mass $0.5 < M/M_J < 12$.
2. Imaging very young planets (0–15 Myr) in the process of forming or migrating at > 100 pc distances
3. High-SNR studies of planetary atmospheres and their astrophysics
4. The studies of circumstellar disks ranging from young protoplanetary disks through debris disks to high-density extrasolar zodiacal debris in inner solar systems.

1.2 PFI System Overview

The scientific goals of PFI lead to extremely challenging technical requirements. Current AO systems achieve contrasts on the order of 10^{-5} at angular separations of ~ 1 arc second; TMT PFI requires a three order of magnitude improvement in contrast and a factor of twenty in angular separation. Some of the improvement of course comes from the larger telescope aperture, but equally important is the design of an AO system and instrument dedicated to high contrast imaging, with precise and accurate control of optical wavefronts. The combined requirements of high dynamic range, a wide variety of target brightnesses, very high angular resolution and the need to minimize systematic errors, lead to a multi-stage integrated instrument with each subsystem serving a well-defined role in controlling a particular aspect of the problem.

The high-speed **front AO system** (Section 3) is optimized for searching for planets orbiting nearby field stars. This requires achieving extremely high contrast ($>10^{-8}$) on bright targets, which in turn requires very high update rates (2–4 kHz) to minimize dynamic atmospheric errors and efficient use of the available photons. To achieve this, we have selected as our baseline a variant of the pyramid wavefront sensor run in a quasi-interferometric mode (Section 3.2.3). This takes advantage of the high Strehl ratio at the wavefront sensing wavelength to achieve measurement errors a factor of 2-4 better than a conventional Shack-Hartmann sensor. Combined with a compact, high-order MEMS DM this system will produce *H*-band Strehl ratios above 0.9 on bright stars and 0.84 down to *I*=9 mag. On dimmer stars, it will provide partial wavefront correction while the back IR wavefront sensor produces the final image.

Table 1: PFI Requirements from the OCDD. Speckle suppression processing of IFS data cubes is expected to increase contrast by a factor of ~10 for methane-dominated planets

Property	Requirement	Goal	PFI value
Contrast (I < 8 mag.) @ IWA	10^{-8} @ 50 mas		$2x10^{-8}$ without speckle suppression
Contrast (I < 8 mag.) wide angle		10^{-9}@ 100 mas	
Contrast (H<10 mag.) @ IWA	10^{-6} @ 30 mas	2 10^{-7}@ 30 mas	$1x10^{-6}$ without speckle suppression
Plate scale	Nyquist @ H	Nyquist @ J	5.5 mas
Field of View (radius)	0.7 arc sec	2 arc sec	2 × 2 arc sec
Spectral resolution, full FOV	50	100	70
Spectral resolution, partial FOV	500	1000	700
Wavelength range	1–2.5 μm	1–4 μm	1–5 μm
Imaging polarimetry	Simultan. two channel		SDC
Sensitivity (1 hr., 5-σ)	H = 27 mag.	H = 32 mag.	

Unless controlled, light scattered by diffraction from the telescope pupil would completely swamp the signal from a planet. After considerable exploration of alternatives using simulations, we have selected a dual-stage shearing nulling interferometer or "nuller" as the **diffraction suppression system** (DSS), described in Section 4. This combines four offset and phase-shifted copies of the telescope pupil to remove the uniform component of the electromagnetic field that causes diffraction. This has two major advantages over conventional coronagraph architectures. First, it allows for very small inner working angles (IWA)—as small as 3 λ/D—that are needed both to detect nearby planets in reflected starlight and to image young solar systems at distances as great as 150 pc. Second, it is robust against highly obscured pupils such as currently proposed for TMT.

The large aperture of TMT allows PFI to exceed the contrast of 8-m ExAO systems by an order of magnitude, but does nothing to relax the fundamental requirement of sub-nanometer internal static optical errors; since the effect of internal errors is independent of telescope size, these requirements are even more exacting to reach the 10^{-8} contrast levels that we are targeting. To overcome this, we will use a dedicated, interferometric, infrared wavefront sensor that is tightly integrated with the DSS, and controls a second MEMS DM located inside the nuller. Known as the **back wavefront sensor** (Section 5), this system is a Mach-Zehnder interferometer combining the bright and dark outputs of the nulling DSS. It has two primary purposes. First, operating at the science wavelength and measuring the dark output of the DSS, it will provide sub-nanometer absolute accuracy and correct these errors through feedback to the front AO or through its own MEMS. Second, especially on very red science targets, it will provide additional rejection of atmospheric turbulence, allowing PFI to reach contrast of 10^{-6} to 10^{-7} on $H = 10$ mag. young stars.

PFI's science instrument is a dedicated **Integral Field Spectrograph** (Section 7) optimized for high-contrast imaging, maximum scientific return, and for spectral follow-up of extrasolar planets. This combines moderate spectral resolution ($R = \lambda/\Delta\lambda \sim 70$) with a 2 × 2 arc second Nyquist-sampled field of view and spectral coverage from 1–5 μm. Carefully designed to minimize chromatic errors, this allows planets to be distinguished from artifact speckles by their differing behavior as a function of wavelength. The instrument also includes a $R = 700$ mode that can be used for follow-up of previously-detected planets. A dual-channel imaging polarimeter mode will be available to study circumstellar disks and distinguish disk structure from the planets embedded in them.

Figure 1 shows a block diagram of the PFI system.

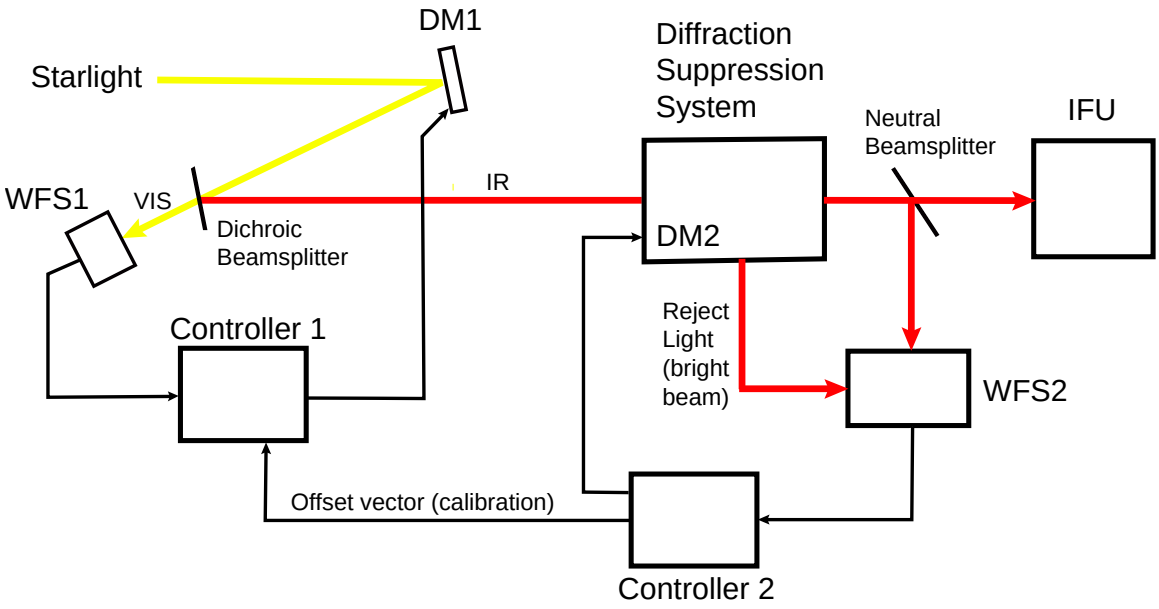

Figure 1: Simplified block diagram of the PFI system.

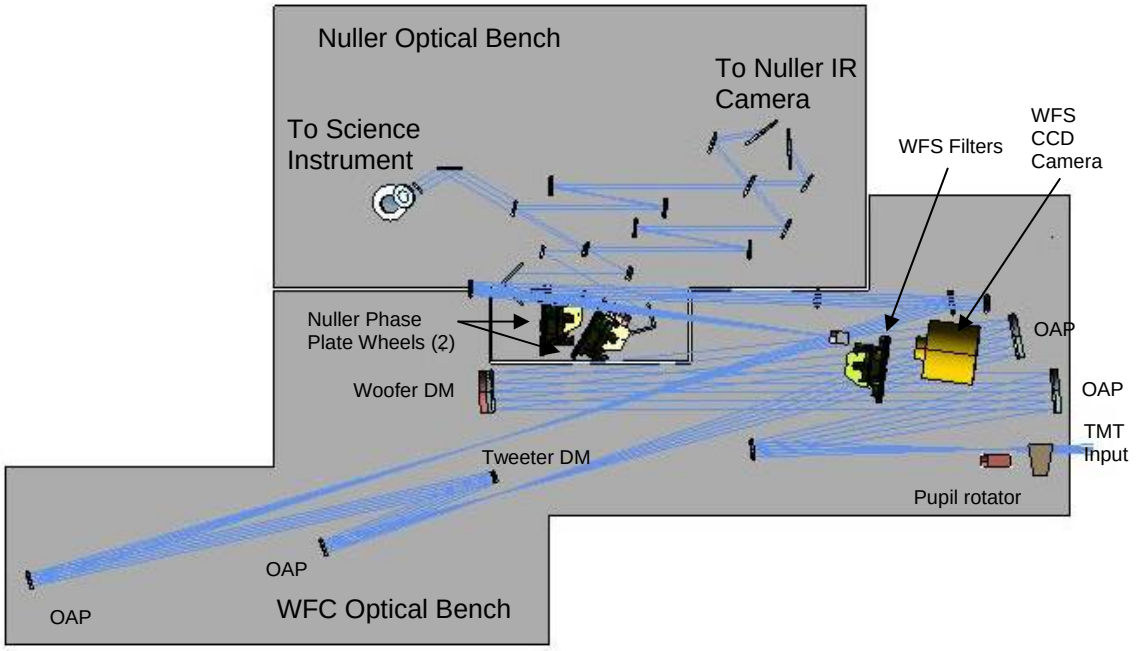

Figure 2: Top-view CAD rendering of the PFI optical bench. Long dimension is 4.5m

One of the most significant areas of concern for PFI—or any ExAO instrument on TMT—is the telescope itself. We have carried out simulations for a variety of telescope configurations, reflecting the evolution of thinking about M1 segments and warping harnesses. Errors on M1 are shown to be the greatest concern, both amplitude errors (which PFI may be designed to correct) and phase errors. Whiffletree, segment warping, and segment piston/tip/tilt errors, as currently specified, will limit PFI performance in hour-long exposures at < 0.1 arc second. We propose a series of upgraded specifications for TMT that will allow PFI and TMT to meet the science goals (Figure 3).

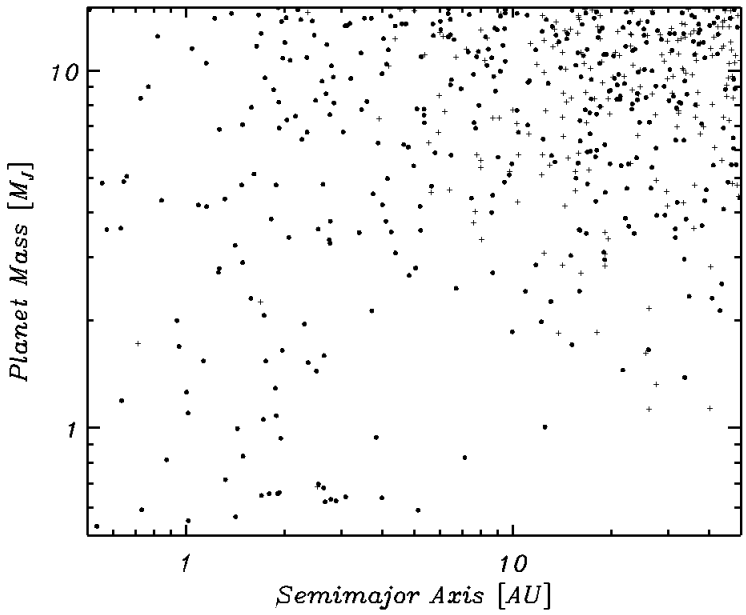

Figure 3: Plot of the planets discovered in Monte Carlo simulation of a PFI field-star survey of the solar neighborhood. Planets discovered in a similar GPI survey (with different Monte Carlo seed) are marked with crosses.

1.3 Feasibility Study Overview

This document represents a feasibility study, not a conceptual or preliminary design. We view it as an existence proof for a planet detection instrument for TMT that meets our science-driven requirements, but by no means has been fully defined or engineered. In the next five years the area of ExAO technology and planet-detection science will progress by almost two orders of magnitude as the first high-contrast AO systems are deployed on the Gemini and VLT telescopes. With the lessons learned from these systems, TMT can expect to progress by almost as much again, but it inevitable that those lessons will lead to modifications to the design presented here.

Given finite time and resources, we have concentrated on those areas where basic questions have to be answered as well as those most relevant to the current state of the telescope design. Other areas are deferred to the conceptual design phase. For example, the supervisory software and interfaces to the observatory are a key part of any science instrument. However, the details of such software are not relevant at this phase—the cost of such can be estimated from previous experience, and the exact details of the software architecture will depend on TMT standards and interfaces that have not yet been defined. Therefore, we do not discuss mid-level software in this document. By contrast, the issue of dealing with diffraction and wavefront control for a segmented telescope are fundamental to ExAO on TMT and hence is treated in considerable detail.

The document has 12 main sections:

Section 1: This introduction.

Section 2: First, we lay out the context for planet detection in the TMT era—as best we can forecast it ten years in advance—and identify four representative science mission, and the requirements they set on our instrument.

Section 3: Given those science requirements we present a trade study between different AO approaches and identify two high-efficiency wavefront sensor options. We lay out a strawman hardware architecture for the pyramid sensor, using near-term components, and an optical design.

Section 4: We present a similar study of the diffraction suppression system (DSS), with two architectures – a band-limited Lyot coronagraph and the nulling interferometer – studied in detail. We also briefly discuss several alternative coronagraphs.

Section 5: We discuss the second wavefront sensor, optimized for high accuracy and IR operation. Analytic modelling of several different approaches to this wavefront sensor shows that these approaches each have advantages for some science cases; we have selected the simplest and most flexible for end-to-end simulation.

Section 6: This section lays out the control algorithms needed for the two wavefront sensors and some preliminary notes on joint control. Efficiently using information from two cascaded wavefront sensors operating at different wavelengths is a complex problem, which will require further study in the context of specific science missions.

Section 7: Discusses the science instrument, including the down-selects and trades between different IFS types and field of view versus spectral resolution, then shows a representative optical and mechanical design.

Section 8: Gives a simple mechanical overview of the whole instrument, used to estimate the instrument envelope and weight, and lists mechanisms and rough power requirements.

Section 9: Discusses end-to-end simulations of PFI performance including all subsystems. Computational requirements limit these to relatively short exposures (< 4 seconds) but they have been used to verify that PFI meets its atmospheric goals

Section 10: Discusses the effect on PFI contrast of the telescope M1 and M3 mirrors. Errors on M1 are shown to be the greatest concern, both amplitude errors (which PFI may be designed to correct) and phase errors. Both whiffletree and segment-warping errors, as currently specified, will limit PFI performance compared to atmospheric errors. Considerable further analysis, in cooperation with the telescope project, is needed to determine how to specify a PFI-compatible primary mirror.

Section 11: Discusses possible alternative TMT ExAO capabilities and PFI architectures, including PFI de-scopes, enhancements, and early architectures based on NFIRAOS.

Finally, Section 12 discusses follow-on research that is being carried out or could be funded by the TMT project to ensure that the final TMT telescope is compatible with PFI's science goals and to prepare for TMT's conceptual design phase.

Information on the budget and schedule of PFI are in a separate document (TMT.IAO.CDD.06.003.REL01). We have also made available several appendices, including a discussion (extracted from GPI) of high-contrast PSFs and speckle attenuation, and the PFI OCDD and FPRD documents.

2 Science Review

The reconnaissance phase of exoplanetary systems is just beginning. Over 150 Jovian mass planets have been discovered in the inner (< 4 AU) regions of nearby F, G, and K stars (Marcy et al. 2006). Neptune mass planets have been found, which is significant, since the presence of these cores is one of the distinctive predictions of the classical core-accretion model for planet formation (Pollack et al. 1996). However, our own Jupiter would not have been found in the current radial velocity surveys, because its orbital period is too long. Of course the kinematic signature of rocky planets is too small to give a reliable detection. It is evident that a large volume of planet mass and orbital element phase space of remains to be investigated. The unexpected diversity of recently detected exoplanetary systems indicates that our understanding of planet formation is primitive. Although we have sufficient knowledge of planets and planet formation to design experiments that will answer well-posed questions, further exploration is also warranted.

Significant progress will be made during the design and construction phase of TMT. The detection of rocky planets will have to wait for improvements in ground-based Doppler searches, or for dedicated photometric eclipse-detection telescopes in space, while the first Jovian planets detected beyond 5 AU will be found by “extreme” adaptive optics (ExAO) systems on the current generation of 8-m telescopes. The ExAO systems such as the Gemini planet imager (GPI) are sensitive primarily to warm (> 400 K), self-luminous, young (< 1 Gyr) planets in wide (> 5–30 AU) orbits. On a 30-m telescope, ExAO systems will advance planet detection by another order of magnitude.

1. TMT can detect planets not just through their self-emission but also through reflected sunlight, opening up the study of mature planetary systems as old as our own.
2. TMT can probe planetary systems at scales less much less than 5 AU. These two factors combine to produce a sensitivity range that runs from the outer edges of solar systems well into the regime probed by Doppler searches, producing a uniform probe of current planetary populations.
3. TMT can detect self-luminous planets down to much lower masses and older ages than 8-m ExAO systems, producing robust statistics on the outer parts of planetary systems far beyond the limits of Doppler techniques
4. TMT can detect planets forming in circumstellar disks. Because these regions are distant (150 pc) only a telescope with TMT’s unmatched angular resolution can make these revolutionary observations.

This last area is particularly exciting. Astronomers are expert in reconstructing the origin and evolution of astrophysical systems from a limited number of examples. While these narratives are often compelling, they frequently fail when confronted with direct evidence. Thus, genuine understanding of the process of star formation only came with the ability to observe these events occurring. Infrared and mm observations, developed in the 1970s and 1980s were necessary to probe the veil of dust that obscures the youngest stars, and reveal the central role played by discs and outflows. We can now claim that a “standard model” describes the formation of single, solar type stars from a collapsing core in a molecular cloud (Shu, Adams, & Lizano 1987). In contrast, no such model exists for massive stars,

because the angular resolution of current facilities is only now approaching the scales necessary to record the physical processes at work in the vicinity of young O and B stars.

Doppler planet searches can only find planets orbiting old F, G, and K stars because young stars (< 2 Gyr) have active chromospheres, which add kinematic noise that swamps the radial velocity signature of planetary mass companions. The planet formation process probably takes less than 10 Myr and much of the dynamical evolution, including planet-planet and planet disc interactions occur over a span of only a few hundred million years (Yin et al. 2002; Ward 1997). As a consequence we currently have no means of watching planets form. Moreover, because of the short lifetimes relative to the Galactic star formation rate, young planet-forming systems are rare and found in significant numbers only in distant (> 150 pc) star forming clouds.

The angular resolution required to study planet formation *in situ* is therefore of order 7 milli arc seconds (1 AU at 150 pc). It is evident that the TMT will be the first facility to enable direct observation of planets emerging from their parent discs. Likewise, to detect nearby mature planets in reflected light, a comparable angular resolution will be needed (10 mas = 0.1 AU at 10 pc). TMT will thus also be the first facility to be able to directly detect a sizable number of reflected light Jovian planets. The unique combination of angular resolution and sensitivity of TMT will thus enable direct images and spectra to be obtained for both young and old planets (Figure 4). The main instrumental capabilities needed to take advantage of the TMT in this regard is high contrast imaging at an angular separation of a few λ/D from bright stars in the near infrared. This goal is the basic driver for the instrument described here, the Planetary Formation Instrument (PFI).

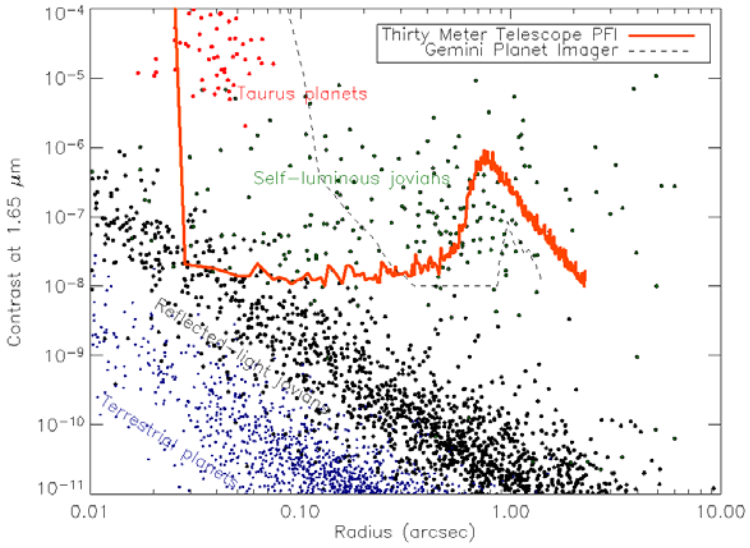

Figure 4: Contrast-separation plot for a Monte Carlo simulation of a variety of targets in the solar neighborhood. Blue dots are rocky planets, beyond the reach of even TMT. Black dots are mature Jovian planets reflecting sunlight. Green dots are self-luminuous Jovian planets, typically those with masses of 3-10 Jupiter masses and ages < 1 Gyr. Red dots are extremely young planets, recently formed or still accreting, in the Taurus starforming region. The expected sensitivity of PFI and the Gemini Planet Imager for a typical target are overlaid. Although the exact sensitivity is of course a function of individual star type, this provides a sense of the unique phase space accessible to TMT.

Planets are not the only emission sources associated with nearby solar systems. In both the young and old phases of solar system evolution, thermal and/or reflected light from dust disks is likely to be present, providing additional avenues for investigation of solar system origins and evolution. In the youngest systems, such as T Tauri stars, accretion disks can still be present, in slightly more evolved systems, dusty debris disks or rings can be found (Figure 5), and in mature systems, analogs to our own solar system's zodiacal disk and Kuiper belt are to be expected. The inclusion of debris disk stars is highly complementary to the planet search. In the ideal case, both a planet and a disk are detected—only in the unlikely case of an exactly edge-on disk could the disk itself hamper planet detection.

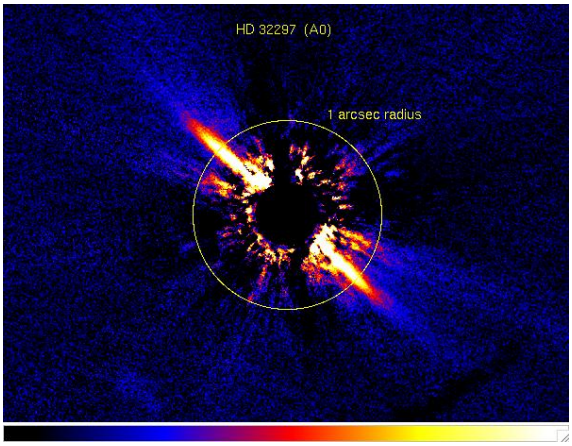

Figure 5: The first ground-based AO images of the edge on debris disk HD 32297. These H-band data were obtained using the Keck AO system (Kalas, Fitzgerald & Graham 2006). The disk extends to a distance of at least 400 AU (3.3″) along its major axis and the disk appears to be inclined at 79±3 degrees from a face-on viewing geometry. The disk exhibits unequal brightness on opposing sides and a break in the surface brightness profile along the NE-side disk major axis. Such asymmetries might implicate the existence of one or more (unseen) planetary mass companions.

2.1 The Astronomical Context for PFI/TMT

The development context for the PFI includes continued progress in ground-based extreme AO (ExAO) on 10-meter class telescopes as well as future space-based missions such as NASA's *Kepler*[1], the James Webb Space Telescope (JWST[2]), and the Space Interferometry Mission (SIM[3]), and potential coronagraphic and interferometric missions, such as the proposed Terrestrial Planet Finder (TPF). Like other indirect searches that depend on detection of orbital motion, *Kepler* and SIM will probe only the innermost regions (< 5 AU) of exoplanetary systems, although it is expected they will reveal the population of terrestrial mass planets in the habitable zone of solar-type stars. The imaging channel of the JWST mid-infrared instrument, MIRI, provides a 1.3 × 1.7 arcmin field of view and coronagraphs at 10.6, 11.4, 15.5, and 23 μm. Thus, JWST/MIRI will be a useful

[1] *Kepler* has a June 2008 launch date and a four-year lifetime. All information on NASA missions is quoted from the FY 2007 President's Budget Request Summary.
[2] The currently anticipated launch date for JWST is 2013, with a 5-year mission lifetime.
[3] SIM is slated for launch no earlier than 2015–16; it has a 5-year operational mission with a 10-year goal.

facility for tracing distribution and composition of hot dust in disks on angular scales of several arc seconds. In the near-infrared, where a 6-m space telescope could compete with ground-based ExAO systems, the wavefront control of JWST will not be sufficient to enable direct detection of exoplanets. While other space-based, coronagraphic or interferometric planet hunting missions have been proposed, such as the Terrestrial Planet Finder (TPF), the start of these missions has been deferred beyond NASA's current planning horizon (2011), and therefore their future is very uncertain.

Ground-based IR and mm interferometers will improve our understanding of circumstellar discs. Current or planned near-IR interferometers, e.g., the Palomar Testbed Interferometer (PTI), Keck Interferometer, or Large Binocular Telescope (LBT) lack the collecting area, baseline, or *uv* coverage for sensitive, high-fidelity, high-resolution imaging. When completed in 2011, the Atacama Large Millimeter Array (ALMA) will be the largest and most capable imaging array of millimeter telescopes in the world. ALMA will be the premier instrument for studying the cool universe, including the molecular gas and dust that constitute the building blocks of planetary systems. One of key science goals of ALMA is the ability to image the gas kinematics in protostars and in protoplanetary disks around young Sun-like stars at 0.1 arcsec angular resolution with a dynamic range of 1000:1 (Figure 6). At a distance of 150 pc (roughly the distance of the star-forming clouds in Ophiuchus or Corona Australis), this angular resolution should enable the study of their physical, chemical and magnetic field structures and perhaps even detect the tidal gaps created by planets undergoing formation in the disks. Thus, ALMA will provide vital information regarding the environment in which planets form. However, it is evident from simulations that the detection of tidal effects on disks caused by a newly formed Jovian planet will push ALMA to its limits (Wolf, Gueth, Henning & Kley 2002). TMT will be highly complementary, allowing detection of the planets themselves that are responsible for the gaps, which in turn enables measurements of mass, accretion rate, and orbital motion.

At a given wavelength, TMT's diffraction-limited angular resolution is three times better than that of an individual Keck telescope and five times better than that of JWST. Thus, in its development and initial operation timeline, the TMT can obtain access to companions much closer to their central stars than any other planned space-based or existing ground-based telescope. The current generation of large, ground-based telescopes has insufficient angular resolution to probe the planet forming environments in the nearest star-forming clouds. Thus, the pathfinder ExAO system, the Gemini Planet Imager, GPI, is designed to find relatively mature (100 Myr–1 Gyr) planetary companions in the solar neighborhood (< 50 pc). Although PFI can study GPI-discovered exoplanets at high SNR, these instruments are complementary because they can probe different phases of exoplanet and exosolar system evolution: PFI can push both to much younger systems at greater distances, and to nearby, fully mature systems (1–10 Gyr).

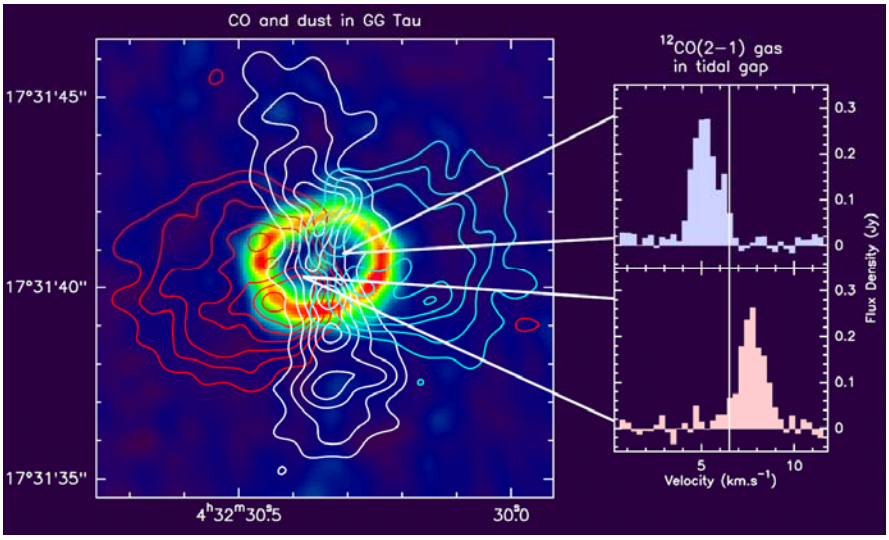

Figure 6: Millimeter observations of GG Tau, which is a 0.26-arcsec T Tauri binary (Guilloteau & Dutrey 2000). The circumbinary disk exhibits Keplerian rotation, with a cavity created by tidal truncation. ALMA will improve the angular resolution of such observations by an order of magnitude and increase the dynamics range to 1000:1. ALMA will provide detailed information on the physical conditions (temperature, composition, and particle size) under which planet formation occurs. Under favorable conditions, the 0.1 arcsec angular resolution may be sufficient to reveal the tidal gaps opened by newly formed Jovian planets.

To take advantage of the unique region of observational parameter space very close to bright stars made accessible by the TMT, PFI will provide a high contrast infrared imager with unprecedented contrast and inner working angles near bright stars, as well as an integral field spectroscopic capability. To provide high contrasts, PFI will provide a wavefront sensor, a diffraction suppression system, and an integral field spectrometer (IFS) for spectroscopic imaging. The core operating wavelength region of the PFI will be 1 to 2.5 µm, although an extension to longer wavelengths is a potential upgrade.

2.2 Science Drivers

Over the last two centuries, observations of our solar system have formed the basis for understanding planet formation within the context of the nebular disk hypothesis (Kant 1755; Laplace 1796). In the past two decades a revolution has unfolded as, first, IRAS detected dusty disks orbiting main sequence stars and then precision radial velocity measurements revealed the first exoplanets, which included an unimagined diversity of planetary systems. Doppler measurements are sensitive to close-in planets, while images of the dusty disks are often most plausibly interpreted as indicating, indirectly, the presence of a planet on a wide orbit. In the past year or so, with adaptive optics (AO), the first direct images have been obtained of objects of planetary mass in orbit around stars other than the Sun.

The core-accretion model for planet formation has been severely challenged by these observations, but remains the dominant paradigm (Pollack et al. 1996). New physical mechanisms, including global gravitational disk instabilities and migration are now important concepts in the planet formation debate. New observing modalities, principally direct imaging of exoplanets, are needed to make the next step in founding a general theory of planet formation and in characterizing the spectra of planets. Direct imaging of exoplanet systems will pose fundamental new challenges to our understanding of the origin and evolution of planetary systems. Detection of radiation from planetary atmospheres will open broad new vistas of the physics and chemistry of planetary interiors and atmospheres.

2.2.1 Direct Detection of Extrasolar Planets

Direct detection of Jovian planets via reflected starlight requires a contrast of $\sim 2\times10^{-9}(a/5\mathrm{AU})^{-2}$ relative to their parent star. Reflected light exo-Jupiters in ~ 5 AU orbits cannot be detected by current space telescopes or ground-based AO systems, but are legitimate targets for TMT. Because of the inverse square law, reflected light searches are impractical way to explore the outer (10–30 AU) regions of solar systems. However, the self-luminous near-IR emission offers the option of detecting energy radiated by a hot planet itself, which is independent of *a*, except when irradiation by the parent star dominates at very small semi-major axis. Old planets are cool and dim, but young planets are hot and therefore bright in the infrared relative to their parent star. For example, at 1.6 µm it is possible to detect a 10 Myr-old 3 M_J planet, or a 100 Myr-old 7 M_J planet, orbiting a G2V star at a contrast ratio of only 4×10^{-6} (13.5 mag., see Figure 7).

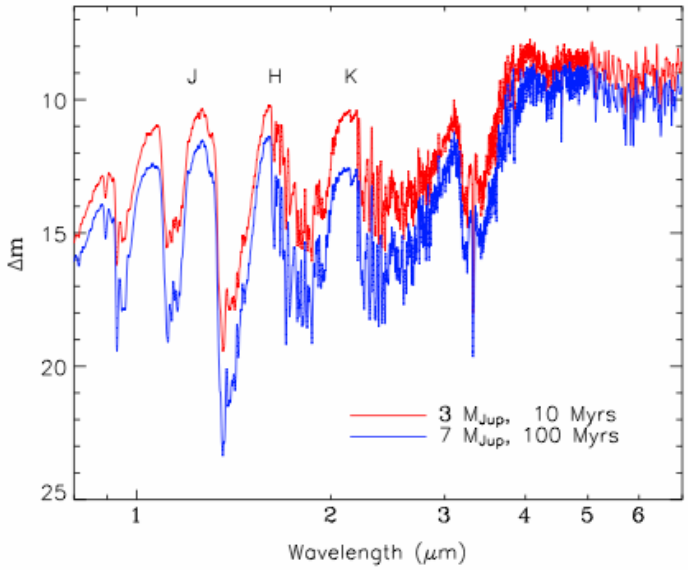

Figure 7: Planet-to-star contrast in magnitudes versus wavelength for a 10-Myr-old 3 MJ planet and a 100-Myr-old 7 MJ planet orbiting a G2V star at 5 AU (Barman et al. 2001). At visible wavelengths, planets shine only by their reflected light. However, young exoplanets are detectable by their intrinsic luminosity. Note the strong emission at J (1.25 µm) and H (1.65 µm) where the low background permits sensitive operation. We calculate these spectra using fully self-consistent models with the PHOENIX atmosphere code and the evolutionary tracks (Baraffe et al. 2003).

The first direct imaging of a companion of planetary mass was the VLT/NACO and HST/NICMOS observations of the brown dwarf 2M1207 and its companion at 0.8 arcsec (47 AU projected separation, Figure 8, left panel, Chauvin et al. 2004, 2005b; Song et al. 2005). At the likely age (8 Myr) of this system in the TW Hydrae Association, near-IR photometry implies a mass of 4 M_J based on evolutionary models of young planets. Second-epoch VLT and HST observations confirm that the companion is gravitationally bound to the primary. This planet was detectable because of its extreme youth and the low mass (25 M_J) of its brown dwarf primary. Given its small mass ratio, the 2M1207 system may have formed by a mechanism that is fundamentally different from that which forms the planets of main sequence stars. Nonetheless, these results give tremendous impetus to direct planet imaging.

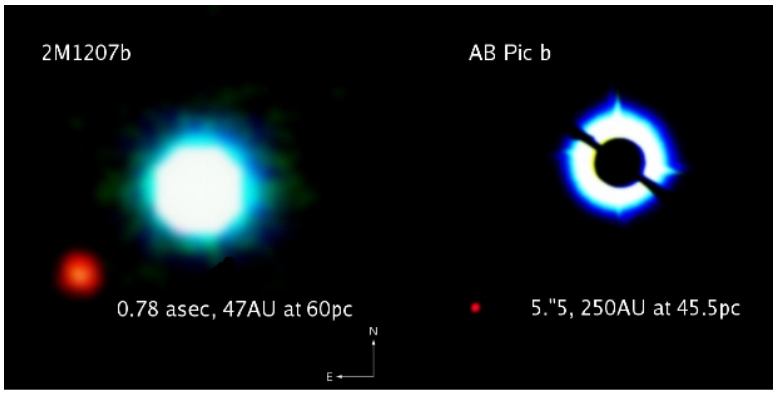

Figure 8: Left: NACO/VLT image of the brown dwarf 2M1207 and its estimated 4 MJ planetary companion (Chauvin et al. 2004; Song et al. 2005. Right: NACO coronagraphic image of AB Pictoris, a 30 Myr old K-type star in the Tucana/Horologium Association (Chauvin et al. 2005a). The mass of AB Pic b, seen in the lower left, lies very near the boundary between planets and brown dwarfs. This image illustrates that young, planet mass objects are bright enough to be detected around young stars and may have large semi-major axes.

2.2.2 The Classical Core-Accretion Theory of Planet Formation

The standard theory for the formation of gas giant planets is the core-accretion model (e.g., Pollack et al. 1996), which begins with dust particles colliding and growing within a protoplanetary disk to form ice/rock planetary cores. If the core becomes massive enough while gas remains in the disk, it continues to grow by gravitational accretion of this gas (Figure 9). Gas giants accrete most of the gas within their tidal reach filling the Hill sphere around them with a hot, extended, gaseous envelope. Further accretion is slowed by the dwindling supply of local raw materials and by the extended envelope, leading to growth times of 5–10 Myr. This predicted planet formation time is uncomfortably long compared to the observed ~ 3 Myr lifetime of protoplanetary disks. Two factors may alleviate this time-scale problem: inward migration can bring giant planets to fresh, gas-rich regions of

the disk; dust sedimentation may reduce atmospheric opacity, which leads to more rapid escape of accretion luminosity and shrinkage of the envelope. Hubickyj et al. (2005) have shown that reducing the grain opacity to a level observed in L dwarfs (Marley et al. 2002) reduces the planet growth time scale to ~ 1 Myr, well within the lifetime of protoplanetary disks. Once accretion stops, the planet enters the isolation stage and the planet contracts and cools to the present state at constant mass.

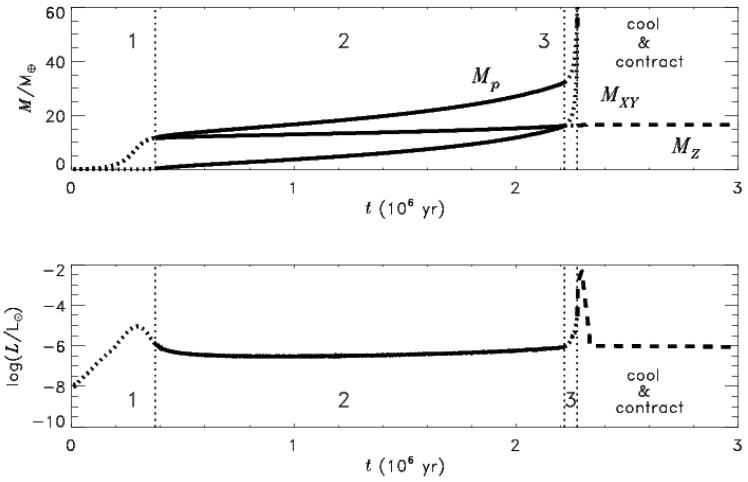

Figure 9: The mass and luminosity vs. time for the formation of a 1 MJ planet (Hubickyj et al. 2005). Phase 1—Dust particles (MZ) form planetesimals that accrete into a solid core surrounded by a low-mass gaseous envelope of H and He (MXY). Initially, solid runaway accretion occurs, and the gas accretion rate is much lower than that of solids. As the solid material in the feeding zone is depleted, the solid accretion rate is reduced; Phase 2—The gas accretion rate steadily increases and eventually exceeds the solid accretion rate. The total protoplanet mass (MP) grows as the gas accretes steadily. The solid core also grows but at slower rate; Phase 3—Runaway gas accretion occurs and the protoplanet grows rapidly. Accretion is limited only by the rate at which the nebula can transport gas to the vicinity of the planet. Accretion is stopped by either the opening of a gap in the disk as a consequence of accretion and the tidal effect of the planet, or by dissipation of the nebula. Once accretion stops, the planet is isolated, and it contracts and cools at constant mass.

The configuration of the protoplanet at the end of phase 3 (Figure 9) represents the initial condition for subsequent cooling and contraction. Fortney et al. (2005, astro-ph/0510009) have conducted preliminary calculations that describe the cooling and contraction of a young planet as it emerges from its disc. The results are shown in Figure 10. The implication of Fortney et al.'s results is that giant planets formed by the core accretion-gas capture mechanism are less luminous *post-accretion* than had been previously appreciated. A significant amount of energy is radiated away in the formation process. The fully formed planet has a smaller radius at young ages than "hot start" models predict (Burrows et al. 1997; Baraffe et al. 2003), leading to a lower post-formation luminosity. There are two significant observational consequences: 1) there is a period of very high luminosity ($\sim 10^{-2}\ L_\odot$) which lasts ~ 40,000 yr; 2) the initial conditions for evolution models are not "forgotten" for tens of millions of years. These factors imply that observations of class II (0.5–3 Myr) and III (< 20 Myr) young stellar objects afford the opportunity the probe the planet formation event. The run-away accretion spike is likely to be broader and fainter

than in these idealized calculations because of gradual accretion across the gap that the protoplanet forms, and the probability of witnessing this event in a typical T Tauri star may be much larger than a few percent.

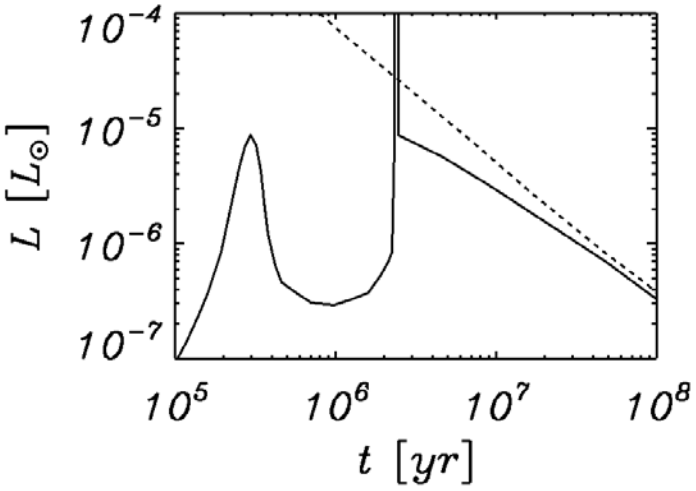

Figure 10: Luminosity vs. time for a 2 MJ planet (Fortney et al. 2005). The thick solid curve includes the effects of core accretion-gas capture. The planet is fully formed at 2.2 Myr. The dashed curve shows the simple cooling track of Baraffe et al. (2003). The full-width at half-maximum of the accretion luminosity spike is ~ 40,000 years. In reality, the spike is likely to be less sharp because of gradual accretion across the gap that the protoplanet forms.

The calculations of the early evolution of planets are provisional. Simulations for more massive cases are underway, and initial results suggest that the timescales for relaxation are longer for more massive planets. Despite the preliminary nature of these results, one message is clear from Figure 10: the luminosity of young exoplanets encodes key information about how they were formed; extrapolating the standard Burrows et al. (1997) or Baraffe et al. (2003) cooling tracks to < 10 Myr may introduce significant (×2–3) errors.

2.2.3 Alternate Pathways to Planet Formation and Planet Migration

Detecting newly formed planets is a key science objective of PFI. However, an equally compelling objective is to image the outer area of solar systems is to sample the zone where Jovian planets are thought to form and quantify the greatest distance out to which giant planets can form. The location of the region of interest depends on at least two competing factors: time-scales for planet building and the availability of raw material. Dynamical and viscous time scales in the disk are shorter at small radii, while for typical surface density laws the amount of mass increases with radius, with a jump in the abundance of solid material beyond the "snow line" where ices condense. The change in the surface density of solid material occurs at 2.7 AU in the Hayashi model (Hayashi 1981). The location of this boundary depends on the disk structure, e.g., shadowing by a puffed-up inner disk and push the snow line outward (Sasselov & Lecar 2000). But for

solar type stars, the zone of interest is beyond that which is readily probed by the Doppler method or other indirect techniques, and falls within the domain of direct imaging. The discovery of giant planets far beyond the snow line would tend to favor theories of planet formation by gravitational instability over solid core condensation and accretion. At larger orbital radii (> 30 AU), gas-cooling times become shorter than the Keplerian shearing time—a necessary condition for runaway gravitational instability (Gammie 2001; Johnson & Gammie 2003; Boss 2002)—while solid core growth by collisional coagulation of planetesimals proceeds prohibitively slowly (Goldreich, Lithwick, & Sari 2004).

A second reason to image the outer regions of extrasolar systems is to probe them for vestiges of planetary migration. Ninety percent of the Doppler sample consists of massive planets with $a < 3$ AU, suggesting that they migrated inwards to their present locations. A variety of mechanisms may drive orbital evolution; the tidal gravitational interaction between the planet and a viscous disk (Goldreich & Tremaine 1980), the gravitational interaction between two or more Jupiter mass planets (Rasio & Ford 1996), and the interaction between a planet and a planetesimal disk (Murray et al. 1998). It is energetically favorable for a Keplerian disk to evolve by transporting mass inward and angular momentum outward (Lynden-Bell & Pringle 1974). Consequently, inward planetary drift appears inevitable, and this is what is found in certain simulations (Trilling et al. 2002; Armitage et al. 2002; Matsuyama, Johnstone & Murray 2003). However, if planets form while the disk is being dispersed, or if multiple planets are present, outward migration can also occur. In a system consisting initially of two Jupiter-like planets, a dynamical instability may eject one planet while the other is left in a tight, eccentric orbit. The second planet is not always lost; the observed Doppler exoplanet eccentricity distribution can be reproduced if the 51 Pegasi systems are formed by planet-planet scattering events and the second planet typically remains bound in a wide ($a > 20$ AU), eccentric orbit (Rasio & Ford 1996; Marzari & Weidenschilling 2002). Divergent migration of pairs of Jupiter-mass planets within viscous disks leads to mutual resonance crossings and excitation of orbital eccentricities such that the resultant ellipticities are inversely correlated with planet masses (Chiang, Fischer & Thommes 2002). Given decreasing disk viscosity with radius and the consequent reduction in planetary mobility with radius, we expect eccentricities to decrease with radius, perhaps sharply if the magneto-rotational instability is invoked (Sano et al. 2000). By contrast, excitation of eccentricity by disk-planet interactions requires no additional planet to explain the ellipticities of currently known solitary planets (Goldreich & Sari 2003). Clearly, observations of the incidence, mass, and eccentricity distributions of multiple planet systems would sharpen our nebulous ideas regarding how planetary orbits are sculpted.

2.2.4 Solar Neighborhood Surveys

The single common science theme for the PFI is the study of the evolution of nearby solar systems, from their formation phases through to their more mature steady-state phases. Figure 11 shows a Monte-Carlo simulation of the planet statistics expected for a simulated solar neighborhood, on a contrast-separation plot, assuming the planet mass and semi-major axis distributions established by Doppler planet surveys. The simulation includes both nearby main sequence stars and less nearby, but still close, associations of young (> 100 Myr) stars. On Figure 11, three types of planets accessible to the TMT PFI can be identified: young (few 10^8 yr), hot, self-luminous Jovian planets, somewhat older Jovian

planets (10^8–10^9 yr) which are slowly cooling and still emit a significant excess of thermal radiation, and mature Jovian planets around older stars (10^9–10^{10} yr), seen in reflected light at small angular offsets from their parent stars. The choice to include only planets with ages > 100 Myr is made to eliminate any uncertainties related to the formation and initial conditions (see Figure 10). Reflected-light terrestrial planets are smaller and fainter than Jovian planets, and so cannot be detected with the PFI project contrast goal of approximately 10^{-8} relative to the parent star.

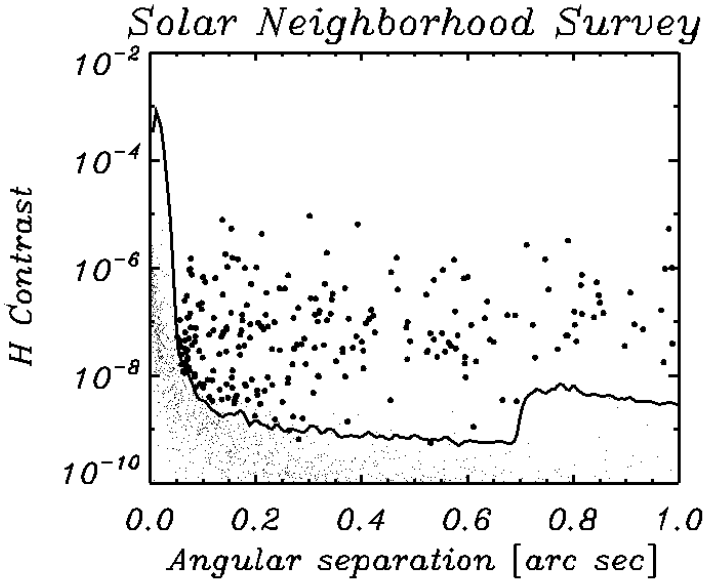

Figure 11: Companion contrast versus angular separation using a semi-analytic estimate of TMT PFI performance for I < 7 mag, showing the direct detection of young luminous planets in a hypothetical survey of field (< 50 pc) stars. The dots represent the planet population: those detected by PFI are filled circles. The dashed line shows the PFI contrast threshold (5 σ) for a 1-hour exposure at 1.65 µm. In this example × 16 speckle noise suppression was assumed.

Table 2 and Figure 11 show that in the field, planet detection can be achieved with good efficiency. In this example the only selection criteria for targets are brightness (for wavefront sensor lock) and distance (< 50 pc). If the input catalog is winnowed using an age selection, e.g., Ca II H and K, which selects for chromospheric activity (< 2 Gyr), the detection rate can be increased by a factor of 2-3.

Table 2: Solar Neighborhood Survey (d < 50 pc; all ages). Integration time: 1 hour; 5 σ detection threshold. Column 5 shows that TMT/PFI finds 2-3 times as many planets as the GPI instrument on Gemini, primarily because of TMT's smaller IWA.

I (mag)	Number of targets	TMT/PFI Number of Planets detected	TMT/PFI Detection Fraction (%)	Gemini/GPI Detection Fraction (%)
< 5	150	62	38	13
< 7	1872	244	13	7

While preliminary versions of such surveys of this kind will have been conduction with the VLT and Gemini planet finders, PFI clearly has the ability to recover these planets, find them in greater quantities (× 2–3; see Table 2), and study them spectroscopically and determine their astrometric properties (astrometric precision scales as D^2). What PFI can do that the 8-m instruments cannot is detect planets in reflected light. These reflected light planets are depicted in Figure 12. The median orbital period of this exoplanet sample is 2.7 years. Thus, this population, which represents about one quarter of the total detectable by direct imaging, provides an invaluable overlap with the indirect searches.

In summary, PFI surveys of the solar neighborhood will yield broad coverage of semi-major axis separations (0.5–50 AU; median 11 AU) and ages (0.15–9 Gyr; median 1.8 Gyr) with a high detection rate. Exoplanets will be imaged at all stages through their cooling history. Roughly one third of these exoplanets are located within 5 AU (Figure 13) and a few percent as close as 1 AU. The transition from the terrestrial region to the domain of the Jovian planets will be covered using a single technique, so it will be possible to measure the semi-major axis distribution without patching together the results from different techniques that have radically different selection criteria (e.g., Doppler and 8-m ExAO imaging), and for the first time, there will be substantial overlap between exoplanets detected by direct and indirect detection methods.

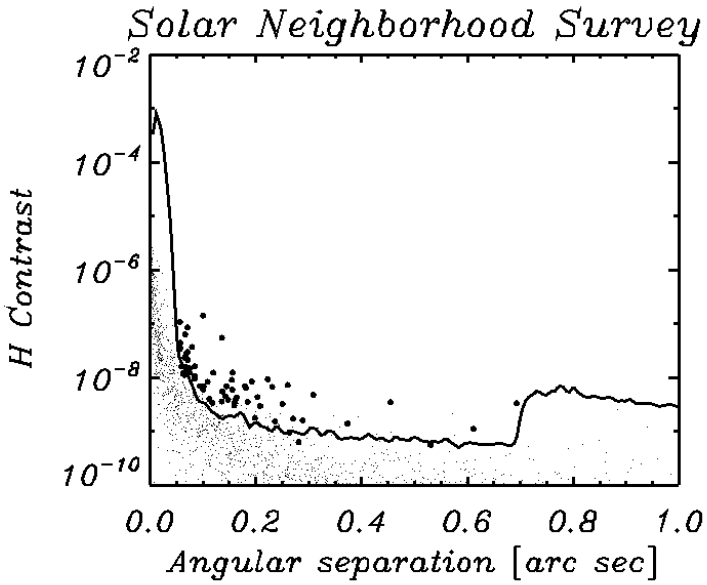

Figure 12: Same as Figure 11, except that self-luminous planets have been excluded. Only those exoplanets detected in reflected light (filled circles) are shown. About 25% of the total detectable exoplanets shine only by reflected light. The median orbital period of these planets is 2.7 years, therefore these are ideal targets for Doppler or astrometric study.

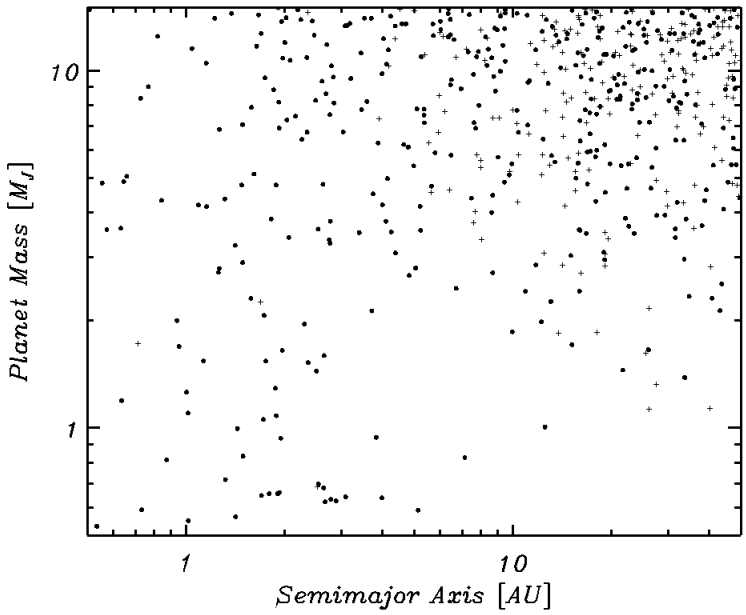

Figure 13: Plot of the planets discovered in a PFI field-star survey of the solar neighborhood. Planets discovered in a similar GPI survey (with different Monte Carlo seed) are marked with crosses.

2.2.5 Using Spectra to Constrain Planet Atmosphere Properties

During their cooling and contraction phase, exoplanets span a broad range of effective temperatures and surface gravities (Figure 14). If the age of the system is know, for example, because the age of the parent star is known, then an approximate mass can be inferred from the effective temperature. Since planets cool at roughly constant radius a measure of surface gravity provides an age-independent mass estimate. Ideally, both $T_{\rm eff}$ and log g should be determined from colors or spectra. For T dwarfs ($T_{\rm eff}$ = 800–1400 K), it is already known that effective temperature and gravity can be estimated from *J-H* and *H-K* colors (Knapp et al. 2004 AJ, 127, 3553).

The near-IR spectra include a variety features that exhibit strong $T_{\rm eff}$ and log g dependence. To emphasize the sensitivity of color and spectra to changes in the fundamental atmospheric parameters, the following figures quantify the predicted difference between a spectrum for a given log g and $T_{\rm eff}$ and a template spectrum with log g = 3.0 or $T_{\rm eff}$ = 300 K. Figure 15 (left panel) shows the effect of varying the surface gravity in magnitudes as a function of wavelength with respect to this baseline. The result is plotted in magnitudes, i.e., these spectra represent flux ratio. The spectra have been convolved with a Gaussian representing an instrumental resolution $R = \lambda/\Delta\lambda = 100$. To obtain a tight color-log g correlation, we simply need to choose one wavelength band where this curve is flat or averages to zero and a second wavelength where there is a large difference between each curve. Evidently the strength of the [1.5] – [2.2] index is a good gravity indicator. While a resolution of 100 is clearly acceptable for measuring log g, a factor of two lower would be satisfactory.

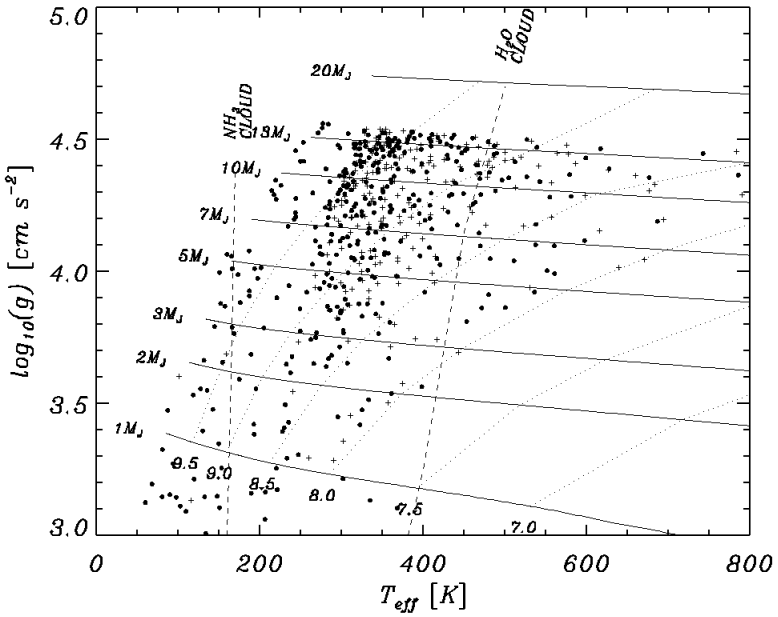

Figure 14: Isochrones [dotted & labeled with ages in log10(t/yr)] and iso-mass contours (solid and labeled with masses in MJ) in the surface gravity/effective temperature space for exoplanet atmospheres (data from Burrows et al. 1997) together with the planets discovered by PFI (solid circles) and GPI (crosses). The water and ammonia cloud condensation lines are shown. The only data points on this plot are Jupiter a 4.5-Gyr and 1 MJ . The coolest known T dwarfs have Teff ≈ 800 K and do not appear on this plot.

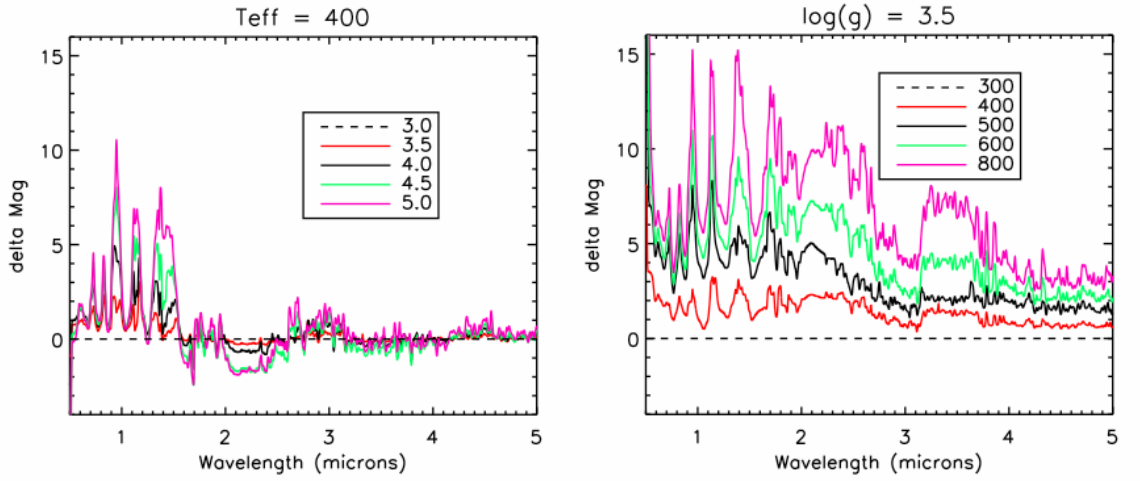

Figure 15: Left: Differential exoplanet spectra in magnitude, i.e., flux ratios, illustrating gravity sensitive indicators in synthetic spectra. The reference spectrum has Teff = 300 K and log g = 3.0. Spectra with log g ∈ {3.5, 4.0, 4.5, and 5.0} are shown. Evidently the strength of the band index [1.5] – [2.2] is a good gravity indicator. Right: Differential exoplanet spectra illustrating effective temperature indicators in synthetic spectra. Spectra with Teff ∈ {400 K, 500 K, 600 K, 800 K} are shown. We

calculate these spectra using fully self-consistent models with the PHOENIX atmosphere code (Barman et al. 2001). The spectral resolution is R = 100.

The sensitivity of the spectra to effective temperature is explored in Figure 15. For effective temperature the change in the differential spectrum diminishes with increasing wavelength as one might naively expect. Figure 15 (right panel) shows that there are plenty good indicators, for example [1.5] – [1.6] within the *H* band. There are suitable indices in the *K* band such as [2.1] – [2.2]. The strong temperature sensitivity of the emergent spectrum in the L band makes this an attractive wavelength regime of establishing atmospheric properties. Again, R ≈ 50 is probably adequate for measuring $T_{\rm eff}$, but the sharpness of the features in the differential spectrum suggest shown in Figure 15 (right panel) suggest that R ≈ 100 should be maintained as an option.

Is there a role for higher resolution observations? One might suspect that for objects that are dominated by broad molecular absorption and lines that are pressure broadened that high resolution is of limited use except for trace constituents and for absorptions formed high above the photosphere. Figure 16 shows a sequence of *R* = 1000 spectra at fixed gravity (log *g* = 3.5), effective temperature ranging from 300 to 800 K, and solar abundances. This figure shows that a number of molecular band features can be resolved, offering the potential to explore chemical composition and abundance variations.

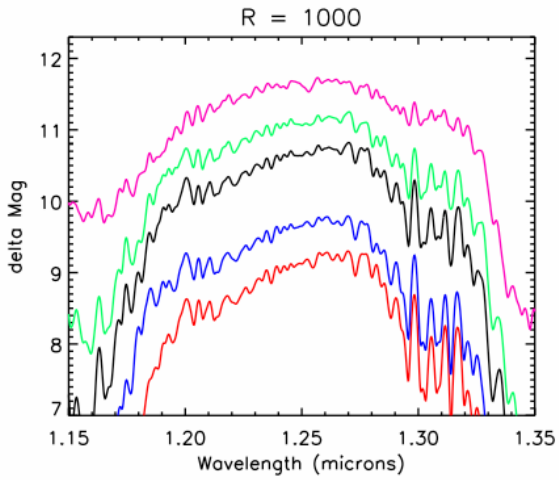

Figure 16: A spectral sequence at R = 1000 in the J band for Teff ∈ {300 K, 400 K, 500 K, 600 K, 800 K} (running from bottom to top), log g = 3.5, and solar abundances.

2.2.6 Imaging Planet Formation and Planet-Forming Disks

Planets are easiest to detect when they are young and hot (Figure 17), because they will be bright at near infrared wavelengths (2.1.2). This would suggest that studying the earliest phases of planet formation should be straightforward. The cooling tracks (Figure 17) also illustrate an important aspect regarding the atmospheres at young ages (< 100 Myr): planetary mass objects in this age range are expected to exhibit T dwarfs spectra. That is, older and more massive brown dwarfs with comparable atmospheric conditions have

already been discovered, studied, and modeled. As a consequence, we have high-fidelity spectral templates for these objects, which can be used to recognize and classify young exoplanets with confidence.

The star formation rate in the Milky Way has been roughly constant for the past 10 Gyr and the luminous phase of planets is brief: a 2 M_J object has $L/L_\odot > 10^{-6}$ for < 30 Myr. Thus only a fraction of a percent of the stars in the Milky Way are young enough to host newly formed planets. Table 3 lists about 300 youthful (< 100 Myr) associations within the solar neighborhood (< 50 pc). To find large numbers of genuinely young (< 10 Myr) stars forces the search for the newly formed planets to more distant star forming clouds, such as Taurus, ρ Ophiuchus, or Upper Sco (see Figure 19). The closest of these regions lie at distances of about 150 pc. The most critical instrumental parameter for this study is thus the inner working angle, as 5 AU at 150 pc translates to 33 mas. This inner working angle translates to about 3λ/D at *H* and 4λ/D at *J*, making these the instrumental requirements.

Table 3: Young stars in local groups and associations.
The approximate number of members refers to the number of stars with *R* < 13 mag.

Name	Age (Myr)	Dist. (pc)	Approx. No.
TW Hya	8	30-60	20
β Pic	8	30-60	20
Tuc/Hor	30	30-50	50
Local Assoc.	60	40	150
AB Dor	70	15-50	20
Debris Disks Stars	10-100	<100	40
Total			**300**

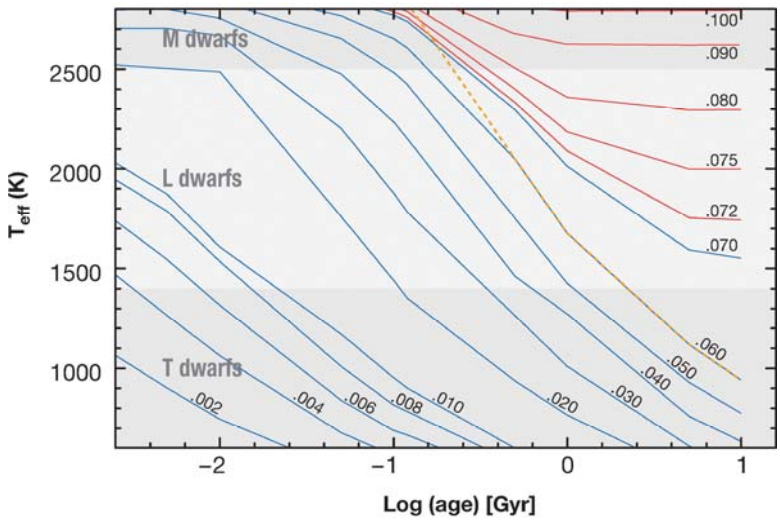

Figure 17: Theoretical tracks showing the evolution of effective temperature with age for low-mass stars (red), brown dwarfs and planets (blue). Tracks are labeled in M⊙. Conventionally, planets correspond to masses less than 12 MJ (0.011 M⊙). Grey areas delineate the approximate temperature ranges for L and T dwarfs (Kirkpatrick 2005). The tracks warmer than 1500 K are from Chabrier et

al. (2000, using the AMES-dusty models of Allard et al. 2001) and the tracks cooler than 1500 K are from Baraffe et al. (2003, using the AMES-cond models of Allard et al. 2001).

We expect that ExAO on 8- and 10-m class telescopes will detect the first handful of such very young, very luminous, hot Jupiters, in the nearest regions of recent star formation—the 8 Myr-old 2M1207b probably is the prototype for such objects (Chauvin et al 2004). Thus, the goal for PFI is to increase the sample space from a few candidate stars into the hundreds, so that a statistically significant sample of planets can emerge. By reaching stellar distances of 150 pc, the potential source list grows to include about ten well known nearby star-formation regions. Many of these are much younger than the nearby associations of – young enough that planets may still be in the accretion phase. This will thus provide an order of magnitude increase in search volume over that attainable in early searches with existing telescopes, with over 100 candidate young stars per star forming region, spanning the whole process of planetary formation.

The contrast needed to find hot, young planets is not overly great (~ 10^{-6}; Figure 7 and Figure 10). However, the targets are relatively remote and even the brightest T Tauri targets have near-IR magnitudes of only about 11, which sets a performance limit for wavefront sensing. Moreover, even though contrast requirements are modest they must be achieved at small inner working angles. The apparent magnitudes of the planets are of order 26, which places stringent requirements on sensitivity. To examine this circumstance in more detail consider T Tauri stars drawn from the young cluster represented in Figure 20 by its color magnitude diagram. Stars have a spread of ages ranging from 5–15 Myr and the distance is 150 pc. The pre-main sequence tracks of D'Antona & Mazitelli (1994) have been used. Figure 18 illustrates that a young hot planet is not confused with scattered light from an optically thick circumstellar disk. However, it does show how the polarization mode of PFI can be used to unambiguously show that a bright source in a disk is a self-luminous planet and not a blob of dust.

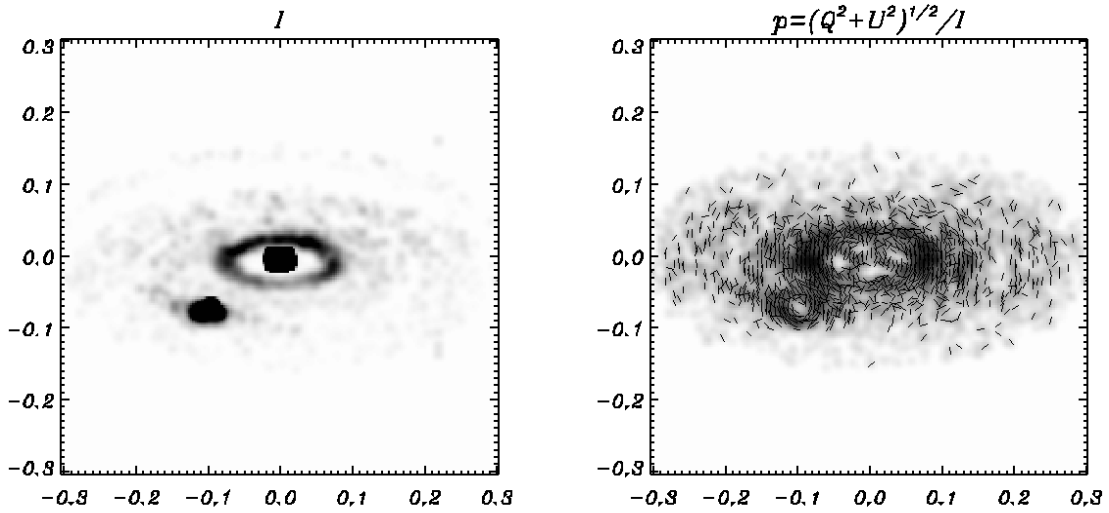

Figure 18: Radiative transfer model of a young planet in an optically thick circumstellar disk. The companion has a contrast of ΔH = 10 mag, representing a <10 Myr planet. This disk has a radius of 50 AU with an inner hole at 5 AU and a Hill-sphere wide gap at the radius of the planet (30 AU). The disk inclination is -66 degrees, such we are viewing the underside of the disk. The distance is 150 pc. The

left hand panel shows the total intensity and the left shows the polarized fraction with the orientation of the electric vector indicated. The concentric polarization vectors and unpolarized core clearly show that the source at 7 o'clock is self-luminous and not just a blob of dust.

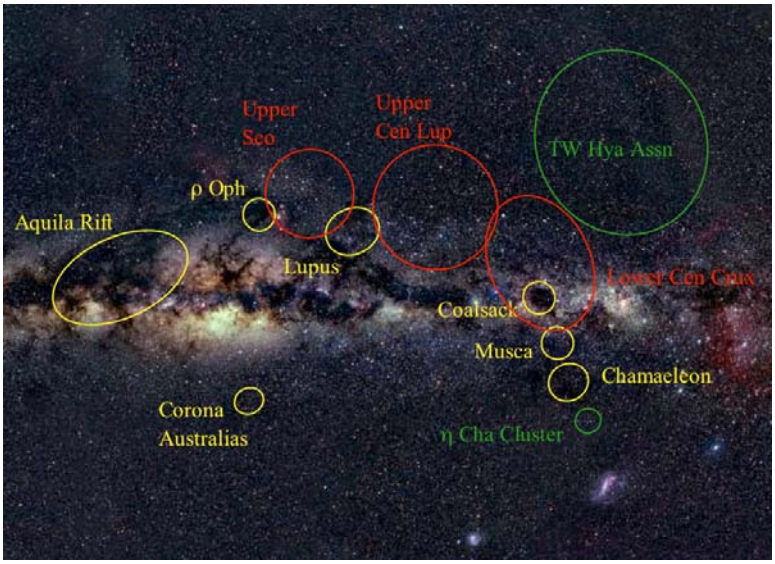

Figure 19: Examples of nearby (150 pc) star forming regions in Sco-Cen with young (< 10 Myr) stars. The Galactic plane runs along the middle of this image and the Galactic Center is to the left of the center.

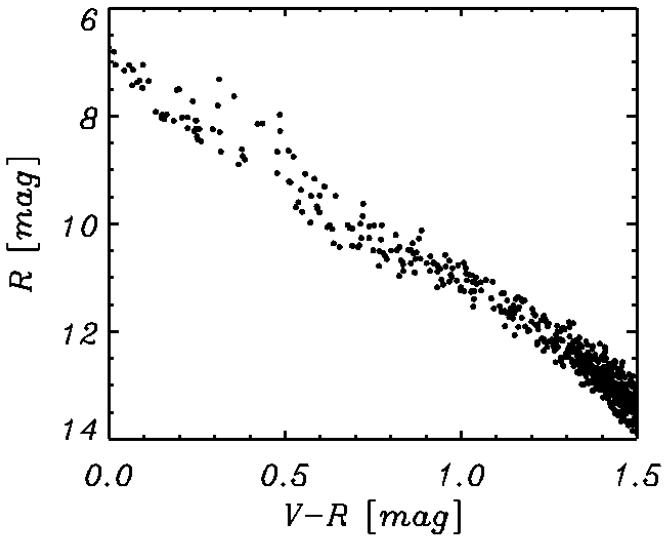

Figure 20: The theoretical color-magnitude diagram of a young young cluster at 140 pc and with stellar ages between 5 and 15 My. The pre-main sequence tracks of D'Antona & Mazitelli (1994) have been used. The mass spectrum is terminated at 0.08 M⊙.

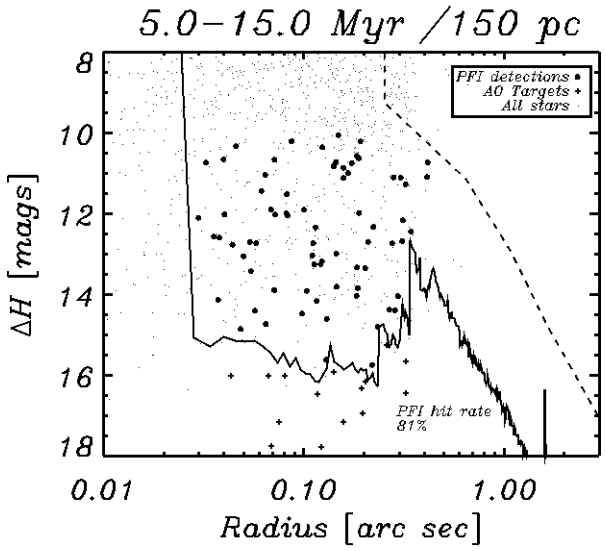

Figure 21: PFI detections of young planets the young (5-15 Myr-old) nearby (150 pc) star forming region represented in Figure 20. The population of T Tauri stars is shown a light dots, those while meet the criteria for wavefront lock I < 14 mag. and H < 9 mag. are represented by crosses. Planets that fall above the detection threshold, 5 σ for a 1-hour exposure at 1.65 μm, are denoted by filled circles. In contrast to Figure 11 the performance of PFI is estimated from a full wave-optics simulation. The detection rate is very high (81%) indicating that an inner working angle of 30 mas is sufficient to obtain a census of this population. The light dashed curve shows the contrast anticipated for the Gemini/NICI, which is designed to search for planetary companions around young stars.

3 Front-end adaptive optics system

3.1 Introduction

This chapter discusses the front-end adaptive optics system for PFI. As discussed in Section 1, PFI has a two-stage AO system. The initial ("front" or "visible") AO is a very high-speed system operating in visible light (0.7–1.0 μm) to provide primary wavefront correction on bright (I<8 mag) targets. On the brighter targets, servo lags will be the dominant source of scattered light within the dark hole region; as a result, the front AO must be capable of very high frame rates and efficient correction with a limited number of photons per timestep. On dimmer targets, the front-end system provides a modest wavefront correction to increase the Strehl ratio enough to enable the post DSS (or "back") adaptive optics to control the correction of the wavefront. Hence it must also be capable of stable behavior at much lower fluxes (I<14 mag). For this purpose we have selected two possible WFS architectures – a pyramid wavefront sensor or a interferometric wavefront sensor – that have much better noise behavior than a conventional Shack-Hartmann sensor and also allow for easy binning to operate on dimmer stars

3.1.1 Instrumentation Concept

AO system key requirements:

- Provide real-time correction of atmospheric and telescope wave front errors up to a spatial frequency of 62 cycles per pupil, without significant aliasing errors
- Operate at frame rates of at least 2500 Hz (Goal: 4000 Hz) for optimal correction on bright stars
- Provide good wave front correction on targets with magnitude I < 9 mag.
- Provide moderate wavefront correction (Strehl >0.1) for targets with magnitude I<14 mag.
- Meet the requirements of the error budget
- Provide diagnostic and telemetry data to the observatory (e.g. telescope offloading) and users
- Provide field steering capabilities over ~3 arc seconds through optical steering of the wave front sensor
- Provide pupil rotation to match the TMT pupil to the DSS

AO system key interfaces:

- Accepts the TMT F/15 beam
- All components mount to the main optics bench
- Accept wave front calibration updates and pointing updates from the post-DSS WFS
- Accept commands and return status to the observatory software
- Produce a collimated beam for input to the DSS with a pupil located well inside the DSS

AO system design summary
- Two-stage wave front correction with conventional and MEMS deformable mirrors
- Spatially-filtered pyramid WFS (operating in interferometric mode) or 4-bin interferometer WFS
- λ/100 - λ/200 optics

3.2 Wave-front Sensor Analysis

3.2.1 Introduction

A number of different wave-front sensors were analyzed to determine which would give the best combination of performance. The choices studied included the Shack-Hartmann wave-front sensor, interferometric wave-front sensors, the pyramid wave-front sensor (operating in interferometric mode) and a Zernike phase contrast wave-front sensor. All of these wave-front sensors, except for the Shack-Hartmann wave-front sensor which measures the phase gradient, measure the phase of the wave-front directly when in the high Strehl ratio regime.

The Zernike phase contrast wave-front sensor converts phase variations across the pupil into intensity variations. This is accomplished by placing a phase-shifting focal-plane filter at focus which provides a relative phase shift of $\pi/2$ between the general field and an on-axis spot roughly λ/D in diameter, where λ is the wavelength and D is the diameter of the telescope [Bloemhof 2004]. The $\pi/2$ phase shift presents some potential problems with chromatic shifts in broadband light, however, achromatic phase shifts can be obtained through the use of multilayer films [Riaud 2001].

The pyramid wave-front sensor was first presented as a slope sensor [Raggazzoni 1996], similar to the Shack-Hartmann wave-front sensor. It has since been realized, however, that the pyramid sensor has properties of both slope and direct phase sensors and that the pyramid sensor can be operated in a direct phase sensing mode at high Strehl ratios [Verinaud 2004]. The pyramid wave-front sensor typically uses a focusing device to form an image/Fourier plane where the apex of a glass pyramid is placed. Four separate expanding beams exit through the four faces of the pyramid and a collimating device is used to form four images of the pupil on a wave-front camera. By comparing the intensities in the sub-apertures between the four images, the slopes in the sub-apertures can be determined. Likewise, in a high Strehl ratio regime, the comparison of the sub-apertures between the pupil images provides the direct phase of the wave-front. In this regime, there may be some advantage to dividing the beam into two rather than four sections, as discussed below.

Interferometric wave-front sensors also measure the wave-front phase directly by interfering reference and signal light. These devices must measure the reference and signal beams within the coherence length of the light, which limits the interferometric designs that can be used with temporally incoherent light from stars. All of the designs are self-referencing in that they derive the reference beam from the signal itself by spatially filtering part of the signal before interfering the two beams on the detector. One such design is based on a Mach-Zehnder interferometer.[J.R.P. Angel 1994] This design must,

however, keep the two arms within the coherence length of the light and provide achromatic phase shifts over a modest, 20%, bandwidth. Recently pixilated interferometers [Millerd et al. 2004] have been designed which utilize polarizers to provide achromatic phase shifts and when placed in a self-referencing configuration automatically maintain the distance between the reference and signal beams forming the interferogram.

Shack-Hartmann wave-front sensors are the most widely used wave-front sensor in adaptive optics systems. The Shack-Hartmann wave-front sensor measures the slope of the wave-front and then reconstructs the phase from the measured slopes. It has been demonstrated to be very robust and does not suffer from chromatic effects. With new techniques applied to these wave-front sensors, such as spatial filtering to prevent aliasing [Poyneer 2004] and matched filtering or weighted center of mass algorithms [Nicolle 2004] to reduce photon noise on the centroiding (provided read noise is negligible), Shack-Hartmann wave-front sensors have been shown to compare favorably to pyramid sensors [Verinaud 2005]. However, the wavefront reconstruction process means that the Shack-Hartmann sensor always measures low-order modes less effectively than the other sensors described here.

3.2.2 Trade Study Summary

A study was performed on several wave-front sensor technologies to determine which technology would be the more appropriate choice as the front-end wave-front sensor. Table 4 summarizes these trades. The wave-front sensors studied, as detailed above, included the Shack-Hartmann wave-front sensor, interferometric wave-front sensors, the pyramid wave-front sensor(operating in interferometric mode) and a Zernike phase contrast wave-front sensor. All of these wave-front sensors, except for the Shack-Hartmann wave-front sensor that measures the phase gradient, measure the phase of the wave-front directly when in the high Strehl regime. This trade study included the derivation of the error due to the inherent signal-to-noise ratio of the wave-front sensor and also implementation of the device to determine effects such as amplitude variations on the wave-front sensor performance.

The root-mean-square (RMS) errors resulting from the signal-to-noise ratio for the different wave-front sensors are given below:

$$RMS\_ZPC = \frac{1}{2}\sqrt{\frac{1.0 + e^2/q.e.num\_phot}{q.e.num\_phot}}, \tag{3.2.2.1}$$

$$RMS\_PYR\_2S = \sqrt{\frac{1 + 4e^2/q.e.num\_phot}{2q.e.num\_phot}}, \tag{3.2.2.2}$$

$$RMS\_4bin = \sqrt{\frac{2 + 8e^2/q.e.num\_phot}{q.e.num\_phot}} \text{ and} \tag{3.2.2.3}$$

$$RMS\_SH = \sqrt{2}\frac{3\pi^2}{4}\frac{\sqrt{1/4 + e^2/q.e.num\_phot}}{\sqrt{q.e.num\_phot}}\sqrt{\left(1 + \left(\frac{d}{r_o}\right)^2\right)}, \tag{3.2.2.4}$$

Where e is the read noise, q.e. is the quantum efficiency, num_phot is the number of incident photons, d is the sub-aperture size and r_o is the Fried parameter, respectively. To compare these sensors it is instructive to simplify the equations by neglecting read noise. In this case, the RMS error for the Zernike phase contrast wave-front sensor reduces to RMS_ZPC = 1/SQRT(4*num_phot). [Bloemhof 2004 oe] The RMS error for a two-sided pyramid wave-front sensor is RMS_PYR_2S=1/SQRT(2*num_phot) in the limit of no read noise. The four-bin interferometric wave-front sensor has an RMS error of RMS_4bin = 1/SQRT(num_phot/2) in this limit [Phillion 1997]. The RMS error for a Shack-Hartmann wave-front sensor may be expressed as RMS_SH~5.2*SQRT(1+$(d/r_o)^2$)/SQRT(num_phot) [Bloemhof 2004 oe]. The derivation of the Shack-Hartman RMS phase error above uses the RMS error associated with the slope measurement [Hardy 1998] σ_{slope}=SQRT(2)(3π/16)(λ/d)/SQRT(num_phot), ignoring read noise and atmospheric effects, and multiples this by (2πd/λ) to get the phase across the sub-aperture. It therefore ignores the increase in the noise propagator for high-order S-H systems due to reconstructing the wavefront. The Zernike phase contrast sensor, the pyramid wave-front sensor (unmodulated) and the four-bin interferometer all act as interferometers and have a nearly flat spectral response to noise, however, the Shack-Hartmann wave-front sensor has a photon noise propagation power spectral density proportional to $d^{-2}f^{-2}sinc^{-2}$ (df), with d being the sub-aperture size [Verinaud 2005]. A reduction of 1.5 in the RMS error of the Shack-Hartmann sensor was observed in certain cases using a weighted center of gravity algorithm [Nicolle 2004]. This algorithm requires a larger number of CCD pixels per sub-aperture and so zero read noise detectors are required to take full advantage of this technique.

A plot of these equations, as a function of star magnitude/photons per sub-aperture, is shown in Figure 22 below. The following assumptions were made; an optical efficiency of 0.5, sub-aperture size of 23 cm, a quantum efficiency of 0.8, a read noise of 1 electrons and the RMS error is given at the science wavelength of 1.65 μm This shows that the Zernike phase contrast wave-front sensor has the smallest root-mean-square error or the best performance for a given signal-to-noise ratio. This is followed by the pyramid sensor, the four-bin interferometer and then the Shack-Hartmann wave-front sensor.

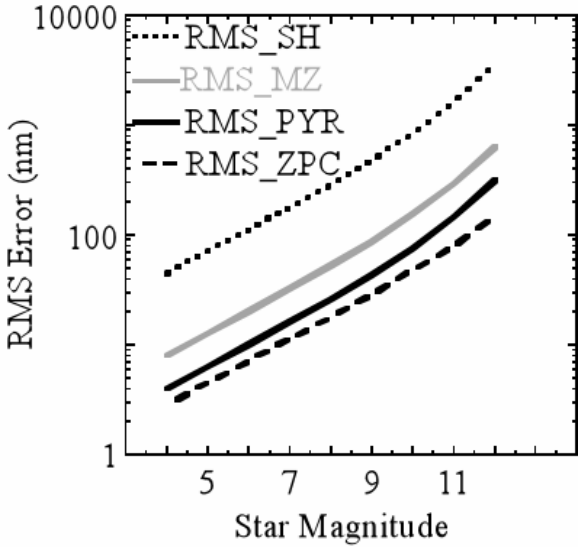

Figure 22: RMS Phase error expected for the different wave-front sensors. SH=Shack-Hartmann; MZ=Mach Zehdner; PYR=Dual-ridge Pyramid sensor; ZPC=Zernike Phase Contrast

Although the Zernike phase contrast wave-front sensor (ZPC WFS) has attractive noise behavior, however, there are several problems practical problems in implementing it. The ZPC WFS is subject to amplitude variations in the pupil showing up as intensity variations on the wave-front sensing camera and hence being interpreted as a phase error. These amplitude variations can occur due to scintillation, due to the presence of the anti-aliasing spatial filter and also due to spatial variations in mirror reflectivities. It was also discovered that if the intensity pattern due to the structure of the phase shifted core is not carefully calibrated, then the first ~16 Zernike modes cannot be accurately measured or controlled, which would make it unusable. If the core is decentered a little from the 90 degrees phase shifting circle, then the reference calibration is no longer valid and again the low and mid-order modes cannot be controlled. That indicated that the wave-front sensor would not be robust against pointing errors. These factors led us to remove the ZPC WFS from consideration as the front-end adaptive optics system.

The pyramid wave-front sensor was also modeled in order to better understand its advantages and limitations. The pyramid sensor is attractive for a number of reasons. Because this wave-front sensor can be run in both slope mode and interferometric mode, it does not require a separate bootstrapping adaptive optics system as would be required for other interferometric wave-front sensors. In addition because of the geometry of the pyramid sensor, the sub-apertures can be binned on the camera to reduce the number of sub-apertures in the case of low photon flux to achieve higher Strehl ratios, which can't be done with a Shack-Hartmann sensor. The pyramid wave-front sensor was found to be only slightly subject to amplitude variations in the pupil showing up as intensity variations on the wave-front sensing camera and hence being interpreted as a phase error.(The amplitude variations for the two-sided pyramid wave-front sensor are reduced by over an order of magnitude from phase aberrations) . The two-sided pyramid wave-front sensor was also found to be very robust against pointing errors in the core.

Four-bin interferometers were also modeled. One of the limitation of these systems is that they only operate well when the Strehl ratio at the sensing wavelength is already high, giving them an inability to initially bootstrap from open-loop into a high Strehl ratio regime, thereby requiring a separate AO system for acquisition. They have the advantage that interferometric designs with three or greater phase bins measure the amplitude and phase and do not have the problem that amplitude variation in the pupil turns into phase measurements at the wave-front sensor, which is a critical problem for Zernike wave-front sensors and to a much lesser degree pyramid wave-front sensors.

Spatially-filtered Shack-Hartmann wave-front sensors were also studied to provide a benchmark from which the other wave-front sensors could be compared. These have the highest RMS error for a given photon flux of the four considered, however, they are the most robust of the designs. They can start from very low Strehl ratios and bootstrap their way into the high Strehl ratio regime, unlike the Zernike phase contrast and four-bin interferometric wave-front sensors. They are not strongly affected by amplitude variations in the pupil plane appearing as phase fluctuations for which the pyramid sensors are slightly subject to.

Table 4: Summary of wave-front sensor comparisons.

	Wave-Front Sensors			
Metrics	Shack-Hartmann (S-H)	Zernike Phase Contrast (ZPC)	Pyramid	Four-Bin Interferometers
SNR performance with noise	Fourth best performance	Best performance	Second best performance	Third best performance
Robustness against intensity errors (scintillation, prim. segments, phase)	Robust for small amplitude changes (neutral)	Intensity errors will appear as phase errors(bad)	Robust for small amplitude changes (neutral)	Directly measures amplitude changes
Self-bootstrapping / low Strehl operation	Yes	No (problem with reference through pinhole)	Not in interferometric mode but yes in slope mode	No (problem with reference through pinhole)
Chromatic issues	Not significant	Chromatic issue with 90 deg. phase shift.	Use mirrors, achromatic prisms/ lenslets	Chromatic issue with phase shifts.
Vibration and flexure robustness	Robust against vibration	Must be recalibrated if core moves	Registration issues if different cameras used	M-Z config could have problems/ pixellated robust
Constant sensitivity per ang. Sep. (Guyon β_π)?	no	yes	yes	yes
Non-common path errors	Depends somewhat on optics quality	Similar to SH	Similar to SH	Mach-Zehnder config has more/ pixellated approach similar to SH
Flat-field errors	Miscalibration within a sub-pixel will change the spot location	Miscalibration. within a sub-pixel will change the phase	Miscalibration. within a sub-pixel will change the phase	Miscalibration. within a sub-pixel will change the phase
Accuracy operating off null due to NC path errors	Will work slightly off null	Same as 4-bin, but in addition perf. based on linearization	In interf. mode, pyr. must operate close to null	Very accurate unless NCP errors impact WFS strehl
Ease of building	Relatively simple to build	More difficult to build	Relatively simple to build	More difficult to build
Ease of operation	Easy to operate	Easy to operate	Easy to operate unless images separated between cameras	Pixellated easy/MZ more difficult

3.2.3 Pyramid Sensor

3.2.3.1 Introduction

Of the wave-front sensors considered for the front-end AO system, the pyramid wave-front sensor is the newest [Ragazzoni 1996] and its operation as a direct phase measuring device has only recently been realized.[Verinaud 2004] Pyramid sensors inherit properties of both slope sensors and direct phase sensors. They can in principle be used in slope mode to bootstrap to a high Strehl ratio regime and then operate in a direct phase sensing mode such that a wave-front reconstruction is not required. The pyramid wave-front sensor in the direct phase mode acts as a zero-path-length-difference white-light interferometer. As such it is appropriate for the temporally incoherent light encountered in astronomical applications.

Because of the novelty of this device, the principles behind pyramid sensors will be reviewed. As stated above, a focusing device is used to propagate the pupil to the image/Fourier plane at which point a pyramid is placed with the core of the resultant point-spread-function centered on the apex of the pyramid. The pyramid slices up the focal plane into sectors with each sector getting a piece of the core which serves as a reference wave. Each speckle in the point-spread-function corresponds to a plane wave component at the telescope pupil. A small sinusoidal phase fluctuation or a small sinusoidal amplitude fluctuation at the telescope pupil gives rise to an antipodal pair of speckles having the same magnitude. Speckle pairs are sent to different images of the pupil by the pyramid. The light passing through each facet is propagated to a CCD camera which is at a plane conjugate to the pupil plane. By interfering unpaired single speckles with the core, acting as a reference wave, each unpaired speckle creates a sinusoidal interference pattern at the CCD. By looking at the phase and amplitude of the intensity variations for both speckles in a pair, the phase and amplitude sinusoidal fluctuations at the pupil plane for that spatial frequency may be determined.

Classic pyramid sensors operate with a four-sided pyramid producing four separate images. In the direct phase mode, wavefront reconstruction is more accurate if the sensor produces only two separate images, splitting the light with a ridge rather than a pyramid. This is the mode we currently prefer for PFI, but we will continue to refer to it as a "pyramid" rather than a "ridge" sensor.

The equations describing the interferometric mode for pyramid sensing are described below. Phase and amplitude fluctuations in the pupil my be represented by

$$z^{A} = 1 + \sum_{n} 0.5\varepsilon_{n}^{A}\left[\exp\left(i\left(\vec{k}_{n}\cdot\vec{x}+\theta_{n}^{A}\right)\right)-\exp\left(-i\left(\vec{k}_{n}\cdot\vec{x}+\theta_{n}^{A}\right)\right)\right]$$
$$= 1 + i\sum_{n}\varepsilon_{n}^{A}\sin\left(\vec{k}_{n}\cdot\vec{x}+\theta_{n}^{A}\right)\text{ for phase} \quad \text{and} \quad (3.2.3.1.1)$$

$$z^{\mathbf{B}} = 1 + \sum_{n} 0.5\varepsilon_{n}^{B}\left[\exp\left(i\left(\vec{k}_{n}\cdot\vec{x}+\theta_{n}^{B}\right)\right)+\exp\left(-i\left(\vec{k}_{n}\cdot\vec{x}+\theta_{n}^{B}\right)\right)\right]$$
$$= 1 + \sum_{n}\varepsilon_{n}^{B}\cos\left(\vec{k}_{n}\cdot\vec{x}+\theta_{n}^{B}\right)\text{ for amplitude} \quad . \quad (3.2.3.1.2)$$

After passing through a two-sided pyramid, the left-hand-side amplitude(LHS) and right-hand-side amplitude(RHS) pupil images will have

$$\begin{aligned} \mathrm{LHS} &= 1+\sum_n 0.5\varepsilon_n^{\mathrm{A}}\cos\left(\vec{k}_n\cdot\vec{x}+\theta_n^{A}\right) \\ \mathrm{RHS} &= 1-\sum_n 0.5\varepsilon_n^{\mathrm{A}}\cos\left(\vec{k}_n\cdot\vec{x}+\theta_n^{A}\right) \text{ for phase} \end{aligned} \quad \text{and} \quad (3.2.3.1.3)$$

$$\begin{aligned} \mathrm{LHS} &= 1+\sum_n 0.5\varepsilon_n^{\mathrm{B}}\cos\left(\vec{k}_n\cdot\vec{x}+\theta_n^{B}\right) \\ \mathrm{RHS} &= 1+\sum_n 0.5\varepsilon_n^{\mathrm{B}}\cos\left(\vec{k}_n\cdot\vec{x}+\theta_n^{B}\right) \text{for amplitude} \end{aligned} \quad (3.2.3.1.4)$$

The average difference between the left-hand-side amplitude(LHS) and right-hand-side amplitude(RHS) provides the phase fluctuations,

$$\left(\mathrm{LHS}-\mathrm{RHS}\right)=\begin{cases}\sum_n \varepsilon_n^{\mathrm{A}}\cos\left(\vec{k}_n\cdot\vec{x}+\theta_n^{A}\right) & \text{for phase} \\ 0 & \text{for amplitude}\end{cases} \quad (3.2.3.1.5)$$

and the average sum of the left hand(LH) and right hand(RH) sides provides the amplitude fluctuations as shown in Eq. 3.2.3.1.6

$$\left(\mathrm{LHS}+\mathrm{RHS}\right)=\begin{cases}2 & \text{for phase} \\ 2+\sum_n \varepsilon_n^{\mathrm{B}}\cos\left(\vec{k}_n\cdot\vec{x}+\theta_n^{\mathrm{B}}\right) & \text{for amplitude}\end{cases}, \quad (3.2.3.1.6)$$

Comparing Eq. 3.2.3.1.1 for the applied phase to the measurement in Eq. 3.2.3.1.5, one finds that in order to retrieve the phase from the pyramid sensor, a Hilbert transform must be performed to convert the cosine terms present in Eq. 3.2.3.1.5 to sine waves, thereby, reproducing the phase present in the pupil. In addition, the pyramid sensor slices the core of the point-spread-function into different segments, which has the effect of shifting the center-of-mass of the part of the core that passed through a give pyramid face. This shift in the center-of-mass causes a shift in the frequencies and this frequency shift must be corrected in order for the pyramid sensor to accurately reproduce the phase.

3.2.3.2 Implementation

A pyramid sensor, operating as a slope sensor, has been tested on the sky [Ghedina 2003]. This sensor utilized a 4-sided pyramid and was operated in a slope sensing mode. In the case of broadband light, with a large number of sub-apertures, the dispersion in a single element pyramid is unacceptable. The pupil image at the different wavelengths through the prism should be sheared by less than a tenth of a sub-aperture, or less than 1/1000 of the pupil size for 128 sub-apertures. For a 20% bandwidth around 1650 nm, the index of refraction, n, for fused silica changes by $\Delta n = 0.00375$, $n_{1500} = 1.44462$ and $n_{1800} = 1.44087$. This would lead to a pupil shear of almost half a sub-aperture. This can be corrected by making the prism from multiple optical materials so that it has a more achromatic dispersion. Reflective prisms could also serve the same purpose as transmissive prisms, in which case chromatic dispersion would not be an issue. It has recently been proposed to use a lenslet array, in a field lens configuration, to perform the function of a pyramid wave-front sensor as shown below in Figure 23 [Bauman 2002]. For pyramid sensors implemented with lenslet arrays, pupil images on the wave-front camera are separated at the lenslet pitch and the pupil image is demagnified by the ratio of the focal lengths of the lens forming the image plane and the lenslet array. In the case of implementation using lenslet arrays, the focal length difference between the lenslets must be less than 0.1% to restrict the amount of radial shearing between the four pupil images. This also requires the chromatic shift in the focal length across the bandwidth be held to less than 0.1% to restrict the amount of chromatic radial shearing within a given pupil image.

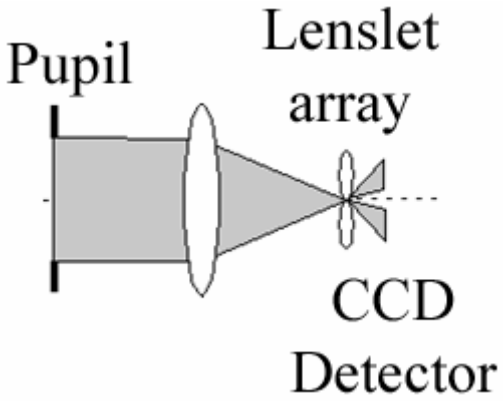

Figure 23: Pyramid sensor based on a lenslet array.

One problem with a pyramid wave-front sensor is that any speckle pair that falls along one of the edges of the pyramid will not be sensed. It is therefore blind to aberrations that only have a pure x or pure y aberration. We propose to use two ridge prisms or lenslet pairs that have their respective ridges orthogonal to one another. Each separate device will be blind to the pure x or pure y aberrations, however, by using both together all aberrations will be sensed. It is possible that one channel could be dispensed with, since the DSS nuller renders PFI blind to planetary companions in the pure +y direction; this will be analyzed in future studies.

3.2.3.3 Simulations

The pyramid wave-front sensor measures the intensities that pass through the faces of the pyramid. In the case of a four-sided pyramid, four images of the pupil are formed on the wave-front sensor camera. To determine the phase of the wavefront, the pupil images must be processed as detailed by the flowchart in Figure 24. This includes normalizing the measured intensities and taking their square root to determine the amplitudes. Differencing the pupil images provides the Hilbert Transform of the frequency-shifted phase. To retrieve the phase, the difference of the amplitudes of the sides must be taken and the left-hand side multiplied by –i and the right-hand-side multiplied by +i. Because the core is split into different pieces by the faces of the pyramid, the center-of-mass of the core passing through each of the pyramid faces is shifted away from the apex of the pyramid. This shift must be corrected in order for the pyramid wave-front sensor to reproduce the phase correctly. As such the next step in the algorithm is to correct the shifted frequencies before inverse Fourier transforming back to arrive at the frequency across the pupil.

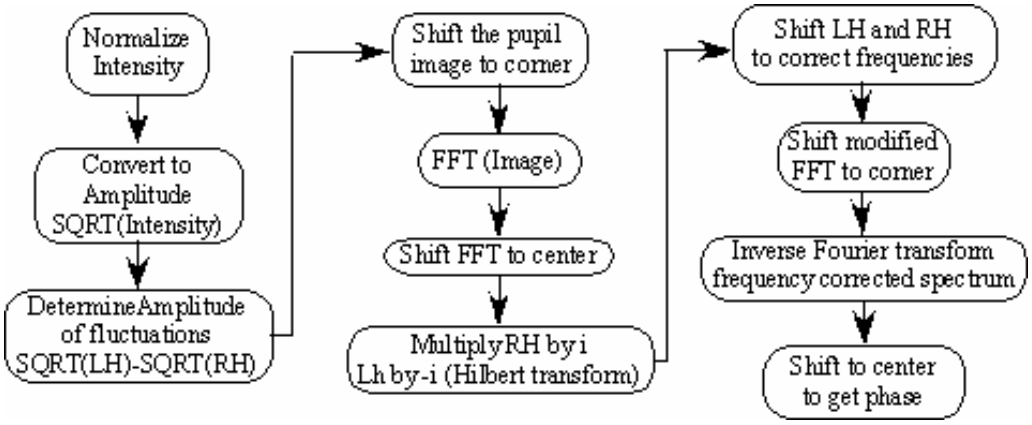

Figure 24: Flowchart of computations for the Pyramid's interferometric reconstruction of the phase.

The pyramid WFS simulator runs from a movie of phase screens generated using the ARROYO code as discussed in Section 9.2.3. The simulator first applies the telescope aperture to the phase screen and then runs the phase screen through a square spatial filter which selects spatial frequencies having k_x and k_y values up to the Nyquist values. There are two orthogonal two-sided pyramid wavefront sensors. The LH-RH sensor has left and right halves, whereas the TH-BH sensor has top and bottom halves. The LH-RH WFS measures the phase except for the y variation and the TH-BH WFS measures the phase except for the x variation. Their two measurements must be combined. The x variation is obtained from the LH-RH WFS and the y variation is obtained from the TH-BH WFS. The x and y variations are then removed from both and they are averaged together. The x and y variations are then restored.

To look at the noise for the pyramid wave-front sensors, simulations were performed with no phase aberration. A zero phase aberration signal with the appropriate pupil was sent through a spatial filter that was set up to remove those components that could not be measured by the interferometer/corrected by the DM. The spatially-filtered signal was then passed through the pyramid wave-front sensors with the appropriate photon noise and read

noise added to the simulated CCD measurements. The residual root-mean-square error for the simulations, with a read noise of 1 e-, was found to be 0.013 rad, 0.039 rad, 0.098 rad for the three cases of 2029, 225 and 51 photons/sub-aperture respectively. That is in good agreement with the analytic values expected from Eq. 3.2.2.3 above of 0.016 rad, 0.048 rad and 0.1 rad for the pyramid wave-front sensor with the photons/sub-aperture reported above. A Blackmann apodizer was used to evaluate the achievable contrast ratio in these simulations. The apodized pupil along with the residual phase was Fourier transformed to form the field at the Fourier plane. The intensity was found from this field and the results are shown in Figure 25. The figure shows the $\log_{10}$(intensity) for the three-photon levels. This plot has been normalized by the maximum of the perfect PSF.

Radially-averaged point-spread functions were then evaluated for comparison to the interferometers discussed in the following section. The results for the three cases above are shown in Figure 26. In each of these graphs, the Mach-Zehnder configuration is displayed as the solid gray line and the pyramid wave-front sensor is displayed as the dashed black line. The pyramid wave-front sensor gives significantly better performance than the Mach-Zehnder interferometer.

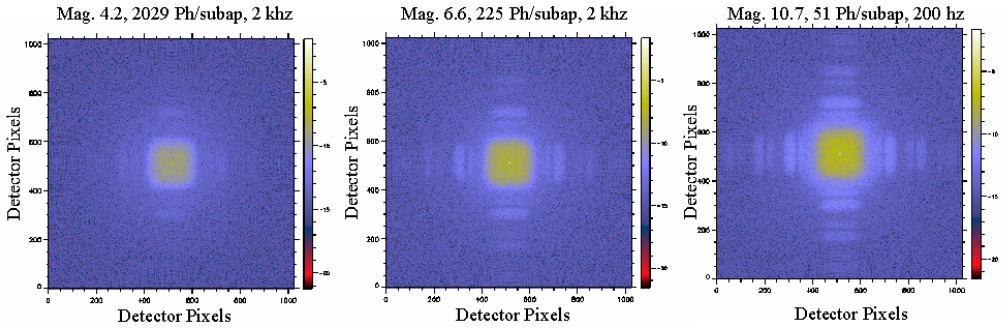

Figure 25: Log10(psf) for three different fluxes incident upon the pyramid wave-front sensor.

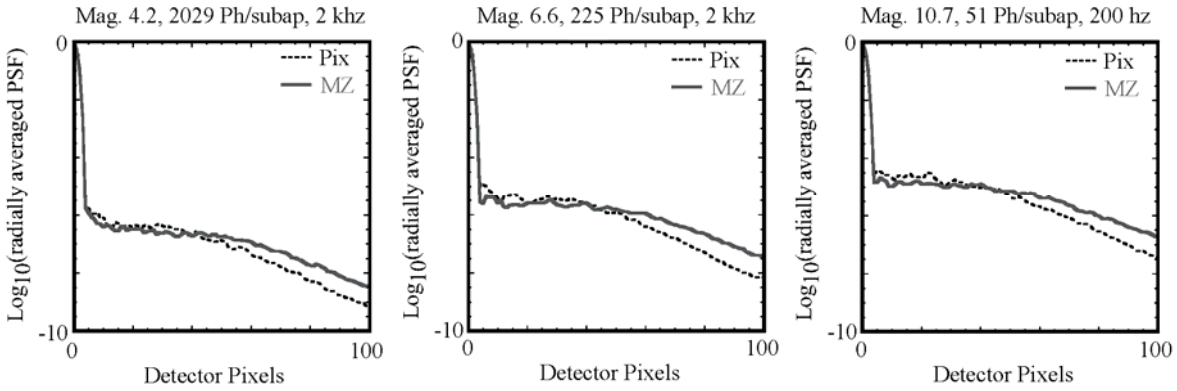

Figure 26: Log10 (radially averaged psf) for three different cases comparing the Mach-Zehnder configuration to the pixellated configuration. No spatial filter was used, so the PSF does not form a dark hole.

A comparison was made between spatially-filtered versions of the three wave-front sensors; the Shack-Hartmann, the pyramid and the Mach-Zehnder interferometric wave-front sensors. The wave-front sensors were compared for the case of G5 star at 10 pc, which had an apparent magnitude of 4.2 in the I/Z band. The TMT standard atmosphere was used and the simulations were carried out to 0.1 seconds. Again, far-field PSFs were constructed with a simple apodizer. The results for the three wave-front sensors are shown in Figure 27 below. In this graph, the Mach-Zehnder configuration is displayed as the dashed black line, the Shack-Hartmann sensor as the solid gray line and the pyramid wave-front sensor is shown as the solid black line. The spatially-filtered pyramid sensor provides a higher contrast than either the spatially-filtered Shack-Hartmann or Mach-Zehnder wave-front sensors. However, it should be noted that the pyramid sensor simulation implemented a lower delay and higher gain than the SH or MZ sensor; hence the difference is actually dominated by lower servo lag rather than the properties of the sensor itself. Still, it shows that the overall PSF structure/performance on bright stars is similar for all three sensors.

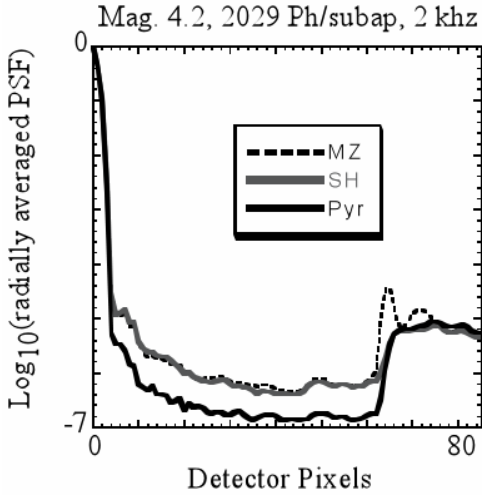

Figure 27: Comparison in instanteanoues contrast ratios for the three different wave-front sensors. These simulations use a simple apodizer. The Pyramid simulation had a lower latency and higher gain than the other two, so the effects on this bright star are dominated by different servo lag rather than the sensor properties.

3.2.4 Interferometers

3.2.4.1 Introduction

Interferometers measure the phase of the wavefront directly and as such do not require reconstruction of the wavefront when operating in the high Strehl ratio regime. The self-referencing interferometers appropriate for temporally incoherent light are not suited for self-bootstrapping and would therefore require a separate AO system to reduce the phase error sufficiently that enough light enters the pinhole for the reference beam which adds to the complexity and cost of such systems. Interferometers have been considered previously for use in adaptive optics systems aimed at detecting extra-solar planets.

3.2.4.2 Implementation

One such interferometric system was proposed by J.R.P. Angel and is a self-referencing interferometer in a Mach-Zehnder configuration [Angel 1994]. The interferometer is show below in Figure 28. In this interferometer the two detectors collect interferograms with the reference beams out of phase by 180 degrees between the detectors. By phase stepping the reference beam by 90 degrees, four channels can be obtained, each of the channels separated by 90 degrees relative to one another. This configuration, however, has non-common path errors between the signal and reference arms of the interferometer, is susceptible to noise vibrations due to the separate paths and requires careful coatings to achromatize the phase stepper.

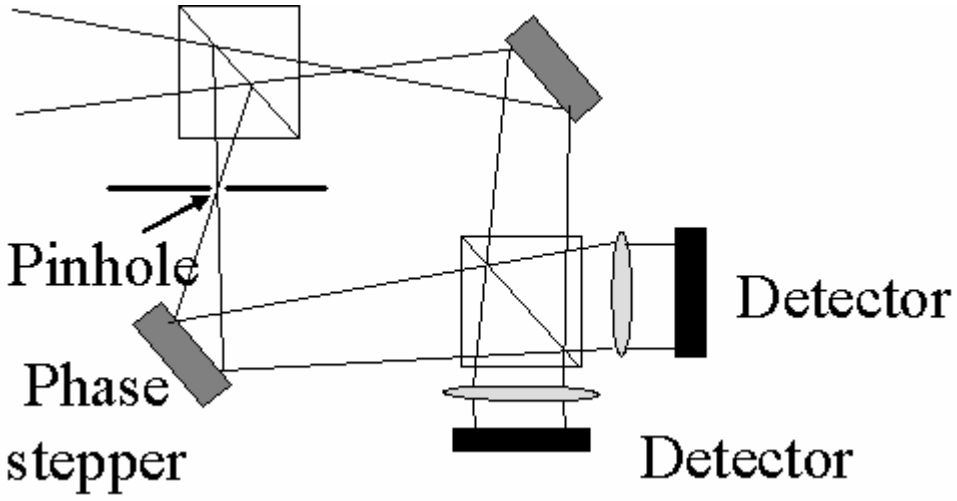

Figure 28: Self-referencing Mach-Zehnder configuration.

Recently a pixelated interferometer design based on polarization-induced phase shifting was developed (see for example the commercial interferometers marketed by the 4D Technology Corporation).[Millerd 2004] A version of this interferometer, in a self-referencing configuration, is shown below in Figure 29. In this interferometer, unpolarized light is incident upon a polarizing plate which allows vertically polarized light to pass through the center circle, ~ λ/D in diameter, and horizontally polarized light to pass through the outer annulus.(In this interferometer a square aperture is placed about the center and outer annulus to prevent aliasing of the measurement) The vertically polarized light which passes through the center forms the reference beam of the interferometer and the horizontally polarized light which passes inside the square aperture, aliasing filter, but outside the central λ/D diameter circle represents the aberrated signal passing through the atmosphere and the telescope optics. These two linearly polarized beams are then collimated with a focusing optic and passed through an achromatic quarter wave-plate. As the two linearly polarized beams pass through the quarter wave plate they become circularly polarized, one right circularly polarized and one left circularly polarized. These two orthogonally-circularly-polarized beams are then passed through a polarizer whose polarization angle with respect to vertical is at an angle, α. When left and right circularly polarized beams pass through a polarizer, their phases are advanced and retarded by the angle α. After passing through the polarizer, the relative phase shift between the two left

and right circularly polarized beams is 2α, independent of the wavelength of the electromagnetic wave. This represents an achromatic method of supplying a spatially dependent or pixelated phase shift between the reference and the atmospherically aberrated electromagnetic waves. An array of polarizers is constructed with the individual size of each polarizer matched to the CCD detector pixel size, which in turn represents the size of a sub-aperture in the pupil. The individual polarizers in this array have their polarization placed at the appropriate angle such that the desired pixelated phase shift is obtained(See Figure 31). This design does not have non-common path errors between the reference and aberrated beams and hence is less susceptible to vibrations and phase stepping errors. It does, however, incur a loss of 50% of the light due to the use of polarizers. This loss can be partially recovered by lower sampling of the pupil as discussed below.

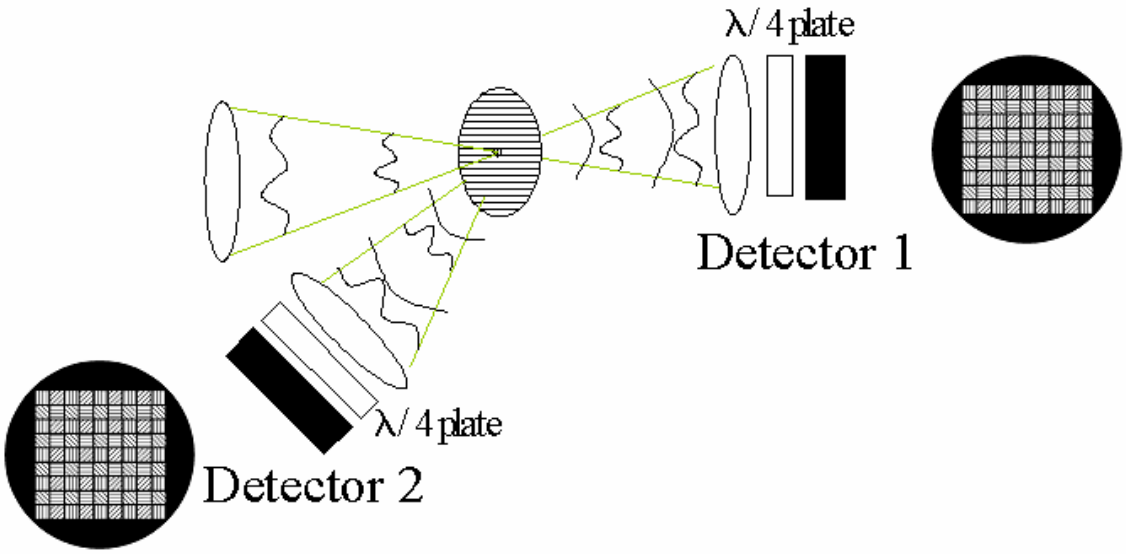

Figure 29: Pixellated self-referencing interferometer. In transmission, the circular beamsplitter plate horizontally polarizes all the light except for a central circle, which is vertically polarized; the opposite is true in reflection. The quarter waveplate circularly polarizes the beams that interfere in the detector plane. A grid of polarizers on the detector selects different phase shifts in each pixel of a group of four.

3.2.4.3 Simulations

In the case of a four-bin interferometer errors in the phase steps added to the reference beam will lead to errors in the reconstructed phase. In the case of the Mach-Zehnder approach these errors in the phase step would arise because of the imperfect chromatic correction of the multilayers placed on the phase stepper and on errors in the overall piston of the phase stepper itself (although the latter would not randomly distributed). In the case of the pixellated interferometer, these errors would arise from mechanical tolerance of how accurately the angle of the polarizers were set. This error source was evaluated analytically by adding random perturbations of α, β, γ and δ to the phase steps. The effective change in the measured phase, ζ, can then be found by taking the standard formula for the reconstructed phase, $\text{Tan}(\phi+\zeta) = (I_4-I_2)/(I_1-I_3)$, from the four channels and making the assumption that the random phase perturbations added (α, β, γ and δ) to the channels (I_1, I_2, I_3 and I_4) are small. Keeping terms of the order of the random phase perturbation, the effective change in the measured phase is found to be $\zeta=(\delta+\beta)/2$. The phase perturbations δ and β are assumed to be random such that the magnitude of the change in the measured phase would be expected to be $\zeta \sim \delta/\text{SQRT}(2)$. Assuming that the errors in the phase measurement are uniformly distributed spatially across the measurement, then the variance, σ^2, may be expressed by

$$\sigma^2 \approx \frac{1}{\zeta}\int_0^{\zeta}\left(\phi-\frac{\zeta}{2}\right)^2 \mathrm{d}\phi = \frac{\zeta^2}{12} \qquad (3.2.4.3.1)$$

The variance as a result of random phase perturbations in the phase steps would then be given by $\sigma^2 \sim \delta^2/24$, where δ represents the magnitude of the random phase perturbations added to each of the phase steps. This was also checked in simulations, the results of which are shown below in Figure 30.

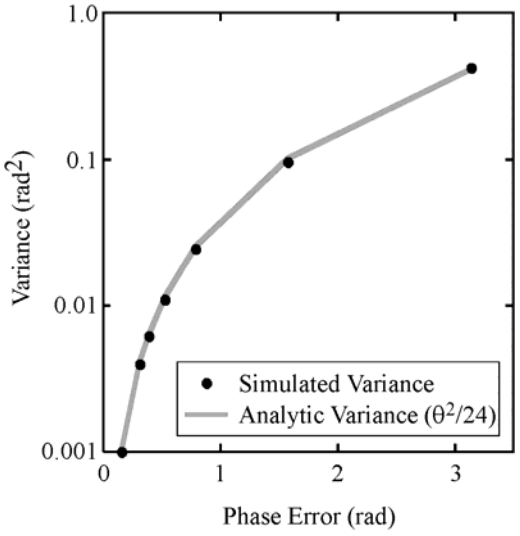

Figure 30: Variance as a function of random phase errors in the phase step arising from miscalibration of or dispersion in, for the case of broadband illumination, the phase stepping device.

The phase shifting scheme for the pixellated self-referencing interferometer is shown below in Figure 31. In this case, detector 1 measures the 0 and π phase shifts between the reference and signal beams and detector 2 measures the π/2 and 3π/2 phase shifts between the reference and signal beams. The phase can then be measured on a grid as shown in the right image of Figure 31, where the phase measurement is on a pitch that is smaller than the sub-aperture size. The phase measurement can also be reconstructed on the face of each sub-aperture using a five pixel cross pattern centered on the sub-aperture. This interferometer would require the construction of a polarization/filter mask

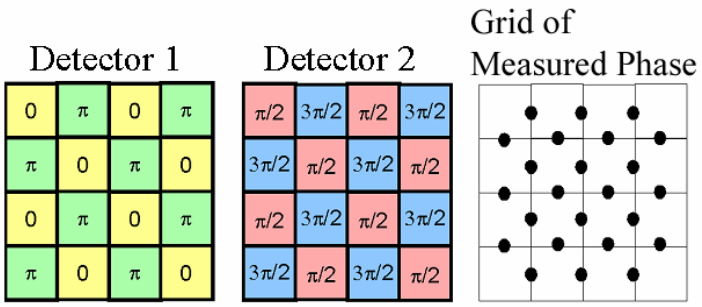

Figure 31: Pixellated phase shifts and wave-front measurement grid.

The interferometric wave-front sensor would require four CCD pixels/phase measurements per sub-aperture in the case of the Mach-Zehnder configuration and would have essentially two CCD pixels/phase measurements per sub-aperture for the pixellated interferometer. In the case of the Mach-Zehnder interferometer, the phase is determined by using the reconstruction formula, $\mathrm{Tan}(\phi+\zeta)=(I_4-I_2)/(I_1-I_3)$, where the I_1, I_2, I_3 and I_4 represent the measured intensities in the four CCD pixels. In the case of the pixellated interferometer, the reconstruction formula takes on the form

$$\mathrm{Phase}(I_{4a}) = \mathrm{ATAN}\left\{\frac{I_{4a} - 0.25(I_{2a} + I_{2b} + I_{2c} + I_{2d})}{I_{1a} - 0.25(I_{3a} + I_{3b} + I_{3c} + I_{3d})}\right\}, \qquad (3.2.4.3.1)$$

as detailed below. The intensities/phase steps are taken from a five pixel cross with the sub-aperture in the center of the cross.

The 4-bin pixellated interferometer can reconstruct the phase at less that the pitch of individual pixels composing the sub-aperture as seen in Figure 31. There is, however, a loss of effective resolution of this pixellated interferometer because of the need to use multiple pixels in different spatial locations. As detailed below, an effective spatial resolution of 1 sub-pixel with high sampling to 1.2 sub-pixels when the phase is sampled with just four sub-apertures per wavelength is achieved.

In the case of a pixellated spatial phase shifting interferometer, a macro-pixel consisting of multiple pixels with different phase delays is used to reconstruct the phase and the spatial resolution of the interferometer is reduced from that achievable by an individual pixel. For a four-bin phase mask with two detectors, the phase can be reconstructed at the centers of adjacent pixels where all of the four separate phase pixels comprising a macro-pixel touch each. The phase reconstructed in this manner has a pitch equivalent of 0.71 pixels within the macro-pixel. To obtain the phase in the center of the pixels, an averaging of the phase, which involves phase information from five pixels, including the reconstructed pixel, in a cross pattern, must be performed as quantified by Eq. 3.2.4.3.1.

The spatial frequency that the phase can be reconstructed was evaluated by reconstructing a sinusoidal phase profile using the algorithm above. The sinusoidal phase profile was 1024 by 1024 simulation pixels across. An individual sub-pixel on the simulated detector was nominally 8 simulation pixels across or 128 by 128 sub-pixels. The residual variance of the reconstruction algorithm, given a sub-aperture size of 8 simulation pixels, was compared with the residual error calculated using a test four bin interferometer in which the phase was reconstructed at the macro-pixel level. This test four bin interferometer calculated the phase using four different reference phases at each macro-pixel and incremented the macro-pixel size discreetly in simulation pixels from 2×2 to 32×32 to determine the relative spatial resolution achievable by the algorithm given above. The effective spatial resolution of the algorithm as a function of the spatial sampling of the sinusoidal phase is shown in Figure 32. The solid black line represents the algorithm above and ranges from an effective spatial resolution of 1 subpixel with high sampling to 1.2 subpixels when the phase is sampled with just four sub-apertures per wavelength.

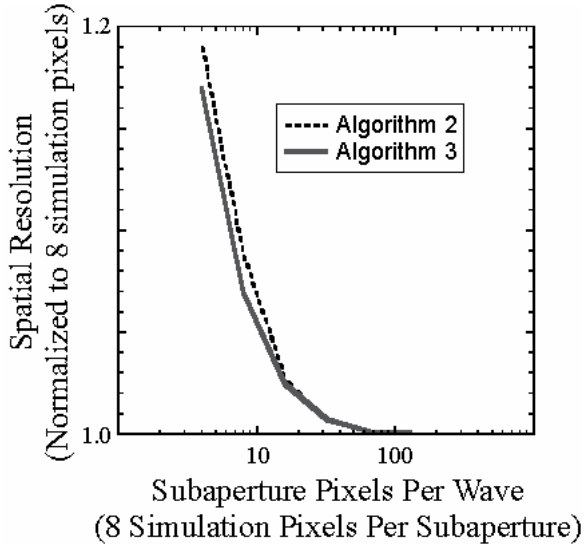

Figure 32: Effective spatial resolution of the pixellated phase shifting interferometer.

To look at the noise for the two interferometer configurations, simulations were performed with no phase aberration entering the interferometer. A zero phase aberration signal with the appropriate pupil was sent through a spatial filter that was set up to remove those components that could not be measured by the interferometer/corrected by the DM. The spatially-filtered signal was then passed through the interferometric wave-front sensors with the appropriate photon noise and read noise added to the simulated CCD measurements. The residual root-mean-square error for the simulations, with a read noise of 1 e-, was found to be 0.0308 rad, 0.093 rad, 0.210 rad for the three cases of 2029, 225 and 51 photons/sub-aperture respectively. That is in excellent agreement with the analytic values expected from Eq. 3.2.2.3 above of 0.0314 rad, 0.095 rad and 0.206 rad for the Mach-Zehnder interferometer with the photons/sub-aperture reported above. A Blackmann apodizer was used to evaluate the achievable contrast ratio in these interferometer simulations. The apodized pupil along with the residual phase was Fourier transformed to form the field at the Fourier plane. The intensity was found from this field and the results for the two interferometers are shown in Figure 33. The top row shows the $\log_{10}$(intensity) for the Mach-Zehnder interferometer with the three photon levels and the bottom row shows the same quantity for the pixellated configuration. Both of these plots have been normalized by the maximum of the signal with zero phase.

The radially averaged point-spread functions were then found to provide a more detailed comparison between these configurations. The intensity was radially averaged and then the log base ten of the radially averaged intensities were taken. The results for the three cases above are shown in Figure 34. In each of these graphs, the Mach-Zehnder configuration is displayed as the solid gray line and the pixellated configuration is displayed as the dashed black line. In general, the Mach-Zehnder give slightly better performance closer to the core of the psf with the pixellated design giving better performance further away.

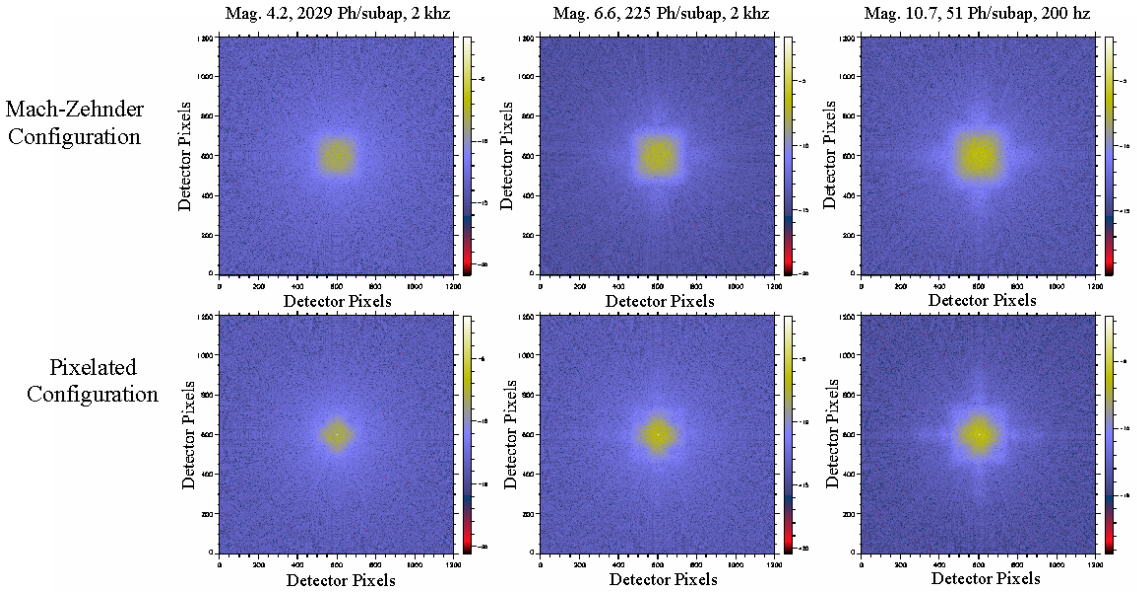

Figure 33: Log10 (psf) for three different cases comparing the Mach-Zehnder configuration to the pixellated configuration.

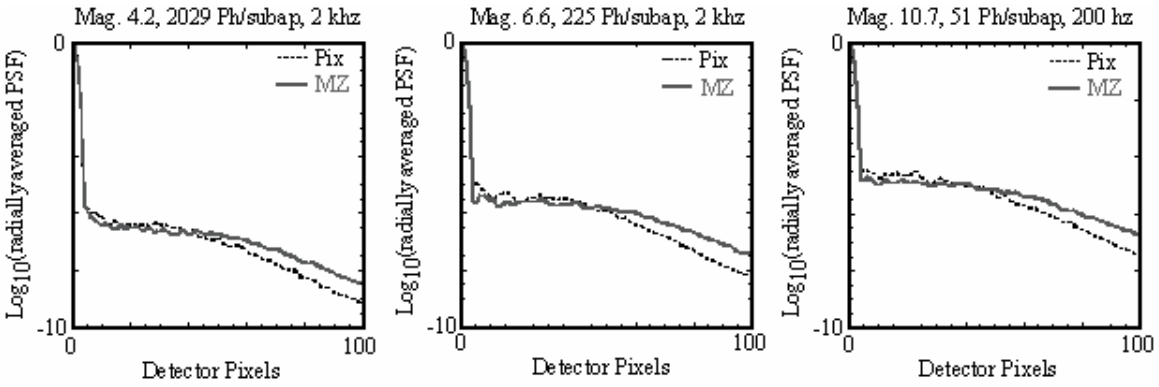

Figure 34: Log10 (radially averaged psf) for three different cases comparing the Mach-Zehnder configuration to the pixellated configuration.

3.2.5 Conclusions

Of the sensors presented, the pyramid (or ridge) sensor is the most attractive, with very good noise behavior, high dynamic range, and no moving parts, and is the baseline for PFI. The pixellated interferometer is our second choice; it also has very good noise behavior and robustness in implementation, though it has never been tested in an astronomical context. Over the next several years, we will prototype both sensors. A lower-risk descope option (at the cost of performance) is a spatially-filtered Shack-Hartmann sensor.

3.3 Real-time computer for the front AO system

3.3.1 Introduction

The dual-ridge interferometric WFS and 128 × 128 DM being proposed for the PFI instrument will require approximately an order of magnitude more real-time processing power than even the most aggressive high-contrast adaptive optics being developed today. There was, therefore, some concern about the feasibility of a computer to support these real-time requirements. To allay these concerns, we have done a strawman design for a computer using currently-available or near-future components. This computer nearly meets the PFI frame rate requirements, being able to process ~1630 frames per second (with one frame latency) even with current components. It does, however, require custom printed-circuit board development, 1 GHz FPGAs (500 MHz is currently available) and advances in DM integrated drive electronics (either on chip or on the DM carrier board). As we get closer to the actual development of the PFI instrument, with more sophisticated processors or FPGAs we expect we will easily meet the full frame rate requirements and the need for custom circuit design will decrease.

The assumptions that guided the real-time AO computer design are given in Table 5.

Table 5: AO real-time computer requirements.

	Assumption
	WFS
1	the WFS will be a dual-ridge configuration, operating in interferometric mode the two ridges will be considered to be perpendicular, with the X leg measuring everything except pure Y components of the phase and the Y leg measuring everything except pure X components. Both phase and amplitude will be calculated
2	the four pupil images will be imaged onto 4 separate 144x144 pixel CCDs or larger, giving an 8 pixel buffer around the nominal pupil; if larger arrays are used, it is assumed that the 144x144 ROIs can be read out the 4 CCDs will be designated XL (left), XR (right), YT (top), and YB (bottom) if so desired, the four pupil images could be combined into 1-2 256x256 CCDs
3	each 144x144 pixel CCD will have 36 amplifier taps able to operate at up to 6 MHz each (or, fewer taps able to operate faster)
4	the CCDs will have 14-bit resolution
	DMs
5	Tip/tilt will be provided by mounting the woofer mirror on a tip/tilt stage
6	there will be a ~900 actuator woofer with about 700 actuators illuminated
7	there will be a 128x128 MEMS tweeter with about 13k actuators illuminated
8	there will be drive electronics integrated into the MEMS mounting board or MEMS chip, with a 28-bit wide communication channel able to run at 128 MHz (the 28 bit wide data channel includes 14-bits of address and 14 bits of data)
9	the DM drive electronics will have 14-bit resolution
	Algorithm
10	the required loop update rate is 2 kHz (goal 4 kHz)
11	boundary conditions can be handled algorithmically rather than with significant padding for this discussion, the boundary is 8 pixels

3.3.2 Computer Design

The real-time computer consists of several basic modules connected on custom PCBs. Figure 35 shows the strawman real-time AO computer design. The computer is parallel in nature so that data from the CCDs can be processed in parallel to the degree possible. This becomes less possible as the algorithm progresses. As this happens, the processors will work in cooperation rather than in parallel. The computer is also distributed, meaning that processing is distributed across several specialized processors: CCD Processors for preprocessing CCD frames, main X and Y processors for the two channels of the ridge sensor, and high-speed FFTEngines.

CCD Processors will, in all likelihood, be FPGAs. These processors will perform rudimentary pixel pre-processing in real time and pass the results onto the main Processors by DMAing into the main processors' RAM.

The main **XProcessors and YProcessors** will be general-purpose processors such as Intel 64-bit XEONs. (We have assumed a 4 GHz Xeon for this document; 3.4 GHz is currently available.) XProcessor 1 will be the master of the subsystem, coordinating overall activity when required. When pixel data is received from the CCD Processors, the main processors will run through the algorithm steps listed in Table 5. The main processors will work in parallel when possible and in cooperation when necessary, offloading to the FFTEngines as appropriate. When the four main processors are working on their respective X and Y data and must work cooperatively, they will normally each work on half a matrix. When data has to be shared (such as when transposing a matrix), the processors will access each other's RAM. When the X and Y data have been combined, the XProcessors will perform the remainder of the processing (as the design progresses, this processing can be better balanced so that the YProcessors do not sit idle during this time).

The **FFTEngines** will be high-performance FPGAs running high-speed FFT cores, such as Xilinx Virtex-II Pro FPGAs running rfengines' HyperSpeed FFT cores. Because FPGAs can perform FFTs in a massively parallel manner, processing-time becomes negligible and data-transfer time becomes the limiting performance factor. This design takes advantage of that fact by setting up an FFT pipeline, DMAing data in and out of the FPGA as fast as possible. Once the pipeline has been primed, by transferring in data for the first FFT, FFTs can be performed as fast as the data can be moved. Additional processing required for 2d FFTs will be done by the main processors as described in Table 7.

The final results of the algorithm processing will be DMAed to the DM's integrated drive electronics. It is possible that the DM electronics could take on more processing responsibility (i.e., applying linearization corrections) but that is not included in this design.

The core software architecture for the main processors for this design will follow other recently developed adaptive optics computers, which are modular in nature. This modularity lends itself well to distributing functionality across the components of the computer. The FPGA cores will be purchased and tailored to the specific needs of the AO computer.

3.3.3 Processor Requirements

The estimated processing requirements for the PFI real-time AO computer are given in the tables below. The processors and data paths referenced in the tables are shown in Figure 35.

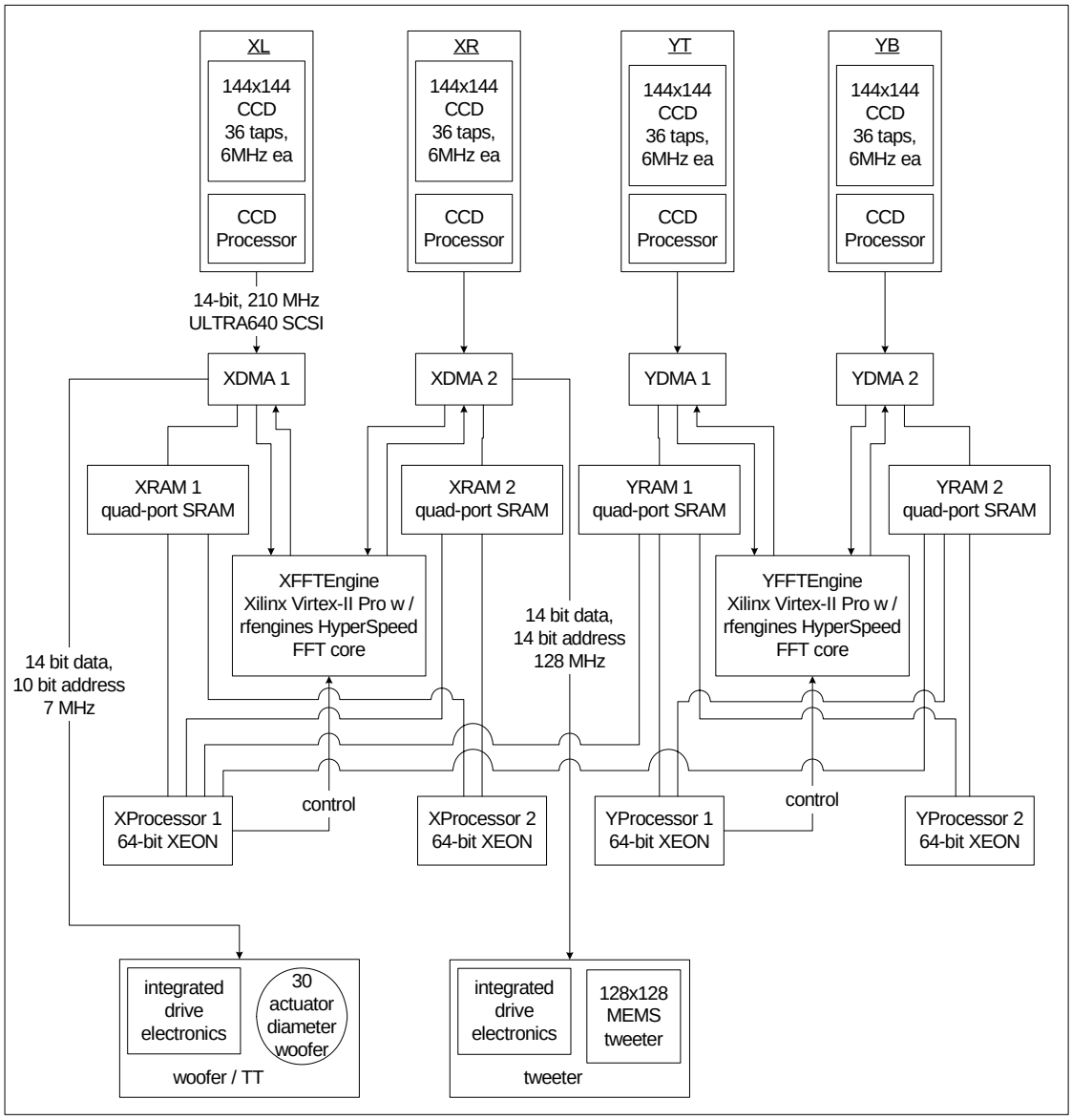

Figure 35: Simplified block diagram of hardware architecture using near-commercially-available components, delivering ~1630 frames per second performance with 1 frame latency from end of CCD integration.

Notes that go along with the tables are:

Algorithm (Table 6):

1. all data given in the tables are for a single 500 microsecond cycle
2. most values given are approximate (most rounded up, a couple rounded down)
3. it is assumed that the XProcessors and YProcessors can execute 1.5 GFLOP per second each
4. only pixels in the illuminated portion of the CCD will be processed, where appropriate
5. in the tables, 'real-time' means that an operation can be performed while data is streaming with virtually no additional cost in time
6. refinements, such as automatic gain optimization, are not included
7. no telemetry or diagnostic data transfer or storage is included; in the final design, additional processors (or more powerful processors) will be used to pick diagnostic data off

FFT (Table 7 and Table 8):

1. all FFTs will be performed on high-performance FPGAs
2. FPGAs can perform FFTs in a massively parallel manner, meaning that an FFT can be performed in log(n) time steps
3. this makes FFT processing-time negligible and makes transferring data into and out of the FPGA the limiting performance factor

Table 6: Algorithm steps with estimates of data transfer and processing requirements (note that rates and times are given for two processors combined).

	Step	Transfers	Path	Rate, MHz	Ops	FFTs	Processor	Time, s
	Pre-process pixels							
0	Subtract backgrounds from pixels in XL and XR				21k x 2		2 CCD processors	real-time
1	Multiply pixels in XL and XR by flat-field values				21k x 2		2 CCD processors	real-time
2	Transfer pixels from X CCDs to RAM	21k x 2	DMA over 2 ULTRA640 SCSI buses	210 per bus				100
3	Sum the intensities in LH and the intensities in RH, compute the ratio LH_Sum/RH_Sum, and multiply all the intensities in RH by this ratio (to even out the energy in the two sides)				48k		XProcessors	16
4	Normalize LH and RH, combined, to 1				32k		XProcessors	11
	Compute Amplitude							
5	Calculate sqrt(LH) – sqrt(RH) for each pixel				48k		XProcessors	16
	Compute Phase							
6	Calculate LH–RH for each pixel				16k		XProcessors	6

	Step	Transfers	Path	Rate, MHz	Ops	FFTs	Processor	Time, s
7	Apply Hilbert transform to the result:							
7a	Perform 144x144 FFT					one 2d 144x144	XFFTEngine	34
7b	Shift FFT result to center	16k	XProcessors	8000				2
7c	Shift the left side a quarter-pixel to the left and the right side a quarter-pixel to the right:							
7ca	Perform 72-element 1d FFTs for each row of the left side and each row of the right side					288 1d 72 element	XFFTEngine	11
7cb	Apply phase terms to the left rows to shift left and the right rows to shift right				96k		XProcessors	32
7cc	Perform 72-element 1d inverse FFTs for each row of the left side and each row of the right side					288 1d 72 element	XFFTEngine	11
7d	Multiply the left side by –I and the right side by i				32k		XProcessors	11
7e	Perform 144x144 inverse FFT					one 2d 144x144	XFFTEngine	34
7f	Shift to center	16k	XProcessors	8000				2
	Process Y And Combine							
8	Repeat steps 1 through 7f for Y, substituting Y for X, YT for XL, YB for XR, top for left, bottom for right, column for row, YProcessors for XProcessors, and YFFTEngine for XFFTEngine						YProcessors & YFFTEngine	in parallel with X
9	Calculate means of X and Y amplitudes				32k		XProcessors	11
10	Remove pure X phase components from X and pure Y phase components from Y (these will go to the TT stage)				34k		XProcessors	12
11	Calculate means of remaining X and Y phases				32k		XProcessors	11
	Apply Influence Functions And Convert to DM-space							
12	Perform 144x144 FFT					one 2d 144x144	XFFTEngine	34
13	Shift to center	16k	XProcessors	8000				2
14	Divide between woofer and tweeter (spectrally, for now (more thought required, here))	16k	XProcessors	8000				2
15	Multiply woofer frequencies by woofer influence function Fourier filter				8.4k			3
16	Multiply tweeter frequencies by tweeter influence function Fourier filter				87.6k			30

	Step	Transfers	Path	Rate, MHz	Ops	FFTs	Processor	Time, s
17	Perform 144x144 inverse FFT for woofer					one 2d 144x144	XFFTEngine	34
18	Shift to center	16k	XProcessors	8000				2
19	Perform 144x144 inverse FFT for tweeter					one 2d 144x144	XFFTEngine	34
20	Shift to center	16k	XProcessors	8000				2
21	Apply DM actuator calibration curves to convert from phase to Volts				70k		XProcessors	24
22	Send commands to TT mirror and DMs (both address and data)	28k	DMA over 2 ULTRA640 SCSI buses	210 per bus				124
	Total:							611

Table 7: two-dimensional 144x144 FFT steps with estimates of data transfer and processing requirements (note that FFTEngine is fed by two data paths).

	Step	Transfers	Path	Rate (MHz)	Ops	FFTs	Processor	Time (s)
0	Transfer all 144 rows from RAM to FPGA	21k	DMA	2000				11
1	Perform FFTs on each row as it is received					144 1d 144 element	FFTEngine	real-time
2	Transfer rows from FPGA to RAM as they are completed	21k	DMA	2000				real-time
3	Transpose matrix	42k	XProcessors	8000				6
4	Transfer all 144 rows from RAM to FPGA	21k	DMA	2000				11
5	Perform FFTs on each row as it is received					144 1d 144 element	FFTEngine	real-time
6	Transfer rows from FPGA to RAM as they are completed	21k	DMA	2000				real-time
7	Transpose matrix	42k	XProcessors	8000				6
	Total:							34

Table 8: 288 one-dimensional 72-Element FFTs steps with estimates of data transfer and processing requirements (arrays must be contiguous; note that FFTEngine is fed by two data paths).

	Step	Transfers	Path	Rate (MHz)	Ops	FFTs	Processor	Time (s)
0	Transfer all 288 arrays from RAM to FPGA	21k	DMA	2000				11
1	Perform FFTs on each array as it is received					144 1d 72 element	FFTEngine	real-time
2	Transfer arrays from FPGA to RAM as they are completed	21k	DMA	2000				real-time
	Total:							11

3.3.4 Conclusion

The strawman design presented here, that nearly meets PFI real-time AO processing requirements with slight extensions of currently available components, gives us a very high degree of confidence that, when the time comes to build the PFI instrument, we will be able to meet its AO processing requirements. The amount of custom hardware development that will be required will have to be assessed when the time comes.

3.4 Deformable Mirror Specifications

3.4.1 Deformable Mirrors

3.4.1.1 Mirror Technologies for the Tweeter

Several technologies exist for the tweeter technology to achieve the actuator number densities required. These technologies include a modular high-density approach to conventional mirror technologies being developed at Xinetics and MEMS technologies being developed at Boston Micromachines Corporation (BMC).

The "Photonics Module" mirrors being developed at Xinetics consist of individual modular units of 32-40 mm. The DMs have been demonstrated with both 1.0 and 2.5mm pitches. This technology has been used at the High-Contrast Imaging Testbed (HCIT) at the Jet Propulsion Laboratory to achieve contrast ratios of better than 10^{-9}. In this case, the DM had 64 by 64 actuators with a 1mm pitch and an inter-actuator stroke of ~200 nm. This is insufficient for ground-based AO applications even in a woofer-tweeter configuration. The DMs with 2.5 mm pitch have over 2 microns of stroke and are better matched to ExAO. This technology is scalable to 128 by 128 actuators and beyond and can reach the required speeds of >5 kHz. However, the 2.5 mm device would result in a pupil size of 32 cm, requiring large optics and path length. Purchasing the nm-level optics PFI requires for its science mission (section 9.1.10) would be prohibitively expensive.

Continuous face sheet MEMS mirrors currently being sold at Boston Micro-machines Corporation (BMC) have inter-actuator spacings of 340 μm and inter-actuator strokes of ~1 μm, full stroke of ~ 2.0 μm. These devices have 32 by 32 actuators and they have been used to achieve corrected in-band RMS errors of ~0.5 nm at the Laboratory for Adaptive Optics in Santa Cruz. BMC is currently under contract to produce 64 by 64 devices in the next couple of years with 4 μm of stroke and an inter-actuator stroke of 2 μm for the Gemini Planet Imager. This architecture could in principal be extended to 128 by 128 actuators, possibly by butting four of the 64 by 64 devices together, with moderate wiring routing changes.

In current MEMS devices, each individual actuator has its own input connection on the chip carrier, circuit board, and cables leading to rack-mounted drivers. This approach would work (marginally) for a 128 × 128 actuator device but is cumbersome. A simple intermediate step would incorporate multiplexing on the MEMS carrier board; since MEMS require little driving current this seems practical. On-chip multiplexing is an even more attractive option. Segmented MEMS mirrors with actuator counts of 256 by 256 are being developed at Lucent technologies with on-chip electronics under a DARPA contract. These devices are being developed with modular 64 by 64 electronics and are capable of 6 μm of piston motion and $\pm$ 10 degrees of tilt motion. That indicates that MEMS devices are scaleable to the required number of actuators with on-chip electronics, however, the segmented MEMS mirror are not desirable for this particular application due to the diffraction from the segment gaps. Several MEMS manufacturers are working on other on-chip multiplexing approaches and it seems likely that on TMT timescales such devices will be available.

3.4.2 Tweeter requirements

3.4.2.1 Adjacent Actuator Stroke Requirements due to Atmospheric Turbulence

The stroke requirements for the deformable mirrors are determined by the turbulence profile expected for the thirty meter telescope or more specifically the overall optical path difference. The assumed Von Karman turbulence parameters for the TMT telescope are as follows: the Fried parameter, r_o, at 1.65 μm is r_o=0.63 m (r_o=0.15 m at 500 nm), the outer scale is $L_o = 25\ m = 2\pi/\kappa_o$, the TMT telescope pupil is 30 m and the sub-apertures are assumed to be 0.23 m apart. This is also done for a Kolmogorov turbulence spectrum in which the outer scale is much greater than the pupil diameter. A phase screen is generated with the above specified parameters to determine the overall required stroke. The generated phase screen is shown below in Figure 36.

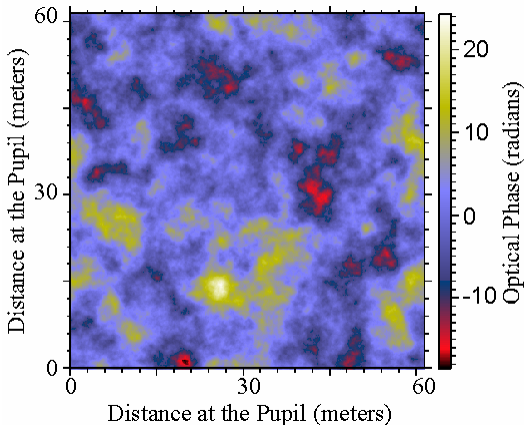

Figure 36: Von Karman phase screen generated with the expected atmospheric conditions for the TMT telescope. The overall peak/valley phase over the TMT aperture is about 30 rad at 1.65 μm.

The structure function for the Von Karman phase screen shown in Figure 36 above was calculated and compared with analytic formulas for the Von Karman turbulence to verify that the phase screen contained the expected parameters. The figure below, Figure 37, has the measured structure function from the simulated phase profiles drawn as the solid black line. A Kolmogorov structure function, with r_o=0.63 m, is drawn as the dashed black line. The analytic structure function for a Von Karman turbulence profile, with r_o=0.63 m and $\kappa_o = 2\pi/ L_o$ (L_o = 25 m), is drawn as the dashed red line and a first order expansion of the Von Karman turbulence profile drawn as the dashed green/yellow line. The good agreement with the measured, solid black line, and the analytic Von Karman line, dashed red line, shows that the phase screens exhibit the correct turbulence parameters.

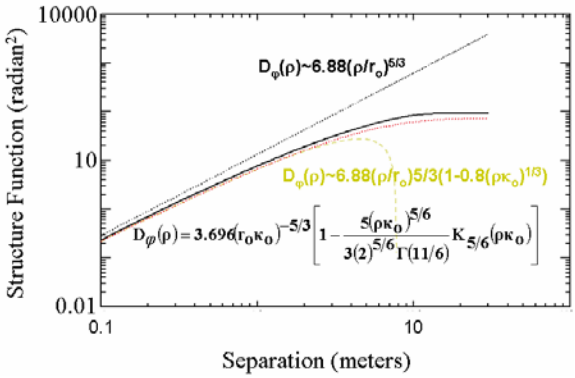

Figure 37: Structure function calculation for the Von Karman turbulence screen shown in Figure 3.4.2.1.1 above. The y-axis is the structure function in rad2 and the x-axis is the spatial separation between the shifted phases in meters. The assumed parameters are ro=0.63 m at 1.65 μm and κo = 2π/ Lo (Lo = 25 m).

To obtain the required stroke for the tweeter mirror, the phase is averaged over a sub-aperture and each sub-aperture in the pupil is compared with the surrounding eight sub-apertures to determine the largest difference between that sub-aperture and its adjacent neighbors. This was repeated for both a Von Karman and Kolomogorov spectrum with the Kolmogorov turbulence spectrum giving the larger excursion between adjacent actuators. The resultant adjacent actuator throw requirement is shown in Figure 38 below. This results in a maximum optical path difference of ~3.3 microns indicating that the tweeter mirror will require an adjacent actuator stroke of ~1.6 μm for this simulation. An adjacent actuator stroke of ~2 μm would allow for cases of stronger atmospheric turbulence to be treated.

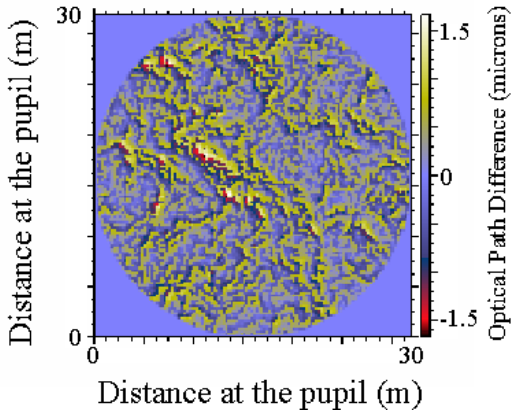

Figure 38: Phase jump between sub-apertures when the original phase screen is averaged over eight pixels such that the pixels in the averaged sub-aperture have a 0.23 m pitch (30 m/1024 px)(8 px). The maximum optical path difference is ~3.3 μm.

3.4.2.2 Number of Tweeter Actuators Required (Wave-front Fitting Error)

Detailed evaluation of contrast performance vs subaperture size requires end-to-end simulations, which are prohibitively difficult for a dense grid of subaperture sizes, frame rates, and star brightness. To first order, increasing the subaperture count does not increase the achievable contrast – the depth of the dark hole – which is dominated by servo lag errors (section 9.1); instead, it merely increases the width of the dark hole λ/d. As a result, contrast is a weak function of subaperture size, and the optimal subaperture size is a balance between flux of science targets, performance of the spatial filter in the wavefront sensor, and plausible technology. Analytic performance calculations show that for all but the brightest stars the optimal subaperture size for most targets is in the 15-25 cm range; we have assumed a conservative d=23 cm (128 actuators across the TMT mirror) for this study, but can revisit this if higher actuator count mirrors appear likely to be available. One option would be to use a higher actuator count DM for the front AO (which provides fast correction on bright stars) than the post-DSS AO system.

3.4.2.3 Actuator Speed Requirements (Temporal Error)

The mean square error due to a time delay, assuming a Von Karman power spectrum, may be expressed as

$$\sigma^2 = \frac{2}{\pi}\left(\frac{f_g \pi \tau'}{y}\right)^{5/3}\left[\frac{1}{2}\beta\left(\frac{1}{2},\frac{5}{6}\right) - \frac{y^{5/6}\Gamma(-1/3)\cos(-5\pi/6)K_{5/6}(2y)}{\sqrt{\pi}}\right], \qquad (3.4.2.3.1)$$

where y is equal to $\kappa_o v_\perp \tau'/2$. The mean square error, in the limit that y is much less than 1, reduces to $\sigma^2 \sim 28.44(f_g\tau')^{5/3}\{1\text{-}0.8(v_\perp\tau'\kappa_o)^{1/3}+0.14(v_\perp\tau'\kappa_o)^2+\cdots\}$. Assuming $v_\perp\tau'\kappa_o << 1$, the mean square error may be expressed as $\sigma^2 \sim 28.44(f_g\tau')^{5/3}\{1\text{-}0.8(v_\perp\tau'\kappa_o)^{1/3}\}$ or $\sigma^2 \sim 6.9(v_\perp\tau'/r_0)^{5/3}\{1\text{-}0.8(v_\perp\tau'\kappa_o)^{1/3}\}$ given our assumption that the transverse wind velocity is constant along the propagation path such that the Greenwood frequency reduces to $f_g = 0.43 v_\perp / r_0$. In the limit that $\kappa_o = 0$, these results reduce to those obtained for a Kolmogorov power spectrum.

The optimal speed desired from the tweeter mirror can be found by optimizing the overall system variance taking into account both the time delay error and the mean square error due to signal-to-noise in the detector. Taking the brightest star of interest, a G5 star at 10 pc, the initial AO system would be expected to operate with approximately 4000 photons per sub-aperture. For a fixed sub-aperture size of 0.232 m, the optimal frequency would be approximately 4.5 kHz which would set the upper limit on the speed required from the tweeter mirror. Studies using the analytic error budget (Section 9.1)support this conclusion. The next generation of Boston Micromachines high-stroke DMs have a effective bandwidth of ~2500 Hz, due primarily to air damping at sea level. Performance at observatory altitudes may be higher; if PFI retains the 4 kHz requirement we will evaluate this.

3.4.2.4 Surface Finish Requirements for High-Contrast

Boston Micromachines MEMS devices have surface print-through on the actuator scale, as well as features such as small regular etch holes. To first order, any periodic actuator structure scatters light to a scale of λ/d_{act}, outside the dark hole region. To second order, PSF "beating" effects can produce small static speckles inside the dark region. Numerical simulations for a circular aperture show that this effect is acceptable for actuator-scale surface finish of <5 nm RMS.

3.4.3 High-stroke (Woofer) deformable mirror

The high-stroke DM has the responsibility of removing low-frequency wavefront errors outside the stroke capabilities of the MEMS tweeter. The exact balance between woofer actuator density and tweeter stroke depends on the technologies used. For the current design we have set a woofer with ~30 actuators across the primary mirror at ~5 mm pitch, simulations carried out for GPI indicate this is more than sufficient to reduce wavefront phase to the control range of a MEMS tweeter. This large woofer mirror dominates the size of the whole initial AO relay, however, and a more compact woofer would be attractive. If the TMT facility adaptive secondary is available at PFI first light, the woofer could be replaced by a smaller tip/tilt only mirror, reducing the instrument size.

3.5 Wave-front Sensor Camera

The WFS CCD has the following requirements:

- High quantum efficiency in the 0.7–1.1 μm range (spatially-filtered interferometric wavefront sensor performance improves with longer sensing wavelength)
- On-chip binning for low light level cases such as dim T Tauri stars
- Size >128 by 128 pixels for interferometers, 256 × 256 for Shack-Hartmann
- Frame rate up to 4 kHz
- Readnoise < 3 electrons.

The German company PN sensors offers a device, originally designed for X-ray applications, that is well matched to these requirements. The quantum efficiency of their detectors is greater than 95% from 700 to approximately 1000 nm for their 500 μm thick devices. The device operates at extremely high bias voltage, so that charge diffusion is minimal even for this thickness. The size of the focal plane arrays are currently 256 by 256 pixels and can be binned in 1D at ×2, ×3 and ×4. The devices currently have a read noise of 2.3 e- at 1100 frames-per-second (fps) with no binning and a read noise of 2.3 e- at 2200 fps with 2x binning. The current devices can be upgraded to 8 analog outputs, 1 analog-to-digital converter each, to obtain a read noise of 2.3 e- at 4000 fps unbinned. The transfer time for their devices is 25 μs, which is an order of magnitude shorter than the shortest integration time of 250 μs, running at 4 kHz. Both the Gemini Planet Imager and the Palomar PALM3000 upgrade are considering these devices, so operational AO experience should be available in the next several years.

3.6 AO Module

The optical requirements for the AO module are given in Table 9 below. Many items are TBD at this early design phase, but the items are included in the list below as placeholders for the next phase of the design. Table 9 details some of the features of the AO relay.

Table 9: Optical requirements for the adaptive optics module.

Characteristic	Description	Comments
General optical requirements	Wavelengths	0.7-1.0μ WFS band 1.0-5μ science band aggregate science band (about 20% bandwidth at any one time); allow refocus after changing filter
	Field size	2.0x2.0 arcsec square (Goal: 4x4 arcsec)
	Throughput, uniformity for warm surfaces	>80% throughput, 0.9-2.5μ; >60% throughput, 0.5-0.9μ; <1% non-uniformity of throughput across pupil, 1.0-2.5μ
First-order specifications	• Input: f/15 output from telescope • 2 DM positions (at pupils): 150 mm diameter for woofer; 51.2 mm pupil diameter for tweeter • No image plane or exit pupil plane tilt	• Woofer DM size set by dimensions of feasible conventional DM technology (5 mm * 30 actuators) • Tweeter DM size set by dimensions of available MEMS DM (400μ * 128 actuators)
Tip-tilt correction	± TBD arcsec on sky (dominated by expected TMT windshake) tip/tilt correction performed at pupil plane or meet pupil wander requirement	Ideally, one DM would be mounted on a tip/tilt platform and positioned at a pupil, but that may not be possible while meeting angle and frequency response requirements; see text
Pupil wander	< 0.1 % of the pupil	~1/8 sub-aperture
Rotation	Field or pupil rotation	Rotator to maintain pupil orientation fixed in output to DSS
Common path wavefront error	< 10 nm rms integrated for aberrations with spatial frequency between 2λ/D and 64λ/D; < 20 nm rms integrated over spatial frequencies outside this range	
Non-common path wavefront error	< 5 nm rms integrated for aberrations with spatial frequency between 2λ/D and 64λ/D; <20 nm rms integrated over spatial frequencies outside this range	

The AO relay is shown in Figure 39. This is conservative design, requiring simple OAPs and with each section self-contained. A K-mirror derotates the pupil. In the current design this is placed near focus to reduce size, but this presents a concern for phase-induced optical errors (9.1.10); this will be examined in future design phases. The next section is the woofer module, which contains a 30 by 30 actuator conventional DM with 5 mm inter-actuator spacing. The module consists of an OAP (OAP1) that collimates the beam to a 150 mm diameter, the DM, a focusing OAP (OAP2), and a fold mirror. The two OAPs are identical; therefore, the design aberrations are very small across the field. The output of the module is telecentric with a focal ratio (f/#) identical to the input. The fold mirror before OAP1 separates the Naysmith focus from the focus in the output of the woofer module. Separating the input and output image planes of the module is desirable for packaging and eases integration and testing of the modules. The next section is the tweeter module, which contains a 128 by 128 actuator MEMS with 400μ inter-actuator spacing; therefore, the pupil is 51.2 mm in diameter. The module similarly consists of an OAP (OAP3) that collimates the beam to a 51.2 mm diameter, the MEMS, a focusing OAP (OAP4) which focuses the beam and another fold mirror to separate the input and output images, and a fold mirror to separate the input and output images. The light then reflects off a collimating OAP (OAP5) that produces a pupil for the nuller optics. (Section 5.5). Since one of the nuller arms contains an identical MEMS (128 actuators with 400μ spacing) the pupil size is the same as above: 51.2 mm.

Following OAP5, there is a dichroic which reflects the I/Z band light to the wave-front sensor (WFS). The dichroic and a subsequent fold mirror form a steering pair that aligns the guide star and the telescope pupil onto the wave-front sensor. In Figure 39, an OAP is shown that focuses the light to an f/30 beam. This is a configuration consistent with a pyramid WFS.

One advantage of the matched OAP pair approach is that it is very modular so that a change in DM does not necessarily cause the entire relay to change, but the price is a larger number of optics than theoretically necessary to accomplish the optical task. It is possible to reduce the mirror count at the cost of larger, more complex optics and a larger instrument envelope by using off-axis ellipses rather than off-axis parabolas. More care is required to avoid tilted pupil or image planes in the design, but it is possible to do so, and in fact it has been done in one version of the Gemini Planet Imager design. The off-axis ellipse approach results in DM's that are in non-collimated space, but having the DM's in collimated space is not strictly required in an AO system. The path length between optics necessarily grows, and so the advantage of fewer surfaces may be negated if additional fold mirrors are required. This trade-off would be conducted in the next phase of the design.

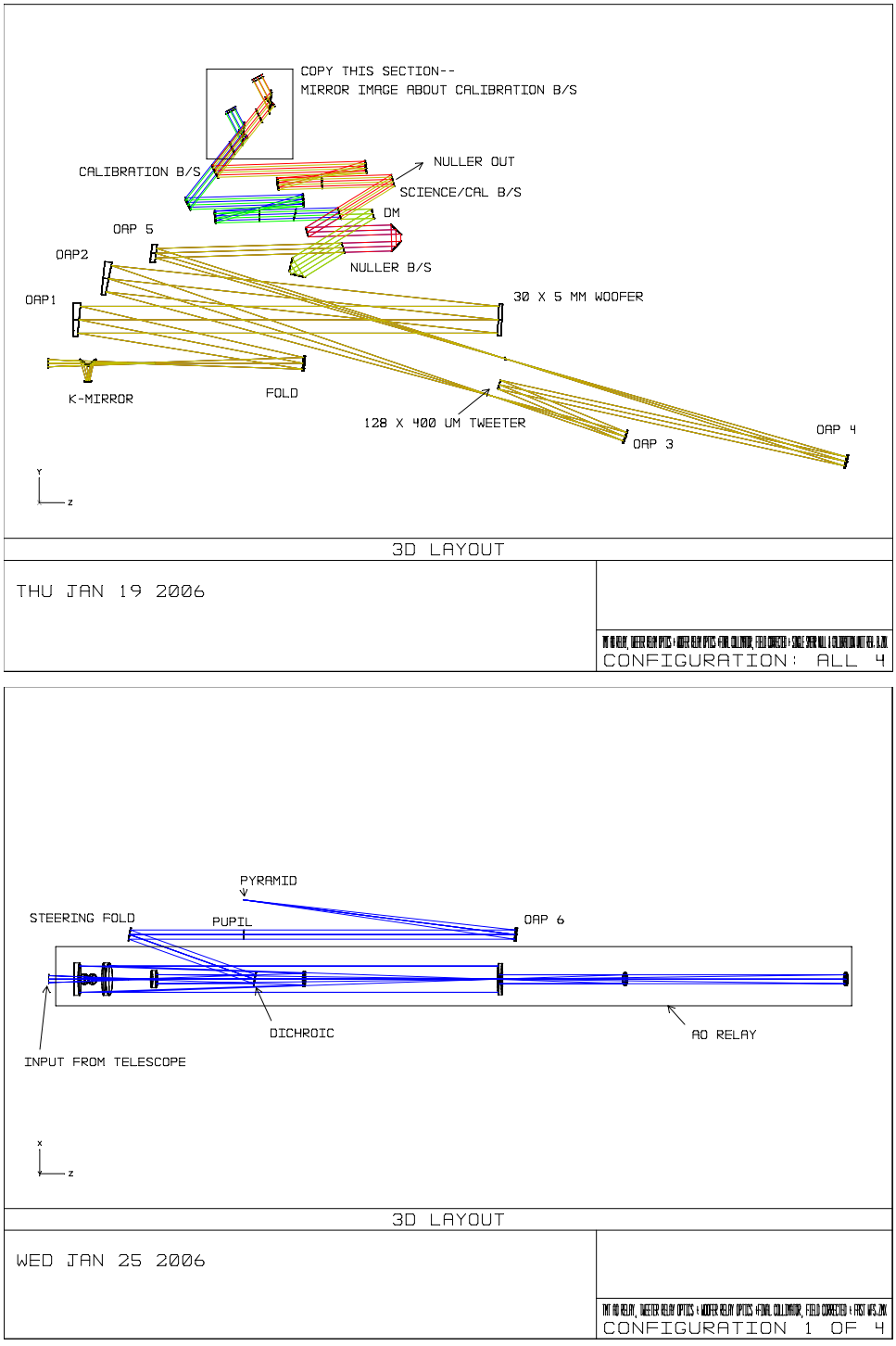

Figure 39: (top) Top view of AO relay, including nuller optics. The light from the telescope enters at the bottom left of the figure. See text for details. (bottom) Side view of AO relay showing the WFS path. The dichroic reflects the beam away from the table to avoid interference.

3.6.1 Tip/tilt mirror

There are several possible approaches to implementing tip/tilt correction:

- Use a DM located on a tip/tilt platform;
- Use a "woofer" DM that has enough stroke and speed to accommodate the tip/tilt requirements (~100μ mechanical stroke).
- Use a separate tip/tilt mirror at a pupil
- Use a separate tip/tilt mirror away from a pupil, but not so distant so as to cause excessive pupil

Using a separate tip/tilt mirror at a pupil is the most conventional approach, but would require additional optics in order to create an additional pupil in the system. The baseline design has a DM mounted on a tip/tilt platform. This approach would obviate the need for a separate pupil for a tip/tilt mirror, and so would minimize the number of surfaces in the relay. However, the mechanical stage tilt and frequency response requirements may make this approach insufficient. While this baseline design is sufficient for purposes of a feasibility study, we have flagged the method of tip/tilt correction as an important conceptual design task. A high-stroke woofer DM such as a bimorph could provide the high-speed portion of the tip-tilt correction, particularly if the high-stroke portion (e.g. from windshake) is removed by an adaptive M2.

3.6.2 Optical Specifications

The small error budget for mid-spatial frequency errors places a premium on obtaining high quality optics. Certainly not every optics vendor is experienced with mid-spatial frequency errors, which will be a crucial cost and performance driver. Fortunately, other projects such as Terrestrial Planet Finder and Extreme Ultraviolet Lithography have paved the way for our requirements. Our specifications for the relay optics will have surface figure requirements in the form given below:

> Wave-front errors: After removal of piston, tip, tilt, and focus, the wave-front aberrations (i.e., reflected wave-front errors, not surface errors) shall be:
> - < 5 nm rms integrated for spatial frequencies less than 2 cycles across the nominal beam diameter
> - < 1 nm rms integrated for spatial frequencies between 2 and 64 cycles across the nominal beam diameter
> - < 1 nm integrated for spatial frequencies greater than 64 cycles across the nominal beam diameter

4 Diffraction Suppression System

4.1 Introduction

Light is scattered within the PSF of a telescope both by wavefront errors and diffraction. Even in the absence of wavefront errors, e.g. a perfect circular telescope in space, diffraction will produce an Airy pattern much brighter than any potential planets. For TMT, the maze of segment and secondary structures within the pupil will produce a brighter and more complex pattern. Although in principal these patterns are static and could be removed through PSF subtraction, the added light results in added photon noise. More significantly, the Airy pattern amplifies speckles caused by dynamic and static wavefront errors (e.g., Soummer and Claude 2004) through the "pinned speckle" cross-terms (Perrin et al. 2003) and, since these speckles are particularly sensitive to PSF shears, reduces the stability of the PSF. This effect is even more significant for TMT, where telescope primary mirror phase errors have the same characteristic λ/d_{seg} spatial frequency characteristics as the telescope diffraction pattern.

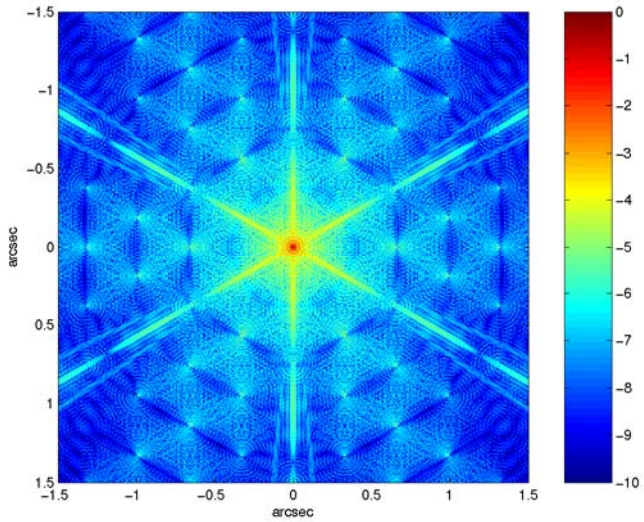

Figure 40: Monochromatic diffraction pattern (PSF) from the Thirty Meter Telescope reference design

PFI therefore requires a Diffraction Suppression System (DSS) that removes diffraction down to levels comparable to the final contrast goals (Residual diffraction average intensity ~10^{-9} of the peak stellar intensity). It must also allow observations at inner working angles as small as 3 λ/D. For a segmented telescope with a large secondary obscuration this is an extremely challenging requirement.

This section will describe the analysis of the Diffraction Suppression System (DSS). The purpose of the DSS is to extinguish the signal from the on-axis source while maintaining an acceptable throughput for off-axis sources. Typically, DSSs exploit the fact that optical systems are linear systems. In a two stage DDS, which is a common architecture, the first stage acts as a high-pass filter while the second stage acts as a low-pass filter. The two stages are designed so that in combination, they filter-out all the spatial frequencies for the on-axis source. The two architectures studied in this analysis are a band-limited Lyot

coronagraph (Kuchner and Traub 2002) and a visible-nuller architecture (Mennesson, et.al. 2004). Most other architectures – such as a classic Lyot coronagraph - perform poorly on the TMT architecture.

4.1.1 DSS requirements

Diffraction Suppression System key requirements:

- Reduce average diffraction structure to a level of 10^{-8} (goal: 10^{-10}) of the peak intensity of the star at radii > 3 λ/D
- Block PSF structure caused by secondary supports and segment gaps
- Provide an inner working angle of 3 λ/D
- Operate over an entire single atmospheric bandpass (*YJHKLM*) at a time
- Provide a reference beam for the post-DSS wavefront sensor
- Reduce throughput by no more than 50%
- Allow observations of at least 50% of the field of view
- Operate in the TMT vibration environment
- Includes a location for the back AO system's deformable mirror

DSS system key interfaces:

- Accept a collimated beam from the front AO system with pupil located well inside the DSS and beam size matched to the DSS
- Produce a collimated (may be changed to f/16 converging) beam for the science instrument with enough back focal distance and proper pupil location
- Shear and piston controlled by the post-DSS WFS

DSS system design summary

- Two-stage shearing nulling interferometer with achromatic phase shifts and adjustable shear
- Lyot stop inside science instrument masks out secondary and supports
- Segment gaps nulled out by shifting an integer number of segments

4.1.2 Overview of different coronagraphs

The table below lists several coronagraph types. We briefly discuss the architecture and issues of each. We have chosen to carry out numerical simulations of two of these – the band-limited Lyot and the nulling interferometer.

Table 10: Summary of coronagraph architectures considered

	Band-limited Lyot Coronagraph (4.2.2)	Nuller (4.2.3)	Four-Quadrant Phase Mask (4.2.4.1)	Optical Vortex Nuller (4.2.4.2)	Shaped Pupil	Phase Induced Amplitude Apodization
Practical IWA (Unobscured)	4 λ/D	2-3 λ/D	1-2 λ/D	2-3 λ/D	3-4 λ/D	1 – 2 λ/D
Practical IWA (Obscured)	6-10 λ/D	3 λ/D			4-5 λ/D	
Transmission efficiency	20-40%	~20-50%	> 80%	~20-50%		60%-100%
Interaction with segmented and obscured apertures	May need large blocked area at secondary or segment gaps	Easy to mask obscuration	Segments and secondary cause leakage in the Lyot plane	Same as 4QPM	Secondary can be masked, segments cannot	No obvious approach
Complexity	Moderate	High	Moderate	High	Moderate	High
Other critical issues	Band-limited masks hard to manufacture in IR and may have phase errors	Complex Implement-ation and operation	Fabrication issues related to broadband, sensitivity to pointing	Difficulty in manufacture of masks, unproven.	Fabrication of fine mask structure. Diffraction couples with wavefront errors to fill in dark hole	Extreme alignment sensitivity, requires two back-to-back systems for large field

4.1.2.1 Band-limited Lyot Coronagraph

This coronagraph is derived from the classical Lyot coronagraph with a tailored shape of the transmission function of the focal plane occultor. The classic Lyot coronagraph has a sharp-edged opaque disk to suppress starlight. In the subsequent re-imaged pupil (Lyot) plane, there is significant ringing that leaks light around the geometric edges of the original pupil. The band-limited coronagraph has a transmission function set so that the light that is not suppressed by the band-limited focal plane occultor will be re-imaged to a pupil image with more concentrated features. These then can, in theory, be perfectly suppressed by a simple pupil plane mask.

There are two challenges to this technique: one theoretical and the other practical. First, for small inner working angles the requisite pupil masking becomes large. This is fundamental, and leads to a very low operating efficiency, particularly if the pupil is obscured by a secondary mirror. Second, the continuously varying spatial apodization of the focal plane mask is at this time difficult to manufacture, particularly in the infrared. Fabrication methods have been developed in the visible, but higher optical densities are associated with phase effects. Ultimately, there may be an engineering solution, but currently presents a challenge. Still, this is the second-most promising approach for TMT, especially if IWA requirements are relaxed, so we have analyzed its performance.

4.1.2.2 Shearing Nulling Interferometer ("Nuller")

This coronagraph is relies upon spatial coherence to destructively interfere and remove the starlight from the final image plane while leaving the companion flux only partially suppressed. Since it is "occultorless", the Lyot pupil can be imaged sharply for very efficient masking. There is no fundamental challenge to fabrication of such a nulling interferometer. The adjustable shear also makes the nuller amenable to different science targets of interest.

The response patter of the nuller on sky presents a challenge if one is accustomed to more traditional coronagraphs. The response of the nuller on the sky is a fringe pattern that creates a black picket fence pattern with alternating bands of deep suppression and high transmission. For narrow optical bands, this makes half the sky unobservable at any given instant. For very high contrasts, the path length fluctuations internal to the nuller must be extremely small (sub-nanometer). Experimental work at JPL is currently directed at making the nuller very rigid and stable.

4.1.2.3 Four Quadrant Phase Mask

A two by two phase mask with alternating pi phase changes along one diagonal comprises the four quadrant focal plane phase mask. When an obscured pupil forms an image that is centered on this phase mask, all the light in the subsequent Lyot pupil is scattered outside the geometric image of the input pupil. Thus, the on-axis light can be completely suppressed. The suppression in this coronagraph therefore occurs primarily in the Lyot plane.

The elegant simplicity of this coronagraph is complicated by a few technical details. First, the PSF must be centered exactly on the intersection of the phase steps of the FQPM. If not, there will be no perfect cancellation. It is highly sensitive to these pointing errors. The second is that this coronagraph requires an un-obscured pupil to work well. Light from a central obscuration, for instance, will lead to leakage internal to the Lyot pupil. Finally, these devices have been made and have been used on sky, however they require narrow band operation due to the chromatic nature of the phase change.

This device is briefly analyzed later in the text.

4.1.2.4 Optical Vortex Coronagraph

This coronagraph, like the FQPM described above, is also a transmissive element in the image plane. This element creates an azimuthally linear phase spiral from 0 to 2π radians in a 360 degree rotation about the focal plane vertex. This phase change creates a dark hole that nulls on axis light in the subsequent image plane. The advantages of this coronagraph are that it's less sensitive to pointing errors than the FQPM and it can also be made to operate over a wider optical bandwidth.

This coronagraph is also briefly described briefly in the text, and some simulations that predict its performance are included. This coronagraph was not studied in detail because it is still seen as a maturing technology.

4.1.2.5 Shaped Pupil Coronagraph

In this technique a shaped pupil is located in a re-imaged telescope pupil plane. This shaped mask preferentially diffracts light away from optical axis, and thus creates areas in the focal plane with little starlight from an on-axis object. It is fundamentally extremely easy to implement. Theoretical contrast that is consistent with the requirements for TMT-PFI is possible. However, this coronagraph suffers from sensitivity to residual wave front errors and pupil amplitude effects such as obscurations and segment gaps. It was therefore ruled out straight away and was deemed worthy of including in this report.

4.1.2.6 Phase Induced Amplitude Apodization

It is obvious that amplitude apodization in the pupil plane can be employed to modify the point spread function in the image plane. However, this usually comes at the cost of lower telescope resolution and also lower efficiency for the faint companion. However, amplitude apodization can also be induced from phase changes in the pupil plane. If this phase induced apodization can be done with reflecting optics, then it is intrinsically broadband and can be done with high efficiency. Thus the phase induced amplitude apodization (PIAA) coronagraph has the advantages of improved chromaticity and higher efficiency when compared with other pupil apodization techniques.

However, this coronagraph has no obvious approach to dealing with obscured apertures, is challenging to manufacture and align, and has not yet been experimentally demonstrated at contrast levels needed for PFI.

4.1.3 Modeling context definition

The DSS analyses undertaken in support of this study have focused on two types of problems. The first type of analysis has concentrated on the performance and trade studies used to down-select to a favored architecture. The second analysis is an end-to-end simulation study that has enabled an integrated analysis of the performance of the PFI instrument with the selected DSS architecture. This section will describe the performance and trade studies. The DSS component of the end-to-end simulations will be described in Section 9.2.4.

4.1.4 Modeling paradigm

The modeling paradigm used to assess the performance of the DSS is based on the Exit Pupil approach developed by J. Green at JPL. The input pupil included in the model is the segmented TMT pupil, which includes the segment gaps, the Secondary Mirror (SM) obscuration, the obscurations from the three SM support spiders, and the obscurations from the six cables (omitted in the Lyot simulations.) The Grey Pixel approximation algorithm developed by Chanan and Troy (Troy and Chanan 2003) is used to generate the input pupil. This algorithm is intended to mitigate the problems associated with segment gap widths that are smaller than the resolution element size of the array used to hold the pupil. Pupil phase and amplitude maps can be incorporated into the pupil model to study the impact of static and dynamic perturbations to the system, as discussed in Section 10. Figure 41 displays the input pupil used for the coronagraph simulations. Figure 42 shows the input pupil used for the nuller simulations.

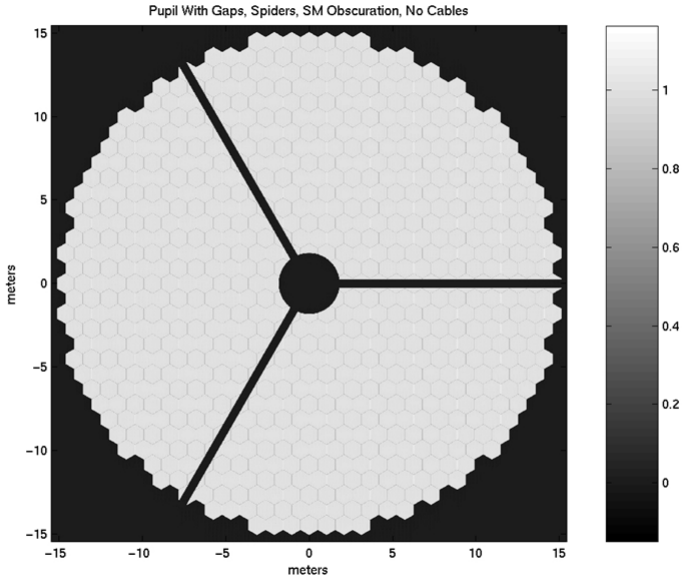

Figure 41: TMT input pupil for Lyot coronagraph simulations.

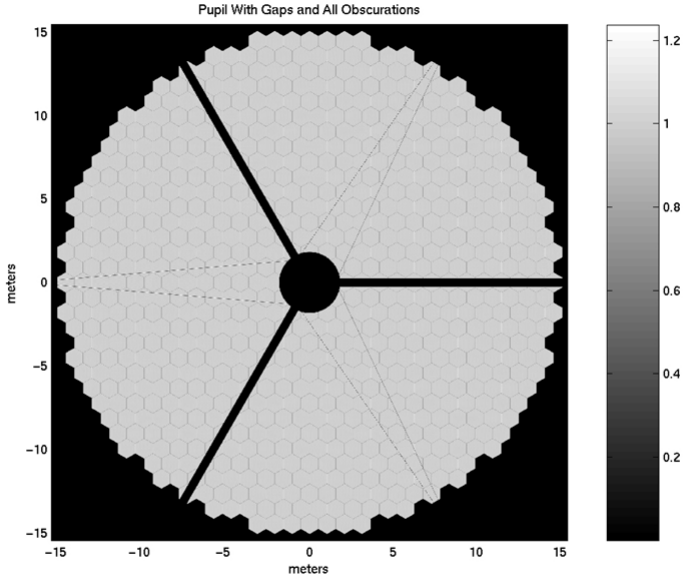

Figure 42: TMT input pupil for nuller simulations.

The Fourier Transform (FT) of the complex field in the input pupil yields the complex amplitude of the field in the image plane (Goodman 1996). In a traditional Lyot coronagraph, the occulting mask is applied to the field at this location.

The simulation for the band-limited Lyot coronagraph employed a radial $sinc^2$ transmission mask. The functional form of this mask places it in the family of band-limited masks proposed by Kuchner and Traub (Kuchner and Traub 2002). The advantage of a band-limited occulting mask is that it has finite support in the Lyot pupil; diffraction from any input pupil structure will be completely confined to regions in the re-imaged Lyot pupil. The use of this type of mask means that there are locations in the Lyot pupil where the field from an on-axis source is cancelled exactly. With the choice of a suitable Lyot stop, the field from the on-axis source can be completely suppressed.

The visible-nuller is the other DSS architecture that has been studied. This architecture is described in Menneson et.al. (Mennesson et. al. 2004). The visible nuller is a variation of the sheared pupil interferometer concept. The specific implementation that has been studied is a linear two-stage nuller. In this implementation, there are two stages of pupil shear with the direction of the shear the same in both stages. The interferometer transmission pattern on the sky is a sin^2 pattern with a null centered on the on-axis source. Since the first image plane in the model is conjugate with the sky, a wavelength-dependent sin^2 transmission mask applied to complex field at this location can simulate the interferometer.

The Fourier Transform of the complex field in this image plane is the complex amplitude in the Lyot plane, which is conjugate with the input pupil. The functional form of the diffracted field from the central source in this plane is the convolution of the pupil support function of the occulting mask (or nuller beam pattern) with the support function for the input pupil. Since the radial $sinc^2$ occulting mask for the coronagraph and the sin^2 transmission pattern for the nuller are band-limited, this convolution will produce regions in the Lyot plane in which the cancellation of the on-axis source is perfect. A mask, the Lyot stop, is used to isolate these regions.

The field at the final image plane, where the contrast performance is evaluated is the Fourier Transform of the field transmitted by the Lyot stop. This field is converted to intensity for the contrast calculations. The resulting intensity map is the nulled Point Spread Function (PSF). The algorithms that are used to evaluate the contrast from an architecture require the calculation of the intensity at the final image plane from an on-axis source that is not occulted by a mask in the first image plane. The resulting intensity map is the bright PSF. The algorithm for calculating contrast from these PSFs is described in Section 9 of this report.

4.2 Trade study summary

4.2.1 Performance metrics

There are a number of performance metrics that are useful for comparing different DSS architectures for PFI. The most significant metric, of course, is contrast, which describes the ability of an implementation to suppress light from the parent star. The Inner Working Angle (IWA) at which a specified contrast is reached is another important metric that is

driven from the science requirements imposed on the instrument. The required strength of the signal from an off-axis source, which is again driven by the science requirements, is directly proportional to the efficiency of the DSS. In this study, efficiency is described by the occulter transmittance of the Field-Of-View (FOV) and the Lyot stop size and shape. Finally, instantaneous FOV, which is determined by the fraction of the FOV for which a particular DSS architecture in a particular viewing orientation has a non-negligible sensitivity, is also important because the observing efficiency of the instrument is directly proportional to it.

While all of these metrics are significant, in practice contrast has been the only metric that has been given serious consideration because, as will explained later, it proved to be the discriminator in establishing the that the band-limited coronagraph was not a viable option for a DSS with an obscured Primary Mirror and a requirement to work at 3 λ/D.

4.2.2 Classical band-limited Lyot coronagraph

4.2.2.1 Description

The Lyot coronagraph architecture that has been considered is a variant of the original concept proposed by Lyot (Lyot 1939). The band-limited occulting mask that has been used in this study is the radial sinc^2 mask, which has the following functional form:

$$1 - \text{sinc}^2\left(\frac{r}{\sigma}\right)$$

where

r is the distance from the center of the mask and

σ is a scaling parameter that depends on the required IWA at a particular wavelength

Since the size of the PSF is a function of wavelength, the physical mask is optimized for a particular wavelength in the bandpass. The value used for this analysis is 1.65μ.

This occulting function is band-limited because the FT of a sinc^2, in one-dimension, is a triangle function. A plot of the transmittance of the occulting mask versus position in the FOV is displayed in Figure 43. This mask is designed to have a transmittance of 50% at the 3 λ/D field point for a wavelength of 1.65μ. A plot of the pupil support function for this occulting mask is presented in Figure 44. Note that the scaling of the dimensions in this pupil has been chosen so that the 30-meter TMT Primary Mirror (PM) in the input pupil maps to 30 meters in the Lyot plane. The fact that the support of the FT of the occulting mask is smaller than the diameter of the PM means that if the pupil were unobscured there is some region in this pupil where the field from the host star is exactly cancelled. For the real TMT pupil, things are not so simple.

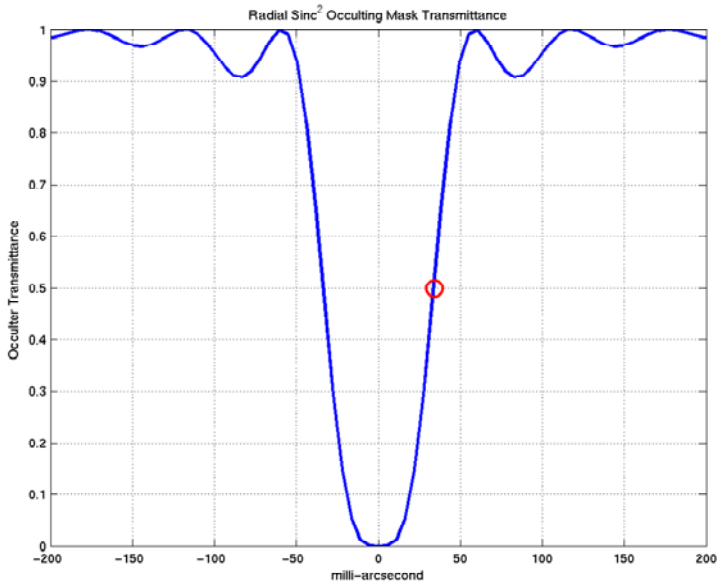

Figure 43: Sinc^2 Occulter Transmittance for an occultor with good transmission at a IWA of 30 mas.

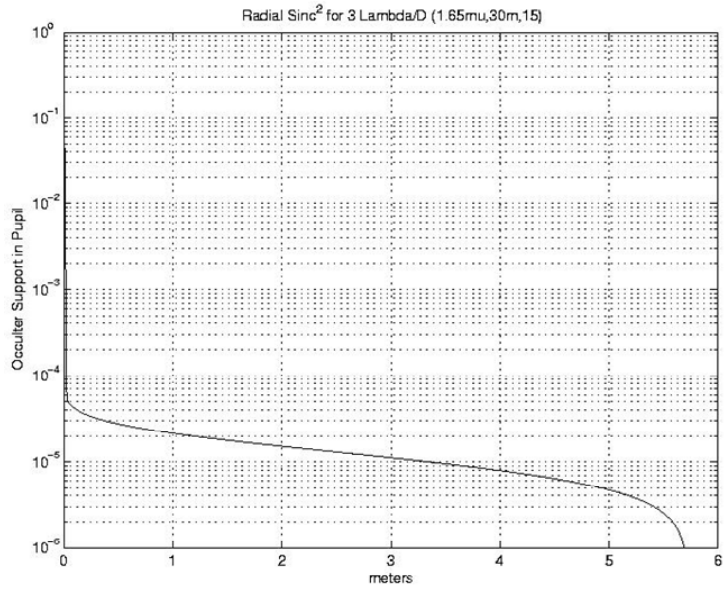

Figure 44: Support in Lyot pupil for the sinc^2 occulter mask shown above. Any obscuration in the input pupil will be convolved with this function.

The Lyot stop is a mask that is used to exclude the area in the pupil where the field from the on-axis source is not perfectly cancelled. The design of the Lyot stop is driven by the apparently conflicting requirements of minimizing the transmitted field from the parent star while at the same time maximizing the throughput from the off-axis source. Minimizing the signal from the on-axis source drives the Lyot stop design to mask out more of the pupil while maximizing the signal from the off-axis source drives the design to transmit more of the pupil.

The Lyot stop used in the analysis of the band-limited coronagraph is shown in Figure 45. Note that the input pupil does not contain the obscurations from the cables, which means that the derived contrast results will tend to be optimistic. The outer diameter of the mask is 18 meters, which is consistent with the pupil support diameter of the radial sinc^2 occulting mask (~12 meters). The shadow of the SM (3 meters) is masked by an obscuration with a diameter of 5 meters. The spider shadows (0.5 meters) are masked by obscurations with a width of 0.7 meters. The resulting mask has an efficiency of 38%. As will be emphasized in the next section, even with 62% of the pupil masked off, these obscurations are undersized and the coronagraph contrast performance is inadequate.

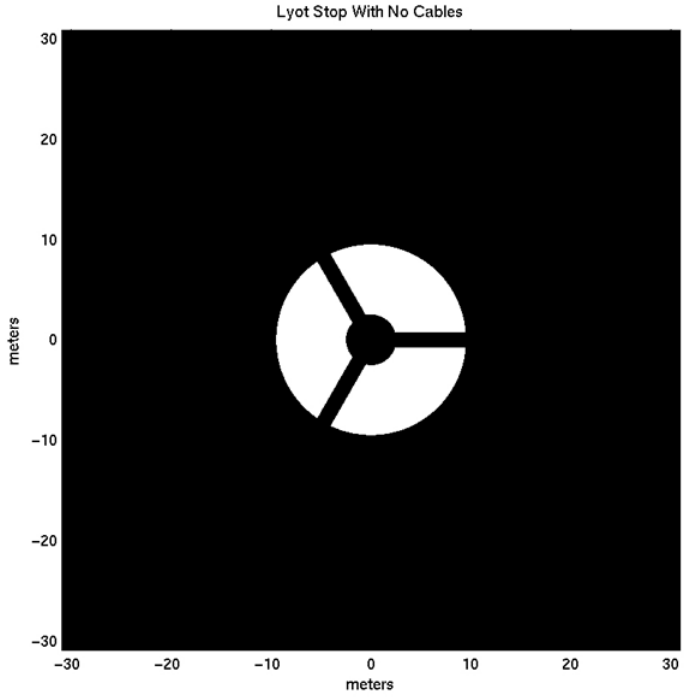

Figure 45: Lyot stop used for initial coronagraph simulations. Outer diameter is matched to the pupil support function given above but secondary and spider obscurations are undersized to increase throughput.

4.2.2.2 Performance

The system parameters used to model the performance of the coronagraph architecture are:

- Input pupil: TMT pupil without cable obscurations (Figure 41)
- Occulting mask: radial sinc^2 optimized for 50% transmittance 3 λ/D for $\lambda = 1.65\mu$ (Figure 43)
- Lyot stop: 38% efficiency mask based on TMT input pupil with cable obscurations (Figure 45)
- Bandpass: monochromatic (1.65μ)

The contrast results are presented as plots of radial averaged contrast versus field angle.. The contrast calculated for the band-limited coronagraph is displayed in Figure 46. These plots were generated with an older version of the radial averaging algorithm, which include all the values in an annulus in the average. Since the same algorithm was used for all of the plots, comparisons across different architectures are self-consistent.

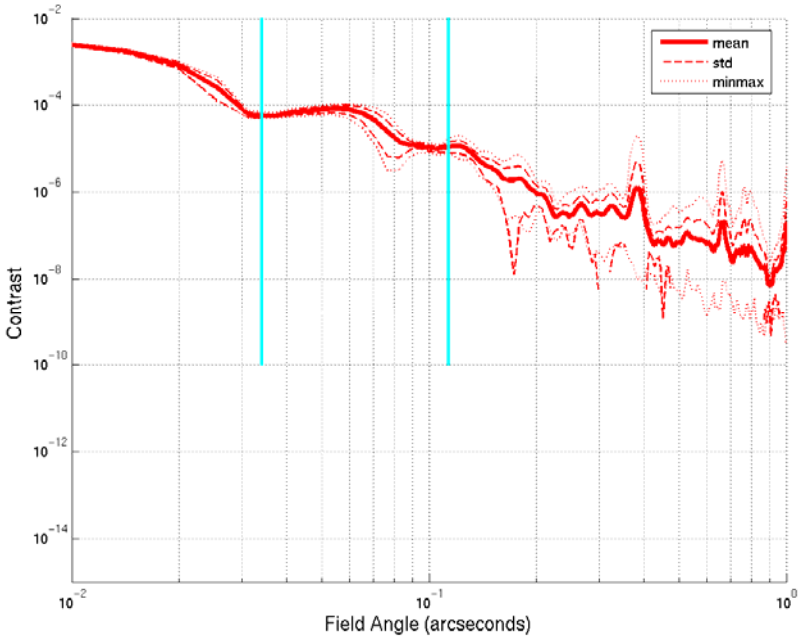

Figure 46: Radial average of contrast for Lyot coronagraph with sinc^2 occulter mask from Figure 43

The important conclusion that can be inferred from the results in this plot is that the contrast performance of this DSS architecture is three to four orders of magnitude worse that the requirement from 3 λ/D to 10 λ/D. The explanation for the poor performance can be understood by considering the pupil support function for the radial $sinc^2$ occulting mask displayed in Figure 44. This function has a significant level of support out to distance of 5.5 to 6 meters from the center. Recall that the field in the Lyot plane is the convolution of the input pupil with the occulting mask pupil support. The convolution of the this function with the TMT SM obscuration (3 meter diameter) defines a region with a diameter of approximately 14 meters in the Lyot plane where the field cancellation is not perfect. The same argument can be made about the required size for the spider masks. Since the spiders have a width of 0.5 meters, the region in the Lyot plane that needs to be masked for each spider is about 12 meters. Clearly the Lyot stop displayed in Figure 45 is inadequate. The residual field after the Lyot stop has been applied is displayed in Figure 47. Note the significant level of field amplitude outlining the SM and spider masks. Improving the contrast performance requires suppressing the field in these regions.

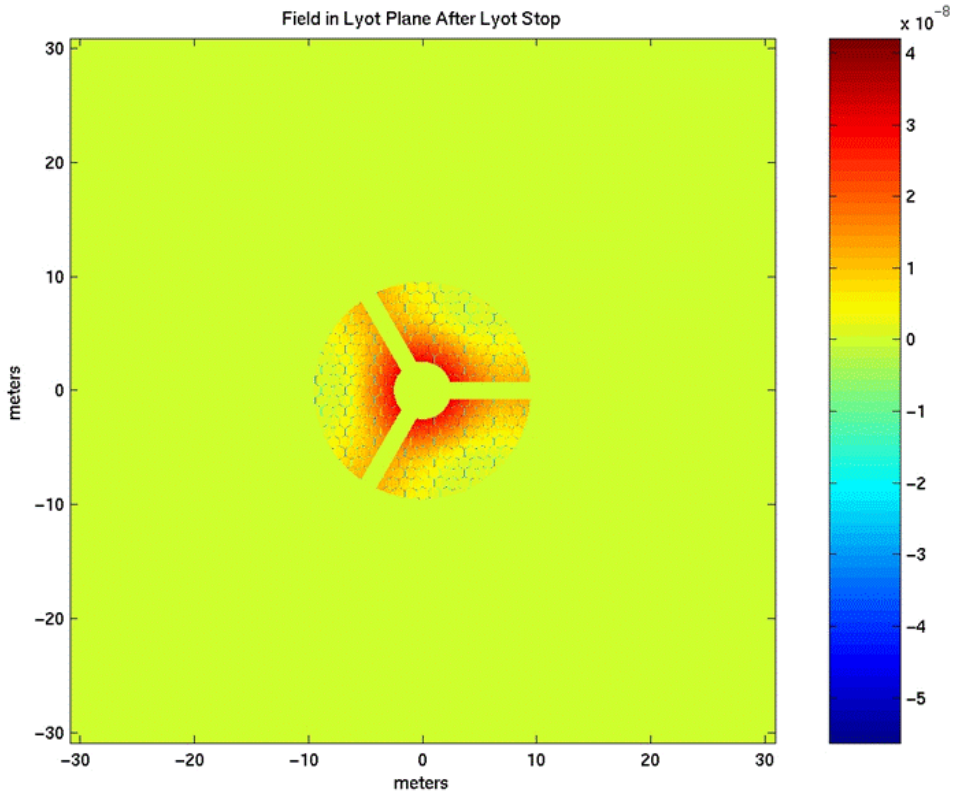

Figure 47: Residual field in Lyot plane after Lyot stop for $sinc^2$ occulter mask from Figure 43. Note strong leakage around the secondary and moderate leakage at the secondary supports and segment gaps.

The most straightforward way to improve the contrast performance is with a more aggressive Lyot stop. A simple way to examine this hypothesis is to generate a mask by applying a threshold to the field amplitude in the Lyot pupil. A mask generated by using a threshold value that is about an order of magnitude less than the maximum amplitude in Figure 47 is shown in Figure 48. The residual field after this mask is applied in displayed in Figure 49. Consistent with our expectations, the radial averaged contrast, displayed in Figure 50, shows an improvement of one to two orders of magnitude in contrast from 3 λ/D to 10 λ/D. This improvement, however, comes with a price. The efficiency of the thresholded Lyot stop is 10% compared with the 35% efficiency of the original Lyot stop. Performance at small angles is still two to three orders of magnitude worse than the requirement. This problem is fundamental to coronagraphs with obscured pupils and very small inner working angles. We have concluded, therefore, that a band-limited Lyot coronagraph is not a viable option for PFI.

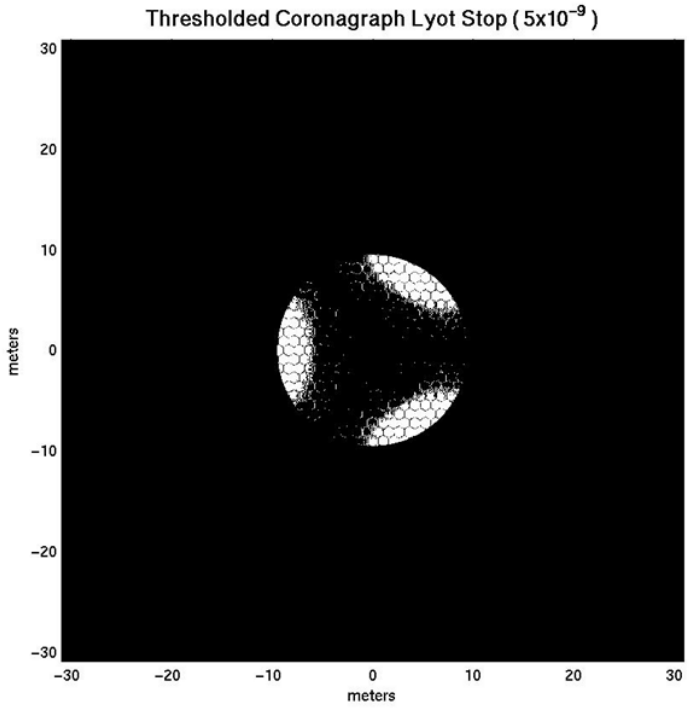

Figure 48: Thresholded Lyot stop based on Lyot plane in Figure 47. Throughput is extremely low.

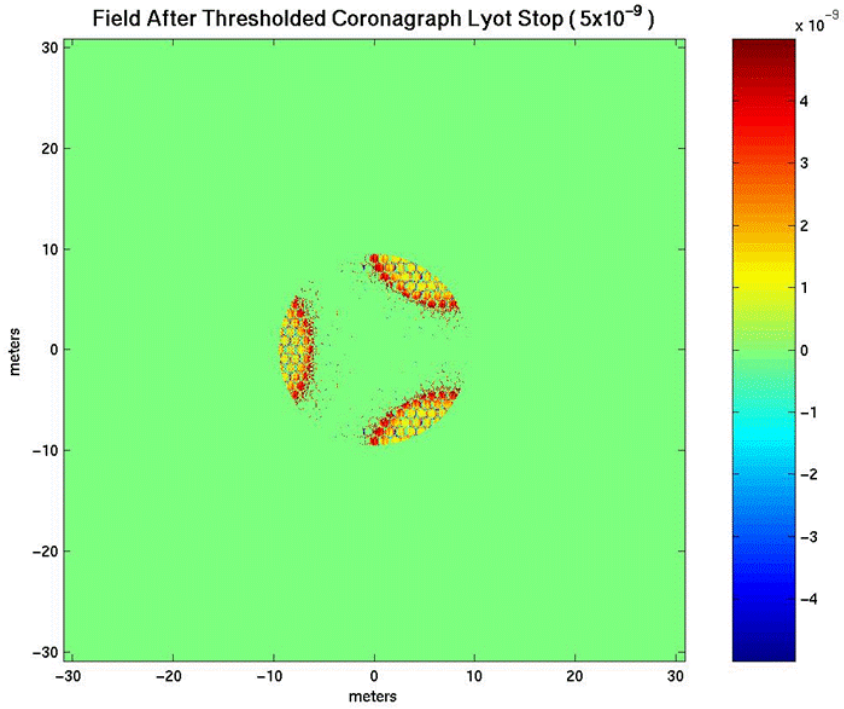

Figure 49: Residual field after thresholded Lyot stop above. Leakage is primarily due to segment gaps but throughput has become extremely low.

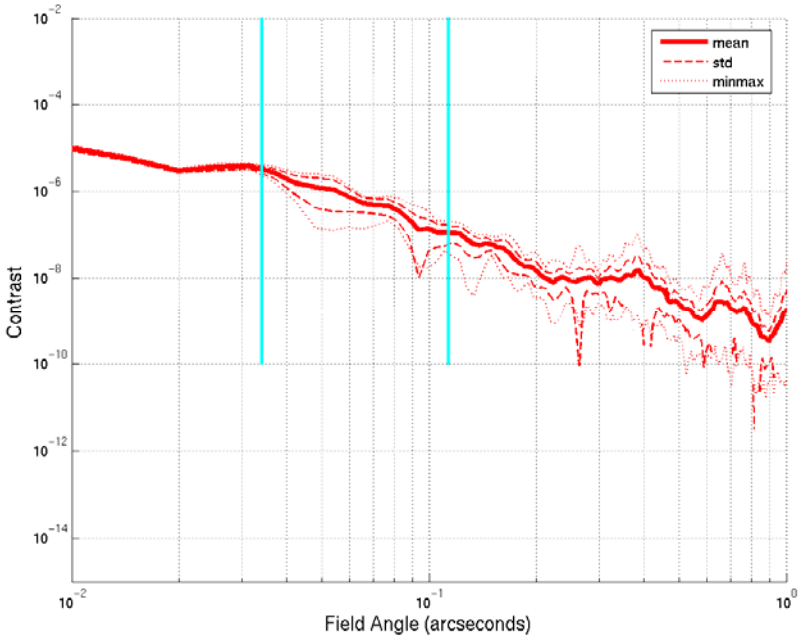

Figure 50: Radial average of contrast with thresholded Lyot stop.

This conclusion, however, is for a design with an IWA of 3 λ/D. If the IWA requirement is relaxed to 9 λ/D, then the pupil support function for the radial sinc^2 mask decreases in diameter by a factor of three (from 12 meters to 4 meters). The Lyot stop matched to this new occulter could have a significantly larger outer diameter (up to 26 meters) and a significantly smaller central obscuration (about 7 meters). This mask may be efficient enough to make the coronagraph a viable option. This hypothesis was tested with a radial sinc^2 occulting mask tuned for 9 λ/D (Figure 51) and the Lyot stop displayed in Figure 52.

The radial average contrast plot is shown in Figure 53.

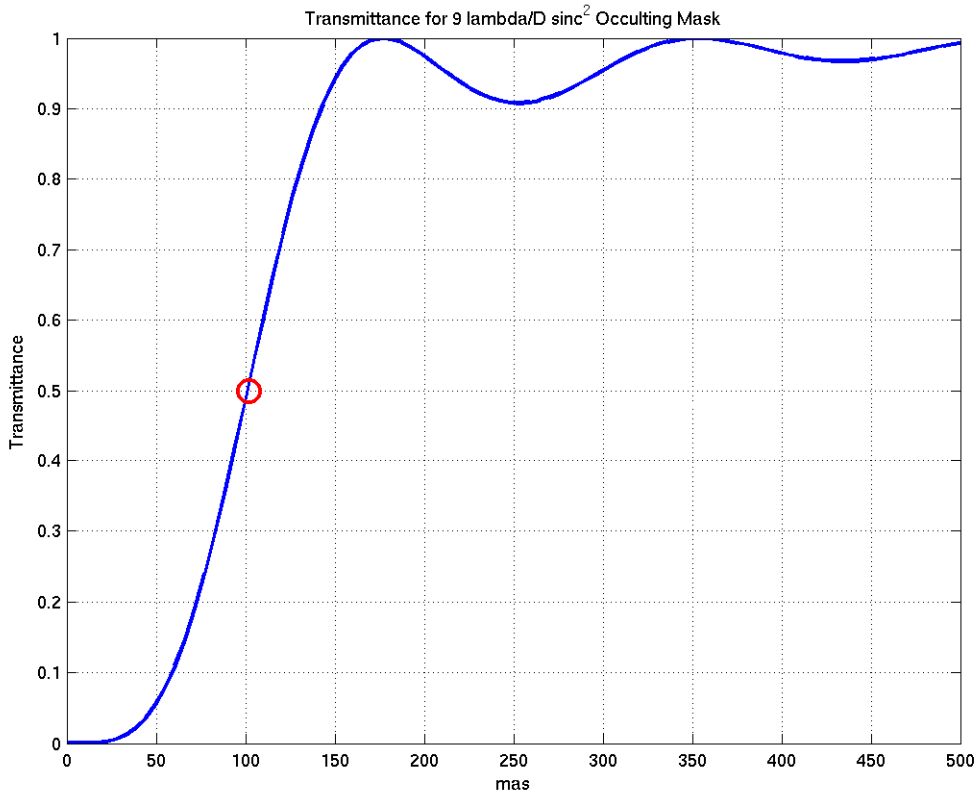

Figure 51: Transmittance for 9 λ/D radial sinc^2 occulting mask, tuned for better throughput in the Lyot plane but a much larger inner working angle.

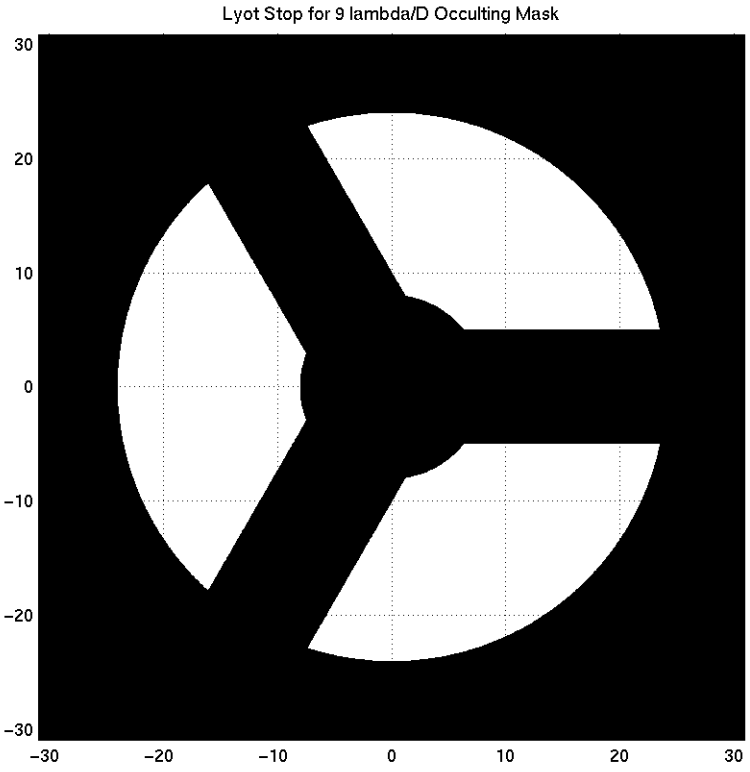

Figure 52: Lyot Stop tuned to completely mask secondary and supports for 9 λ/D radial sinc^2 mask.

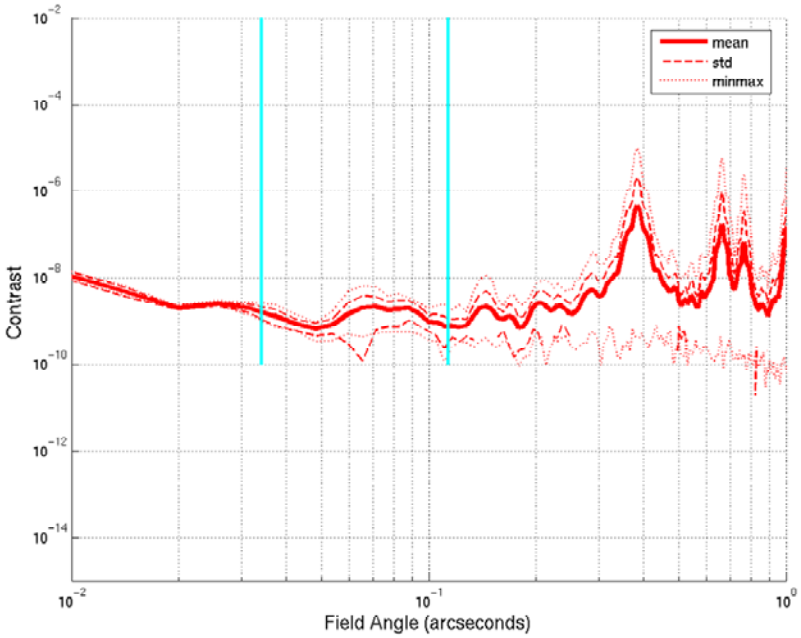

Figure 53: Radial averaged contrast for 9 λ/D sinc^2 occulting mask. Within 0.07 arcseconds the transmission is so low that practical contrast would be much lower than shown.

From this plot, we can conclude that a contrast level of ~10^{-9} can be achieved for field angles greater than 0.1 arcsecond. (The relatively good contrast from 0.01 to 0.1 arcseconds should be discounted because the mask transmittance is small in this region – the planet would be too attenuated relative to the photon noise.) If the PFI inner working angle requirement is relaxed, this band-limited coronagraph is a viable option. However, the small inner working angle is the greatest benefit of TMT compared to other planet imaging systems – PFI science performance would be significantly impacted by an 0.1 arcsecond IWA. (see for example Figure 4). In addition, these masks have no effect on the structure from the segment gaps, producing the contrast bumps at 0.4" in the above figure. Nonetheless, this remains a backup option for PFI, especially if combined with apodization (Section 4.2.4.3 and Soummer 2005).

Manufacturability of the band-limited mask remains a technical concern – though one that TPF-related projects are attempting to address. Current approaches such as electron-beam irradiated glass introduce significant wavefront errors and cannot produce enough attenuation (~10^8) in the infrared.

4.2.3 Visible Nuller / Interferometer

The visible nuller is the other implementation for the DSS that has been considered for this study. The concept is described in Mennesson et.al. 2004. A brief discussion that highlights the particular implementation of this architecture will be provided here.

4.2.3.1 Description

The nuller is a variant of a sheared pupil interferometer. In this type of interferometer, the pupil is interfered with a shifted copy of itself. The resulting interferometer beam pattern is a harmonic with a fringe frequency of B/λ (cycles/radian) where B is the shear. The nuller implementation of this type of interferometer adds an achromatic π phase shift to the field in one of the pupils before recombination. In principle, the field from an on-axis source is perfectly cancelled in the overlap region of the two pupils.

The implementation studied for this report is a linear, two-stage nuller. After the first, where the null is generated, the recombined pupil is split again, sheared, and recombined without a relative phase shift. The shear in both stages is along the same direction and the size of the shears is the same. The central null of the resulting beam pattern has a θ^4 transmittance dependency on field angle. An example of the beam pattern on the sky for the nuller is displayed in Figure 54. This beam pattern is designed for a 50% transmittance at 3.04 λ/D with $\lambda = 1.65\mu$. This corresponds to a pupil shear of 3.12 meters, which is equal to the width of three hex segments. An integral number of hex segment widths were chosen because this guarantees that the segment gaps in each pupil image will lie on type of each other when the sheared pupils are recombined.

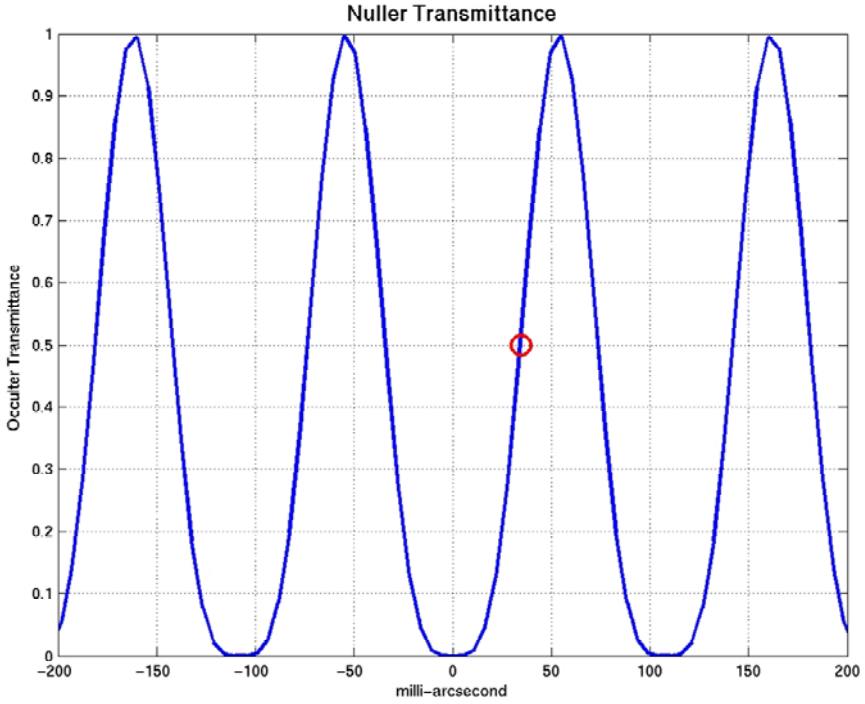

Figure 54: Nuller transmittance.

4.2.3.2 Performance

The model parameters for the nuller simulation are as follows. The central wavelength was 1.65μ. Both monochromatic and broadband simulations were done. The broadband simulations use a value of 5 for R. The single stage pupil shear was 3.12 meters. The input pupil, displayed in Figure 42, is the fully obscured TMT pupil with segment gaps, SM obscuration, spider obscurations, and cable obscurations. The Lyot stop used for this analysis is displayed in Figure 55. It has an efficiency of ~38%. Note that this simulation remains idealized in the sense that no error sources were incorporated into the model. This comment applies as well to the simulations reported for the band-limited coronagraph. This is not a fundamental limitation of the analysis approach. Static and dynamic telescope errors as well as the effects of amplitude and phase errors introduced by the atmosphere could be incorporated into the model. Static telescope errors were studied independently in the analyses of errors on M1and M3. The results for M1 errors are reported in Section 10.5 and the results for M3 errors are reported in Section 10.6. Atmosphere effects were incorporated in the end-to-end simulations reported in Section 9. The results of the simulations reported in this section should be interpreted as establishing the contrast floor achievable with this architecture.

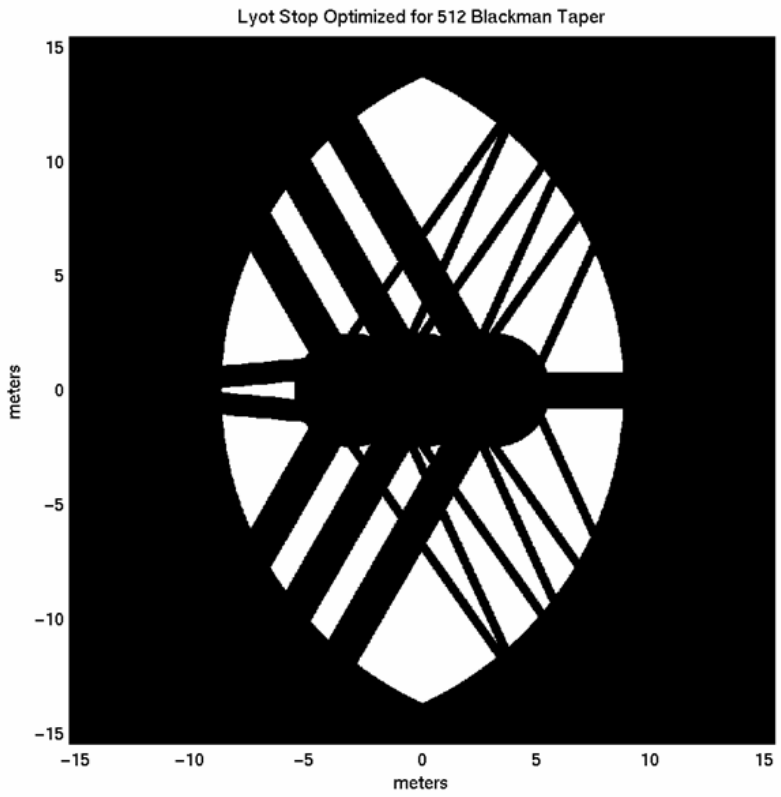

Figure 55: Lyot stop for two-stage nuller.

4.2.3.3 Contrast performance

The baseline contrast performance was evaluated using the TMT input pupil displayed in Figure 42 and the Lyot stop shown in Figure 55. The field amplitude in the Lyot plane, showing the +3.12 meter shear introduced by the nuller, is displayed in Figure 56. The residual field after the Lyot stop is show in Figure 57.

The monochromatic performance is summarized in the radial averaged contrast field angle plots shown in Figure 58. The contrast values from which this plot has been derived have not been normalized by the antenna response. The mean curve in this plot is the relevant curve. The contrast performance is several of orders of magnitude better than the PFI requirement.

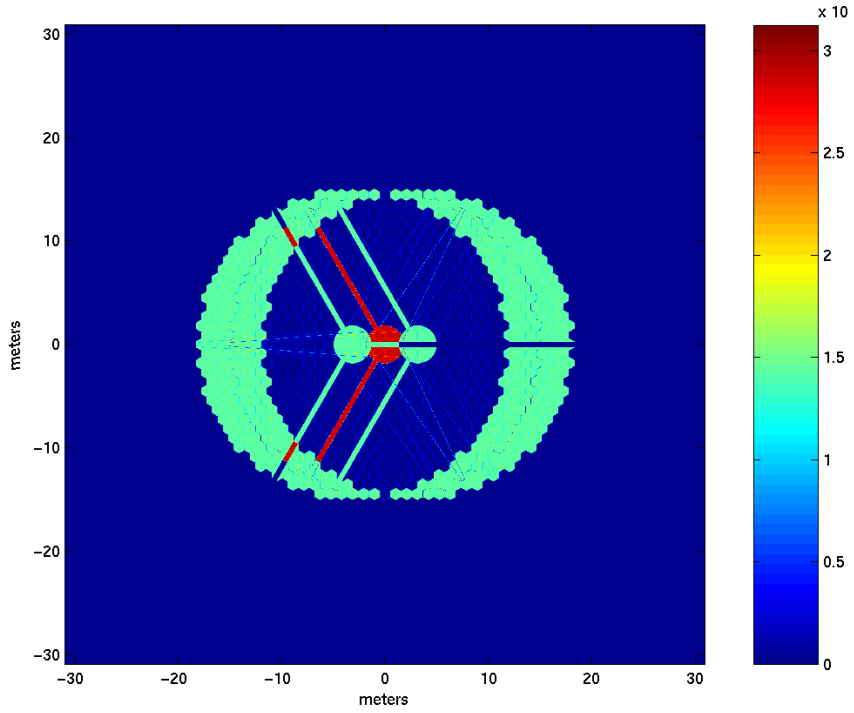

Figure 56: Field in Lyot plane after the DSS.

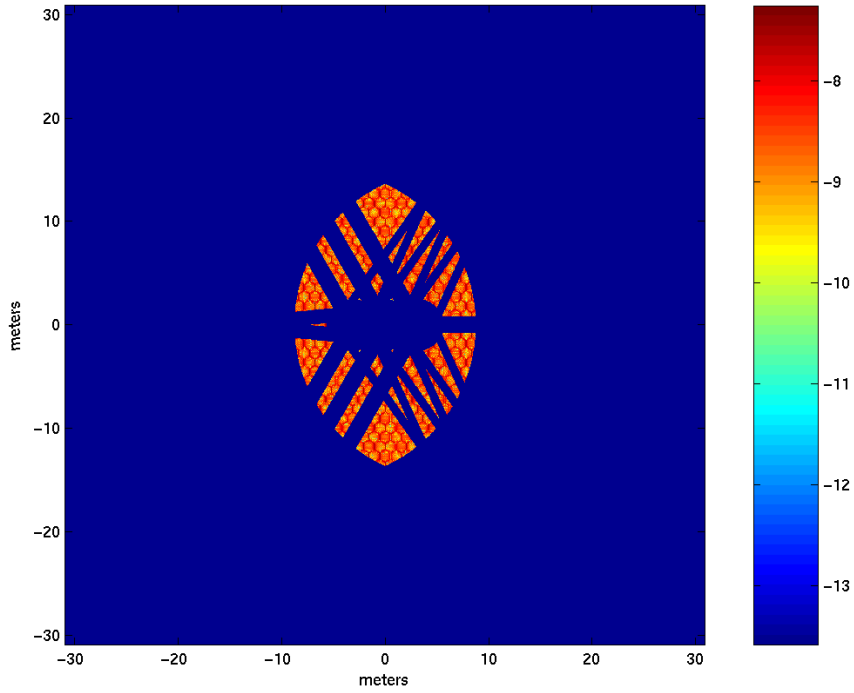

Figure 57: Figure in Lyot plane after the Lyot mask (Logarithmic scale).

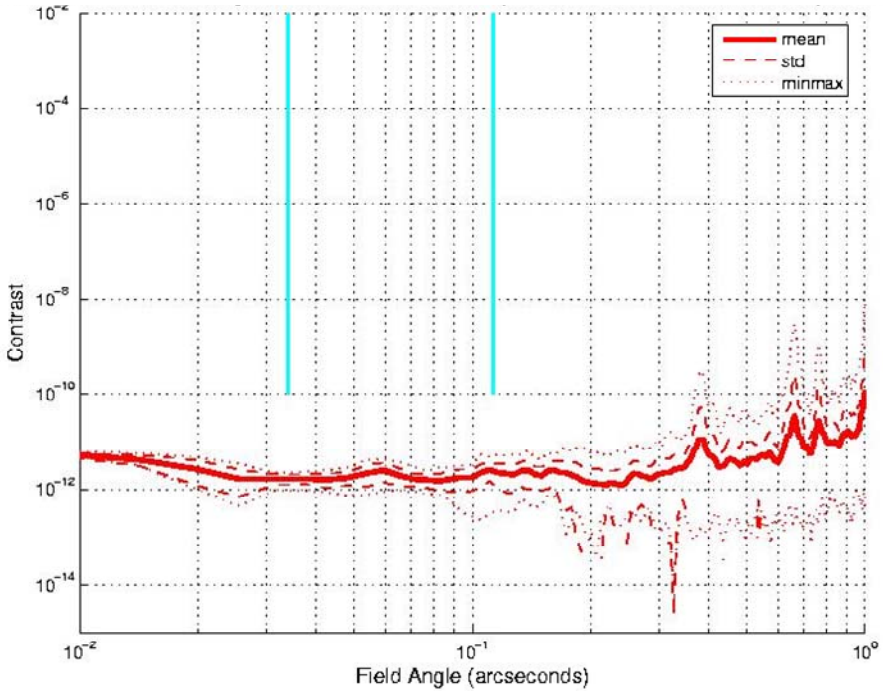

Figure 58: Radial average of contrast for two stage nuller.

The broadband contrast performance of the nuller was evaluated for a bandpass corresponding to R = 5 with a central wavelength of 1.65μ. The spectrum was assumed to be flat across the bandpass. The radial average contrast is plotted in Figure 59. The salient feature of this plot is that the contrast is insensitive to a finite bandpass. This result is consistent with the way the null is generated. The pupil shear is obviously independent of wavelength. If the π phase shift is truly achromatic, then the field from an on-axis source is perfectly cancelled in the pupil overlap region for all wavelengths.

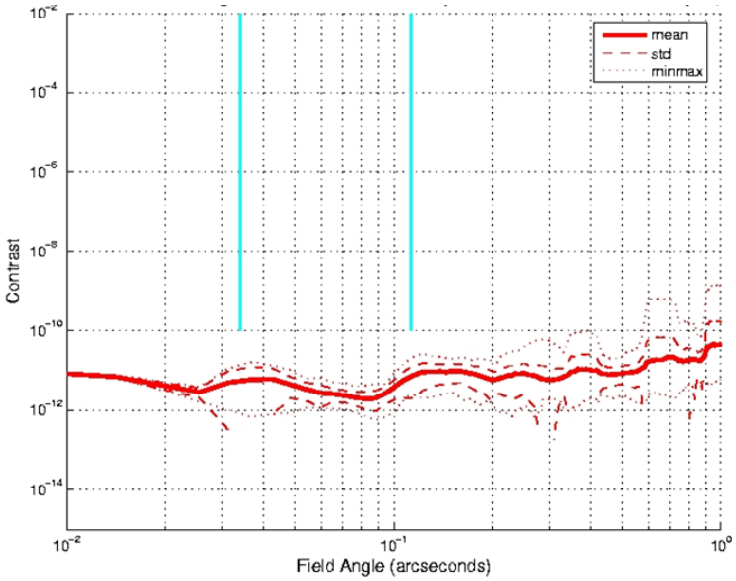

Figure 59: Radial average of contrast for broad-band (R=5) simulation of two stage nuller.

The contrast floor in Figure 59 is about 10^{-12}. The nuller architecture should in principal be able to perfectly null the field from an on-axis source in the overlap region in the Lyot pupil. The Lyot stop shown in Figure 55 was explicitly designed to transmit only that part of the pupil in which the cancellation is perfect. This implies that we should expect a contrast floor of 10^{-15} to 10^{-16}. The hypothesis for the failure to meet the expected contrast floor is that the artifacts introduced into the input pupil by the Grey pixel approximation result in imperfect cancellation of the field. This hypothesis was tested by applying a Blackman taper to the field in the first image plane in the model. This has the effect of applying a low pass to the input pupil and suppressing ringing in the Lyot plane caused by high spatial-frequency artifacts introduced by the Grey pixel approximation. With a suitable choice for the width of the tapering function, we found that we could lower the contrast floor to 10^{-14} to 10^{-16}, which is close to the expected numerical limit. In practice, of course, the contrast floor will be limited by residual tip/tilt, and chromaticity and errors in the π phase shift.

4.2.3.4 Shear sensitivity study

The baseline model summarized in the previous section used a single stage shear of 3.12 meters, which is equal to the width of three hex segments. This shear guarantees that the replicas of the input pupil in the Lyot plane are offset from one-another so that the image of the segment gaps overlay each other exactly. This is required to produce exact cancellation of the field from the on-axis source. We wished to understand how the contrast performance degrades as the shear deviates from a non-integral number of segment widths.

To answer this question, the contrast was calculated for the following set of pupil shear errors (in segment widths): {±0.2 ±0.1 ±0.003}. A shear of 0.003 segments corresponds to about one segment gap width, which translates into a shear of 10μ in a re-imaged pupil that is 0.1 meters in diameter. A shear of 0.1 segments translates to a shear of 300μ in the same re-imaged pupil.

The simulation was monochromatic and used a wavelength of 1.65μ. The Grey pixel input pupil is displayed in Figure 42. The Lyot stop was redesigned to handle the range of pupil shears. This Lyot stop is displayed in Figure 60. Nominal results with a shear of three segments and this Lyot stop were calculated. These results are virtually unchanged from the results reported in Figure 59, which indicates that the new Lyot stop did not impact the results about to be reported for finite shear errors. The results for the relatively large shear errors (±0.2 and ±0.1 segment widths) indicate a significant compromise in performance. The result for the case where the shear was equal to 2.9 segments is displayed in Figure 61. The degradation in performance with respect to the baseline case is about one to two orders of magnitude and performance is below the PFI goals from 0.3-1 arcsecond where segment gap diffraction is most significant.

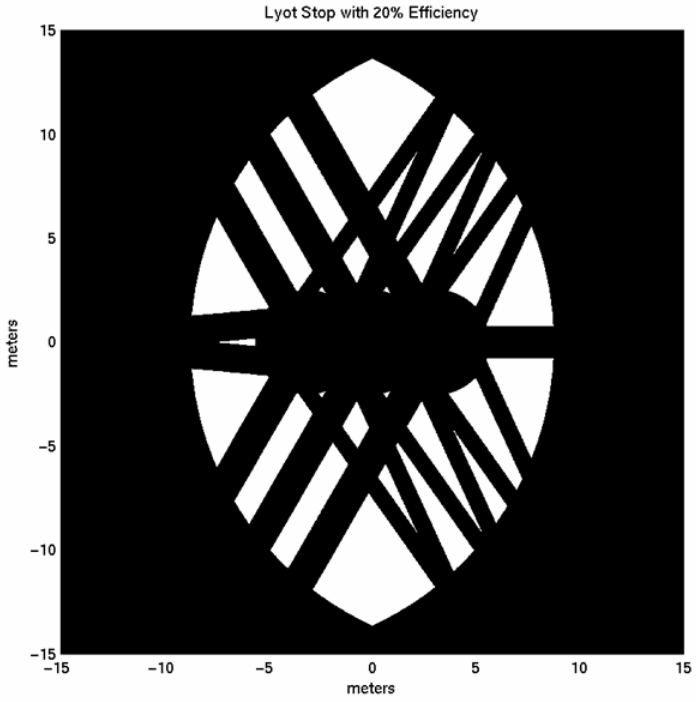

Figure 60: Lyot stop for shear sensitivity study.

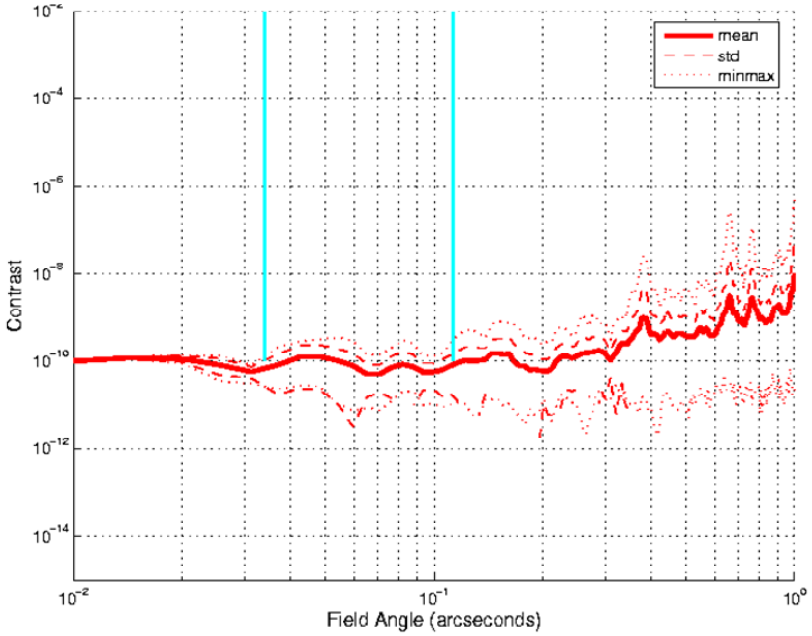

Figure 61: Radial average contrast for 0.1 segment shear error.

The results for a shear of one gap width (0.4% of a segment size) are more encouraging. This result is displayed in Figure 62. The performance degradation appears to be minimal. This sets a tolerance on the pupil shear of ~0.4% of the segment size. If this level of control is impractical, then a more aggressive Lyot stop, with lower efficiency, will need to be used to mask out regions in the vicinity of the segment gaps.

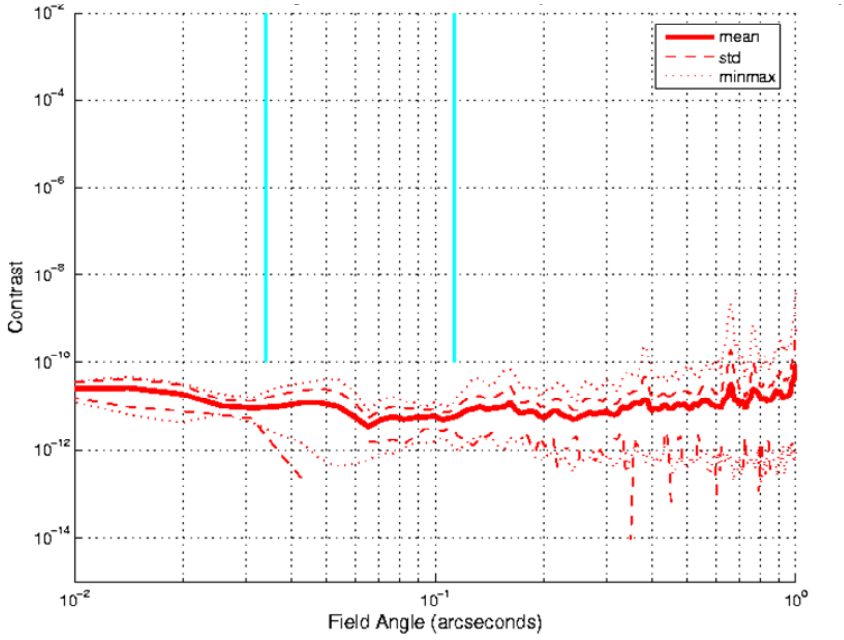

Figure 62: Radial average contrast for contrast for shear error equal to one segment gap width.

4.2.3.5 Nuller contrast sensitivity to tip/tilt errors and finite stellar diameter

The functional relationship for nuller contrast as a function of telescope tip/tilt errors can be shown to be (Mennoson et. al. 2004):

$$N = 0.25\left[\frac{2\pi^2}{\lambda^2}\left(s^2\theta^2\right)\right]^2$$

where

N = contrast at the center of the PSF

λ = wavelength

s = shear

θ = telescope tip/tilt.

For $\lambda = 1.65\mu$, s = 3.12 meters, and $N = 10^{-6}$, this equation predicts a value for θ that is approximately 1 milliarcsecond (mas). It is important to note that the contrast at the third Airy ring will be about 0.01 times the value at the peak. This means that the contrast in the region of interest will remain better than 10^{-8}.

For a star of finite angular diameter, D^*, one obtains (from integrating the previous expression over the stellar diameter) , the following equation (Mennenson et. al. 2004):

$$N = \frac{1}{32}\left(\pi s\frac{D^*}{\lambda}\right)^4$$

For $\lambda = 1.65\mu$, s = 3.12 meters, and $N = 10^{-6}$, this expression predicts a stellar diameter of ~2.6 mas. Again, at the third Airy ring, the contrast will be on the order of 10^{-8}.

The advantage in tip/tilt sensitivity of a two-stage nuller, which has a fourth order sensitivity to θ, over a single-stage nuller, which has a second order sensitivity to θ, is shown in Figure 63. For the same case described above, achieving a contrast of 10^{-8} at the third Airy ring for a single stage nuller requires controlling or correcting the tip/tilt to about 0.035 mas.

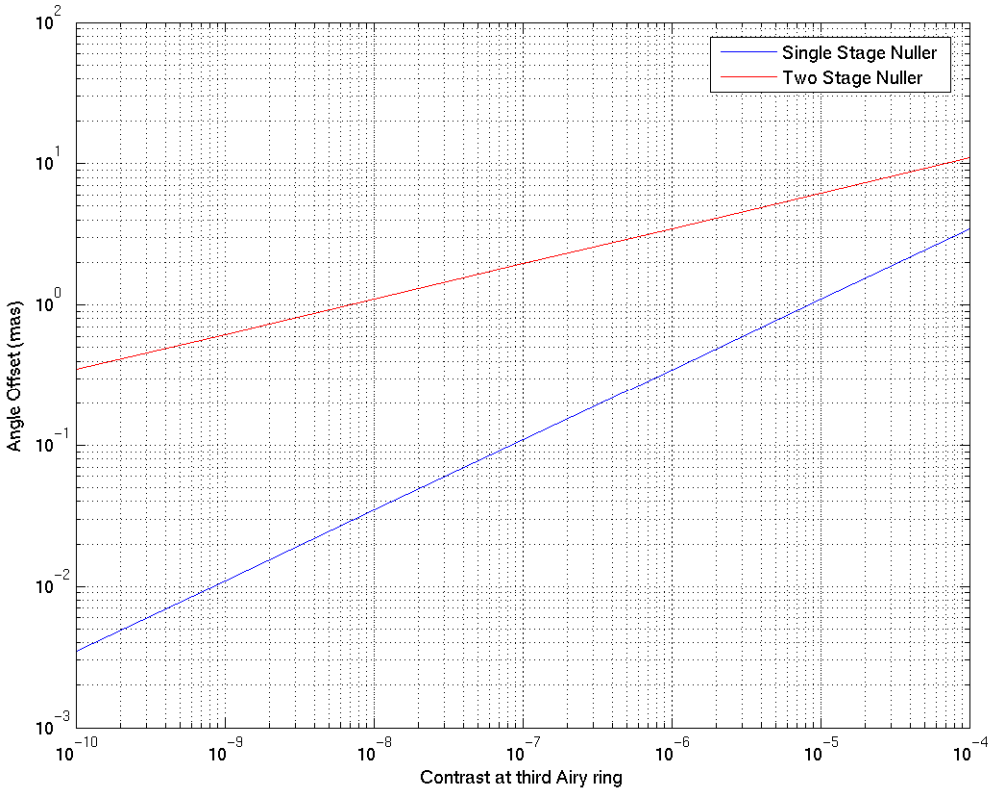

Figure 63: Pointing offset/error allowed versus desired contrast.

4.2.3.6 Nuller sky coverage and operational scenarios

The nuller transmittance as shown in Figure 33 results in incomplete field coverage for any single exposure. It is therefore important to develop an observing scenario to maximize the chance of completing the desired science with the maximum efficiency. This optimization and indeed the technique used will vary depending on the science case and is outside the scope of this feasibility study. We have, however, constructed a baseline observing scenario, the details of which are in the PFI OCDD. In short the observing scenario is to observe a target at a given shear for 10 min, change shear and observe for 10 more minutes, then rotate the pupil (and field) 120 deg and repeat the 20 minute observation block, then rotate another 120 deg and repeat the 20 minute observing block again. The 120 degree rotations allow the use of the same Lyot stop. One of many possible optimizations is to allow a 2nd (or 3rd) pupil stop to enable additional rotation angles of the pupil. The resulting focal plane coverage is show in Figure 64. Figure 65 shows the fraction of the field covered with a minimum efficiency per pixel. As an example, 80% of the field is covered with at least 25% efficiency per pixel.

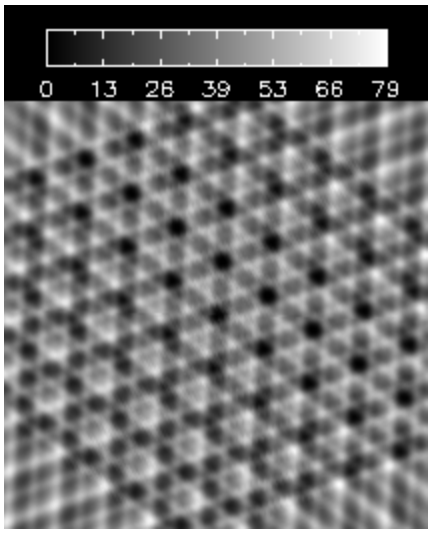

Figure 64 Resultant spatial coverage for three roll angles and two shears. The field is 1.2 x 1.2 arcseconds and the pixels are Nyquest sampled at H.

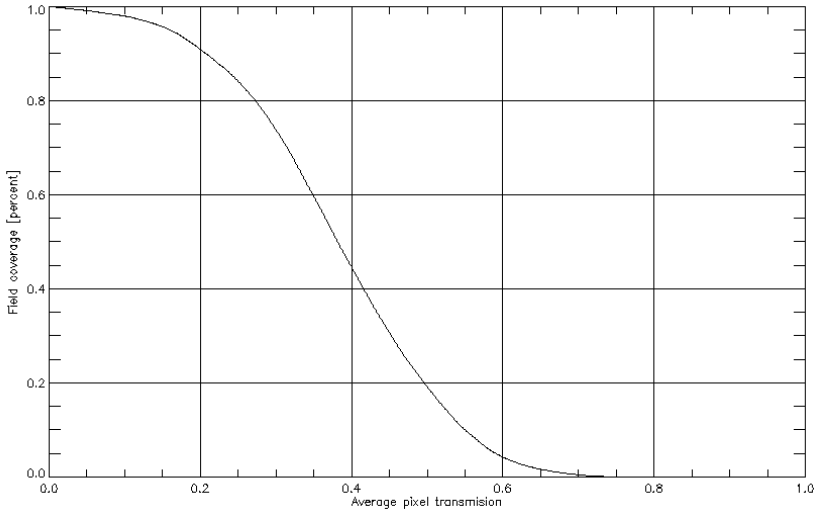

Figure 65 Plot of coverage of the 1.2 x1.2 arcsecond field versus minimum pixel transmission

4.2.3.7 Segment projection and non-uniform spacing

A related problem is the fact the segments on the PM may be uniform in size. However, because of the curvature of the PM, the spacing of the projection of the segment gaps onto the Lyot pupil will not be uniform. The projected gap spacing will decrease from the center to the edge of the pupil. A nuller shear designed to be an integral number of hex segment widths will produce pupil replicas in the Lyot plane for which the gaps no longer line up exactly. This problem will be worse along the direction of pupil shear.

In order to gain some insight into the impact of this effect on contrast performance, we solved for the projected location of the gaps onto the Lyot pupil along the direction of shear. The Lyot stop that we used for a three segment shear is approximately 18 meters

wide. At 9 meters from the pupil center, a segment width of 1.04 meters projects to 1.0285 meters. This gives a difference of 0.0115 meters per segment. For a three segment shear, the gap offset becomes ~0.035 meters.

The contrast was recalculated using a shear of 3.155 meters. Surprisingly, these calculations did not show any degradation in contrast. This result is probably due to the fact that the diffraction spikes in the Lyot pupil at the locations of the gaps are approximately 0.05 meters in width, which means that the shear offset is effectively smaller than one gap width. Our working hypothesis is that the large diffraction spike widths are an artifact of the Gray Pixel approximation used to construct the model of the TMT pupil.

We expect that shear offset errors that are larger than one gap width (0.003 meters) will produce similar contrast degradation. This is supported by the results quoted earlier where shear offsets of 0.1 and 0.2 gap widths produced similar contrast results. The contrast results plotted in Figure 38 indicates that the PFI contrast requirements are not quite met. However, this is probably a worse-case result because there will be parts of the pupil (near the center and orthogonal to the direction of shear) where the offset will be less than one gap width. This issue will be re-examined at a later date with a more accurate representation for the TMT pupil. If these results still indicate a potential problem, we will examine whether the problem can be partially mitigated by an optimal choice of shear or by masking out the offending gaps.

4.2.4 Other DSS techniques

Several other DSS techniques have been evaluated, though with less detailed modeling than described above. Two – the Optical Vortex and the Four Quadrant Phase Mask – appear to provide very small inner working angles but only on unobscured apertures. Apodozation of the segment boundaries, described in 4.2.4.3, is a viable approach to controlling segment diffraction that will be further studied, but does not help with the secondary and large secondary supports. A hybrid combining this apodizer with a band-limited coronagraph is an interesting possibility for further study, particularly if the TMT secondary supports become smaller.

4.2.4.1 Four Quadrant Phase Mask

One way of obtaining small IWAs is by using an image plane phase mask that modifies the phase, rather than the amplitude, of the focal plane field. The optical vortex (OVC) described below is an example of this. A simpler example is the four-quadrant phase mask (FQPM) coronagraph, which has alternating phases of 0 and π radians in the four image plane quadrants (Rouan, et.al. 2000, Riaud, et.al. 2001, Gratadour, et.al. 2005). Like the OVC, it can in principle achieve perfect point source rejection with an unobscured circular pupil, its IWA is small ($\leq\lambda/D$), and the mask's lateral structure (but not its thickness) is achromatic. However, performance is degraded by beam obstructions such as an on-axis secondary mirror and secondary support legs, and so only modest levels of starlight rejection have been achieved to date with an on-axis system.

The high degree of starlight rejection possible in theory with a FQPM coronagraph leads to consideration of its use with an unvignetted circular pupil. Such a pupil can be obtained simply by using an off-axis section of a large on-axis telescope. While the current generation of telescopes would allow an off-axis sub-aperture no larger than about 3 - 4 meters, the large size of the TMT would allow off-axis sub-aperture diameters of up to about 13 m, much larger than any existing on-axis telescope.

The approach is then very simple to implement: a set of relay optics magnifies the pupil prior to the AO system, and a set of appropriate masks (a FQPM in an image plane, and a circular stop in a subsequent pupil plane) is inserted into a coronagraphic camera (Haguenauer et al. 2005). The main question is then whether a smaller IWA and better wavefront correction can compensate for the sub-aperture's larger diffraction beam width, and current modeling indicates that there is a small central region of the focal plane where this appears to be the case. This approach is currently being investigated on the Palomar 200 inch telescope, where it has already produced Strehl ratios as high as 94% (Figure 66), and stellar rejections of 0.01, on a 1.5 m off-axis aperture (Figure 67). Again, this would not be able to null out structure from segment gaps, which would limit contrast in the region around $\lambda/1.2$m. This architecture is extremely sensitive to the pointing of the star to the center of the FQPM. In addition, stopping down the telescope increase sensitivity to speckle and photon noise and reduces the final contrast by a factor of $(13/30)^2$.

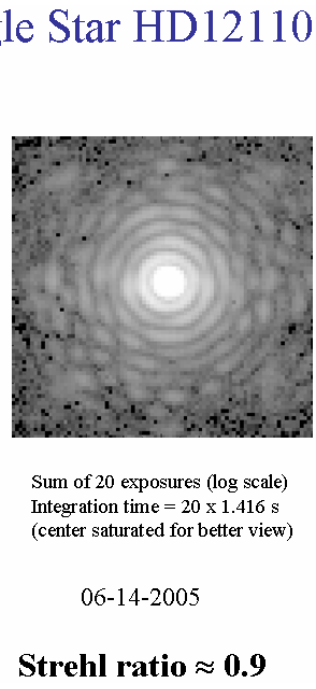

Figure 66: PSF with Palomar AO system matched to a 1.5-m off-axis aperture.

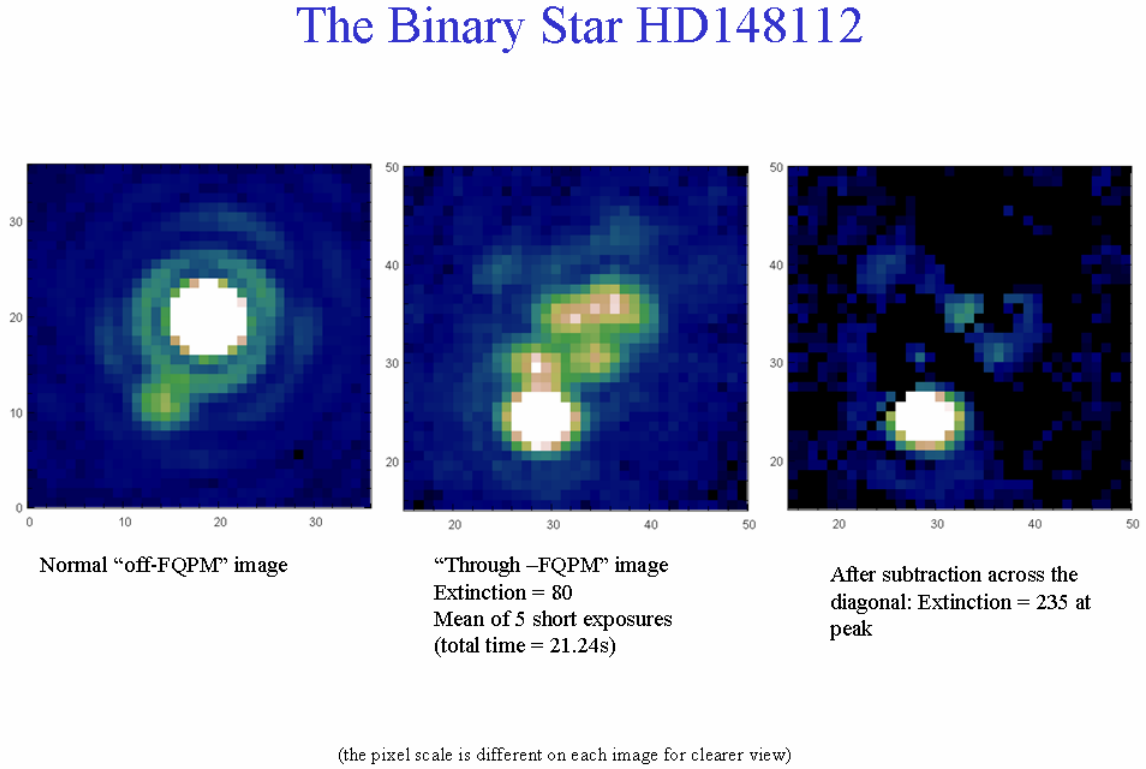

Figure 67: Suppression of on-axis source by FQPM.

4.2.4.2 Optical vortex

A key design feature of a conventional coronagraph is an occulting mask in the image plane of a telescope. The mask obstructs starlight, allowing faint, nearby objects to be imaged. However, the performance of a conventional coronagraph degrades with increased spectral bandwidth and increased wave-front aberrations. These limitations can be alleviated through the use of an optical vortex coronagraph (OVC).

Unlike the conventional coronagraph, an OVC makes use of a transparent optical vortex phase mask rather than an opaque amplitude mask (Figure 68). The vortex mask is created by etching a helical relief pattern onto an optical surface. The azimuthally varying surface of the mask transforms the impinging phase front into a helix. Total destructive interference occurs along the axis of the helix and this may be used to null out an on-axis star without significantly affecting the images of other nearby bodies. It has been shown in numerical simulations that an OVC is less sensitive to aberrations, obtains higher planet light throughput, and a higher planet/star contrast than a conventional coronagraph (Palacios 2005). In addition, the mask may be designed to create higher order modes that also contain a dark null for wavelengths spanning the visible spectrum. This effect may be exploited to increase the operating spectral bandwidth of an OVC (Swartzlander 2005).

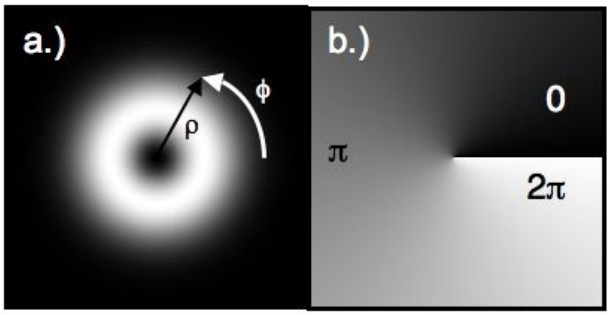

Figure 68: The amplitude (a) and phase (b) profiles of an m=1 optical vortex embedded in a Gaussian beam. At r=0 a dark null of destructive interference, known as a vortex core, forms at the center of a helical phase ramp. In a single revolution about the helix the phase varies from 0-2π radians.

For an unobscured aperture, an OVC achieves very good performance (Figure 69). However, the OCV is essentially unable to control diffraction from secondary or segment obscurations, and is hence impractical for TMT. As with the FQPM, this is an interesting option for deployment on a subsection of the input pupil, but is not a PFI baseline.

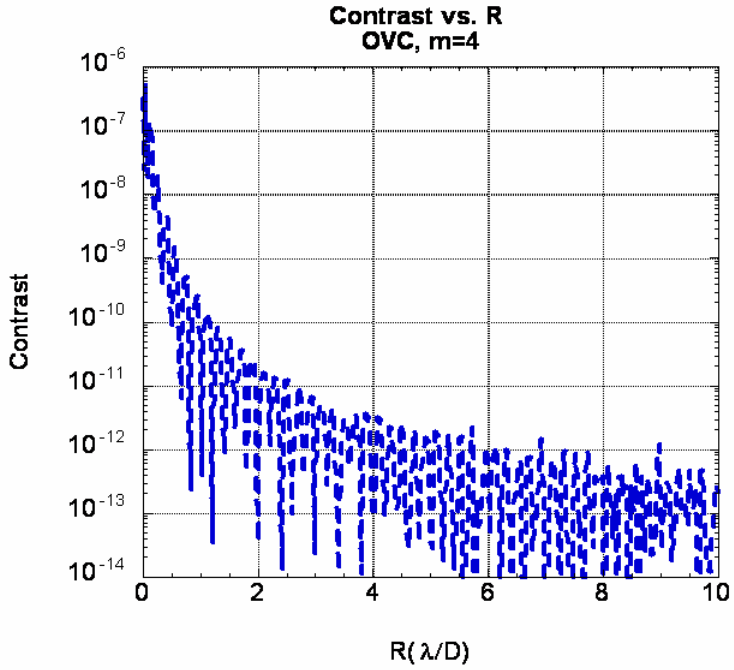

Figure 69:Contrast Performance of Optical Vortex Coronagraph for an unobscured aperture.

4.2.4.3 Segment Apodizer

Apodization of the telescope pupil segment gaps can also significantly attenuate their scattered light. Remember that the PSF of a unaberrated telescope $T(x,y)$ is given by tt^*, where lower-case letters denote Fourier transforms of upper-case pupil-plane quantities. In general, a pure apodization multiplies the original telescope transmission function by an apodizer $A(x,y)$ for a effective transmission in the pupil plane given by $P(x,y)=AT$. To suppress diffraction, we want a target pupil function $AT = P_{targ}(x,y)$ selected so that its Fourier transform p rapidly approaches zero at large angles.

For a uniform unobscured aperture, so that $A=P_{targ}$, an example would be the Blackman function, given by

$$P_{targ}(x,y) = 0.42 - \tfrac{1}{2}\cos(2\pi r) + 0.08\cos(4\pi r)$$

where $r=(x^2+y^2)^{0.5}/D$. More sophisticated apodizers such as the prolate spheroidal wavefunctions can produce contrast $>10^{10}$ at 4 λ/D.

Apodizers have three major limitations. First, their throughput is relatively low –70% or more of the starlight is blocked, and the angular resolution is also degraded. For ground-based AO, especially on a 30-m telescope, this is not an insurmountable issue, since photon noise will rarely dominate speckle noise. Second, they are difficult to manufacture – their behavior depends crucially on high levels of attenuation occurring smoothly at the edge, requiring optical density >5 without significant chromaticity or phase errors. Finally, of course, simple apodizations such as the Blackman window do not significantly help telescopes with transmissions T containing complex internal obscurations such as segment gaps or even secondary supports.

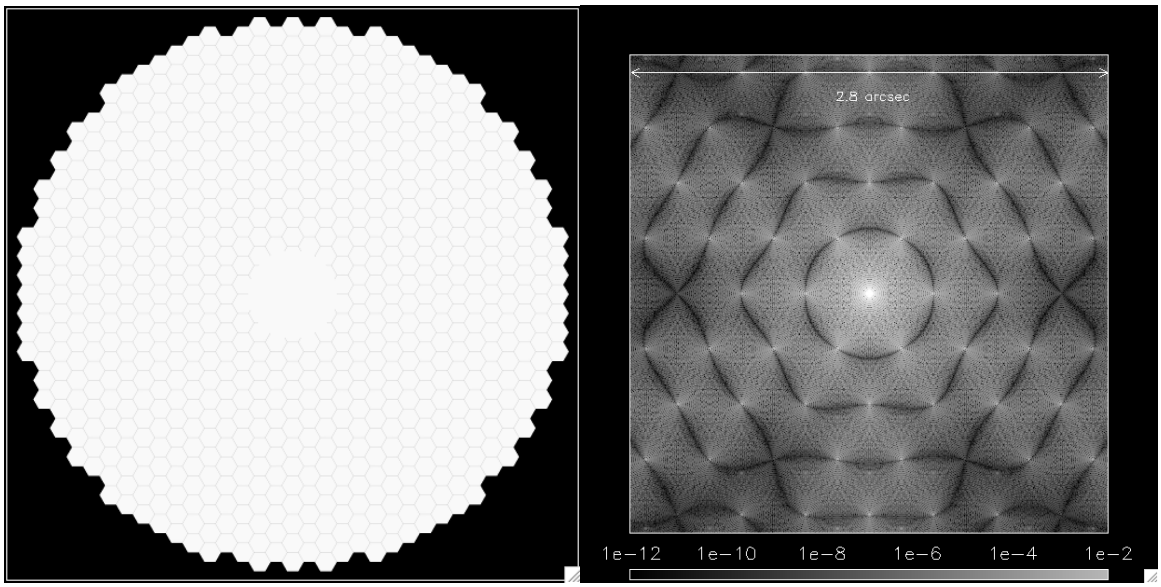

Figure 70: Pupil (left) and PSF (right) of a segmented 30-m telescope with 1-m segments. The secondary mirror and supports have been removed to emphasize the effects of gap diffraction.

The trivial solution $A=P_{targ}/T$ is of course impossible since $T=0$ at some locations. If we simply set the apodizer function to the target function $A=P_{targ}$ the apodization suppresses

the effects of diffraction from the outer edge, but not of the internal structure. Figure 70 and Figure 71 compare the pupil and PSF for a massively-segmented telescope to that of the same telescope apodized with a simple Blackman window.

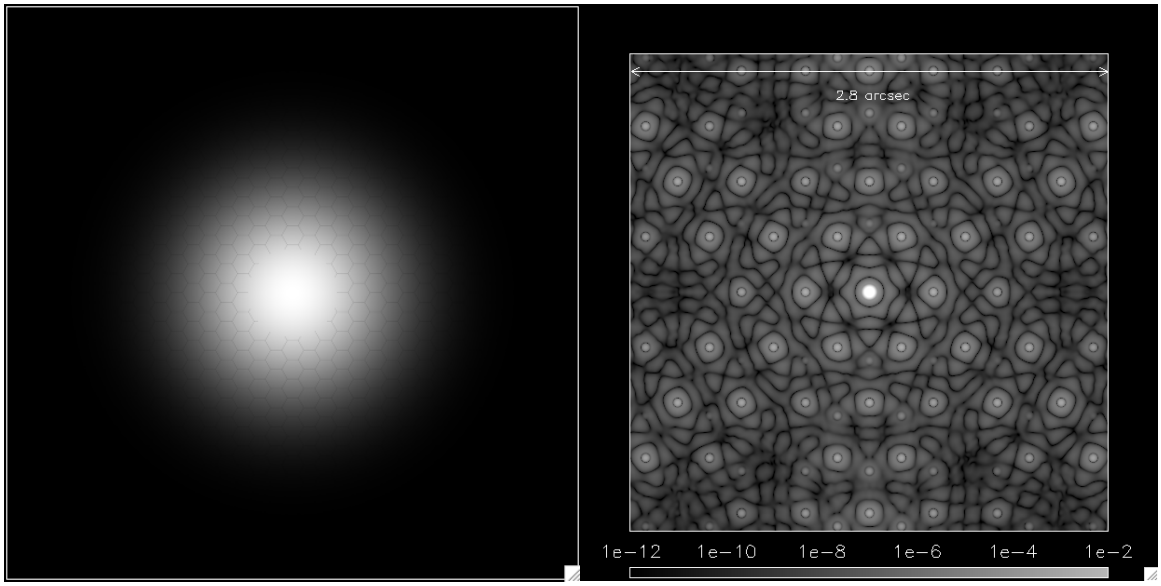

Figure 71: Pupil (left) and PSF (right) of the same telescope as above, multiplied by a prolate spheroidal wavefunction.

However, if we only wish to suppress diffraction over part of the focal plane, say for example out to a maximum angle θ_d, more effective solutions exist. For *P* to have the same PSF as P_{targ} at requires that the Fourier transforms are equal up to a spatial frequency $f_d = \theta_d/\lambda$, i.e. $\mathrm{FT}(AT)h_{f_d} = \mathrm{FT}(P)h_{f_d}$ where h_{fd} is a tophat function of radius f_d. We refer to these as band-limited apodizers, since they solve the apodization equation only over a specific frequency range. This is equivalent to solving $(AT) * H_{f_d} = P * H_{f_d}$ subject to the constraint that *A* must be real and <=1. In neral this equation is not exactly solvable but we can find many families of approximate solutions through numerical techniques. Since the AO system itself produces a square dark hole, it could be more efficient to use a square range of spatial frequencies than a circular tophat. Figure 73shows such a solution for the pupil given in Figure 72 and its corresponding PSF.

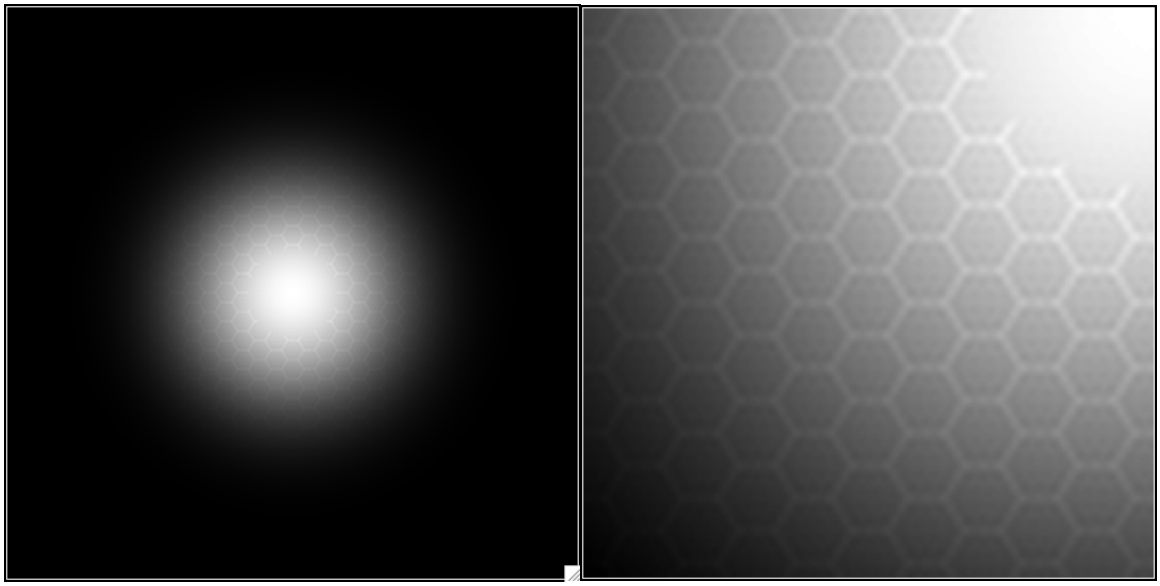

Figure 72: Band-limited apodizer: whole pupil (left) and close-up (right.) The quasi-periodic structure inside each segment acts to suppress the diffraction over a specific range of spatial frequencies, producing the PSF shown below. The individual features have contrasts of 20-30% relative to their immediate neighbors, so the total throughput is dominated by the target Blackman or prolate wavefunction envelope.

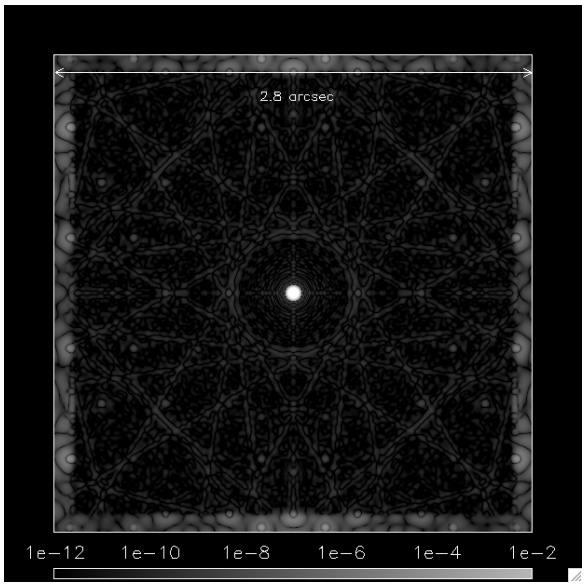

Figure 73: PSF corresponding to the apodizer in Figure 72 multiplied by the telescope transmission from Figure 70. The star pattern is the underlying PSF of the numerical approximation to the prolate wavefunction. The bright edge corresponding to the bandlimit can just be seen.

Although this is a powerful technique, it does have several limitations in common with more classic apodizers. Although the segment features themselves only reduce throughput by 20-30%, if the overall apodization envelope is a high-contrast function such as a Blackman window the total transmission is <20%. Second, in common with most

apodization functions, the inner working distance is 4 λ/D. Third, for θ_d~1", this approach works well for features smaller than 10 cm but not for larger features, such as the TMT secondary supports. (Interestingly, this means that TMT is actually easier to apodize than a telescope with large segments and large gaps such as the GMT reference design.) Third, manufacturing issues remain. The fine structure on the apodizer can be manufactured using techniques such as electron-beam irradiation of sensitized glass, but the overall taper at the pupil edge still requires attenuation of 10^4 or more. A hybrid approach may be more effective, in which the band-limited apodizer is used to attenuate the effects of segment gaps in combination with a more classic Lyot-type coronagraph (or its variants, e.g. Kuchner and Traub 2002 and Soummer 2005) that suppresses diffraction from the primary and secondary mirror edges and the secondary supports. This approach will be analyzed in a future paper and remains worth research for TMT applications, especially if the TMT secondary supports can be reduced in size to 10-20 cm.

4.2.5 Trade study summary

The visible nuller appears to be the only DSS architecture that is both practical with current technology and meets PFI's requirements, especially the small IWA on an obscured aperture. We have therefore selected it as the baseline. The band-limited Lyot coronagraph could be used in an early implementation with larger IWA, and the segment apodizer will continue to be studied.

4.3 PFI Nuller Optical Layout

4.3.1 Overview

The nulling interferometer employed as a diffraction suppression system for PFI is shown above in Figure 74. Before we begin the discussion of the optical layout, a brief description is in order. A nulling interferometer suppresses starlight by causing the broadband light to completely, destructively interfere with itself while at the same time allowing a significant fraction of light from the companion object to pass through. A couple of technical challenges to this technique are: 1) creating the broadband pi phase change internal to the interferometer that produces the deep cancellation of starlight and 2) maintaining the stability of the interferometer during the science observation.

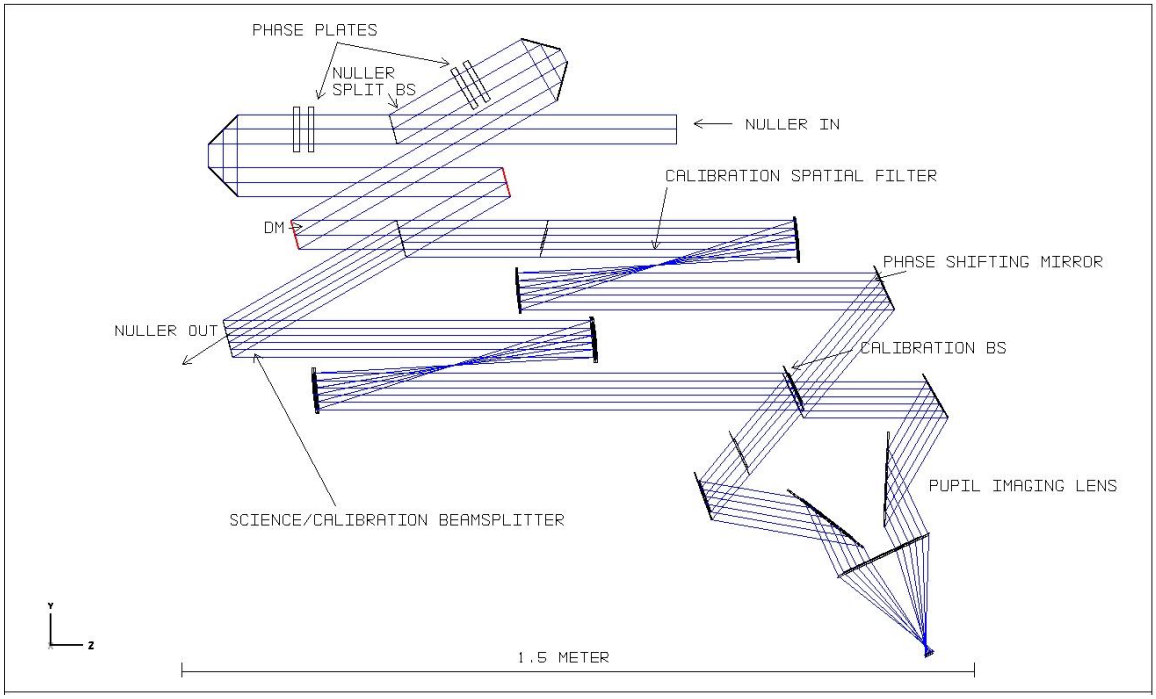

Figure 74. Shown above is a diagram of the PFI nuller and calibration unit highlighting the location of key optical elements in the system. Light enters the system from the top right and exits to the science camera at the bottom left. This design shows only a single-stage nuller; for the dual-stage nuller we prefer, a beamsplitter would bring the beam out of the plane into a duplicate nulling interferometer.

There are a couple of important advantages to this technique worth highlighting. First, as opposed to classic coronagraphs this one does not rely upon a focal plane mask for removing starlight. In the canonical coronagraph, this focal plane mask acts as a high spatial frequency pass filter, which creates a blurring of sharp features in the re-imaged (Lyot) pupil plane. A simple mask equivalent to the geometric image of the original pupil would not suffice to suppress the starlight. In order to achieve very high contrast levels, this plane needs aggressive masking; the primary impact is less system efficiency. Since the nulling interferometer does not have such an image plane occulter, the Lyot plane will be essentially a geometric image of the original pupil thus permitting higher throughput efficiency for the science object.

Second, the pupil shear that creates the scale of the nuller fringe pattern can be continuously adjusted. Although in practice this may not be desirable, due to challenges such as masking of segment gaps of the primary mirror, in principle it is more easily adjusted than an image plane occultor. In a similar vein, the contrast of the nuller can be modified by changing the phasing of internal the interferometer. One can imagine situations when intermittent knowledge of the stellar position would be valuable (i.e. astrometry), detuning the internal nuller path length makes this possible.

4.3.2 System Description

A collimated input beam is split into two separate paths at the nuller split beam splitter. Phase plates internal to the nuller create a pseudo-achromatic π phase difference between the two arms of the interferometer so that at the nuller recombination beamsplitter, the symmetric port has a destructive interference in the pupil overlap region (nulled output) while the other side of the beamsplitter produces a constructive interference (bright output). These phase plates cannot produce a π phase change over an arbitrarily large science optical bandpass, and they will therefore need to be set for each one.

Subsequent to the phase plates is a pair mirrors at 90 degrees to each other: the rooftop optics. In this configuration, the pathlength and shear adjustments are orthogonal to the degrees of freedom for the rooftop. Pathlength adjustment is done by translation of the rooftop along the input beam direction while shear is done by a transverse motion. Since these are orthogonal adjustments, searching for the coherent fringe for a given shear size is trivial. There are nulling architectures where this is not the case and some care must be taken in determining the system requirements and instrument geometry.

A deformable mirror is located in one arm of the interferometer. The DM elements are 400 μm in size and it is 128×128 actuators in size for a pupil size of 51.2 mm. The angle of incidence on this element is a moderately small 15 degrees. Since the DM has limited stroke, the flat optical mirror in the other arm of the nuller can be used to de-saturate the piston term of the DM to stay on the white light null.

The nuller recombination beamsplitter completes the diffraction suppression system. Do to the extreme symmetry required at recombination; this beamsplitter must be as identical to the first, split beamsplitter as possible: same substrate, thickness, coating, wedge, angle of incidence and orientation.

4.3.3 Contrast floor set by phaseplate chromaticity

Optimization of the air delay and differential thickness for a pair of glass allows one to create a phase difference between the two beams at the point of recombination that is very nearly pi radians over the H band as shown above in Figure 75. The equivalent null depth, or contrast, for this solution is given below in Figure 76. The average contrast over this bandpass is 4.2 10^{-9}.

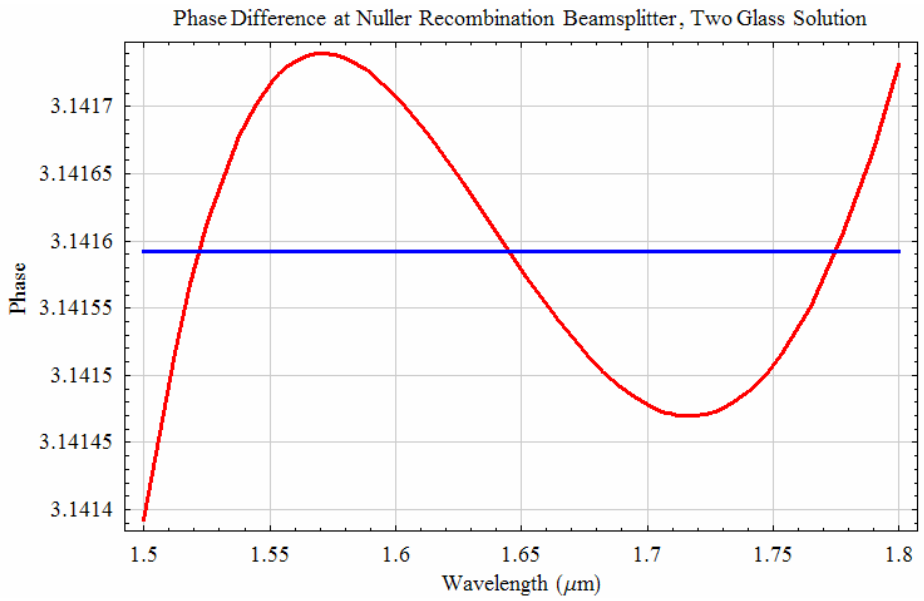

Figure 75: Phase difference at the nuller recombination beamsplitter as a function of wavelength.

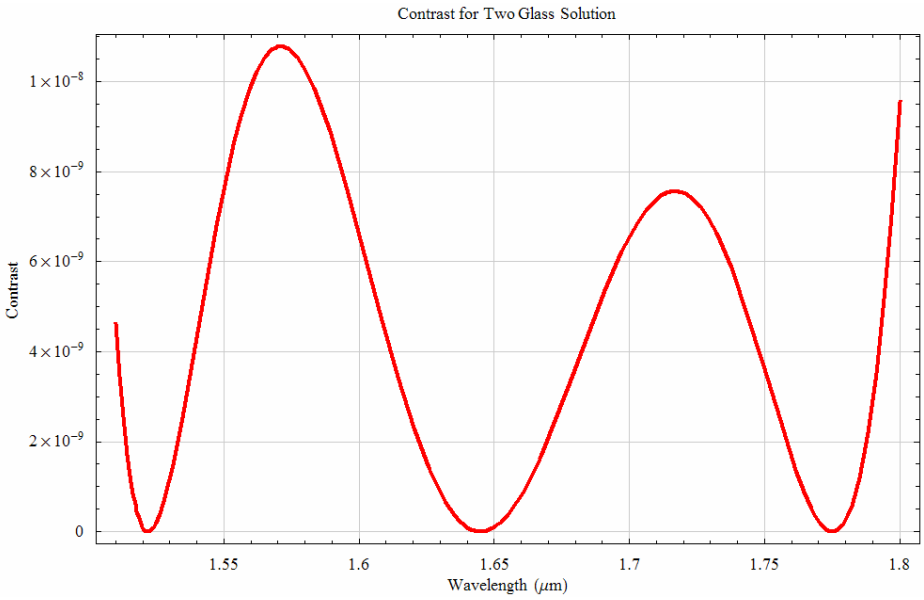

Figure 76: Contrast performance of phase plate design as a function of wavelength.

5 Post-DSS IR WFS

5.1 Introduction

The next major subsystem in the PFI optical beamtrain is the back-end wavefront sensor and its associated AO system. The backend wavefront sensor performs aberrometry on the ac-coupled wavefront exiting the diffraction suppression system or coronagraph; the DSS output is ac-coupled in the sense that the coronagraph or nuller removes the average (DC) intensity, leaving phase aberrations only. By measuring and compensating for these residual aberrations the backend AO can help PFI approach photon-limited contrast performance. Current AO coronagraph systems are all limited in contrast performance due to intense speckle noise, primarily from quasi-static sources (Racine et al. 1999). The dual in-series AO systems approach presented here, with the coronagraph located in-between (as in Figure 77), offers some clear advantages that can help PFI suppress speckle noise — especially at the low- to mid-spatial frequencies (30–100 mas) within the AO control radius.

Back AO system key requirements:

- Measure the wavefront output from the DSS
- Cover a wavelength range of 1-2.5 μm
- Sample the wavefront at scales matching the front WFS (124 samples per pupil) without significant aliasing errors
- Operate at frame rates of at least 500 Hz (Goal: 1000 Hz)
- Achieve sub-nm levels of accuracy over all controllable spatial frequencies
- Provide correction of quasi-static wavefront errors
- Correct residual atmospheric wavefront errors in regimes where the correction from the front-end WFS can be augmented
- Provide good wavefront measurement and correction on targets with magnitude $H < 10$ mag.
- Meet the requirements of the error budget
- Feed back time-averaged wavefront and pointing and centering information to the system
- Provide wavefront telemetry for diagnostic purposes and subsequent PSF reconstruction

Back AO system key interfaces:

- Accepts collimated outputs from the bright and dark output ports of the DSS with stable relative delays
- Mounted on the same optics bench as the DSS
- Accept commands and return status to the observatory or PFI software

Back AO system design summary

- Near-IR pupil-plane interferometric wavefront sensor
- Controls a single DM in one arm of the DSS

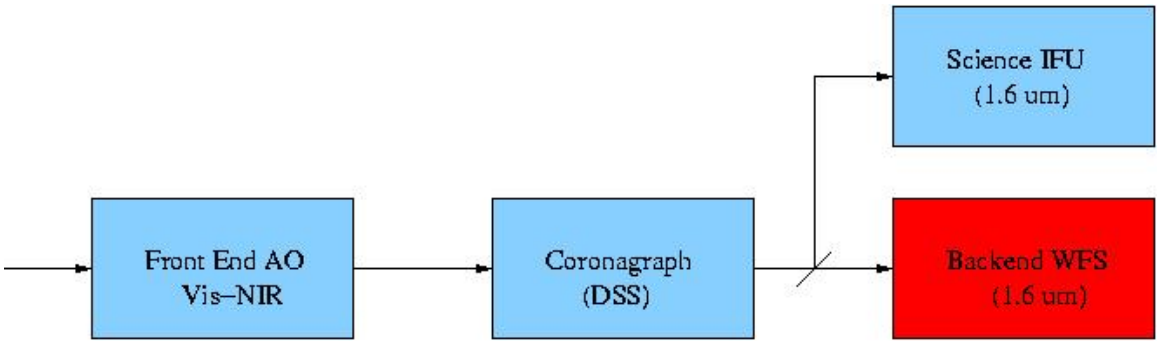

Figure 77: A top level functional schematic of PFI showing the optical location of the post DSS WFS relative to other major subsystem in the beamtrain. Light enters from the left from the telescope subsystem. The WFS will be located in the vicinity of the Science FSU in order to minimize non-common path. All non-common residual errors between the two would have to be calibrated.

5.2 Overview

It is relatively straightforward to conceptualize this approach (see also Guyon 2005 for a broad treatment of wavefront sensors in high contrast imaging). A basic requirement is that the front-end adaptive optics delivers a high Strehl-ratio to the coronagraph. The remaining small aberrations in the "AO-corrected" wavefront, leak through the DSS. The role of the backend wavefront sensor is to efficiently detect and correct for these remaining wavefront errors. In this it offers some clear and unique advantages: it is located behind the coronagraph and views the entire optical system, the effects of non-common path errors before the science channel are much reduced when imprinted onto the AC wavefront, and the detector can be tailored to a particular wavelength and update rate.

The backend wavefront sensor works at the same wavelengths as the science IFS in order to eliminate or minimize chromatic effects. Essentially, a beamsplitter is placed near the exit pupil of the DSS splitting the light between the WFS and the IFS. At this stage the wavefront has several sources of error that should be summarized here:

1. Wavefront amplitude errors or scintillation due to Fresnel propagation in the atmosphere. In addition there will be static or quasi-static amplitude errors due to reflectivity variations on optical surfaces, and Fresnel effects within the optical system.
2. Quasi-static wavefront phase errors due to temporal drifts or misalignments that leak through the front end AO system, or those that occur in the optical beamtrain downstream of it; e.g. thermal drifts and resulting misalignments within the coronagraph.
3. Dynamic residual phase errors from imperfect correction of the atmospheric phase due to any number of reasons – detectivity limitations, finite bandwidth effects and the chromaticity of propagation effects.

These errors collectively result in speckles of varying lifetimes in the science image, thereby degrading the ultimate contrast. As a result of its location in the optical system, the post-DSS WFS views the entire optical beamtrain at the same wavelengths as the science light. Summarily, its functions in the general scheme are twofold:

A. Measure the wavefront exiting the coronagraph (or nuller) and provide real-time corrections to a tweeter deformable mirror inside the nuller. We envision that the post-DSS WFS will run slower than the primary WFS allowing increased detectivity for nanometer-class wavefront errors even with noisy IR detectors.

B. Record the residual wavefront at full rate for post-processing of the science image to improve its contrast and quality; i.e. PSF subtraction or speckle subtraction.

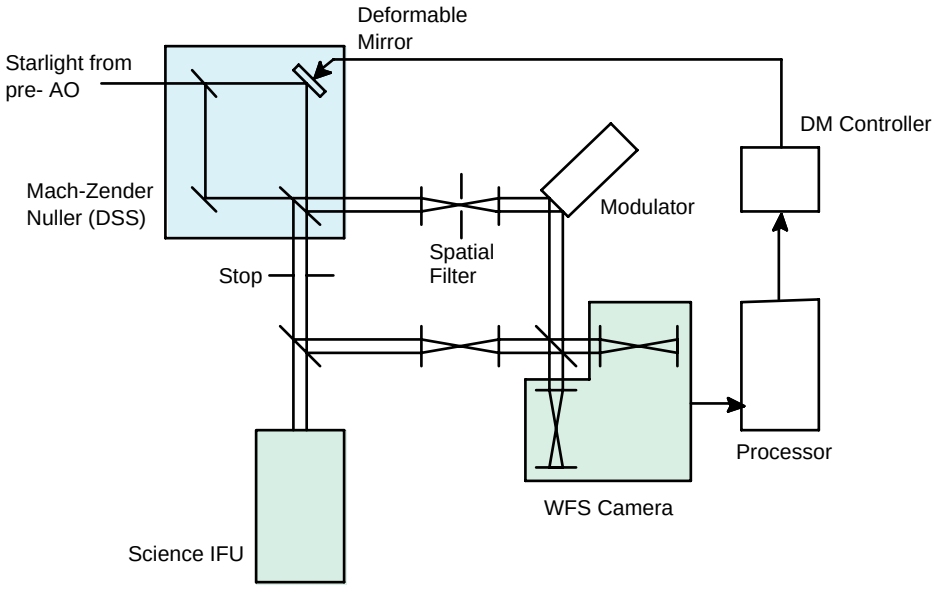

Figure 78: A functional schematic of the backend AO system. A shearing MZ Nuller provides a nulled and a bright output to the backend. The nulled output is split 50:50 between the Science Camera and the post-DSS AO. The bright output is spatially filtered, modulated and interfered with the nulled output at the main beamsplitter. In this case, the combined pupil emergent on both sides of the beamsplitter is reimaged onto a focal plane array. The processor acquires the WFS data and generates a phase map error, which is handed of a controller to close the loop with a DM which is placed in one arm of the nuller.

The backend wavefront sensor uses phase shifting interferometry to measure the complex science wavefront. The basic idea is to interfere the starlight rejected by a classical coronagraph or equivalently the bright output of a nulling Mach-Zehnder, with the science wavefront. This is done in a Michelson configuration and interference from both sides of the main beamsplitter is imaged onto separate focal plane detectors. The rejected wavefront, once spatially filtered, provides the necessary homodyne gain for improved sensitivity. Any putative stellar companion, being extremely faint and incoherent with the starlight, does not affect the measurement. The complex components of the wavefront are measured by employing standard phase-shifting techniques; e.g. the bright wavefront can be modulated by one or more waves using an optical modulator scanning pathlength in a triangle or sawtooth wave, preferably the latter. Integrated measurements taken along the scan can then be inverted in a least squares sense to provide estimates of the real and imaginary parts of the wavefront.

Three separate possibilities for sensing the interference were examined as part of this work; i.e. (1) detection in the pupil plane (2) detection in the image plane (3) mixed detection with homodyne light in pupil plane and science light in the image plane. The three techniques have their own merits. We believe that for survey-style observations the pupil plane WFS has a clear advantage, whereas for scenarios where the existence of a planet or companion is suspected at a certain location in the image, there is a great advantage to redirecting all the homodyne gain to that location in the image plane. We employ at pupil plane WFS in all simulations and optical designs presented in this document, but it is possible to conceive of a WFS that could offer all of the above modes.

5.3 An idealized analysis of the post DSS wavefront sensor

In order to understand the observables of the post-DSS WFS it is instructive to consider a somewhat idealized situation involving an ideal coronagraph. The Lyot pupil of the DSS or the input to WFS may be described as

$$P_L = (1 + \varepsilon(\vec{r}))e^{j\phi(\vec{r})} - 1$$

where $\varepsilon(\vec{r})$ and $\phi(\vec{r})$ capture the spatial distribution of amplitude and phase errors respectively. The amplitude term has both static and dynamic components. The dynamic component is due to atmospheric turbulence, while the static component may be due to reflectivity variations across the primary mirror or due to phase errors on mirrors that are far from the pupil plane e.g. M3. Similarly, $\phi(\vec{r})$ captures phase leakage from the front end WFS, internal propagation effects, and the effects of instrumental alignment drifts during the course of the observations (quasi-static phase errors). For now we drop the dependence on the spatial coordinate $\vec{r}$ for the sake of brevity.

For an input with high Strehl *S*, ϕ<<1 rad., a Taylor expansion of P_L with preservation of terms to second order yields,

$$P_L = \varepsilon - \frac{\phi^2}{2} + j(\phi + \varepsilon\phi)$$

The post-DSS WFS is a homodyne interferometer, viz. it measures the complex pupil P_L by beating it against the coherent starlight rejected by the coronagraph; or equivalently, in the case of the nuller, by beating it against the bright output of the nuller. It provides estimates of the real and imaginary parts of P_L,

$$\hat{X} = \varepsilon - \frac{\phi^2}{2}, \quad \hat{Y} = j(\phi + \varepsilon\phi)$$

For $\widehat{X}$ two separate regimes may be considered: (1) when $\varepsilon << \phi^2$, which is the case when residual phase errors dominate static amplitude and scintillation residuals, $\widehat{X}$ essentially measures second order terms in phase, or (2) when $\phi^2 << \varepsilon$ that is residual phase errors are small enough for the second order term in phase to be ignored. In the latter case, $\widehat{X} \approx \varepsilon$ i.e. the amplitude or scintillation term. We omit any further discussion of amplitude measurement and correction, although it is eminently feasible to build a system that corrects for amplitude effects.

Further make the simplification $\widehat{Y} \approx \phi$ since the second order cross term $\varepsilon\phi$ and higher order terms can be ignored in relation to ϕ at all spatial frequencies. The measure $\widehat{Y}$ therefore provides an estimate of the phase map. Note that in the case of the nuller the measured phase map is a somewhat different relation to ϕ than in the case of a classical Lyot coronagraph.

For a single shear nuller with θ^2 rejection on-sky the backend wavefront sensor measures in the overlapped region,

$$\widehat{Y} = \phi(\vec{r} - \vec{s}/2) - \phi(\vec{r} + \vec{s}/2)$$

where $\vec{s}$ is single shear vector. Similarly, for a double shear nuller with a θ^4 response on-sky,

$$\widehat{Y} = 2\phi(\vec{r}) - \phi(\vec{r} - \vec{s}) - \phi(\vec{r} + \vec{s})$$

where $\vec{s}$ is the shear vector. for this reason it is easiest to place the DM that the post-DSS WFS uses for wavefront correction along one arm of the nuller rather than sharing control on a single tweeter deformable mirror along with the front AO system (see schematic in Figure 78).

5.4 Choice of wavefront sensor

As mentioned earlier in section 5.1 three separate interferometric wavefront sensors were evaluated as options for the PFI backend AO. The three main configurations are --

A. Pupil Plane WFS: the two beams are interfered in the pupil plane and the combined pupils on the two sides of the main beamsplitter are reimaged onto separate focal planes.

B. Image Plane WFS: interference occurs in exactly the same way as above except sensing is done in the image plane.

C. Mixed Plane WFS: in this case the interference on the main beamsplitter takes place with the homodyne light in the image plane and the nulled light in the pupil plane. The combined beams are imaged onto the detector such that homodyne power is distributed uniformly (in a pupil sense) out to the control radius of the AO systems. However there is no strict restriction -- homodyne illumination can be reformatted to an outer radius of ones choosing. For example it is possible to consider a scenario where the homodyne power is redirected to a smaller zone within the AO control radius, say out to 10 λ/D. This can allow for improved

signal-to-noise performance for faint speckles within this angle at the expense of spatial frequencies beyond the illuminated zone (i.e. > 10 λ/D). Within the illuminated zone the SNR increases by κ_{AO}/κ where κ and κ_{AO} are the radii of the illuminated and AO controlled zones respectively.

It is easiest to compare performance by using a signal-to-noise metric for individual speckles. Phase residuals in the Lyot pupil can be cast in modal form with discrete Fourier amplitudes $\phi_{\vec{\kappa}}$ such that,

$$\phi(\vec{r}) = \sum_{\kappa} \phi_{\vec{\kappa}} \cos\left(\vec{\kappa} \cdot \left(\vec{r} - \vec{r}_{\vec{\kappa}}\right)\right)$$

where $\vec{\kappa}$ is the spatial frequency vector and $\vec{r}_{\vec{\kappa}}$ is an arbitrary spatial offset vector at that spatial frequency. Each mode results in a pair of speckles at locations $\pm\vec{\kappa}$ in the image plane. If we further specify that the arbitrary offset changes with time in the manner of frozen flow with wind speed $\vec{v}$ then $r_{\vec{\kappa}} = \vec{v}t$.

It this then easy to show that each speckle in the image plane will be modulated on a timescale of $\tau \cong \frac{1}{\vec{v} \cdot \vec{\kappa}}$. This timescale specifies the minimum speed that any backend WFS (in all of the above case) must operate at in-order to properly measure out to a spatial frequency of $\mathrm{mod}(\vec{\kappa})$. Note that image plane WFSs have no real speed advantage --- even though the intensity of most speckles only varies on timescales D/v, the phase of the speckle will contain structure on timescales of $\tau \cong \frac{1}{\vec{v} \cdot \vec{\kappa}}$ which must be sensed in order for a full correction to be applied.

Case A: Pupil Plane WFS

In the pupil plane WFS the scalar electric field at any location in the combined pupil, ignoring amplitude errors and considering just phase errors for simplicity,

is given by

$$E(\vec{r}) = e^{-j\psi} + j\phi(\vec{r}),$$

where ψ is an angle induced by the modulation of the homodyne signal. The corresponding intensity in the plane of the detector is $I(\vec{r}) \propto EE^*$.

We omit the expression for intensity explicitly, but it suffices to say that the signal $\phi_{\vec{r}}$ is directly proportional to the amplitude of the induced modulation. An amplitude error would similarly be measured as the sine component of the modulation. Once the "observable" and all noise sources are known, it is easy to write down the expression for the signal-to-noise. We choose the signal-to-noise ratio for each Fourier mode as this allows easy comparison between equivalent expressions for the other WFS cases.

Once again, omitting intermediate steps, it can be shown that for a 4-bin detection the signal-to-noise is for a Fourier mode of amplitude $\Phi_{\vec{\kappa}}$ is

$$SNR_A = \frac{\left(\frac{\sqrt{2}}{\pi}\right) N\Phi_i \sqrt{\Delta t}}{\sqrt{N + 4m^2 f\sigma^2 + B\Omega}},$$

where N is the target photon rate per second, f is the camera read rate or frame rate, σ is rms read noise per read, B is the background per unit solid angle and Ω is the solid angle corresponding to the instrument field of view (from here on we assume $B \to 0$ for H band operation), m is the number of sensors in 1D, and S is the Strehl of the system with $S \to 1$. For a photon limited detection the SNR simplifies to, $SNR_A = \left(\sqrt{2}/\pi\right)\Phi_i \sqrt{N\Delta t}$.

This holds when $N > 4m^2 f\sigma^2$ i.e. a stellar electron rate of

$$N > 4\times 10^7 e/s \left(\frac{m}{100}\right)^2 \left(\frac{f}{100Hz}\right)\left(\frac{\sigma}{3e^-}\right)^2$$

or an photon rate of 5x10^{-7} e/s for a quantum efficiency of 0.8.

The one obvious advantage of this scheme is that SNR_A is independent of the spatial frequency, making it useful perhaps for survey style observations.

Case B: Image Plane WFS

In this case the scalar field at any location κ_i in the image plane is given as

$$\hat{E}(\vec{r}) = \hat{1}e^{-j\psi} + j\hat{1} * \hat{\phi}\left(\vec{k}\right),$$

where $*$ denotes a convolution and the $\hat{1}$ denotes a Fourier conjugate. A Fourier mode in the pupil plane is then imaged into a pair of speckles with location $\vec{\kappa} = \pm\vec{\kappa}_i$, with each speckle forming a 2x2 pixel^2 footprint on the detector.

The signal-to-noise ratio for a top-hat shaped combined pupil can be shown to be

$$SNR_B = \frac{\left(\frac{\sqrt{2}}{\pi}\right) N\left(\pi D_\lambda \kappa\right)^{\frac{-3}{2}} \Phi_i \sqrt{\Delta t}}{\sqrt{N\left(\pi D_\lambda \kappa\right)^{-3} + \frac{N\Phi_i^2}{2} + 16f\sigma^2}}$$

with a photon limited SNR identical to the previous case

$$SNR_B = \left(\sqrt{2}/\pi\right)\Phi_i \sqrt{N\Delta t}.$$

This speckle becomes self-noise limited when $(\pi D_\lambda \kappa)^{-3} < \Phi_i{}^2 / 2$ or on angular scales of

$$\theta > 30 \frac{\lambda}{D} \left(\frac{\Phi}{0.01 rad} \right)^{-2/3}$$

This should in general not be problematic since the average amplitude of a residual atmospheric speckle should fall as κ^{-1}.

The sensor becomes read limited when $N(\pi D_\lambda \kappa)^{-3} \approx 16 f \sigma^2$ i.e. detector noise creates a angular cutoff (which is well within the AO control radius) of

$$\theta > 13 \frac{\lambda}{D} \left(\frac{N}{5x10^7} \right)^{\frac{1}{3}} \left(\frac{f}{100 Hz} \right)^{\frac{-1}{3}} \left(\frac{\sigma}{3e^-} \right)^{\frac{-2}{3}}$$

beyond which the SNR plummets rapidly with angular separation; we assume here a star with a photon rate identical to that in Case A. Obviously, the image plane sensor has a clear homodyne power (and SNR) advantage at small separations which is coupled with a limited field-of-view. For example at an angular separation of 4λ/D the image-plane sensor has an advantage of almost 4 magnitudes over the pupil-plane WFS.

Case C: Mixed Plane WFS

The motivation for this hybrid scheme is to distribute the homodyne gain across the image to overcome some of the field restrictions of case B. The scalar electric field at any location κ in the image plane can be written as

$$\hat{E}(\vec{r}) = \hat{1} e^{-j\psi} + j\hat{1} * \hat{\phi}(\vec{k}).$$

The SNR for a speckle of size Φ_i radians is

$$SNR_C = \frac{\left(\frac{\sqrt{2}}{\pi} \right) \left(\frac{N}{m} \right) \Phi_i \sqrt{\Delta t}}{\sqrt{\frac{N}{m^2} + \frac{N \Phi_i{}^2}{2} + 16 f \sigma^2}}$$

which can be reformulated as

$$SNR_C = \frac{\left(\frac{\sqrt{2}}{\pi} \right) N \Phi_i \sqrt{\Delta t}}{\sqrt{N + \frac{N m^2 \Phi_i{}^2}{2} + 16 f m^2 \sigma^2}}.$$

Although the signal-to-noise ratio is independent of field, it is no better or worse than performance in case A.

5.5 Conclusions

We have described the role of the backend AO system and analysed the performance of three related schemes for the wavefront sensor. The underlying design is that of a phase shifting interferometer with homodyne gain derived from rejected starlight. We choose to implement the pupil plane WFS for all simulation work in this report because its performance is neutral vis-à-vis spatial frequency. The image plane schemes have a clear signal-to-noise advantage but this comes at the cost of spatial frequency or angular coverage.

5.6 Optical design

The basic design of the post-DSS WFS is a second interferometer connecting the two outputs of the DSS nulling interferometer. Since it is also a white light interferometer, the two arms must be matched in path for all wavelengths which leads to another symmetric optical layout.

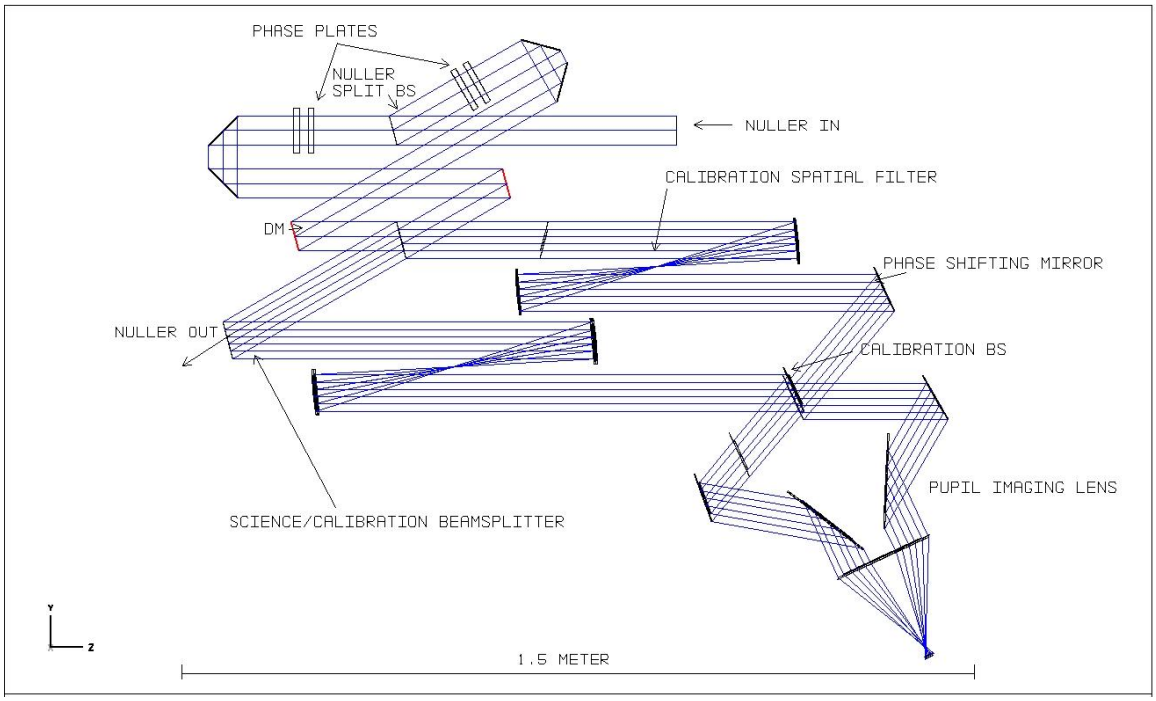

Figure 79: Optical layout of the combined nuller (showing only one stage) and back WFS. In this design, both images are put onto a single detector.

The bright output passes through a couple of off-axis parabolas and is spatially filtered at the intermediate focus. This beam then reflects off a phase shifting mirror that introduces the phase diversity to solve for the amplitude and phase of the nuller output wavefront.

The nuller dark output is sampled by a spectrally neutral beamsplitter that passes some fraction of light to the calibration interferometer. This beam then passes through a pair of OAP's identical to those in the other arm of the calibration interferometer to match path length and pupil polarization properties of the two beams (as much as possible).

These two beams recombine at the WFS recombination beamsplitter. Again, we are concerned about matching beam properties of both arms so this element must also be matched to the nuller beamsplitter. After this recombination beamsplitter, light from both the front and back sides are relayed to the calibration camera via a pupil relay lens. This lens forms a real image of the DM onto the infrared array.

To operate at the target rate of 1000 Hz, the system requires an advanced infrared detector optimized for high framerate and low speed. The PFI requirements for this detector match the general TMT AO requirements for IR tip/tilt and wavefront sensors. TMT has begun discussions with manufacturers and other observatories on development of these detectors.

One potential development effort is the Rockwell SPEEDSTER FPA, which is currently under development. This is a 256x256 near infrared ROIC with integrated ADCs allowing a purely digital interface. The unit cell is derived from the experience of the Calico "striped" multiplexer and will be able provide 256x256 at 1 kHz and 128x128 at 4 kHz respectively with low (5 e) read noise. If we use one detector per channel, we could achieve 128x128 pixels at 1 kHz with 2-3 electrons noise.

If we are limited to current technology, the WFS could be built around a HAWAII-1RG small pixel pitch array with a 10 electron noise performance, but significantly lower frame rates – in this case the sensor would only serve to measure quasi-static aberrations. In this case the DM, 128x128 in size is reimaged to a similar format on the focal plane. There is sufficient room on these arrays (quadrants) to accommodate these pupil images. Independent quadrants can be read out synchronously for preselected positions of the dither mirror. The image plane focus before the ultimate pupil plane image is a convenient location for an anti-aliasing spatial filter which also serves as a cold stop.

The dither mirror rides on an optical delay line that is zeroed at the start of operation and kept at zero against drifts in the DSS (nuller case) and downstream optics; relative optical path between the bright and faint arms is easily sensed in the interferometer. The dither strokes OPD through one full wavelength – during this period the modulated fringe pattern at each pixel in the detector is integrated and sampled in four bins. Demodulation of the fringe, and estimation of the wavefront error at each location in the pupil is a task of the real-time computer. If a sawtooth waveform is used for dithering, faster updates for the phase map may be obtained using a sliding average. This can allow for a small but perceptible increase in the servo bandwidth of the backend AO.

In operation, the WFS continuously acquires phase shifted images and from these reconstructs the amplitude and phase of the wavefront in the dark nuller output. The phase term is used to actively correct a DM in the DSS. Both the amplitude and phase are also recorded to be used to estimate the science camera image in post processing.

6 Integrated systems controls/operation

6.1 Introduction

The main subsystems of PFI are, a Front-end AO System as described in Section 3, a Diffraction Suppression System, from Section 4, and a Post-DSS or back-end AO, described in Section 5 and this section. Figure 80 shows how these pieces combine from the viewpoint of a control architecture.

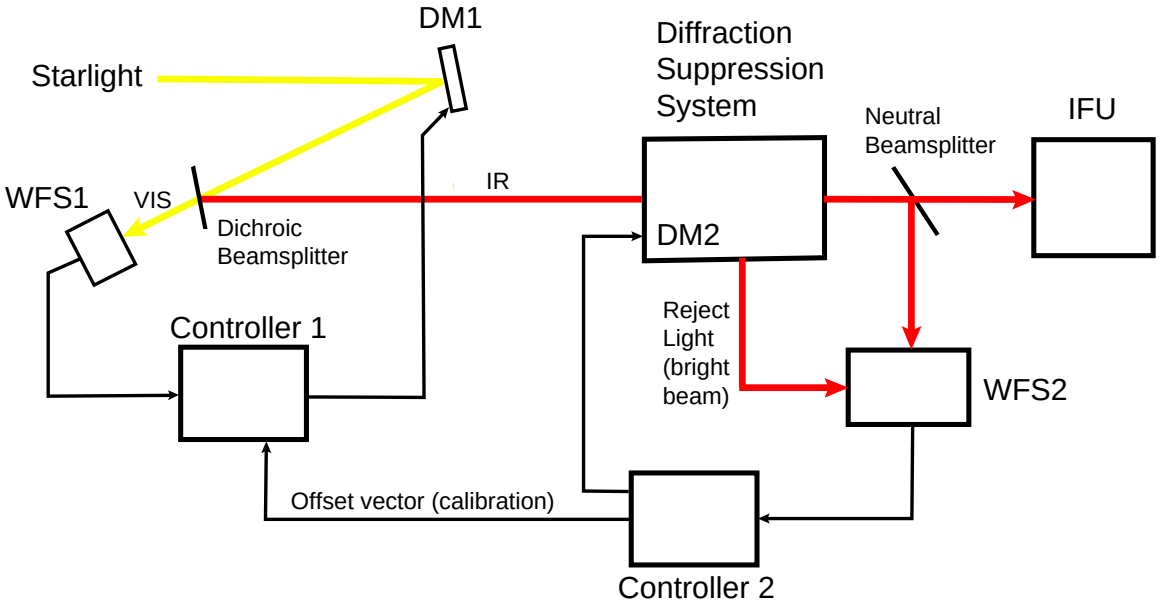

Figure 80: PFI Generic Control Architecture.

Some common features of the architecture are:

- The front-end AO is fed with wavelengths shorter than about a micron, while longer wavelengths are passed on to the Diffraction Suppression System, back-end AO and Integrated Field Spectrograph (IFS).
- The Diffraction Suppression System has two optical outputs – one to the IFS, called the 'null' beam in this section, and one used by the back-end AO, called the 'bright' beam.
- A portion of the null beam is picked off to drive the back-end AO. For this study, this is a neutral 50/50 beamsplitter.
- The two AO systems have independent controllers, and are designed and built by different institutions. The AO systems will operate independently, with a link from post- to pre-AO passing a wavefront error or calibration vector to form an offload loop to send time-averaged static wavefront errors to the front AO system and pointing and centering commands to the optomechanical system

- Preliminary analyses of the control properties of the pre- and post-DSS AO systems are presented below, along with a few words about integrated operation and operational modes.

6.2 Control systems

6.2.1 Front-end AO (Poyneer)

The front AO system carries out wavefront sensing and reconstruction as discussed in Section 3, most likely with a pyramid/ridge wavefront sensor. This control modality includes a Fourier/Hilbert transformation. This includes a Fourier/Hilbert transformation in which additional Fourier filters could be applied to the reconstructed signal. For example, the DM influence function could be corrected (for the assumption of uniform influence functions) by applying an inverse of the filter imposed by the influence functions. More broadly, this architecture allows us to run a version of the optimal Fourier modal reconstructor proposed for the Gemini Planet Imager. (Poyneer and Veran 2005)

This optimal Fourier controller is a Fourier-domain version of the modal controller used in instruments such as ALTAIR or ESO AO systems. As with any modal controller, the system has independent gains for each mode in its basis set. The Fourier modes provide a very natural basis set for a square wavefront sensor, especially in the ExAO context, where each Fourier component corresponds to a different part of the point spread function. The controller uses telemetry from the wavefront residuals to independently adjust the gain of each Fourier mode. The method can be called adaptive because it uses real-time information on the operating conditions (i.e., atmosphere, noise levels) to modify the control algorithm. The method is optimal because it has a formal metric for performance that is optimized, producing the best control algorithm given present conditions

The behavior of the control-loop and the optimization process is, at its essence, quite simple. For a single mode, the AO system is attempting to control a real signal (the modal coefficient through time) in the presence of noise. When the system uses an integrator-gain control law, increasing the system gain decreases the residual error due to compensating the signal. However, this high gain results in more error due to noise propagating through the system. A very low gain would produce very low levels of error due to noise, but substantial amounts of error due to the lag in compensating for the signal. Our formal metric is the total squared-error as seen in the residual phase for a single mode. This metric can be solved for in frequency space as the integral of the close-loop frequency response of the control system times the input temporal PSDs of the signal and noise modes. This metric is explicitly a function of the gain of the control law g. The optimal value of g is the one where the value of the integral and hence the total error is minimized.

shows the modified block diagram of the control system, with the gain optimization included. The gain optimization has two computationally intensive blocks. Given a time-series of modal coefficients, the temporal PSDs are estimated using FFTs and the averaged periodogram method. These PSDs are then sent to the gain optimizer, which finds the optimal g given the estimated PSDs

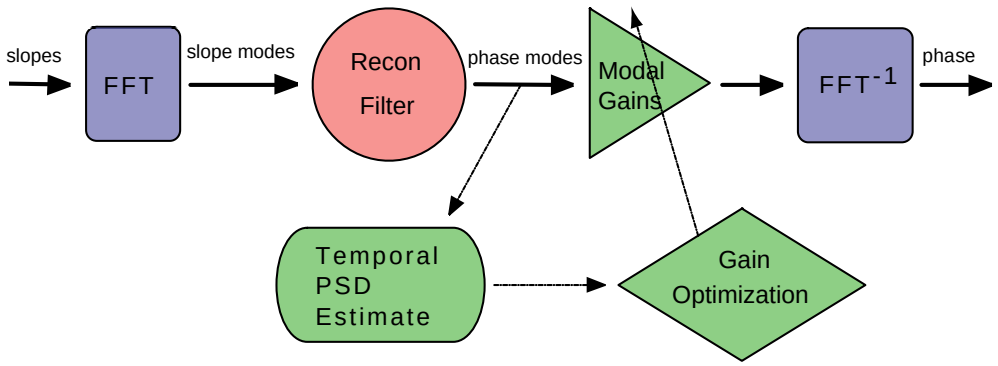

Figure 81: Block diagram of optimal gain FTR process.

This optimization is Strehl-maximizing, since it minimizes the residual error. This may not be exactly what we want for high-contrast imaging, as errors from the slowly varying atmospheric signal reduce contrast much more than noise errors. However, we have a solid understanding of what the optimizer is doing and the equations used in optimization can be weighted to favor one component over the other, to produce a contrast-optimizing controller. We are currently studying contrast-optimal controllers for GPI (Figure 82).

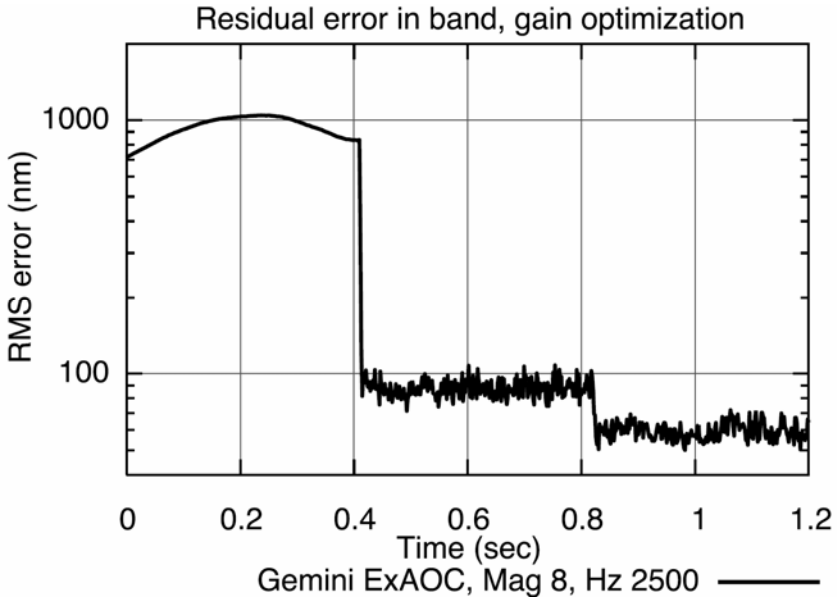

Figure 82: Residual wavefront error (within the controlled spatial frequency band, i.e. neglecting fitting error) for the Optimal Fourier Controller running on Gemini ExaOC. At time t=0.4 seconds the main control loop closes; at t=0.8 seconds the optimizer first applies the optimal modal gains.

The Fourier basis sets also provide a natural framework for implementing predictive controllers, an area in which LLNL intends to carry out significant research over the next 2-3 years. If predictive control is practical it should provide significant improvement in controlling low-frequency modes, in turn improving the PSF in the crucial 0.1-0.2 arcsecond range.

6.2.2 Back-end AO control

6.2.2.1 Architecture

A primary feature of the Post-DSS AO, or back-end AO, is tight integration with the DSS itself, both optically and mechanically. To be effective, the DSS and its associated AO system must be designed and analyzed as a unit.

Figure 83 shows a concept with a Lyot-coronagraph DSS, optically integrated with a pupil-pupil type of back-end wavefront sensor. This coronagraph-plus-WFS combination can either drive a second DM (not shown), or provide only calibration, or outer-loop, data to the front-end AO.

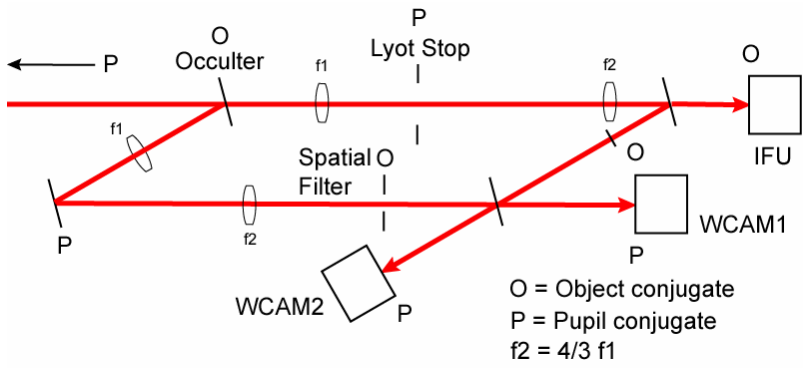

Figure 83. Schematic of Integrated Back-end WFS for a Lyot architecture.

The figure below shows a concept with a nuller DSS, optically integrated with a pupil-pupil type of back-end WFS. This scheme does have a second DM (DM2). This DM is inside one arm of the nuller, which gives rise to a somewhat different control analysis. This is the PFI baseline we have chosen to analyze.

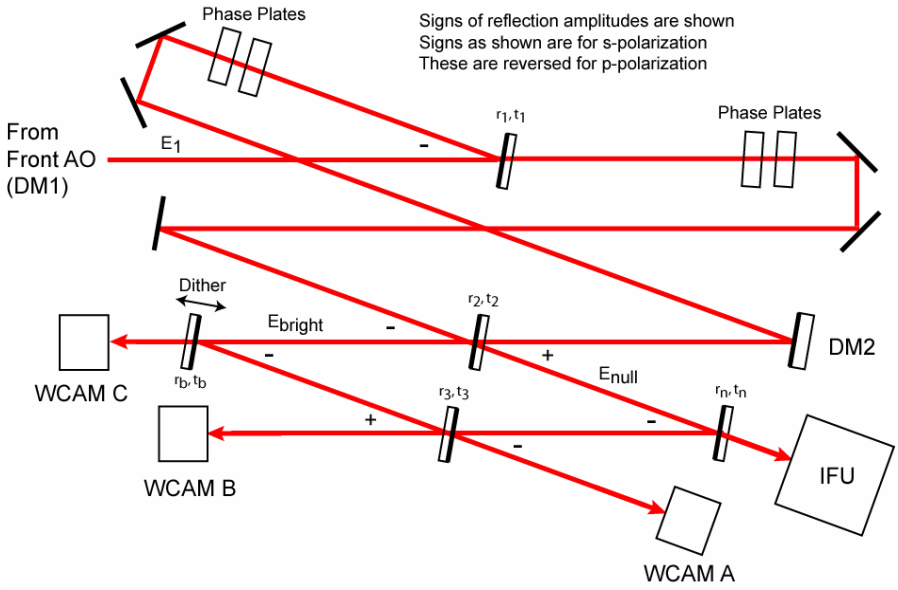

Figure 84: Two-arm nuller with integrated back WFS

To provide a bit of generality, two cameras, WCAM A and B, are shown for the interferometric beams in their pupil planes, while a third camera, WCAM C, looks at the 'bright' beam focal plane. In some implementations, cameras A and B might be a single Dewar. WCAM C is optional, but may be useful in practice for alignment and tip/tilt sensing.

A simplified model of the beamsplitters assumes that transmission amplitude coefficients are real and positive, and that reflection coefficients are real. The model corresponds to s-polarization Fresnel reflection from a material with constant index n and a perfect AR coat on the second surface.

6.2.2.2 Back-end AO Interaction Matrix

An interaction matrix for the back-end AO may be derived by assuming a flat input beam, a small phase perturbation on the DM, and a four-phase dither on the dither mirror, then calculating complex field amplitudes, field intensities, and demodulated signals in the WFS A, and WFS B beams.

Call :

E0	Undistorted field from star
E1	Corrected field from Front AO
Enull	Field out of null port of SSS
Ebright	Field out of bright port of SSS
EA, EB, EC	Field into each WCAM

Also, define the corrected field from the first AO as E_1 :

$$E_1 = E_0 \otimes (Atmosphere + Telescope + DM1)$$

$$E_1(\vec{x}) = A_1(\vec{x})e^{i\varphi(\vec{x})}$$

We will invoke an assumption of high strehl at a later time

Describe the effect of DM2 and the dither mirror by :

$$\Phi_2 = e^{i\varphi_2}$$

$$\Phi_d = e^{i\varphi_d}$$

The field amplitudes of the null and bright beams, and the A, B, C and IFS beams, are then:

$$E_{null} = E_1(\vec{x} - \frac{\vec{s}}{2})t_1t_2 - \Phi_2 E_1(\vec{x} + \frac{\vec{s}}{2})r_1r_2$$

$$E_{bright} = -E_1(\vec{x} - \frac{\vec{s}}{2})t_1r_2 - \Phi_2 E_1(\vec{x} + \frac{\vec{s}}{2})r_1t_2 \qquad (6.2.2.2\text{-}1)$$

$$
\begin{aligned}
E_A &= -E_{bright}\Phi_d r_b t_3 + E_{null} r_n r_3 \\
E_B &= -E_{bright}\Phi_d r_b r_3 - E_{null} r_n t_3 \\
E_C &= E_{bright} t_b \\
E_{IFU} &= E_{null} t_n
\end{aligned}
\qquad (6.2.2.2\text{-}2)
$$

The corresponding field intensities of the A, B, C and IFS beams are:

$$
\begin{aligned}
I_A &= \left| - E_{bright}\ \Phi_d\ r_b t_3 + E_{null}\ r_n r_3 \right|^2 \\
I_B &= \left| - E_{bright}\ \Phi_d\ r_b r_3 - E_{null}\ r_n t_3 \right|^2 \\
I_C &= \left| E_{bright}\ t_b \right|^2 = I_{bright}\ T_b \\
I_{IFU} &= \left| E_{null}\ t_n \right|^2 = I_{null}\ T_n
\end{aligned}
\qquad (6.2.2.2\text{-}3)
$$

For I_A and I_B, form signals with 4 dither phases, and label these signals with indices 0,1,2,3:

$$i = 0,1,2,3$$

$$\varphi_d{}^i = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$$

Now invoke the assumption of high strehl from the first AO :

$$
\begin{aligned}
\left|E_{bright}\right|_{\mathrm{Re}} &>> \left|E_{bright}\right|_{\mathrm{Im}} \\
\left|E_{bright}\right| &>> \left|E_{null}\right|
\end{aligned}
\qquad (6.2.2.2\text{-}4)
$$

The intensities at phases 1 and 3 from WCAM A are then :

$$
\begin{aligned}
I_{A1} &= I_{bright} R_b T_3 + 2\left|E_{bright}\right| r_b t_3 r_n r_3 \left|E_{null}\right|_{\mathrm{Im}} \\
I_{A3} &= I_{bright} R_b T_3 - 2\left|E_{bright}\right| r_b t_3 r_n r_3 \left|E_{null}\right|_{\mathrm{Im}}
\end{aligned}
\qquad (6.2.2.2\text{-}5)
$$

Similar expressions can be written for the intensities at phases 0 and 2 in terms of the real part of E_{null}. These would be useful in a system which corrects amplitude as well as phase errors.

Likewise, expressions can be written for the signals from WCAM B, which have interaction terms with the opposite sign from those for WCAM A. For the rest of the controls analysis, only WCAM A will be considered; the effect of WCAM B is to improve the signal-to-noise by $2^{1/2}$, at the cost of an increase in complexity.

Express the DM phase function as a sum of actuator influence functions,

$$\varphi_d = \frac{2\pi}{\lambda} \sum_j a_j u_j(\vec{x}), \qquad (6.2.2.2\text{-}6)$$

where

u_j = Normalized influence function, with peak height = 1.00, and

a_j = Displacement in microns.

Define a normalized influence matrix F_{jk} to describe the overlap geometry of the j^{th} DM actuator and the k^{th} WFS subaperture :

$$F_{jk} = \iint_{Subap_k} u_j(\vec{x}) d\vec{x} \qquad (6.2.2.2\text{-}7)$$

The following table may help give a picture of F_{jk} as calculated for the numerical modeling done in this study :

0.0247	0.0906	0.0351
0.0906	0.5435	0.1368
0.0351	0.1368	0.0500

(6.2.2.2-8)

The central number is the value of F_{jk} for all j=k, and the surrounding numbers are for nearest neighbors as viewed on the pupil. All more distant values are 0.

Assume a high-strehl, loop-locked case and neglect scintillation, to get:

$$\frac{\partial E_{null}}{\partial a_j} \cong -i \frac{2\pi}{\lambda} u_j |E_1|$$

$$\frac{\partial E_{bright}}{\partial a_j} \cong 0 \qquad (6.2.2.2\text{-}9)$$

$$E_{bright} \cong -E_1(t_1 r_2 + r_1 t_2)$$

Integrate over the k^{th} subaperture to get the directly measured WFS signal S^k, which is the number of detected photons per subaperture per integration time for one dither slice. This includes detector quantum efficiency η :

$$S^k = \eta \iint_{Subap_k} (I_A) d\vec{x} \qquad (6.2.2.2\text{-}10)$$

Similarly, define $I_1{}^k$ and $I_{bright}{}^k$ as photons per subaperture per integration time in the front-end AO and bright beams.

Let us take the point of view that updates to the DM are computed twice per dither cycle, and that f_s, the sample frequency for control purposes, is twice f_d, the dither frequency. Let us number the measured signals with index i, forming a time series. Thus :

$$S_i^k \cong \eta I_{bright}^k R_b T_3 + (-1)^i \frac{4\pi}{\lambda} \eta I_1^k (t_1 r_2 + r_1 t_2)(r_b r_3 r_n t_3) F_{jk} a_j \qquad (6.2.2.2\text{-}11)$$

This gives an expression for a sort of interaction matrix :

$$A_{jk} \equiv (-1)^i \frac{\partial S_i^k}{\partial a_j} \cong \frac{4\pi}{\lambda} \eta I_1^k (t_1 r_2 + r_1 t_2)(r_b r_3 r_n t_3) F_{jk}, \qquad (6.2.2.2\text{-}12)$$

where the sign alternation is pulled out of the interaction matrix definition and kept track of in the control law.

6.2.2.3 WFS camera dynamic range

In order to extract the interaction term corresponding to optical wavefront error, the most recent WFS signal and it predecessor in the time series must be subtracted. The sensor dynamic range needed to support the subtraction is set by the modulation depth M :

$$M = \frac{E_{null}}{E_{bright}} \frac{4 r_n r_3}{r_b t_3} = \sqrt{\frac{2 R_n R_3}{R_b T_3} \frac{I_{null}}{I_{bright}}} \qquad (6.2.2.3\text{-}1)$$

Notice that an intensity ratio of 10^7 between null and bright beams results in a WFS modulation depth of about $10^{3.5}$. This improvement in visibility is sometimes referred to as the "heterodyne gain" or "homodyne gain" of this WFS approach.

6.2.2.4 Control Law and Frequency Response

A significant feature of this AO configuration is that the interaction matrix is so close to diagonal that each WFS subaperture and its corresponding DM actuator can be treated as a simple feedback system independent from all others. That is, no reconstructor, matrix multiply or FFT step is needed in the control law. Note that this may not be true for DMs with broad influence functions, in which case a more sophisticated control law (probably a Fourier filter) will be applied.

Accordingly, we will drop the k index for subaperture and j index for actuator number. The following derivations are then valid for any fully-illuminated subaperture.

Let us call x_i the atmospheric disturbance, d_i the DM displacement, and y_i the WFS signal, at the i^{th} sample time, and express them all in the same units, say, microns. With this normalization, the WFS output, as it sees the atmosphere as modified by the DM, is :

$$y_i = x_i - d_i \qquad (6.2.2.4\text{-}1)$$

Let us approximate the effect of the integration time of the WFS by saying that its current readout corresponds to the atmosphere one sample time back. Then it is possible to write a normalized equation describing the back-end AO control law, and generate a universal frequency response plot :

$$d_i = d_{i-1} + k\left(y_{i-1} + y_{i-2}\right) \qquad (6.2.2.4\text{-}2)$$

Here k is a parameter that sets the frequency response of the loop, with respect to the sample frequency. Figure 85 shows the error correction of this loop for three values of k, as a function of frequency. The chart is plotted for an update rate of 500 Hz, which is a dither rate of 250 Hz. It can be interpreted as a universal chart with the right limit being half the update rate.

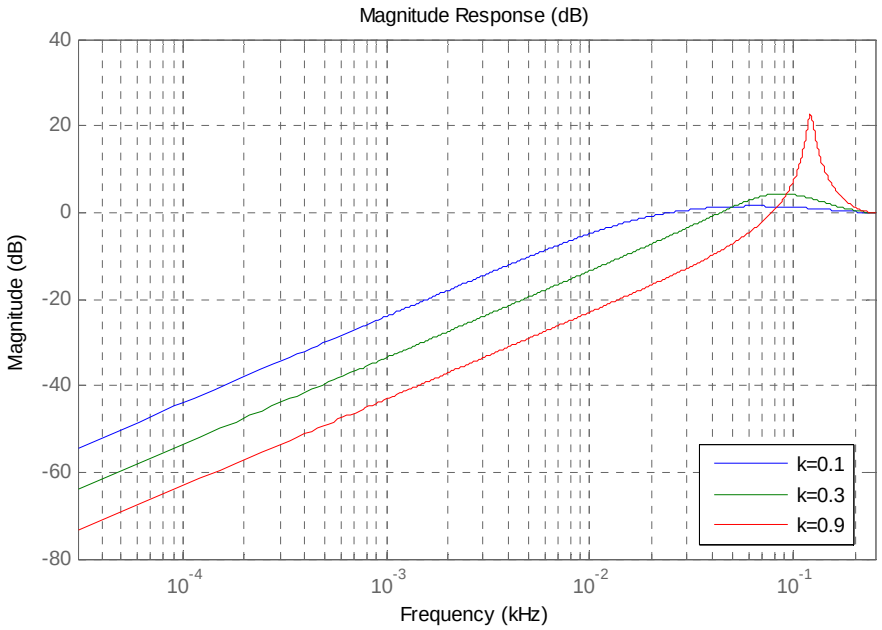

Figure 85 Error Rejection of Back-end AO (Fs = 500 Hz)

As k increase above about 0.3, loop stability becomes marginal, and if the normalization of the DM and WFS signals is accurate, the loop goes into oscillation at k=1.00. If the DM commands are expressed in microns, it remains to scale the WFS signals to match, using equation 6.2.2.2-11:

$$y_i = (-1)^i \frac{\lambda}{4\pi F_{jj}} \left(\frac{r_3 r_n}{r_b t_3} \right) (t_1 r_2 + r_1 t_2) \frac{S_i}{\langle S \rangle} \qquad (6.2.2.4\text{-}3)$$

Here $\langle S \rangle$ is the DC average WFS signal in the subaperture, and the sign alternation is synchronized with the dither. For the numeric simulations in this study, the r's and t's are all 0.707, and, per Table 6.2.2.2-8, Fjj is 0.5435. This leads to a useful approximation :

$$y_i \cong (-1)^i \frac{\lambda}{2\pi} \frac{S_i}{\langle S \rangle} \qquad (6.2.2.4\text{-}4)$$

6.2.2.5 *Noise Performance*

The noise performance of the back-end AO can be estimated from 6.2.2.4-4 and 6.2.2.2-11:

$$n_{DM} = \frac{\lambda}{2\pi} \frac{2f}{N} \sqrt{G(k)\left(N/2f + \sigma^2\right)} \qquad 6.2.2.4\text{-}5$$

Here, λ the wavelength, n_{DM} is the rms noise in a subaperture of the DM in the same units as λ, f is the update frequency (twice the dither frequency), σ is the read noise in electrons, N is electrons/subaperture/second on the WFS, and G(k) is a noise propagation factor. For the following figure, k=0.3 and G(k) = 0.6547.

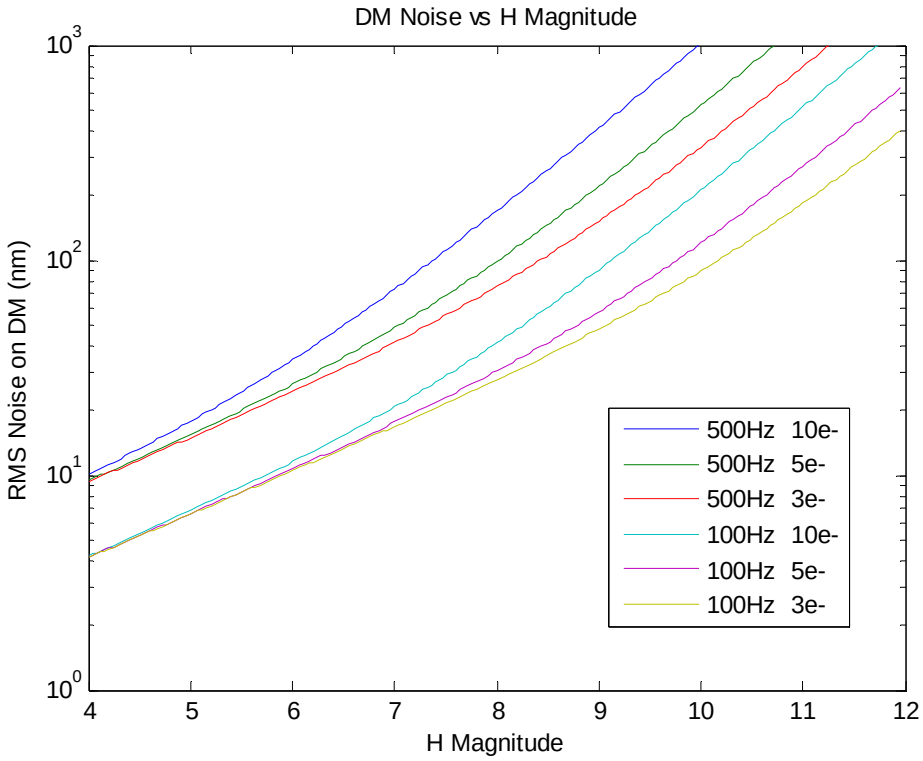

Figure 86 DM Noise (rms nm) vs H Magnitude

Figure 86 plots the rms noise in one subaperture of the DM due to WFS read noise and the photon noise in the bright beam, but does not include noise due to atmospheric effects or the front-end AO. The effect of the back-end AO on the wavefront it receives from the front-end AO, telescope and atmosphere is summarized in the error suppression curve of Figure 85.

It is clear from Figure 86 that getting the back-end WFS read noise down to the 3 to 5 electron level is key to PFI performance. This will push IR sensor technology.

6.2.2.6 Real-time Processor Requirements

Because of the one-to-one mapping of the WFS signals to the DM, the processing requirements of the back-end AO are modest. Assuming a 128x128 WFS and DM, and two-byte transfers per element, and 1000 updates/second, the IO bandwidth needed is 32MB/s each. Assuming about 1/3 of the 128x128 subapertures are controlled, the processing power needed to implement the control law 6.2.2.4-2 is about 15MFlops. Both the IO and the processor requirements are well within the capabilities of a single DSP, FPGA or even a PC.

6.2.2.7 What the WFS cameras see

It is instructive to estimate what the A and B cameras see as at the extremes of the dither cycle.

Combining expressions from 6.2.2.2-1 and 6.2.2.2-2:

$$E_A = -\left(-E_1(\vec{x}-\frac{\vec{s}}{2})t_1 r_2 - \Phi_2 E_1(\vec{x}+\frac{\vec{s}}{2})r_1 t_2 \right)\Phi_d r_b t_3 + \left(E_1(\vec{x}-\frac{\vec{s}}{2})t_1 t_2 - \Phi_2 E_1(\vec{x}+\frac{\vec{s}}{2})r_1 r_2 \right) r_n r_3$$

$$E_A = E_1(\vec{x}-\frac{\vec{s}}{2})\left(t_1 r_2 \Phi_d r_b t_3 + t_1 t_2 r_n r_3\right) + \Phi_2 E_1(\vec{x}+\frac{\vec{s}}{2})\left(r_1 t_2 \Phi_d r_b t_3 - r_1 r_2 r_n r_3\right)$$

For dither phase 0, the left term is large. The right term is small and can be made quite small if the bright and null beamsplitters are identical.

To first order :

$$\begin{aligned} E_{A0} &= E_1(\vec{x}-\frac{\vec{s}}{2}) \\ E_{A2} &= \Phi_2 E_1(\vec{x}+\frac{\vec{s}}{2}) \end{aligned} \qquad (6.2.2.7\text{-}2)$$

So, the WFS A and B cameras alternately see the left pupil, then the right pupil, at the extremes of the dither cycle.

6.2.2.8 Spatial frequencies in the null and bright beams

It is also instructive to estimate what spatial frequencies of the input beam end up in the null and bright beams.

Very roughly :

$$\begin{aligned} E_{null} &= E_1(\vec{x}-\frac{\vec{s}}{2}) - E_1(\vec{x}+\frac{\vec{s}}{2}) \\ E_{bright} &= -E_1(\vec{x}-\frac{\vec{s}}{2}) - E_1(\vec{x}+\frac{\vec{s}}{2}) \end{aligned} \qquad (6.2.2.8\text{-}1)$$

Take the spatial Fourier transform :

$$\begin{aligned} \widetilde{E}_{null} &= \widetilde{E}_1(\vec{k})(e^{i\frac{k_x s}{2}} - e^{-i\frac{k_x s}{2}}) = \widetilde{E}_1(\vec{k})2\sin\left(\frac{k_x s}{2}\right) \\ E_{bright} &= \widetilde{E}_1(\vec{k})(e^{i\frac{k_x s}{2}} + e^{-i\frac{k_x s}{2}}) = \widetilde{E}_1(\vec{k})2\cos\left(\frac{k_x s}{2}\right) \end{aligned} \qquad (6.2.2.8\text{-}2)$$

So, in the x-dimension, low spatial frequencies end up in the bright beam, and high spatial frequencies end up in the null beam. For the y-dimension, there is no filtering action.

6.2.3 Integrated control

An integrated control scheme can help the front-end AO and the back-end AO to cooperate, and give a healthy combined benefit in suppressing the incoming wavefront error and creating a deep null field.

The two items considered for integrated control are frame rate synchronization and an offload loop from the back-end to the front-end AO. Not considered for PFI as a first-light instrument is the possibility of a combined controller operating both AO loops. The systems are quite different – pyramid WFS vs interferometric WFS, visible vs IR, different frame rates, and not the least, integration and test by different institutions.

6.2.3.1 Frame rate synchronization

It is prudent to arrange a WFS frame rate synchronization scheme between the two AO systems. Operating the front-end and back-end camera frame rates in n:1 or, less preferably, n:m frequency synchronization will reduce the likelihood of periodic systematic error ("herringbone" effects) and considerably simplify joint telemetry and data display. Frame rate synchronization should also be considered a prerequisite for nested loop operation between the front-end and back-end AO systems. It is easy enough to implement and heads off many potential headaches.

6.2.3.2 Bright stars: Nested Loop / Calibration Vector

A likely scenario for joint operation of the two AO systems on bright stars is that the back-end AO operates more slowly than the front-end AO, at an integer submultiple of the front-end frame rate. The back-end AO operates at the infrared science wavelength, and sees the entire optical train of the telescope plus PFI instrument. In this scenario the back-end AO is in a good position to report to the front-end AO on systematic error and persistent speckles. The data passed would be the most recent post-AO DM2 command vector, or some version of it in zernike space. The front-end AO would treat this as a calibration input, or centroid offset vector.

A realistic outer loop would have the usual servo gain, loop compensation, and loop on/off controls. The real-time portion of this could reside in either AO system, while the operator interface would presumably be in some joint GUI for the PFI instrument as a whole.

As the back-end AO as currently envisioned does not provide information over the whole pupil, some splicing or extrapolation will have be provided. Some techniques for doing this have been proposed, and will be tested on the next available Palomar AO engineering run.

For the most part, the error rejection vs frequency curve of the combined system is simply the product of the curves of the individual AO systems. Hence, if the front-end and back-end AO systems are both type I (error rejection inversely proportional to the first power of the frequency), the combined system is type II. This is true even if no link is provided between the systems. The main effect of adding the calibration loop is the improvement in static and slow errors.

Full optimization of these control loops is a subject for study during the Conceptual Design phase. The exact modes used are highly dependent on detector technology. If the back WFS's IR detectors are significantly noisier than visible detectors (10-20 electrons) it will have to operate at significantly slower frame rates and provide primarily static correction with a low gain; if low-noise WFS detectors are available (Section 5.4) the two sensors are more balanced.

6.2.3.3 Dim star modes

Many of PFI's core science targets, the young T Tauri stars, are extremely red. For these stars many more photons are available to the back IR WFS than the front visible-light WFS. On these targets, the back WFS would actually operate faster than the front WFS, though most likely still in integer multiples. The front WFS would provide a low-order correction (taking advantage of the binning ability of the pyramid or interferometer sensors) to produce a high enough input Strehl ratio for the back-end sensor to operate.

6.3 Summary of Operational Modes

From an AO control standpoint, the operational modes of the PFI instrument are :

- Bright star: front-end AO system running at 2-4 kHz frame rate and correcting atmosphere; back end AO system running at 0.5-1 kHz and providing improved correction of slow atmospheric modes and quasi-static errors
- Bright to medium star: front end AO running at 2 kHz and back-end system running at low frame rate or gain and correcting only static errors
- IR-bright star: front end system running at 100-200 Hz to get enough light into the back end system's pinhole and back-end system running at 0.5-1 kHz to provide most atmospheric and static correction

Additional startup or test modes:

- No AO on, seeing sky or front-end local source
- Front-end AO only, seeing sky or front-end local source
- Back-end AO only, seeing sky, front-end local source or back-end local source

Consideration should be given in the software design to allow independent front-end AO control and back-end AO control, or single PFI operator operation from a joint GUI.

7 Science instrument

7.1 Introduction

The primary purposes of the PFI science instrument are to detect planetary companions by distinguishing them from PSF speckle noise, to record low- and medium resolution 1-5 μm spectra of these planets, and to detect and measure circumstellar disks, particularly through polarization. Two speckle-noise suppression techniques are used depending on the science applications. The first one consists of acquiring multi-wavelength image cubes s*imultaneously* and taking advantage of the deterministic behavior of the speckle pattern with wavelength for discriminating speckles from a true companion signal, i.e. the position of interference speckles in the focal plane scales with linearly with wavelength, and hence broad-band speckles appear elongated compared to a point source companion. The second technique is to take advantage of the polarized nature of the companion/disk to distinguish it from the unpolarized speckle pattern of the central star.

Experience with the three-channel camera TRIDENT on CFHT (Marois et al 2005) has shown that the speckle-suppression performance depends critically on the detailed implementation of the instrument. In particular we must minimize the presence of non-common path aberrations between different wavelength channels (NCA) in the optical system. In the regime where performances are speckle-noise limited, a useful figure of merit for the speckle-suppression performance is simply the speckle-noise attenuation factor $\tau \equiv N/\Delta N$ where N is the input PSF noise on the science instrument and ΔN is the residual noise left after data processing. One is striving to make τ as high as possible to subtract speckles down to the photon noise limit without affecting the signal (spectrum) of the companion.

The science instrument is required not only to achieve the highest speckle-noise attenuation but also to yield the best possible throughput, the largest FOV and spectral resolution. Some compromises are inevitable to meet all these criteria. As argued below, the most attractive solution a TIGER-type integral field spectrograph (IFS).

7.1.1 Instrument Specifications

The science and interface requirements of the science instrument are summarized below.

Science instrument key requirements:

- Field of view 2" x 2"
- Spatial sampling 0.005" (Nyquist sampled at 1.5 μm)
- Low-resolution (R ~ 50-100) spectroscopic or narrowband capability for planet classification and detection
- Goal: Medium resolution (R~500) capability for follow-up studies on specific spectroscopic features
- Spectral coverage from 1-2.5 μm with a single *JHK* band per exposure.
- Goal: extend to 5 μm
- Simultaneous imaging at multiple wavelength for speckle noise suppression with τ >10 within the dark hole
- Simultaneous dual-channel polarimetry

- Rotating cold pupil to act as cold stop and Lyot stop
- Throughput (including detector) >35%
- Pupil-viewing mode for alignment and test purposes

Science instrument key interfaces:

- Accepts a converging f/15 beam.
- Must rotate along the input f/15 optical axis.

Science instrument key design features

- Lenslet-based Integral Field Spectrograph with prism/grism dispersers
- Reflective fore-optics before lenslets and either refractive or reflective optics after
- Hawaii 4-RG-10 4096x4096 pixel HgCdTe infrared array (10 µm pixel pitch)

7.2 IFS Design

7.2.1 IFS Trade-off

The requirement to have a simultaneous MWI capability is satisfied by the following three IFS options: an image slicer, a fiber-fed (coupled with lenslets) IFS and a lenslet-based (TIGER-type) IFS. The three concepts are sketched below.

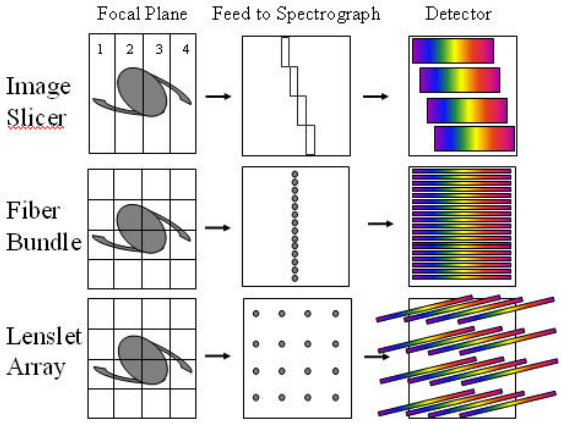

Figure 87: Possible IFS concepts. Diagram from James Larkin.

The slicer yields a very effective packaging of the spectra on the detector but it is not well immune against NCAs since each PSF slice has to go through a different optical path of the spectrograph. NCAs are essentially eliminated with the two lenslet-based IFS since the PSF image is dissected in two dimensions at the lenslet array. The fiber-fed IFS offers good flexibility for spectral packaging but requires the manipulation of more than 150000 cryogenic fibers which does not make it very practical to implement. The TIGER-type IFS offers good immunity to NCAs and simple implementation even though the spectral packaging is not as efficient as the other two solutions. We have selected the lenslet-based TIGER approach for PFI's instrument. This same approach will be implemented in the

integral field spectrograph for the Gemini Planet Imager; extensive participation by our team will provide considerable experience in high-contrast-optimized IFSs.

7.2.2 Baseline IFS concept

A schematic diagram illustrated the baseline science instrument IFS is shown in Figure 88. The instrument parameters are derived in the next section.

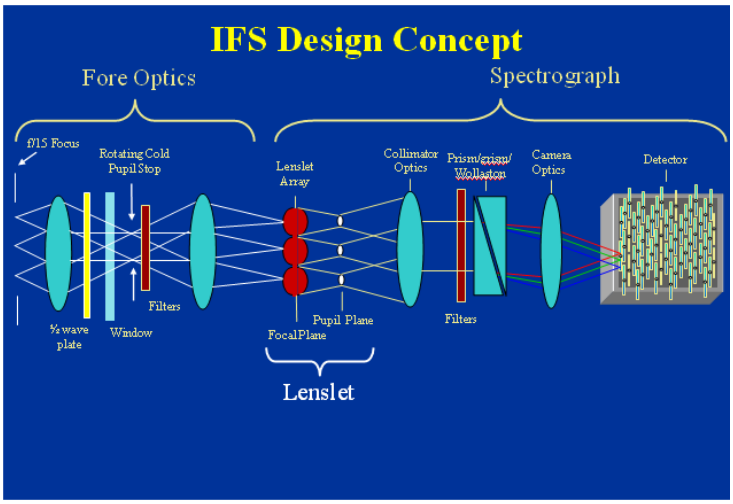

Figure 88: IFS design concept (based on figure from James Larkin).

The input f/15 beam is first collimated by a warm off-axis parabola and enters the cryostat through a CaF_2 cryostat window. A rotary table holding a ½-wave plate on a sliding mechanism is mounted on the warm side of the cryostat window. This modulator is used for the polarimetry mode. The OAP forms a 30-mm pupil where a pupil mask wheel is located. The pupil wheel holds 8 (TBD) masks that can be individually rotated along the optical axis to follow pupil rotation. An 8-position filter wheel follows immediately the pupil wheel. These filters are used for the follow-up medium resolution observations. The remaining of the fore-optics produce a f/126 whose focus is located on the micro-lens array. The micro-lens array plane has two mechanisms: 1) a linear stage to select the micro-lens array, one for short wavelength operation (1–2.5 µm) and another for the 3–5 µm spectral range and 2) a slit mounted a linear stage used for calibration. The PSF image is segmented at the lenslet and light from each micro-pupil is collimated and travels through a filter wheel holding the main broadband photometric filters (*JHKL'M'*). The 120 mm collimated beam goes through the grism wheel holding all "direct vision" dispersing elements and Wollaston prisms and finally through a f/3 camera onto a 4096x4096 Hawaii-4-RG-10 1-5 um detector.

The instrument has two wavelength ranges (SW: 1.1-2.4 µm and LW: 3-5 µm), both sharing the same fore- and spectrograph optics but with different lenslet arrays. In the LW, the detector is effectively binned 2x2. The next section discusses how the main instrument parameters are defined. The reader can skip this section without loss of clarity and can go directly to the last sub-section where all parameters are summarized.

7.2.3 Instrument Parameters

Figure 89 shows the basic parameters defining the TIGER-type IFS. The relevant parameters are:

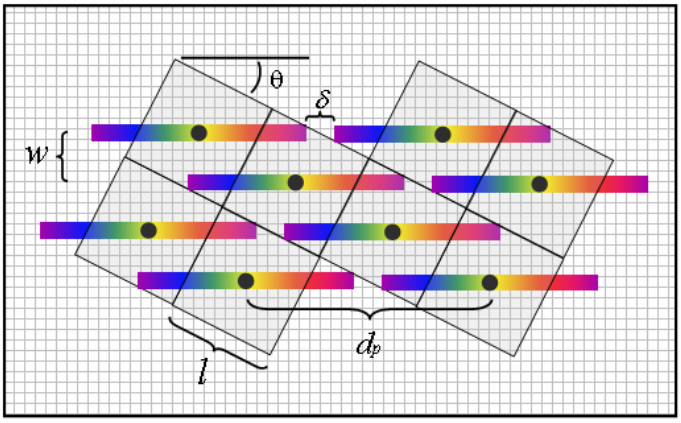

Figure 89: Schematic view of a Tiger-Type IFU. Each square represents one micro-lens as projected on the detector. The grouping pattern illustrated here is 2x2 (*p=2*). Based on figure from James Larkin.

D: telescope diameter in meter.
m: detector size (*mxm*) in pixels
q: detector pixel pitch in µm
FOV: square size of field of view is arcsec
s: image sampling = number of micro-lens per λ/D at wavelength λ_{min} (in µm)
λ_{max}: maximum wavelength
F_{coll}: collimator focal ratio. Same as the lenslet f/#
F_{cam}: focal ratio of detector camera
F_{in}: input focal ratio at the lenslet array
R: spectral resolution (λ/Δλ). Assumes 2 detector pixels per spectral resolution element (70; 500)
B: fraction of the spectral bandpass covered simultaneously (e.g. *B*~20% for the *H* band)
l_p: lenslet pitch in µm
l: lenslet pitch in µm at the detector
w: spectral spacing in pixels at the detector
d_p: separation in pixels between two adjacent micro-pupil along the dispersing axis
d_l: separation in pixels between two adjacent spectra along the dispersion axis
p: lenslet group pattern (integer > 2).
θ: lenslet array angle wrt detector
δ: number or "guard" pixels between adjacent spectra along the dispersion axis
The Tiger-Type IFS satisfy the following equations

$$d_p = w(p^2 + 1) \quad (7.2.3\text{-}1a)$$

$$d_l = d_p - \delta \approx 2 \cdot R \cdot B \quad (7.2.3\text{-}1b)$$

$$\theta = arctg(1/p) \quad (7.2.3\text{-}1c)$$

with the FOV (square size) given by

$$FOV = \frac{0.206 m \lambda_{\min}}{sD\sqrt{2BRw}} \quad (7.2.3\text{-}2)$$

Eqn (2) immediately sets the requirement for the detector size. Assuming the most aggressive spectral packaging (*w=2*), R~70 spectra over a 20% bandpass with a Nyquist (s=2) image sampling at 1.5 µm, ~2900 pixels are required. In practice, a more conservative packaging is required (*w~3*) in order to minimize the cross-contamination between the spectra which could affect the speckle-suppression performance. In this case, the FOV requirement (>2") is met with a 4kx4k detector.

7.2.3.1 Camera focal ratio (F_{cam})

The speckle-suppression performance of the IFS depends to some extent on the level of cross contamination between individual micro-pupil spectra. This contamination can me minimized by making the micro-pupil diffraction spot as small as possible i.e. by making the focal ratio of the camera as fast as possible. The size of the diffraction-limited micro-pupil on the detector (~λf/#) should therefore be less than ~1 pixel at the longest wavelength λ_{max}. Assuming λ_{max}=2.4 µm, a detector pitch of 10 µm and a √2 factor for the square geometry of the lenslet, this yields a camera f/# F_{cam}~3.0. The same camera and detector can be used for LW operation with λ_{max}=4.8 µm and an effective detector pitch of 20 µm after binning the detector 2x2.

7.2.3.2 Collimator Focal Ratio (F_{coll})

Given a fixed camera focal ratio, the collimator f/# is simply related to the optical magnification of the spectrograph given by the ratio of lenslet pitch at the detector (l) to its physical dimension (l_p). The magnification depends on the spectral length on the detector which in turns depends on the spectral resolution via Eqn. 1a and 1b. Given the geometry of the spectra on the detector (see Figure 89), the collimator focal ratio is given by

$$F_{coll} = F_{cam} \frac{l_p}{qw\sqrt{p^2 + 1}} \quad (7.2.3.2\text{-}1)$$

Given a fixed spectrograph optics and since the LW detector pitch is 2 times that of SW, either the lenslet pitch l_p and/or the grouping pattern p must change to accommodate both SW and LW operation. In practice, the lenslet pitch is also constrained by the fore-optics f/#. With *q=10* and *w=p=3*, it is convenient to chose a SW lenslet pitch l_p=95 µm to have a unit magnification which could simplify the spectrograph optics.

7.2.3.3 Fore-Optics Focal Ratio (F_{in})

The MWI speckle-suppression data processing involves shifting and scaling PSFs which requires the images to be properly sampled by at least 2 samples per λ/D at the shortest wavelength, here a sample being one micro-lens. With the parameters defined above, the fore-optics focal ratio feeding the micro-lens array is given by: $F_{IN} = sl_p / \lambda_{\min}$. For a fixed fore-optics, this condition requires the LW lenslet pitch to be approximately a factor two larger than the SW one (λ_{min}=1.5 for SW and (λ_{min}=3 for LW). Given l_p=95 μm and s=2 at 1.5 μm yields F_{in}=*126*.

7.2.3.4 Spectral Resolution vs FOV trade-off

Figure 90 (see also Eq. 7.2.3.2-1) below gives the allowed FOV size for a given spectral resolution. The values are quantified through the grouping pattern parameter *p*. Given a spectral spacing *w=3* and a detector pitch of 10μm, the maximum resolving power that can be achieved with a FOV greater than 2" is *R*~70.

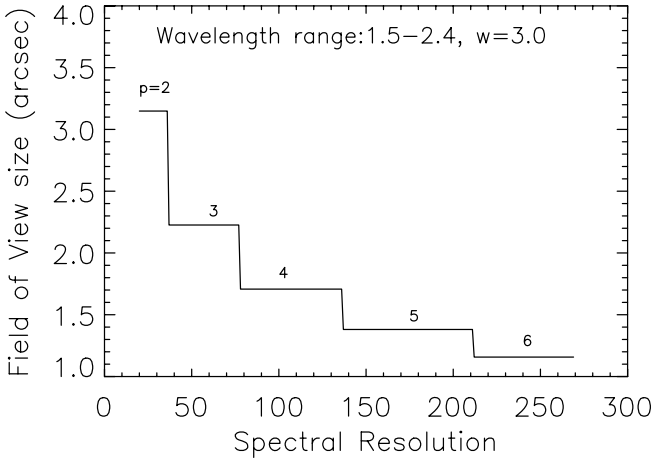

Figure 90: Field of view vs spectral resolution, for different lenslet grouping parameters *p*

7.2.3.5 IFS Specifications

The main IFS specifications are summarized in Table 11

Table 11: Baseline Instrument Parameters.

Detector	4096x4096 Hawaii4-RG	
Detector pitch	10 µm	
Fore optics focal ratio	f/126	
Collimator focal ratio[3]	f/3	
Camera focal ratio[3]	f/3	
	Short wavelength	Long wavelength
Wavelength range	1.1-2.4 µm	3.0-4.8 µm
Spectral Resolution[4]	70; 500	70;500
Field of view	2.2"x2.2"	2.2"x2.2"
Spectral band pass	0.18	0.18
Polarimetric spectral band pass	0.1	0.1
Lenslet sampling (mas/µ-lens):	5.15[1]	10.3[2]
Lenslet pitch (µm)	95	190
Lenslet group pattern (*p*)	3	3

[1] Nyquist (2 lenslet per λ/D) at λ_{min} =1.5 µm
[2] Nyquist (2 lenslet per λ/D) at λ_{min} =3.0 µm
[3] f/# at corner of the micro-lens
[4] The R~500 mode is for restricted spectroscopic features (see section 2.2.5).

7.3 Optical design

This section describes a preliminary optical design for the fore-optics and the IFS. While the latter has been optimized between 1.1 and 2.4 µm, we do not anticipate major show-stoppers in extending the wavelength range to 5 um. Early in this study, we studied a design featuring two separate optical channels, one for the SW and the other for LW. We have produced optical designs for both channels but only the SW is presented here.

7.3.1 Refractive vs Reflective

The fore-optics was chosen to be reflective to minimize non-common aberrations associated with refractive optics. Since this is less an issue past the lenslet array, a refractive design was chosen for its simplicity and relatively low cost. This does not exclude a reflective design for the IFS which could have several advantages given the very broad wavelength involved (1-5 µm). This trade-off study is deferred for the conceptual design phase.

7.3.2 Optical layout

7.3.2.1 Spectrograph

The optical layout of the spectrograph is shown in Figure 91. Light from the lenslet array is collimated by a F/3.1 quadruplet collimator (ZnS, 2 S-FTM16, BaF_2) which creates a 110 mm "false pupil" where the filter wheel is located followed by the grism wheel. The F/2.7camera has four lenses (ZnS, 2 BaF_2, ZnS). All lenses have spherical surfaces. The design is achromatic between 1-2.4 µm. The total optical path from the lenslet to the detector is 1.3m. Two fold mirrors are used to make the packaging more compact.

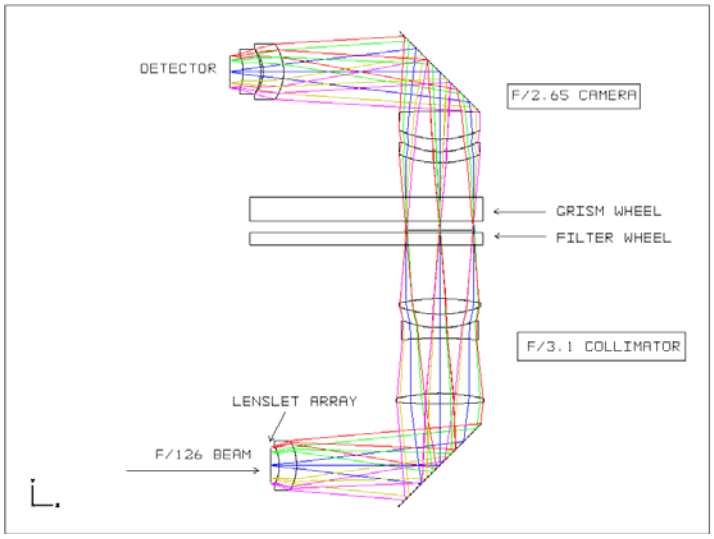

Figure 91: Optical layout of the spectrograph optics.

7.3.2.2 Fore-optics

The optical layout of the fore-optics is shown in Figure 92. The science Dewar was designed to receive a collimated beam with a pupil immediately inside the Dewar. This interface and the optical design was selected early before the decision to implement a NFIRAOS-like f/15 beam as the AO/nuller output, so an additional warm OAP optic is used to collimate the f/15 beam. Once the PFI and NFIRAOS interfaces are formally defined we will redesign to remove this warm optic. The collimated beam enters the cryostat through a CaF2 window to form a cold 30-mm pupil where the pupil mask and another filter wheel is located. The beam is reformatted into f/126 converging beam via two spherical mirrors (M1 and M2). The fore-optics is folded so as to minimize volume with the IFS optics but also to make the incoming f/15 beam roughly aligned with the mechanical axis of the cryostat.

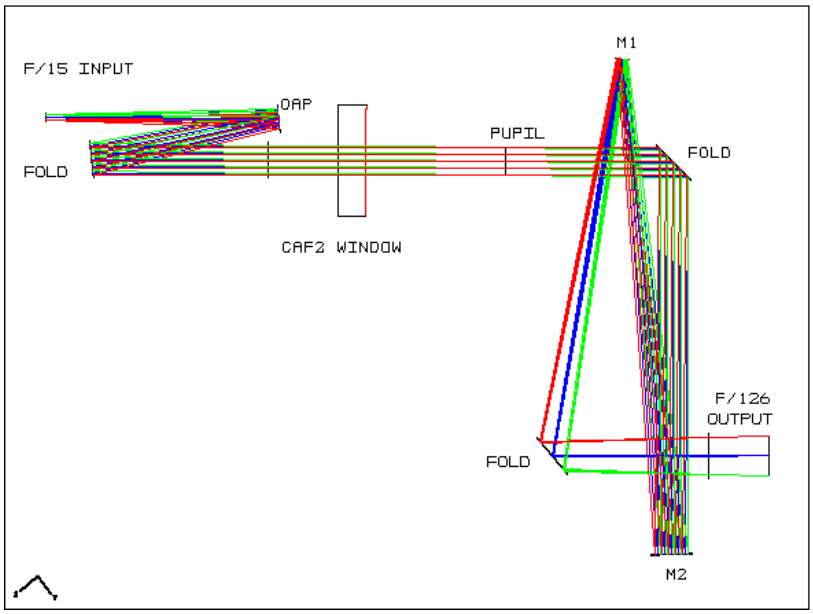

Figure 92: Fore-optics layout.

7.3.3 Image quality

The IFS optics is required to be nearly diffraction-limited to limit the cross-contamination between all the spectra and this criteria is met with the current design (see Figure 93).

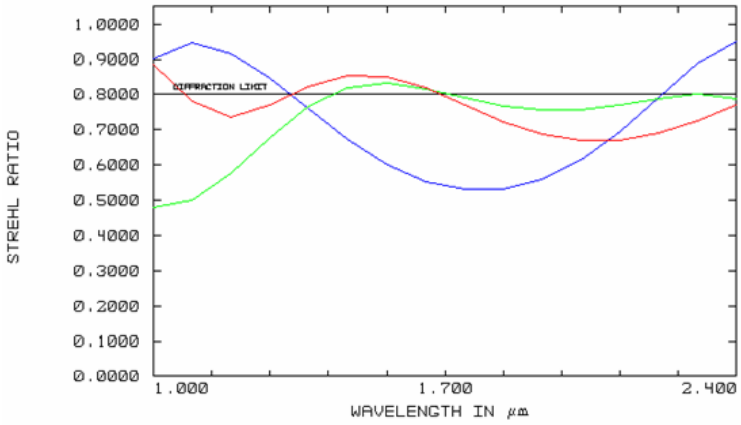

Figure 93: Image quality of the IFS optics as a function of wavelength. The three (colored) curves are for three different field positions.

7.3.4 Filters

The science instrument has two filter wheels: a six-position filter wheel (130 mm each) holding the 5 *JHKL′M′* broad-band filters + an open position and an 8-position holding 40-mm filters used for follow-up R~500 observations around specific spectroscopic features like [FeII]1.644, Brγ etc (see Table 12 below).

Table 12: Science Instrument Filter List.

Filter name	Bandpass (μm)
Filter wheel #1 (Broad band filters), ϕ130mm	
J	1.14-1.36
H	1.49-1.82
K	1.98-2.42
L´	3.40-4.10
M	4.5-4.8
Open	-
Filter wheel #2 (Narrow-band filters), ϕ40mm	
[FeII]1.644	1.62-1.67
HeI2.06	2.04-2.10
H_2 1-0 S(1)	2.09-2.15
Brδ (2.166 μm)	2.14-2.20
2-0 CO (2.29 μm)	2.26-2.32
3.3 μm PAH	3.2-3.4
1-0 CO (4.6 μm)	4.5-4.8
Open	-

7.3.5 Dispersing & polarimetric components

The grism wheel has six positions to hold two prisms, two grisms , one Wollaston prism for the polarimetry mode and an open position. The low-resolution spectroscopic mode (R~70) is achieved with a pair of air-space prisms, one per wavelength range. A very good prism pair candidate is BaF2-$SrTiO_3$ since it offers a good transmission between 1 and 5 μm and a resolving power varying smoothly with wavelength. Another solution restricted to the SW range is BaF2-Fused silica. The most likely solution for the R=500 mode is a direct-ruled grism on CaF2 and/or KRS-5. Resin-replicated grisms, although less expensive, may not be practical because of the large size involved (~150 mm) and their potential risk of delamination of the grating replica due to differential CTE between the prism substrate and the resin.

An air-space MgF2 Wollaston prism is the baseline component for the polarimetry mode. Coupled with a broadband filter, the Wollaston dispersion axis is aligned along the same as the prism/grism, producing two images of every micro-pupil in two orthogonal polarization states. This mode is used in combination with a warm half-wave plate modulator mounted on a rotary table just outside the cryostat window.

7.3.6 Throughput Budget

A throughput budget including the detector is given in Table 13. A reflectivity of 98.5% is assumed for all (gold coated) mirror surfaces and an average reflectivity loss of 2% per lens surfaces for the anti-reflection coating between 1 and 5 μm.

Table 13: Science Instrument Throughput Budget.

Component	Number of surfaces	Throughput (%)
Fore-optics		
Relay	7	91.3
Cryostat window	2	96.04
Spectrograph		
Lenslet	2	96.04
collimator	8	85.08
camera	8	85.08
Mirror folds	2	97.02
Filter	2	90.0
Prism/Grism	4/2	94.1/70
Detector	-	85.0
	Total	**44/33**

7.4 Mechanical design

7.4.1 Cryostat Design

Except for the collimator of the fore-optics, the optics is operated at 40 K inside a hexagonal cryostat approximately 1m high and 1m in diameter. Figure 94 shows a 3-D mechanical sketch of the optomechanical bench. The cryostat has two closed cycle coolers (CCR) mounted on opposite side of the cryostat. A turbo-pump is mounted permanently on the bottom cover of the cryostat.

7.4.2 Mechanisms

The science instrument features 7 mechanisms with the following functions:

- A half-wave plate modulator for the polarimetry mode. A warm rotary table mounted on a translation stage near the cryostat window.

The remaining mechanisms are operated at cryogenic temperature.

- A pupil mask wheel with TBD positions with a rotation capability for each mask.
- An 8-position filter wheel used for the R=500 spectroscopic mode.
- Lenslet array selector. A two-position translation stage.
- A sliding slit used for calibration. Located at the focus of the lenslet array.
- A six-position filter wheel past the IFS collimator.
- A six-position grism wheel.

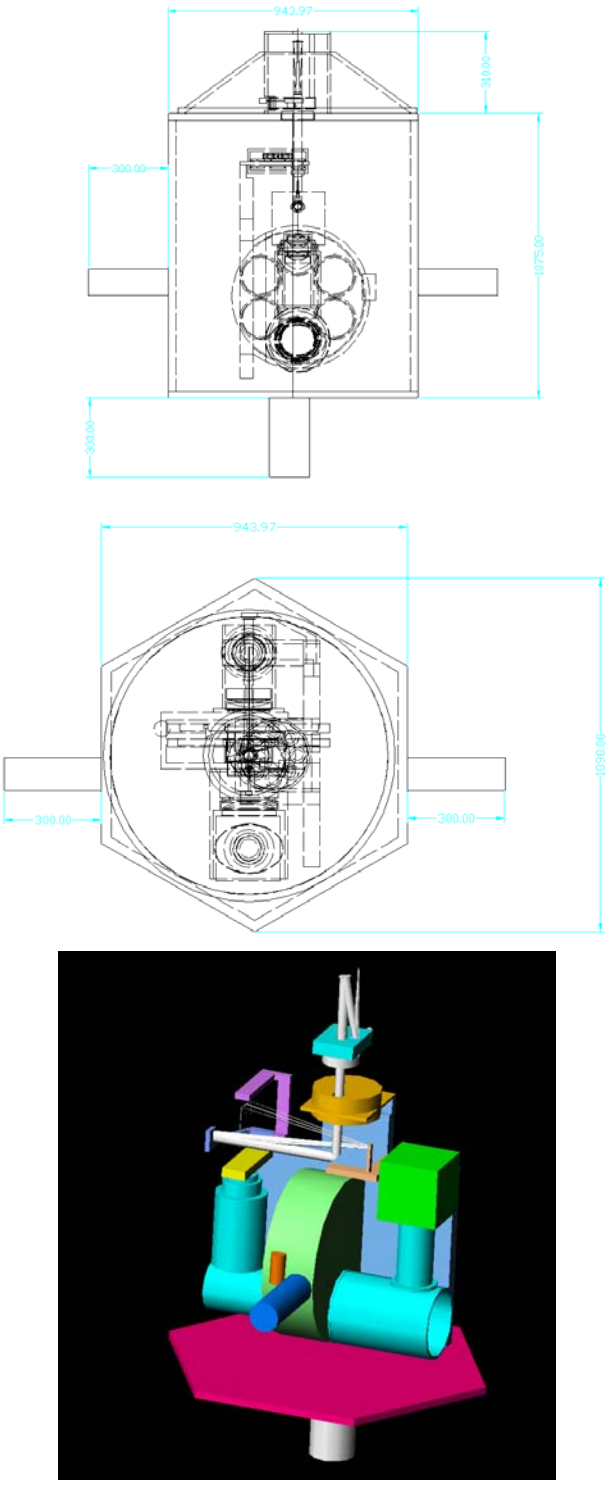

Figure 94: Mechanical layout of the cryostat. Instrument rotator and mechanical interfaces are not shown.

7.4.3 Cool-down time

The cold mass of the cryostat is ~200 Kg requiring ~35 MJ of energy to evacuate from room temperature to 40 K. The heat load (radiation + conduction) is approximately 70 Watts. A single 150 W closed-cycle cooler would cool down the cold bench in approximately 3 days. The propose design includes 2 CCRs. Lenses will cool-down slower than surrounding the cold bench. With 2 CCRs, the camera should be operational in approximately 2 days.

7.4.4 Mass estimate

Table 14 gives a mass estimate of the various components excluding a TBD interface structure. With an estimate of 200 kg for the instrument rotator, the total mass is approximately 900 Kg, 200 of which are cold.

Table 14: Science Instrument Mass Budget.

Component	Kg
Warm mass	
Cryostat shell	410
Electronic hardware	15
Instrument rotator	200
Cryocooler (qty =2)	25
Turbo pump	5
Warm optical bench	5
Polarization unit	10
Sub-total	**670**
Cold mass	
Optical bench	30
Optics	50
Detector module	5
Filter wheel/grism unit	45
Slit/micro lens unit	5
Pupil/filter wheel unit	10
Shield & misc. Hardware	60
Sub-total	**205**
Total	**875**

7.5 Detector

The baseline detector is a Hawaii-4RG-10, a 1-5 µm 4096x4096 HgCdTe detector with a 10 µm pixel pitch. This device is currently under development at Rockwell Science Center and will be the successor of the Hawaii-2RG developed primarily for the James Webb Space Telescope. Quantum efficiencies in excess of 80% over the full wavelength range are expected with read noise less than 10 e- (double correlated sampling) and dark current less than 0.01 e-/s. For comparison, the *H*-band sky-background per pixel expected at R~100 should be ~0.1 e-/s. As shown in Figure 95, the dark current of the 1-5 µm H2RG is varying steeply with temperature. This sets the operating temperature to <40 K in order to achieve a dark current of 0.01 e-/s.

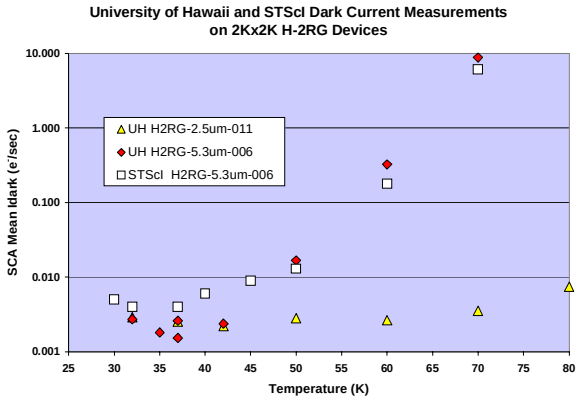

Figure 95: Dark current vs temperature for 2.5 and 5.3 micron cutoff HgCdTe detectors

7.5.1 Readout electronics

The baseline is to use the same or a similar version of the SIDECAR ASIC developed successfully for JWST. The main advantage of the ASIC is to simplify the readout electronics to a single chip. Since the ASIC can be operated at cryogenic temperature, it can be mounted very close to the detector to enable very good read noise performance. The 4-channel ASIC used for driving the H2RG on JWST will drive less than 10 mW of power.

7.6 Data Reduction Pipeline

The IFS data reduction pipeline (DRP) will be very similar to that already developed for the Gemini Planet Imager. The basic data product is a data cube comprising ~30 monochromatic images which can be further processed to apply a speckle-suppression algorithm (MWI or polarimetry) and to reconstruct a broad-band image. One important feature of the DRP is the ability to perform a spectral rectification i.e. to correct for the contamination between spectra. This correction is performed using calibration data obtained by illuminating one lenslet row at time with a sliding slit located at the focal plane of the lenslet.

7.7 Performance

7.7.1 Speckle Suppression Simulations

We have developed IFS simulations tools to assess the sensitivity of the speckle-suppression performance for various spectral spacing *w* with and without spectral rectification algorithm. A single-conjugate aberration spectrum was used to generate a non-coronagraphic PSF (see Figure 96) and the data was processed through our IFS DRP software. The speckle suppression or attenuation factor defined (ratio of input to output speckle noise) was used as figure-of-merit of speckle suppression performance. Figure 97 show some preliminary results for two extreme cases of spectral spacing (*w*=5 and 2) with and without spectral rectification. The conclusions from this preliminary analysis are: 1) the IFS is capable of relatively large speckle suppression (η>100), 2) as expected η is decreasing spectral spacing (small *w*) and 3) spectral rectification do improve (but not drastically) speckle-suppression performance.

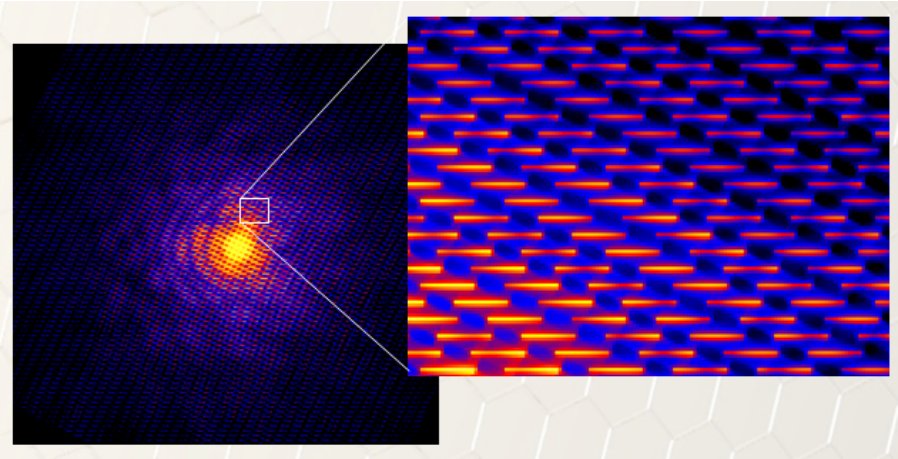

Figure 96: Simulated IFS data with a non-coronagraphic PSF.

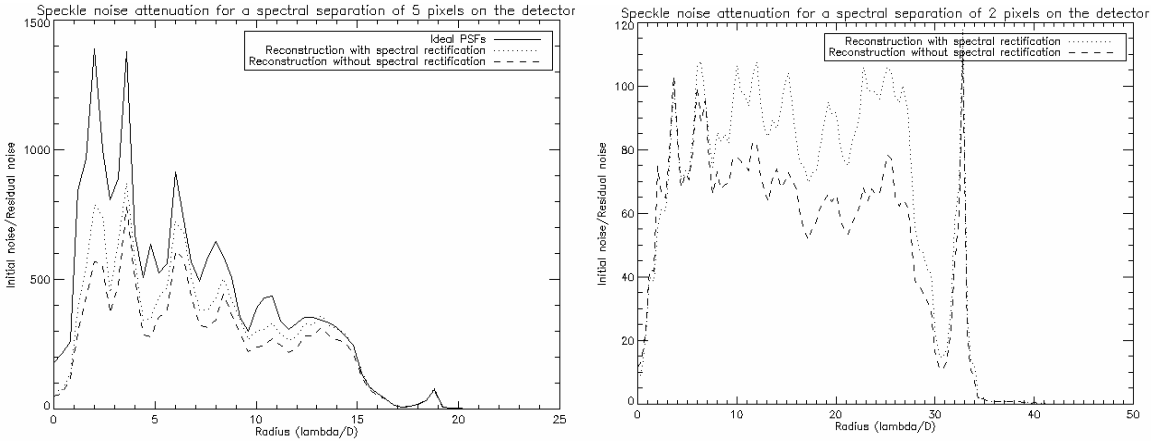

Figure 97: Speckle suppression factor *vs* angular separation in λ/D units for two spectral spacing cases. Left: *w=5*; the solid line is the attenuation assuming a perfect PSF data cube as inout. The dashed and dotted lines are with and without spectral rectification respectively with a data cube extracted from a simulated data set as in shown in Figure 90. Right: same as left with *w=2*.

7.7.2 Speckle suppression with nuller DSS

The nuller transmission pattern varies as a function of wavelength. Fortunately it scales with wavelength in exactly the same pattern as residual speckles do, i.e. magnifying proportional to λ. As a result, a given artifact speckle will be uniformly present throughout the spectral data cube even though the planet itself may not be present in all wavelengths of a single image. We have carried out simulations of speckle rejection on simulated DSS data cubes and shown that they achieve have comparable performance to normal images.

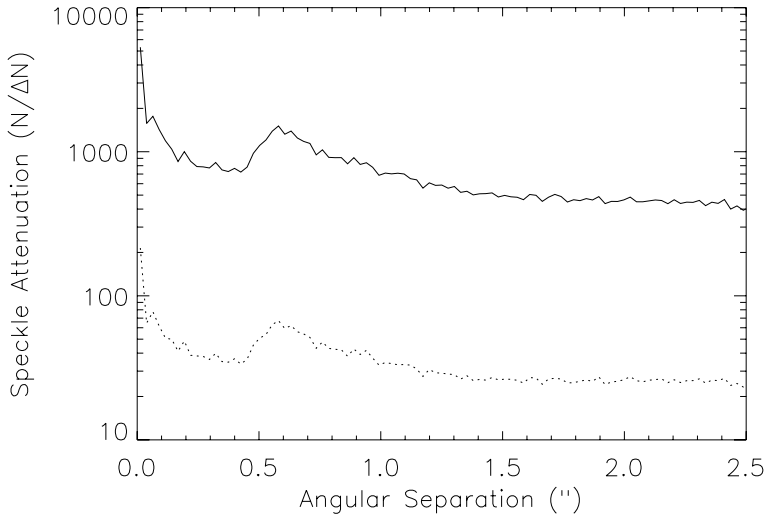

Figure 98: Speckle attenuation as for the nuller DSS architecture. Solid line: speckle attenuation between two adjacent wavelength channels (1.58-1.62 microns.) Dotted line: attenuation from 1.5-1.8 microns.

7.7.3 Preliminary Laboratory Results

We have set up an IFS test bed at UdeM to validate the IFS concept. The set up consists of lenslet array (circular micro-lens on a 100 μm square grid) coupled with a warm uncoated refractive collimator/camera optics designed for *H*-band operation with a Hawaii-1 (1024x1024) detector. The dispersing element is a BaF_2-Fused silica prism pair yielding a resolving power of 40 with a spectral spacing *w* of 4 pixels and a FOV diameter of ~45 λ/D. The IFS is fed with a f/167 artificial star. shows some preliminary results, from raw to processed data. Our preliminary analysis without spectral rectification yields η~15, a very encouraging result given that the 1st generation of speckle-suppression cameras like TRIDENT (Marois et al 2005) achieved η~3.

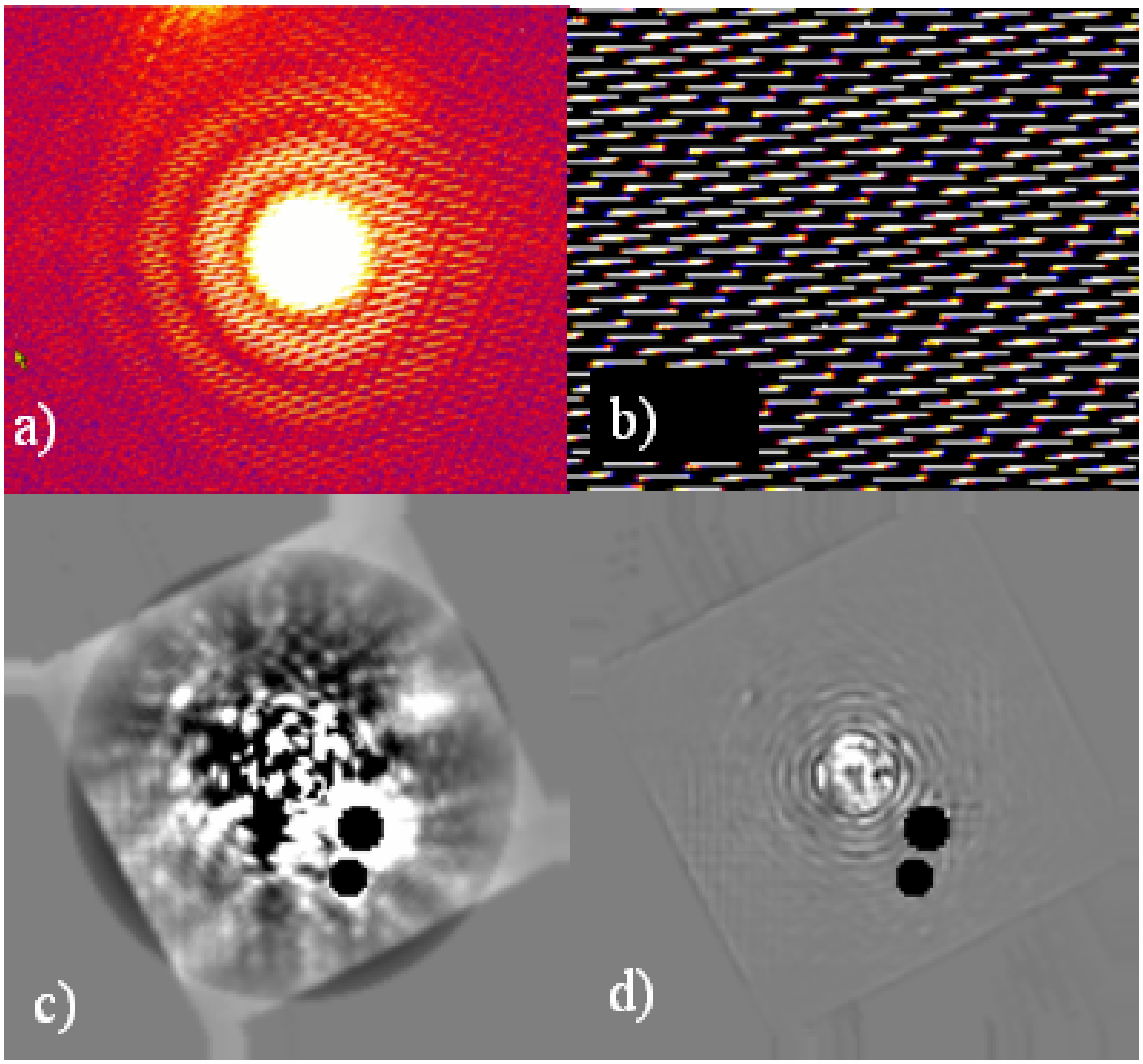

Figure 99: Montage of IFS data from the UdeM testbed. A) Zoom section of a raw PSF data; b) flatfield data; c) Reconstructed monochromatic PSF at 1.6 µm; an average radial profile has been subtracted to show reveil the speckle noise; d) same as c) on the same intensity scale after a MWI speckle-suppression algorithm. The noise in d) as ~15 times smaller than in c). The black areas are masked regions due to optical ghosts

8 Overall mechanical design

8.1 Introduction

At the feasibility study level, it is clearly inefficient to perform a full mechanical design. For this study, ComDev in Canada has carried out a basic design for evaluation of PFI's mechanical envelope, weight, and cost, and enumeration of the mechanisms.

8.2 Structure and envelope

The structure for the TMT Planet Formation Instrument has a number of separable components which are grouped around the major functional elements of the instrument. These are: the coarse (woofer) and fine (tweeter) wavefront correction optics, the front AO pyramid WFS camera, the nuller, the back infrared wavefront sensor and its camera and the integral field spectrograph (IFS).

Correspondingly there are separate optical benches for the wavefront correction optics, the WFS camera, and the nuller / back WFS camera. The nuller optical bench also supports the Dewar containing the science instrument (IFS). In the current architecture the science instrument is down-looking and rotates about its optical axis to provide any image derotation that is needed (though normally PFI will operate with this rotator off, since exposures are short.) This interface could be modified to be a copy of the IRIS/NFIRAOS rotator both to save cost and if it is desirable to swap IRIS and PFI's IFS between PFI and NFIRAOS (See Chapter 11.) If it is desired, a second instrument port could be added down-looking or side-looking, possibly with a PFI providing a warm Lyot stop.

Oblique and top views of the overall PFI instrument are shown in the following figures. In these diagrams and others, all the salmon colored elements are mechanisms. The brown colors represent fixed mounts.

The overall mass of the instrument (not including the enclosure) was estimated at 3490 kg with a large portion of the mass in the optical tables and support legs. The envelope was $2.7 \times 4.5 \times 2.7$ m for the instrument portion. The enclosure was sized at $3 \times 7 \times 3$ m.

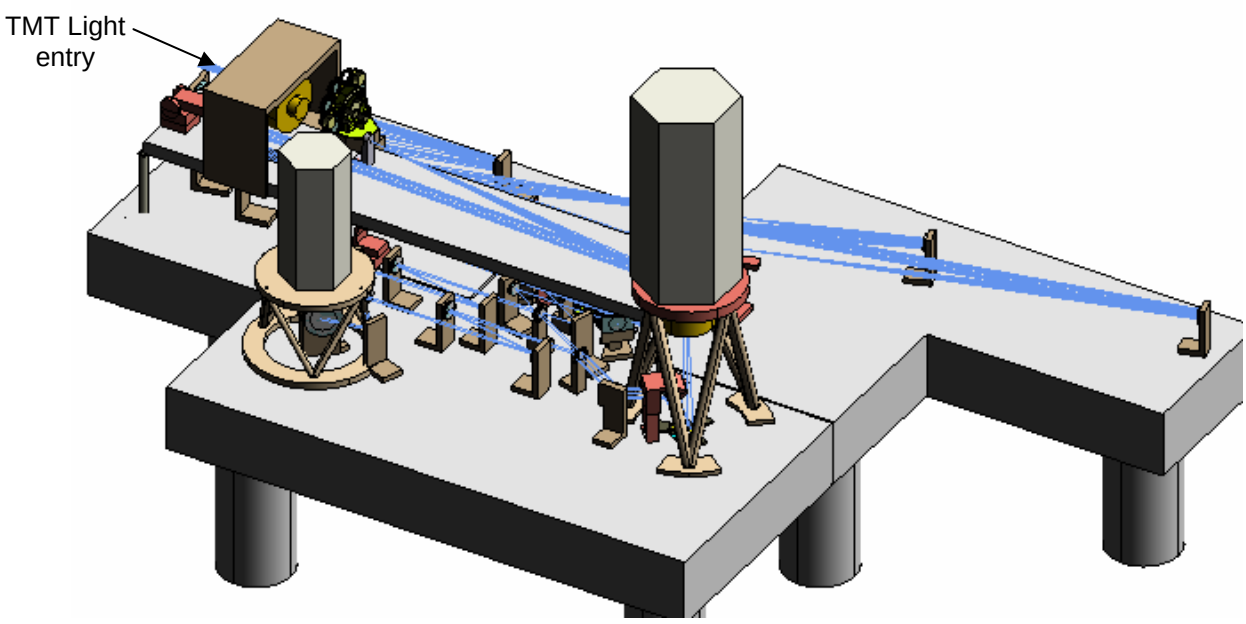

Figure 100: Oblique View of PFI Instrument.

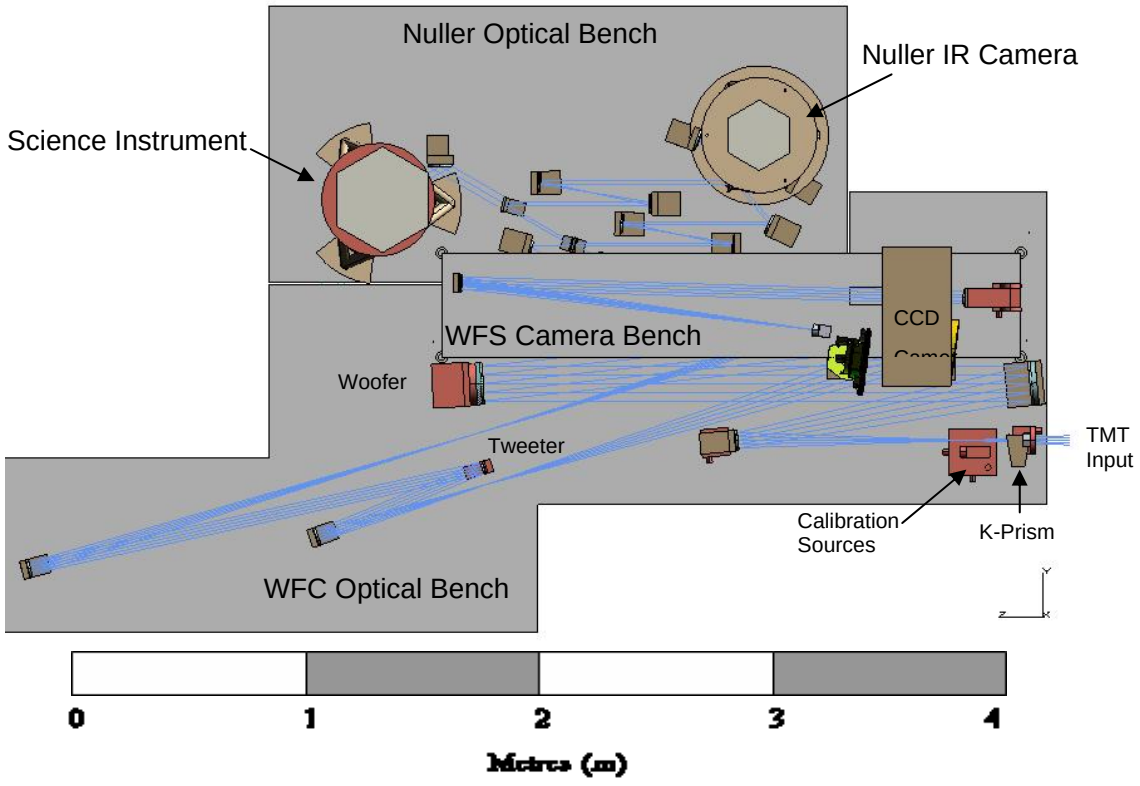

Figure 101: Top View of PFI Instrument.

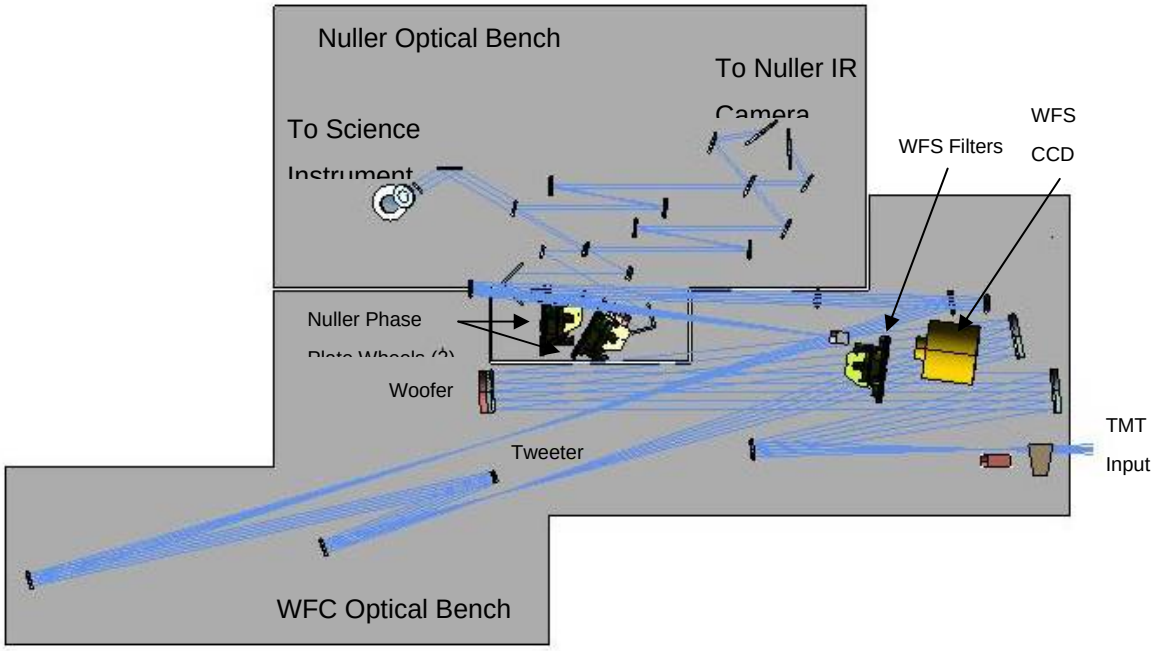

Figure 102: Top View of PFI Instrument with WFS Camera Bench removed.

The optics will be supported with custom designed mounts that will bolt to the optical tables. All of the fold mirrors and off-axis paraboloids (OAPs) will be held from behind, attached at 3 mounting pads to the mount. The three mounting pads can be lapped/faced to provide fine adjustment to the tip/tilt in order to achieve precise alignment. These basic mounts appear throughout the light path, including mirrors in both the front end and nuller areas.

There are a total of four beamsplitters, as well as other refractive optics which require transmissive mounting arrangements, as shown below for example.

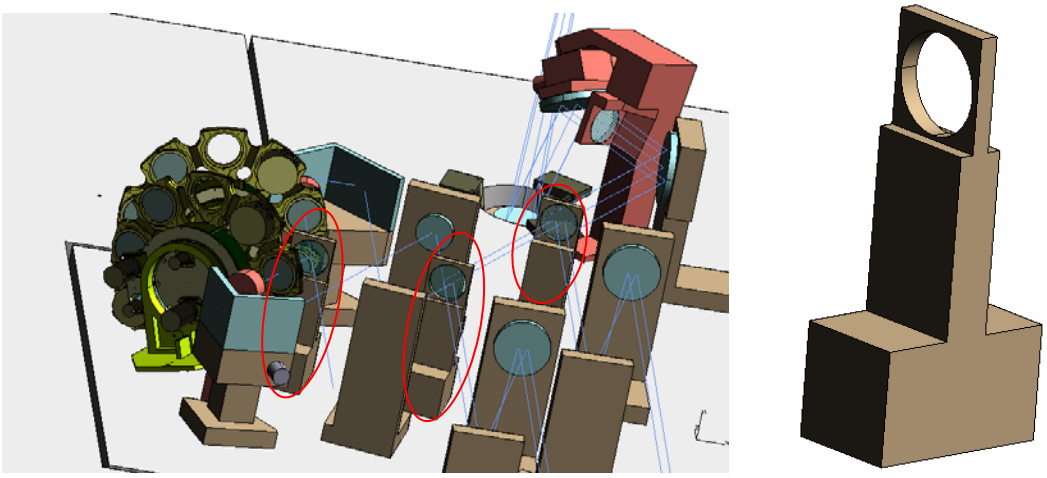

Figure 103: View of Beamsplitter mounts used in Nuller section (left) and close-up of Beamsplitter mount (right).

Besides the common optical component mounts such as the mirror mount and the beamsplitter mount, there are other mounts that will be designed for individual components. These include hexapod mounts for the Dewars, mounts for the cameras and filter wheels, as well as mounts to accommodate the presence of mechanisms for certain optics. Shown below (from left to right) is the hexapod mount for a Dewar, with field de-rotation stage, a filter wheel, the mount for the K-mirror assembly and the mount for the final 3-axis alignment and focus for the science instrument.

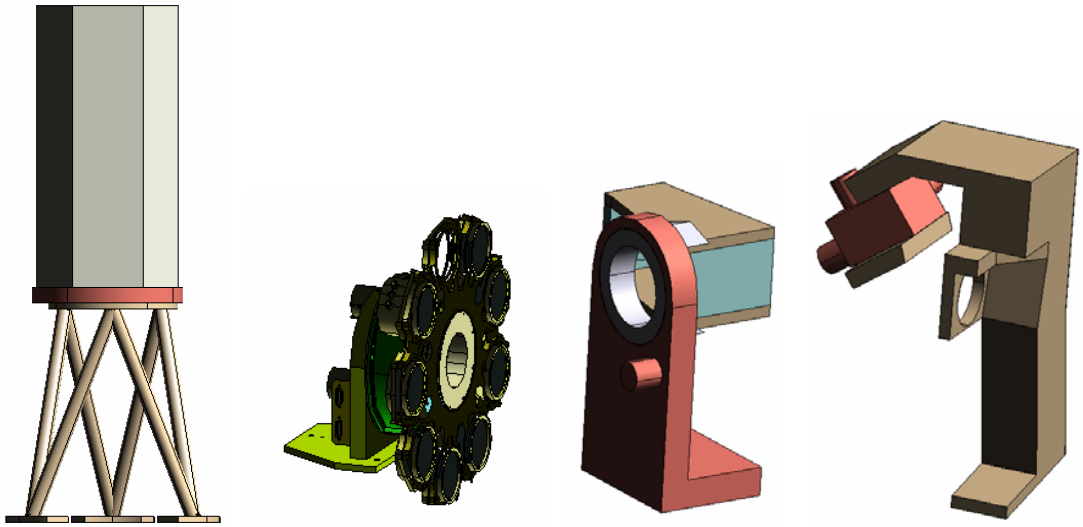

Figure 104: Mounts for IFS, filter wheel, K-mirror image rotator, and science camera alignment mirrors

All of the above mounts will be supported by a pair of optical tables. These types of tables are commercially available but will be custom made to the required shapes. They will be supported by commercially available pneumatic vibration damping legs. As the light path does not require all the available area of a standard rectangular optical table, to decrease mass and increase access some specified portions of the tables will be cut out. The decision to use two tables was based on the sheer size of the optical design and logical grouping of the optical components. One table will be used for the front end of the PFI instrument containing the K-mirror, woofer and tweeter optics, and this bench will also carry the upper WFS camera optical bench which consists of the WFS mirrors, iris, filter wheel and camera. The second table will support the nuller optics and the science IFS. This allows separate assembly, integration and test of these major elements.

The following figures show a view of the optical bench arrangement without any optical components and a top view of the nuller optical bench (with components). Only one stage of the nuller is shown. For a second stage, the nuller optics would be duplicated on the underside of the current bench. Any additional back-side cameras, such as a bright-output sensor, would also be located on the underside.

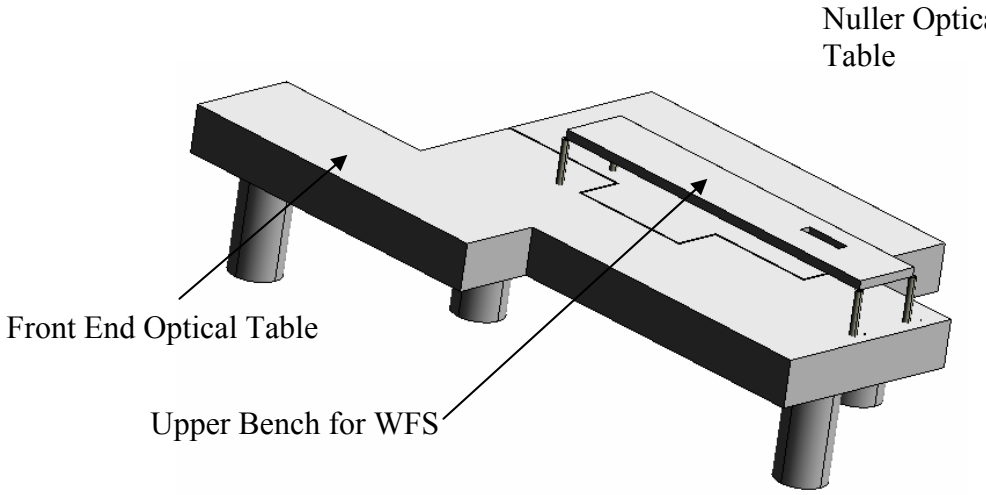

Figure 105: View showing the Optics Tables and Custom Cutouts.

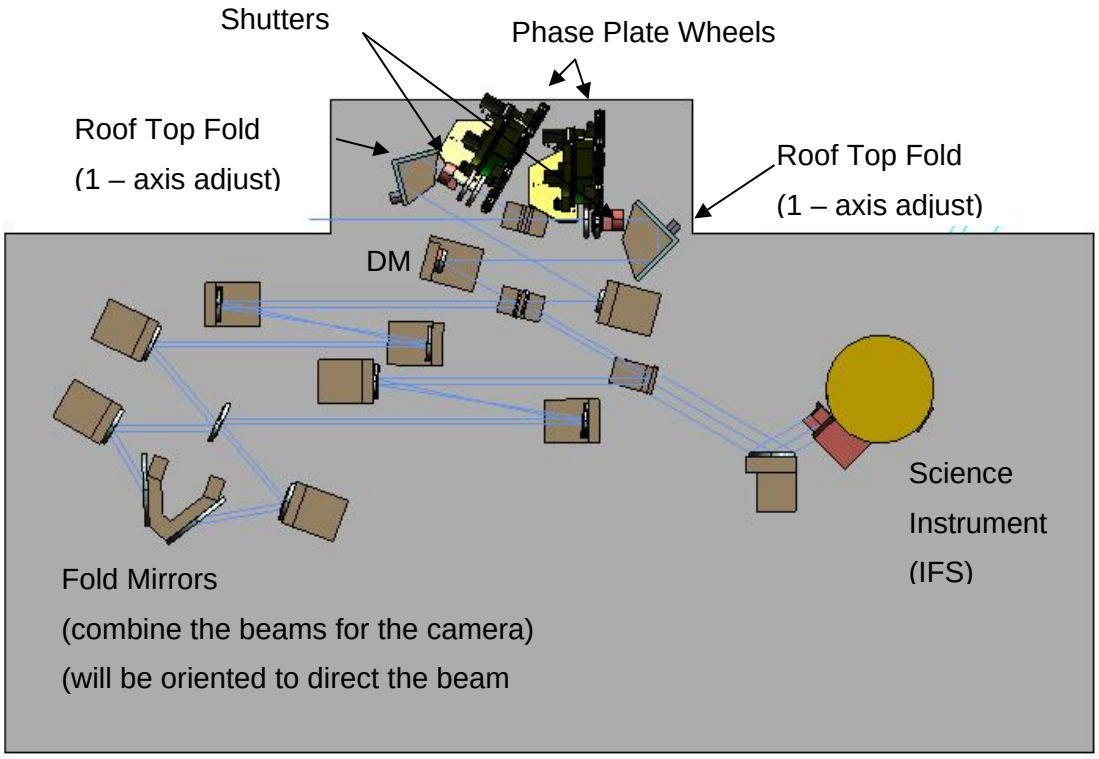

Figure 106: Top View of the Nuller Optical Bench.

There will be an enclosure designed around the optical tables. The enclosure will be required to shield the instrument from dust, wind, and stray light, to provide humidity control – MEMS deformable mirrors are sensitive to humidity – and possibly provide temperature control if the AO optics are chilled to reduce thermal background. This enclosure will be designed cooperatively with TMT, with input from the PFI team pertaining to interface requirements. The enclosure can be mounted on a framework to raise it to the height of the TMT beam above the Naysmith platform. Note that PFI is likely to be extremely vibration sensitive, and solutions involving raising the instrument need careful vibration analysis. Another option for the enclosure is to have the instrument at the same level as the Nasymth platform and use a periscope arrangement of fold mirrors to bring the light from the telescope down to the instrument level.

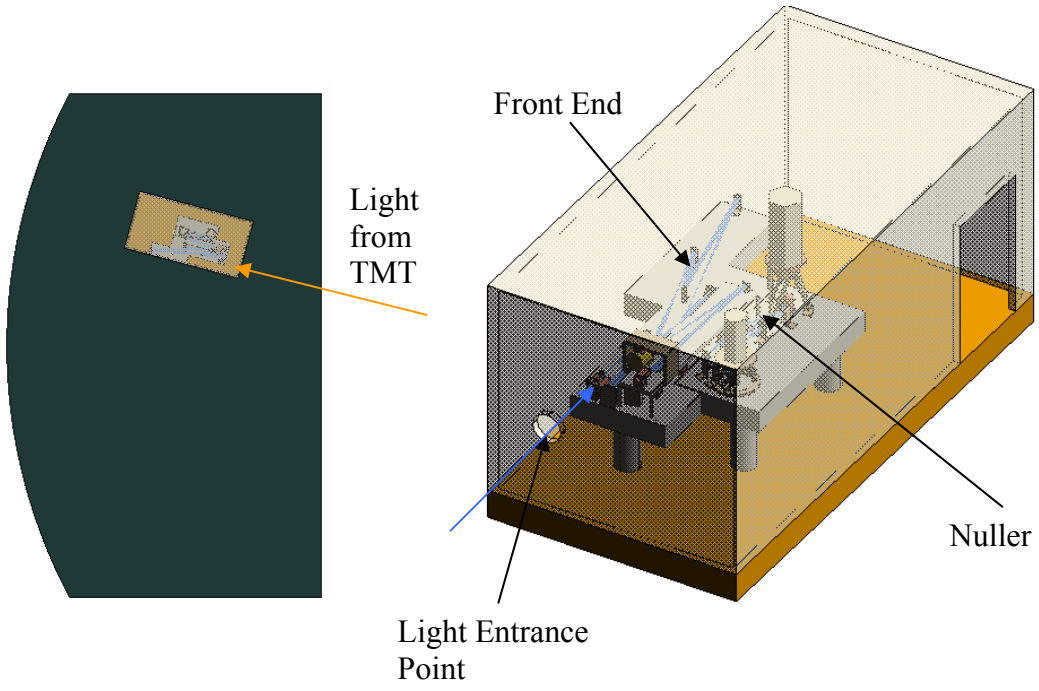

Figure 107: View of the PFI instrument on TMT Nasymth platform (left) and PFI instrument enclosure (right).

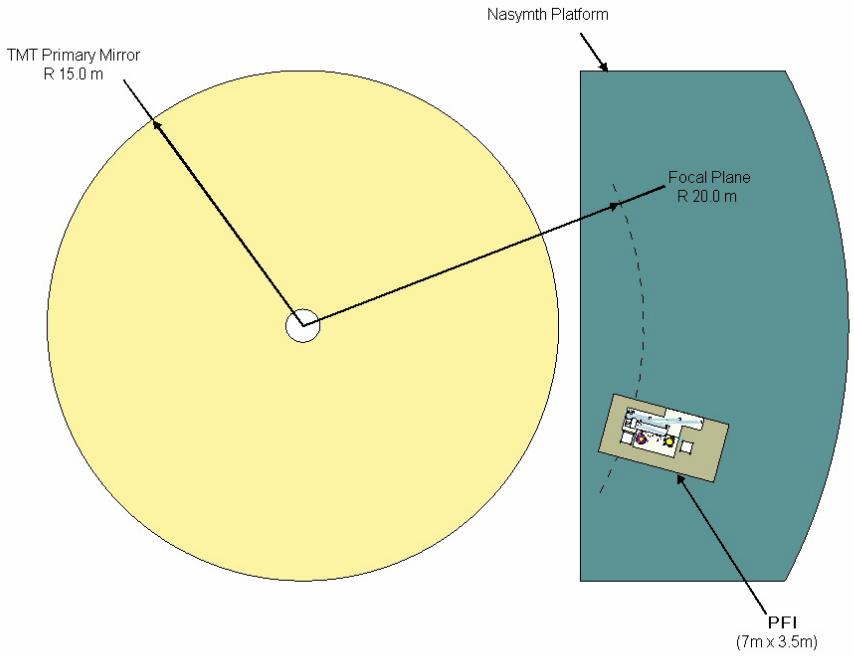

Figure 108: Alternate View of PFI on the Nasymth platform, showing the relative location of the TMT focus and the instrument.

8.3 Mechanisms

This is a preliminary list of required mechanisms for PFI

Table 15: List of mechanisms for PFI

Number	Name	DOF	Type	Freq of Use	Comment
1	K-Mirror	1 - rotation about optical axis	DC Servo	continuous	Pupil rotation control
2	Reference Sources	3 - xyz translation	stepper	setup	stage to insert and align fiber calibration sources (a daytime setup activity)
3	Pupil Alignment Fold	2 - tip, tilt	stepper	continuous	pupil alignment for AO system (1st fold after K-mirror)
4	Woofer	Deformable Mirror	High Stroke	continuous	30x30 actuator bimorph or conventional DM
5	Tweeter	Deformable Mirror	MEMS	continuous	128x128 actuator MEMS
6	Dichroic	tip and tilt	stepper	occasional	operate as a pair to point and center the WFS
7	WFS fold mirror	tip and tilt	stepper	occasional	
8	WFS spatial filter	1 - Iris	stepper	occasional	square, adjustable iris 1 to 5 mm (height & width are linked, only 1 DOF)

9	WFS Filter Wheel	1 - rotation	stepper	occasional	after pyramid
10	WFS CCD Camera Focus	1 - translation	stepper	setup	
A	WFS Atmospheric Dispersion Corrector	3 – rotation plus in/out	servo	continuous	Optional; see below
B	Science ADC	3 – rotation plus in/out	servo	continuous	Optional; see below
11	Nuller A Roof Top	1 - translation	stepper	occasional	variable interferometer delay
12	Nuller B Roof Top	1 - translation	stepper	occasional	variable interferometer shear
13	Nuller A phaseplate wheel	1 - rotation	stepper	occasional	
14	Nuller B phaseplate wheel	1 - rotation	stepper	occasional	
15	Nuller A Shutter	1 - Iris	solenoid	every 10 min	
16	Nuller B Shutter	1 - Iris	solenoid	every 10 min	
17	Nuller C Roof Top	1 - translation	stepper	occasional	variable interferometer delay
12	Nuller D Roof Top	1 - translation	stepper	occasional	variable interferometer shear
13	Nuller C phaseplate wheel	1 - rotation	stepper	occasional	
Number	Name	DOF	Type	Freq of Use	Comment
14	Nuller D phaseplate wheel	1 - rotation	stepper	occasional	
15	Nuller C Shutter	1 - Iris	solenoid	every 10 min	
16	Nuller D Shutter	1 - Iris	solenoid	every 10 min	
17	Nuller DM	Deformable Mirror	MEMS	continuous	
18	IR camera focus	1 - translation	stepper	setup	
19	Instrument Focus & Pupil Alignment	3 - translation, tip, tilt	stepper	setup	Focus & Pupil Alignment for Science Instrument
20	Instrument Rotator	1 - rotation	DC servo	continuous	rotate science Dewar to provide field rotation
21	Modulator	2 - rotation & translation	stepper	occasional	Polarimetry mode - warm mechanism just outside the cryostat

22	Pupil Masks	2 nested rotations	stepper	occasional	ability to select and rotate pupil mask individually
23	Filter wheel #1	1- Rotation	stepper	occasional	"blocking filters" for high-resolution modes
24	Lenslet Array Selector	1 - translation	stepper	occasional	2 lenslets; one SW and one LW
25	Filter wheel #2	1- Rotation	stepper	occasional	JHKLM filters (post lenslet)
26	Grism wheel	1- Rotation	stepper	occasional	Grisms/prisms/Wollaston
27	Focussing stage	1- translation	stepper	occasional	TBD; depends on final optical design (refractive vs reflective)

Color Key

Main AO & WFS System (bench)

Nuller & Calibration (note the entire nuller section is duplicated for the second nuller/calibration)

Science Instrument

8.3.1 Atmospheric dispersion correctors

The PFI nuller requires the star to be centered within ~1 mas to achieve a full null. As a result, PFI is almost certain to require IR atmospheric dispersion correction in the science path – at H band, differential atmospheric refraction for Mauna Kea near zenith is about 1 mas per degree! Similarly, the pyramid wavefront sensor is most effective when the stellar image is a point source, so a visible-light ADC may be required as well. The optimal location and design of these ADCs has not been determined; ideally they would be located at or near pupil conjugates or use a pupil-invariant design (C. Shelton, Priv. Com.) so as to minimize beam decenters, and must have excellent chromatic performance. ADC design could be shared between multiple TMT instruments. If ADC implementation is too challenging, PFI could operate over restricted wavelength ranges using only part of the K and L' atmosphere band and zenith angles.

8.4 Mechanical and environmental summary

Total mass	3500 kg (+ 1500 kg margin)
Volume (enclosure)	3x7x3 m
Environment inside enclosure	Overpressurized dry air + temperature control
Power required	5000 W[1]
Cryogenic cooling requirements	Two CCR heads (150 watts) for science instrument One CCR head (<20 watt) for IR WFS
Location	Naysmith platform

[1] Largest component is AO realtime computer (3 kW).

9 System performance

9.1 Analytic error budget

Classical AO systems – and indeed most optical instruments – are ultimately designed based on an image quality error budget, expressed as enclosed energy, FWHM, total wavefront variance, or Strehl ratio.

Extreme AO systems must be governed by a different philosophy. The ultimate metric of success is the detectability of a faint companion near a bright star, which in turn flows not just from the Strehl ratio of the companion but from the detailed behavior of the halo of scattered light surrounding the star. The overall intensity of that halo at an angle θ is determined, to second order (Perrin et al. 2003 Ap.J. 596, 702), by the wave front error spatial power spectrum evaluated at spatial frequency $\theta/(\lambda/D)$ in cycles per aperture. High spatial frequency errors, scattering light far beyond the region of interest, are therefore much less critical than low spatial frequency errors.

This simple relationship is only true to second order in wavefront error (σ^2.) When high frequency errors are particularly strong and concentrated at specific spatial frequencies, a fourth order term in the PSF expansion (Giveon et al 2004, 2006) that represents a beating between different spatial frequencies becomes significant and can scatter light from high-frequency errors back into the otherwise dark region. This phenomenon explains how high-frequency errors from phase discontinuities at segment edges can scatter light within the PFI region of interest (Section 10)

Equally critical is the temporal behavior of errors. Detectability depends not just on the local intensity of scattered light but on the smoothness of the scattered light halo. Unbiased error sources, such as atmospheric errors, will ultimately (albeit slowly) average out to produce a smooth halo; any bias to this process will produce a pattern of fixed speckles that could swamp the signal of a planet. Figure 109 illustrates this concept. All three images have ~70 nm RMS dynamic wave front error and Strehl ratios that differ by less than 0.0001. Ultimately, in the presence of both dynamic and static errors, the final contrast achieved in an infinitely long exposure will be the same contrast that would have been achieved with only the static errors present.

If these wave front errors were truly static, they could be removed by PSF subtraction. However, if the errors were truly static, they could also be removed through better AO calibration and control. In practice, most of these errors arise from time varying sources—changes in beam position on imperfect optics, changes in gain of a wave front sensor quad cell, due to changing r_0 —that limit PSF subtraction techniques. A truly successful ExAO system must minimize these errors in the design stage.

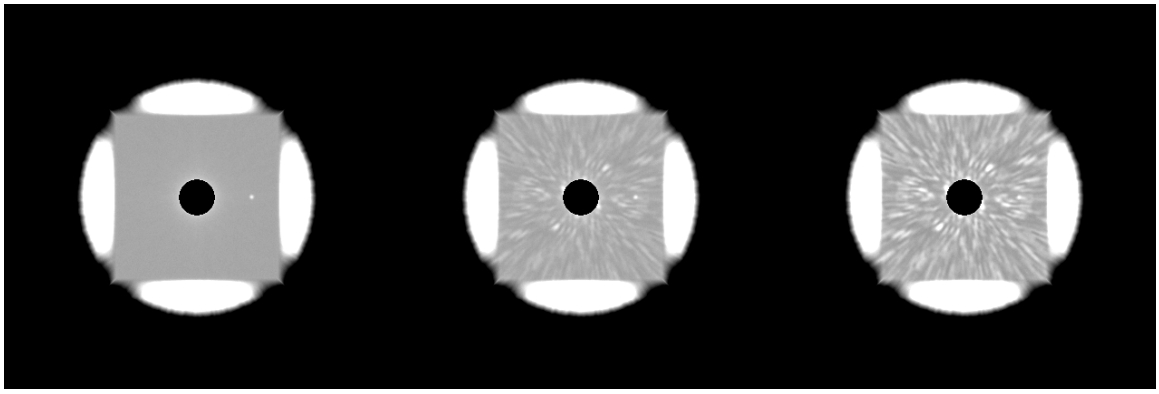

Figure 109: Three simulated 15 minute 8-m telescope extreme AO images with 0 (left), 2 (center), and 4 nm (right) RMS random static wave front error (uniformly distributed in spatial frequencies out to the AO cutoff). This simulation was carried out with low winds (< 5 m/s) to reduce the effect of atmospheric bandwidth errors.

A final key point is to understand how sensitivity scales with telescope size. If a wavefront error is external to the instrument, the final contrast achieved in the presence of that error will scale as telescope diameter D^2. If the error is internal to the AO system, on the other hand, the final contrast will be independent of telescope diameter. TMT's great size compared to 8-10m telescopes allows it in theory to overcome the effects of atmospheric turbulence and in theory achieve contrast 10 times higher than 8-10m telescopes, but to achieve this in practice, internal optical effects must be understood and controlled a factor of 3 better than in instruments such as GPI.

9.1.1 PFI contrast design philosophy

The key optical principals underlying the PFI design are:

(1) Uncorrected static wave front errors must be <0.5 nm over the controlled spatial frequency range.

(2) The point at which the wave front must be most correct is inside the nulling interferometer, where the coherent part of the wavefront is removed through destructive interference. Errors up to that point can scatter light from the PSF core into speckles in the wings, reducing contrast. After that point, the total available intensity has been reduced, and errors are less able to scatter light by a factor greater than 1-Strehl.

(3) Optical errors after the integral field spectrograph lenslet plane are even less significant. After the science beam has been dissected by the lenslet array, it becomes extremely difficult to scatter light from one spatial location to another. The lenslet demagnifies each input pixel into a grid of well-separated pupil images. To first order this decreases residual speckle noise by a factor $(f_{in}/f_{lenslet})^2$. Combined with the fact that the lenslets are after the coronagraph, this makes post-lenslet errors negligible.

(4) For multiwavelength speckle suppression to work effectively, the static optical errors must be identical (at better than the nm level) between different wavelength channels.

(5) The effects of internal wave front errors from imperfect optics will never be perfectly corrected in both phase and intensity. Many other uncorrectable errors – such as chromatic errors from changes in beam location with wavelength – also depend on optics quality. Using the highest quality optics practical will minimize static wave front effects.

9.1.2 Assumed system parameters

The following parameters were used to construct the analytic.

9.1.2.1 Telescope and atmosphere

Clear aperture 30.0 m

The base atmosphere used here is a simplified version of the Gemini CP GLAO Typical/Typical model.

Telescope phase errors are as discussed in section 10. Telescope reflectivity is assumed to be 1% RMS.

Atmosphere layer	Altitude	r_0 (500nm)	speed
Lower	0 km	0.20 m	8.4 m/s
Upper	4 km	0.24 m	14 m/s

9.1.2.2 Adaptive Optics System base parameters

Subaperture size	0.24 m
Update rate	2000 Hz
3dB Bandwidth	200 Hz (With an optimal modal controller (§6.2.1) bandwidth will be a function of spatial frequency; this is a representative goal bandwidth averaged across all controlled frequencies.)

9.1.2.3 Optical transmission (visible/IR)

Telescope	3 optics @ 0.96/0.98
AO common path	11 optics @ 0.96/0.98
Dichroic	1 optic @ 0.95/0.98
WFS optics	4 optics @ 0.96
WFS QE	0.7
DSS optics	5 optics @ /0.98
DSS pupil throughput	0.50
Science camera	0.40 total transmission
Science QE	0.75

9.1.2.4 Optical errors

Spatial frequency distribution for the powered (OAP) optics are based on discussions with Tinsley from the Gemini GPI study. These optics are certainly with the capabilities of modern manufacturing – Tinsley produced large highly aspheric optics with RMS surface error of 0.2 nm for the Extreme Ultraviolet Lithography (EUVL) demonstrator project. However, procuring such optics today would be extremely challenging – these optics are both a factor of two better than the GPI optics and three times larger. If EUVL is used for large-scale semiconductor manufacture, then the price of such optics may come down. Alternatively, a narrow-field PFI design could be explored that uses mostly spherical optics. Finally, internal intensity correction (11.3) would relax the optics specifications.

Other optical surfaces are assumed to have a power spectrum proportional to spatial frequency $k^{-2.5}$ down to a frequency of 1 cycle/aperture. Results here are relatively insensitive to the choice of power law exponent. The assumed quality of each surface type in the pre-coronagraph common path is as follows:

Surface type	Number of surfaces	RMS wave front phase error Per optic, low/medium/high freq.
OAPs/powered	5	1 / 0.4 /2 nm
Flats	7	0.2 / 0.1 / 0.04 nm
Dichroic (trans.)	1	0.8 / 0.5 / 0.2 nm

Optics in the coronagraph and post-coronagraph have relaxed tolerances, but will nominally be manufactured to the same quality. In addition, the deformable mirrors are presumed to have 10 nm RMS of uncorrectable high-frequency errors.

9.1.2.5 Science channel parameters

Science wavelength	1.65 µm
Filter bandwidth	0.3 µm
Detector QE	0.75

9.1.2.6 WFS channel parameters

Wavelength range	700–1000 nm
Detector QE	0.7
Number of pixels per subaperture Per timestep	4
Readnoise	1 electrons RMS

9.1.2.7 Target properties

I	5-7 mag.
I-K	0.6 mag.
Science exposure time	3600 seconds
Parallactic rotation rate	0.1 degrees/second
Angular separation of interest	0.2 arc seconds

9.1.2.8 Nuller DSS parameters

Null depth	1×10^{-9}
Output pupil size	50% of input pupil

9.1.2.9 Image processing

Cases were generated assuming no post-processing of images – the raw contrast delivered by the instrument – and assuming post-processing for speckle suppression from multiwavelength data cubes and image/field rotation. For the former, speckles from wavefront errors assumed to be suppressed by a factor of 16; errors from out of a pupil plane or intensity errors were assumed to be suppressed by a factor of 1-4. Quasi-static errors were assumed to be suppressed by a factor of 2 in subtraction of images from different parallactic angles, appropriate for these small angles. (See Marois et al 2006).

9.1.3 Error budget and PSF components

Using this PSF-based formalism we can produce an error budget for PFI that will allow us to not only predict Strehl ratio but also to estimate final contrast at a given radius, as in Table 16. For each error source, in addition to its magnitude, a speckle lifetime and an estimated spatial frequency distribution has been used; some discussion of those distributions follow. Scattered light terms that do not scale directly from the magnitude of phase errors, such as coronagraph leakage and light scattered by amplitude errors, are also included in the table and will be discussed in more detail in a future document. In the following sections we will briefly discuss the properties of each error source and in several cases illustrate the PSF that would occur if that were the only error source; as discussed above, the final long-exposure PSF will be given by the sum of these separate PSFs, at radii where the diffraction component is negligible.

We divide each error source into three spatial frequency components: low frequencies (< 3 cycles/pupil), corresponding to wave front error sources that primarily scatter light beneath the occulting spot of the coronagraph; high frequencies (> 64 cycles per pupil), which scatter light into the wings of the PSF at large angles; and the crucial mid-frequency range that corresponds to the dark hole in which we wish to detect planets. For each error source, we present the corresponding scattered-light intensity at a radius of 0.1 arc seconds, and the resulting speckle noise in a 3600-second exposure. (Note that since speckle noise occurs from the intensity of the PSF, which is already a squared quantity, the speckle noises add linearly rather than in quadrature.) The contribution to the final noise due to photon noise is also listed.

Table 16: Contrast error budget for *I*=5 star, no speckle suppression. PSF intensity and PSF noise are normalized with respect to the peak intensity of the un-nulled (off-axis) PSF, so that a PSF noise of $1x10^{-8}$ would represent a 5-sigma detection of a companion with a total flux relative to its primary of $5x10^{-8}$.

WFE source	low freq. (nm)	mid freq. (nm)	high freq. (nm)	all freq	PSF intensity 0.2 arcsec	1 Speckle noise 0.2 arcsec 3600 seconds
Atmos. Fitting	0.0	3.3	65.0	65.1	8.2E-08	2.0E-10
M1 phase errors	0.1	0.1	17.3	17.3	1.7E-08	3.2E-09
M1 reflectivity errors					6.9E-09	5.6E-10
Initial calibration	3.0	0.5	0.0	3.0	1.9E-09	1.9E-10
Servo Lag	10.3	10.9	7.6	16.5	5.8E-07	1.4E-09
WFS measurement	0.0	18.8	0.0	18.8	2.7E-06	4.4E-10
Internal optical errors	0.0	0.0	10.6	10.6	3.1E-13	3.2E-14
Post-DSS aberrations	8.7	3.5	1.8	9.6	7.9E-09	8.0E-10
Flexure/Beam Shear	0.0	0.2	0.0	0.0	1.9E-10	2.0E-11
Chromatic beam shear	0.0	0.0	0.0	0.1	1.9E-13	1.9E-14
Intensity scintillation	10.0	0.5	1.0	10.1	1.1E-07	2.3E-10
Scintillation chromaticity					5.4E-08	1.2E-10
Polarization errors	20.0	0.0	0.0	20.0	1.9E-09	1.9E-10
Tertiary intensity errors	0.0	0.0	0.0	0.0	4.0E-11	4.1E-12
Internal static intensity errors	0.0	0.0	0.0	0.0	1.9E-08	2.0E-09
Residual diffraction					1.0E-09	1.0E-10
Flat field errors					1.0E-20	1.8E-10
Photon noise						4.1E-10
Total	26.3	22.2	68.6	76.7	3.6E-06	9.6E-09
Strehl ratio @ *H* band				0.92		

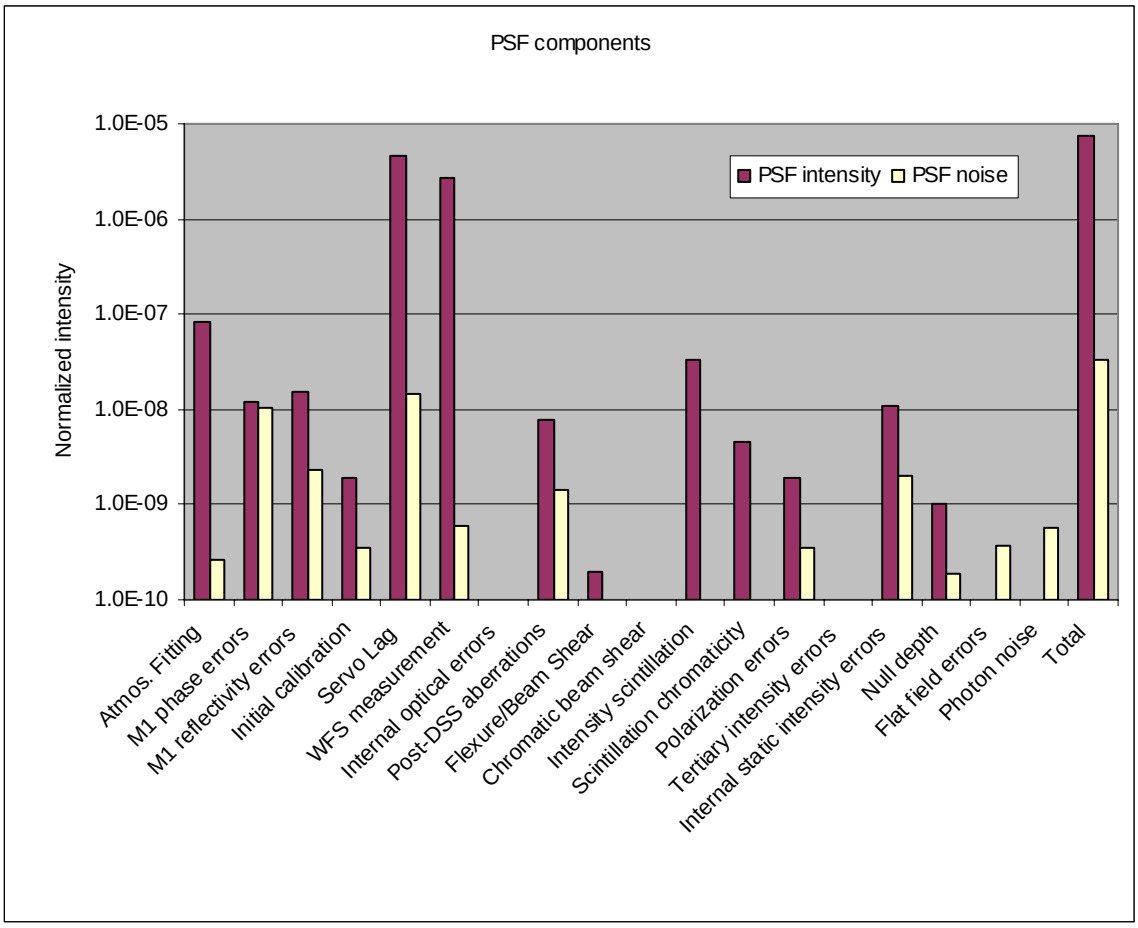

Figure 110: Graphical representation of contrast error budget.

The predicted detectable point source contrast – 5 x10^{-8} at 0.2 arc seconds at the 5-sigma level in a one-hour integration – agrees within a factor of 2-3 with the more detailed simulations presented below, and allows us to see that the system is dominated by servo timelag errors, segment errors in M1, and intensity errors from internal optics. Figure 110 shows this graphically.

Figure 111 shows the same error budget with speckle suppression in place; photon noise in the individual wavelength channels and internal static intensity errors – which occur out of the pupil plane and produce speckle patterns without simple wavelength scaling – dominate.

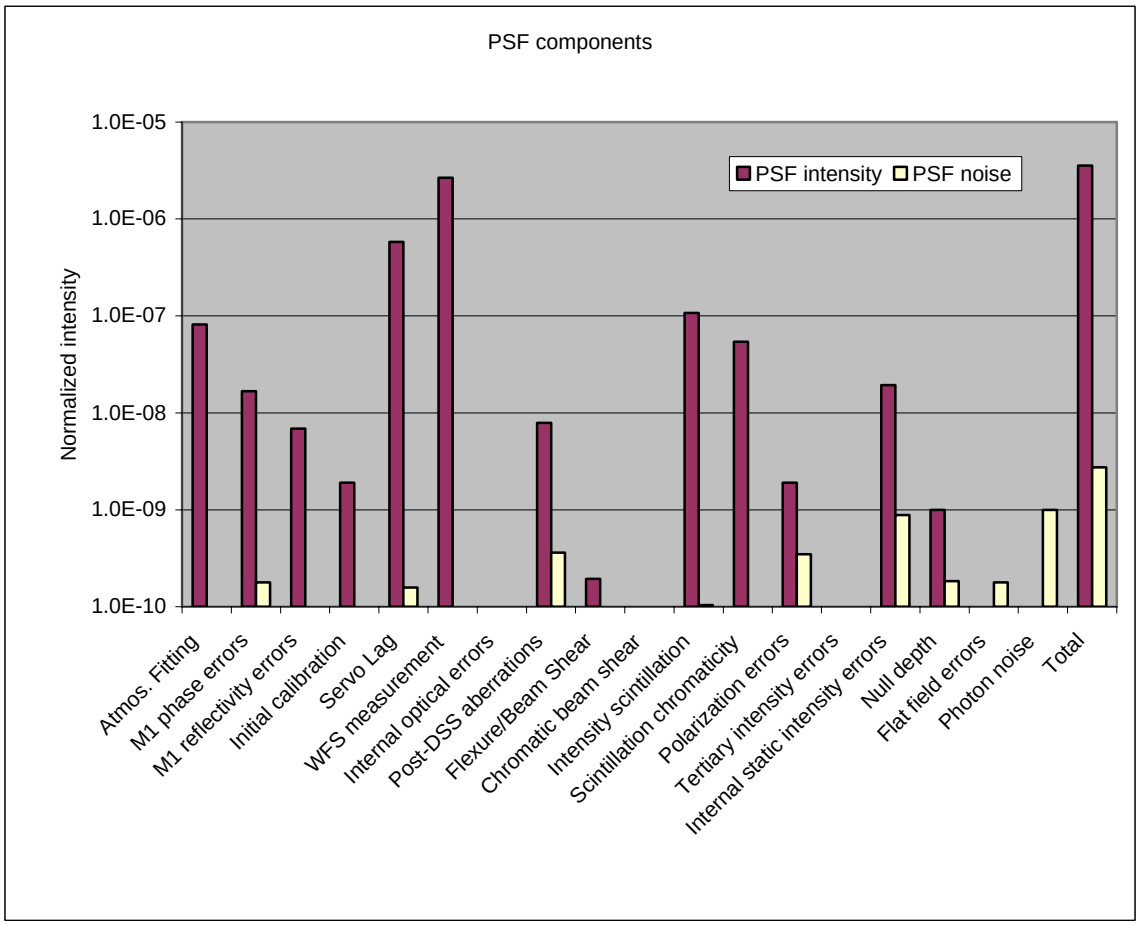

Figure 111: Contrast error budget assuming post-processing speckle suppression.

9.1.4 Atmospheric errors

Classic atmospheric fitting error, i.e., the portion of the atmospheric phase errors that cannot be corrected by the deformable mirror, is almost exclusively a high-frequency error source; a well-behaved DM can reproduce the band-limited components of almost any wave front. This produces a PSF similar to the left portion of Figure 112, bright only outside the AO control region with a small amount of error close to the star caused by fourth-order terms in the PSF expansion. Of course, most AO systems do not produce PSFs of this shape, due to aliasing effects resulting in PSFs similar to the right portion of Figure 112, where the dark hole is filled in. Spatial filters in PFI's wavefront sensors will block aliasing; at high Strehl ratios this can reduce the amount of mid-frequency power by a factor of 100 or more. This produces a significant increase in sensitivity, as speckles due to atmospheric fitting and aliasing errors are long-lived.

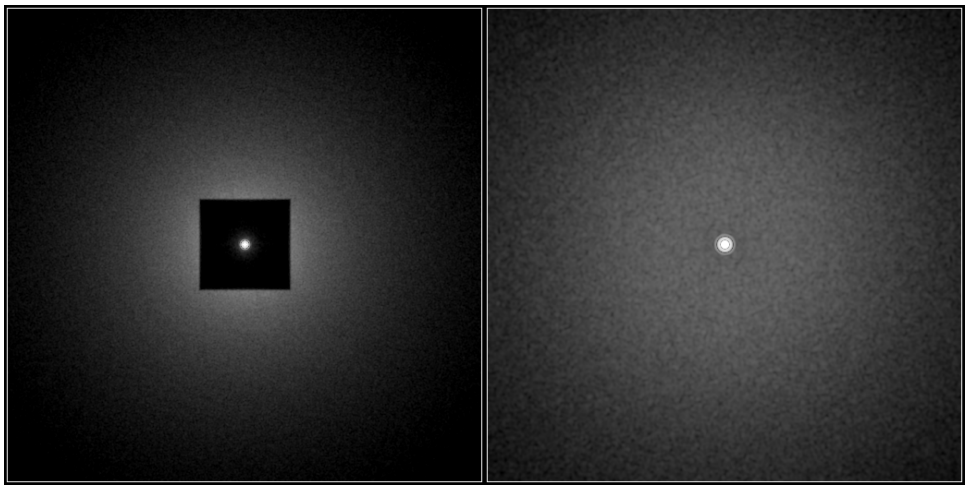

Figure 112: Simulated PSFs due to atmospheric fitting error from an AO system with a spatially filtered wave front sensor (left) and fitting + aliasing errors (right) from a classic AO system. Five-second exposure monochromatic PSF with diffraction suppressed by pupil apodization.

9.1.5 M1 Phase and Intensity errors

These errors are discussed in detail in Chapter 10. Residual errors due to the telescope are assumed essentially static, smoothing out only as the telescope pupil rotates with respect to the sky. Static intensity errors, e.g., non uniformity of the primary mirror coating, are also included in the model.

9.1.6 Initial calibration

In a simple AO system, the wave front sensor produces a flat wave front not at the science focal plane but at the wave front sensor plane. Any optical differences between these two locations – referred to as non-common-path errors – will produce aberrations in the science image. These aberrations can be removed by calibration, using measurements of the science wave front to adjust the control point of the wave front sensor. However, if these aberrations change, e.g., due to flexure or differential atmospheric effects, the science wave front will again be non-ideal. Even small amounts of residual static error can completely dominate final sensitivity.

Our architecture includes two wave front sensors – the fast visible-light sensor and the second infrared wave front sensor integrated into the DSS. Through phase-shifting interferometry the second provides an almost perfect measurement of the nulled wavefront; the nuller itself converts phase errors left over from the AO system into intensity errors, so wavefront errors in the post-DSS sensor itself are almost irrelevant. Using daytime image sharpening techniques similar to those demonstrated on the JPL high-contrast imaging testbed we believe we can reduce residual calibration errors to <5 nm low frequencies and 0.5 nm mid-frequency.

9.1.7 Atmospheric temporal bandwidth errors

One of the two main sources of scattered light within the dark hole is the finite temporal bandwidth of the AO system and its inability to keep up with the moving and evolving atmospheric turbulence. Figure 113 shows the corresponding PSF. For a single Taylor screen, this has a characteristic "butterfly" shape, though of course multiple layers and deviations from frozen-flow will cause the PSF to become more symmetric. This is the main source of scattered light close to the star. As with other atmospheric errors, this produces moderately long-lived (0.2-0.4 second) speckles, so that even with PFIs 2000 Hz sample rate, this remains the dominant external term in the final contrast budget. The optimal AO controller will reduce these errors. The spatial frequency distribution of these errors is currently set by a lookup table based on numerical simulations.

9.1.8 Wave front sensor measurement noise

The second major source of scattered light inside the AO control radius is wave front errors injected by the AO system itself due to the finite SNR of the wave front sensor measurements. By definition, this is a source only of errors within the controlled range of spatial frequencies, and can be significant for dim target stars. Fortunately, since these errors are uncorrelated from measurement to measurement, they produce speckles that decorrelate rapidly – compare the smoothness of the right-hand image in Figure 113 to the left-hand image. We have set $t_{dec} = 1/f_s$ where f_s is the servo bandwidth rate of the AO controller, though for some classes of controller this may be an oversimplification. The spatial distribution of these errors is assumed to be flat within the mid frequency range, a good approximation for interferometric wavefront sensors.

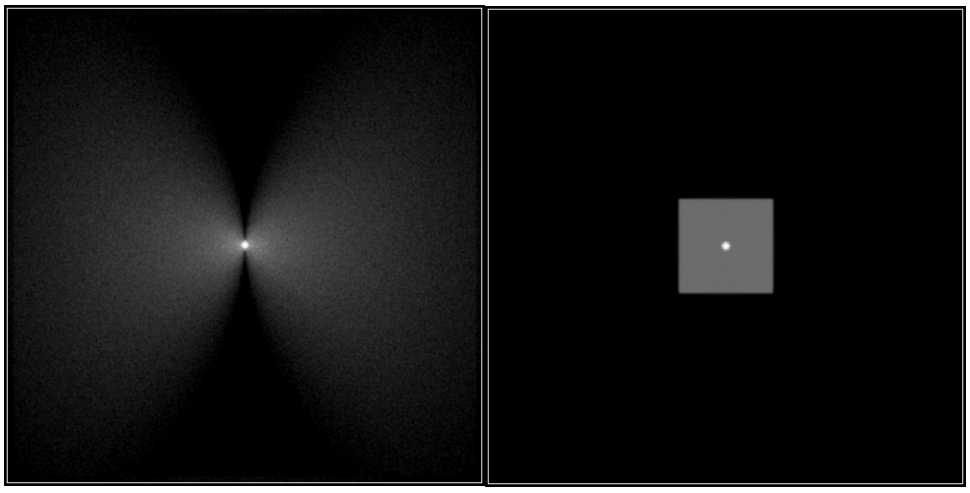

Figure 113: Left: Simulated PSF due to temporal bandwidth errors with the wind moving left to right. Right: Simulated PSF due to wave front measurement noise for a spatially-filtered direct-phase measurement wave front sensor. Five second exposure monochromatic PSF with diffraction suppressed by pupil apodization.

9.1.9 Post-DSS errors

Wavefront errors after the diffraction suppression system or coronagraph occulting spot are often neglected in ExAO modeling. These can still be significant; to first order, the light

scattered after the coronagraph is given by the light scattered before the coronagraph scaled by (1-*S*) where *S* is the Strehl ratio of the system up to the coronagraph spot. Since the pupil after the coronagraph is not uniformly illuminated, correcting for both pre- and post-coronagraph errors with the deformable mirror is similar to correcting intensity errors (ie, possible only with half-dark hole algorithms and in monochromatic light.) We have assumed that these errors cannot be corrected and placed a requirement on post-coronagraph optics as given in above. These errors are currently assumed to be uniformly distributed in radius inside the dark hole, which is a conservative assumption; simulations show that these scatter light primarily near the outer edge of the dark hole.

9.1.10 Internal phase-induced intensity errors

Internal phase errors at planes not conjugate to the DM will result in mixed phase and intensity errors in the pupil plane; when the phase components of these errors are corrected the intensity component will remain. Numeric simulations have been carried out at LLNL to explore the magnitude of these effects for a simplified AO system. This is the main effect setting tolerances on the quality of the internal optics. This model produced the contrast predictions here, but evaluated only the effect of optics in the location of the first OAP. The severity of the intensity errors produced depends strongly on the position of the optic in question – the further an optics is from a pupil, the larger a fraction of the phase error will have converted into amplitude, up to the Talbot length, at which phase has completely converted to amplitude. Optics near focus are particularly significant, and the image rotator in the current GPI design is an area of concern During the construction of the Gemini Planet Imager, LLNL is developing a full optical model of GPI which can be adapted to TMT's PFI and used to set tolerances on individual optics and compare different designs.

9.1.11 Scintillation

Numerical simulations and analytic calculations (see 9.4) have been carried out for the TMT multilayer atmospheres. These show that even for bright targets at high servo rates, scintillation does not dominate the error budget. These results are sensitive to the very highest turbulence layers, however, and should be re-evaluated on the basis of future site survey data.

9.1.12 Scintillation Chromaticity

A related phenomenon is scintillation chromaticity – the difference in wavefront phase between the sensing wavelength and the science wavelength due to propagation from a high-altitude phase layer. Simulations (9.4) show that this effect is much less severe than classic scintillation except at large field angles. Other authors (Guyon 2004) have suggested that this is a more significant effect; this may be due to the smaller wavelength offset between the PFI wavefront sensor, or details of the high-altitude atmosphere used.

9.1.13 Tertiary mirror errors

As discussed in section 9.3, the TMT M3 is sufficiently far from focus that its wavefront errors are well-corrected by the GPI system.

9.2 Numerical Simulations

While analytical error models provide useful first order estimates of system performance, ultimately numerical simulations are needed. Numerical simulations were used to model specific aspects of the system that could not be modeled analytically or to verify analytical calculations. Different subsystems are being constructed by different institutions and require different code bases to analyze their performance. Thus an elegant solution is to write independent codes to analyze the various sub-systems and then to link them together to form a full simulation; the details are discussed below. Results of specific studies using these numerical simulations are described in various portions of this report such as sections 10. ExAO simulations require much longer runs than normal AO simulations, set by the speckle evolution timescale ~D/v (Macintosh et al 2005.) Ultimately, our ability to simulate PFI has been limited by computational resources and time; a second of end-to-end simulation at a single wavelength requires ~1 CPU-day. The results of the full end to end simulations are described in section 9.3 .

9.2.1 Simulation approach

A pipeline approach has been found to be a very efficient method for modeling large problems (Magnuson 1994) . It also allows us to use existing algorithms written in various languages and using different software packages resulting in a significant savings in cost and development time. Each algorithm writes out files (in a specified format) that are then read in by the next algorithm in the chain. This allows individuals to work on, optimize, and analyze the individual portions (algorithms) of the simulations. Figure 114 shows the algorithm pipeline that we developed. The atmosphere algorithm calculates and writes to files the complex pupil at the telescope aperture from the atmosphere as a function of time. The telescope algorithm creates a complex pupil that represents the telescope introduced phase and amplitude errors. The atmosphere and telescope pupils are feed into a simulation of the pre-DSS AO system, which writes out the corrected complex field as a function of time. The DSS algorithm can simulate any of a number of methods for suppressing the starlight, although in this study we concentrated on the nuller. The post-DSS WFS senses the wavefront and controls a DM in the DSS to correct the wavefront. The output of this algorithm is both the corrected complex pupil and an estimate of the complex pupil as measured by the WFS. The corrected complex pupil is read into an algorithm to simulate the science camera that produces co-added PSFs. Finally, a post-processed image can be created using the estimated PSF from the WFS and the science focal plane. The details of each algorithm are discussed below. A limitation of this approach is that it requires the transfer and storage of large amounts of data, but not prohibitively large amounts.

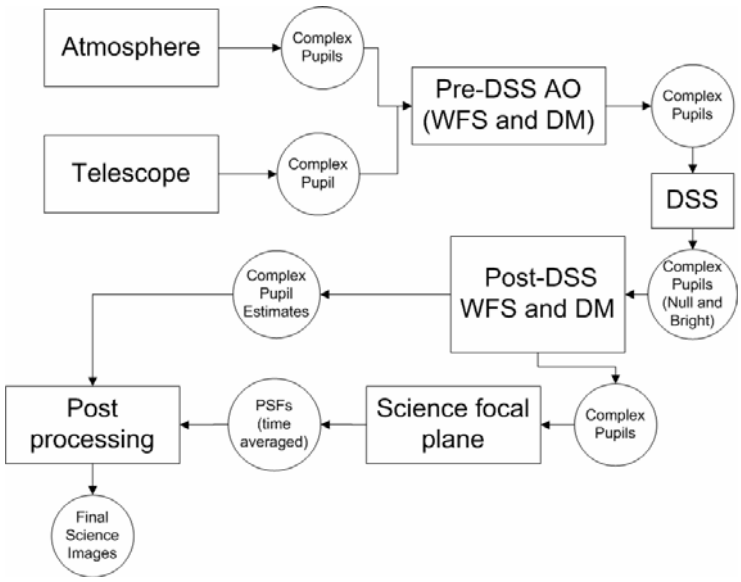

Figure 114: This diagram depicts the algorithms and data flow for end to end simulation of the SSI.

9.2.2 Telescope

We found that a pupil scale of ~3 cm/pixel (slightly zero-padded in a 1024 × 1024 array) gave consistent results that do not diverge with increasing array size. Thus this is the size of the complex pupils used throughout this report. We use two separate tools to generate the amplitude and phase of the telescope's complex pupils.

In calculating the telescope pupil's amplitude map, we had the challenge of accurately representing TMT's 4 mm-wide intersegment gaps in an array of manageable size. Even in an 8192x8192 array, these gaps are only barely resolved; and even at this resolution, constructing a purely logical (0/1) amplitude map will give a poor approximation of the telescope pupil's true shape. Instead, we used a MATLAB implementation of the Gray Pixel Approximation (Chanan 1999) to represent the fractional area of a pixel partially obscured by a gap or other obscuration. In the Gray Pixel Approximation, one computes the analytical inverse Fourier transform of the amplitude function and then performs an FFT on the array of computed values to obtain a representation of the pupil amplitude. We then set the value of all pixels completely outside of the geometrical pupil or completely behind an obscuration to zero.

For the telescope phase maps (due to various segment effects), we applied aberrations (either Zernike modes or phase maps) to each segment. Although exact analytical expressions exist for fitting Zernike modes to hexagonal segments (Jolissaint 2004), they are formidable and we adopt a simpler approach. We constructed a pixel-by-pixel map of the geometrical extent of each segment, and then added a desired phase error to the pixels within that segment. These phase errors took the form of either Zernike modes (in the case of segment surface and alignment errors) or of a directly-applied phase map (in the case of whiffletree print through).

9.2.3 Atmosphere

Numeric atmospheric phase screens were created using the ARROYO library (Britton 2004) in diffractive-propagation mode. Screens were generated for 0.85, 1.65 and 2.20 microns on time steps of 0.5 msec (2000 Hz). A zenith-only, on-axis-only model was used, and the TMT standard C_n^2 profile was used (Table 17) with an integrated C_n^2 of 3.535e-13 $m^{1/3}$ (r_0 of 0.15 cm at a wavelength of 0.5 μm).

Table 17: TMT C_n^2 Profile used in generation of atmospheric phase screens.

h(km)	$\%C_n^2$
0	.646
1.8	.078
3.3	.119
5.8	.035
7.4	.025
13.1	.082
15.8	.015

For wind profile, TMT reports no standard (10/27/05 AO coordination meeting), but Richard Claire is using the values in Table 18 and we have opted the same model.

Table 18: Wind profile used for generation of atmospheric phase screens.

Layer	Height (km)	Gamma	Wind (m/s)
1	0	.6523	5
2	2.577	.1723	13
3	5.155	0.0551	20
4	7.732	0.0248	30
5	12.887	0.0736	20
6	15.464	0.0219	10

Fitting these numbers to the Gaussian model (Hardy 1998) supported by ARROYO, results in:

1. Ground velocity (v_G): 5 m/s
2. Tropopause velocity (v_T): 26 m/s
3. Height of Tropopause (H_T): 8900 m
4. Thickness of Tropopause layer (L_T): 5200 m

The atmospheric layers resulting from this modeling were then projected onto a square telescope pupil, 31×31 m, or 1025×1025 pixels, in size. A typical run was 4 seconds in length.

9.2.4 Pre-DSS AO

The Pre-SSS AO system is arguably one of the most computationally intensive algorithms. We have multiple options for the algorithm depending on the fidelity of modeling needed. These algorithms are discussed in more detail below.

9.2.4.1 Shack-Hartman

The Shack-Hartmann simulator used in this study is a adaptation of the code used in the Gemini ExAOC Conceptual Design project.

The flow of the AO simulation follows the flow of the actual optical system. The first action in any time step is the application of the correction on the DM. The shape of the DM is generated from the command signal using a linear-superposition influence-function model. The specific influence function is that which was determined in the Altair project. The commands which are sent to the DM are determined using a leaky-integrator control law. Tip-tilt is projected out of the actuator commands and placed on a separate linear-phase tip-tilt mirror.

The WFS leg contains the spatial filter, a Shack-Hartmann WFS and a CCD. The spatial filter can be implemented either as an ideal low- pass filter on the phase or with an actual Fourier optics simulation of the propagation to and from the image plane where the field stop is. The latter takes more computation, and in closed loop is not substantially different from the ideal filter, so the ideal filter is normally used.

After the spatial filter, the residual phase is sent to the Shack- Hartmann sensor. The code produces integrated pixel intensities for one-fourth-nyquist sampled quadcells. Both photon and read noise are added based on photometry for the system.

Given this noisy CCD image, the data flow enters the AO control system proper. Given the WFS CCD image, the slopes are calculated with the standard Quadcell algorithm. These slopes are sent to the Fourier reconstructor algorithm. This algorithm was run in non-optimizing mode and with a generic reconstruction filter.

The performance of the simulator is verifiable by running the simulation in modes such as 'no noise', where the WFS is free of noise, and 'no aberration' where no atmosphere is present. This allows characterization of the error sources in the system.

9.2.4.2 Static Shack-Hartmann

In order to simulate the impact of static telescope aberrations (section 10) we choose to use a wave-optics simulation using the ARROYO library (Britton 2004). The AO simulation is monochromatic, but supports two different wavelengths: one for sensing and one for science. The telescope complex pupil is read in and the AO loop is closed and allowed to converge. The wavefront sensor is a Shack-Hartmann sensor in which each sub-image is formed by propagation of the appropriate portion of the wavefront from the pupil to the sensor. In this simulation each sub-image was formed on 32 × 32 pixels. A spatial filter is also simulated in the wavefront sensor focal plane as an image stop of the specified size. A least-squares reconstructor with 127 actuators across the pupil is used to derive the DM residual errors. The DM residual errors are then used in a proportional integral servo loop. The DM is simulated using pyramid influence functions.

9.2.4.3 Pyramid

The pyramid wave-front sensor measures the intensities which pass through the faces of the pyramid. In the case of a four-sided pyramid, four images of the pupil are formed on the wave-front sensor camera. To determine the phase of the wave-front, the pupil images must be processed as detailed by the flowchart in Figure 115. This includes normalizing the measured intensities and taking their square root to determine the amplitudes. Differencing the pupil images provides the Hilbert Transform of the frequency shifted phase. To retrieve the phase, the difference of the amplitudes of the sides must be taken and the left-hand side multiplied by $-i$ and the right-hand-side multiplied by $+i$. Because the core is split into different pieces by the faces of the pyramid, the center-of-mass of the core passing through each of the pyramid faces is shifted away from the apex of the pyramid. This shift must be corrected in order for the pyramid wave-front sensor to reproduce the phase correctly. As such the next step in the algorithm is to correct the shifted frequencies before inverse Fourier transforming back to arrive at the frequency across the pupil.

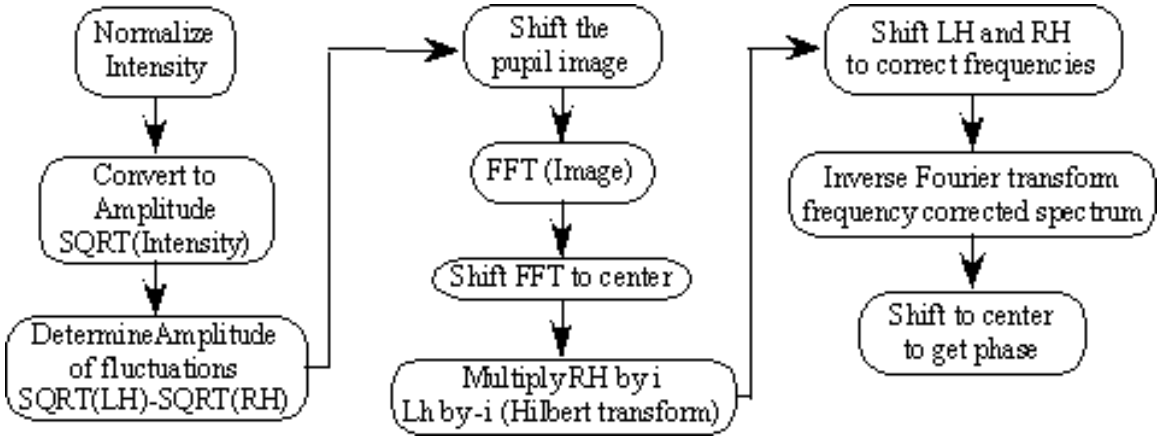

Figure 115: Flowchart of computations for the Pyramid's interferometric reconstruction of the phase.

9.2.4.4 Spatial Filter / Ideal AO

For cases in which the detailed behavior of the front AO system was unimportant, ie those cases where the post-DSS AO system does most of the wavefront correction, we implemented a simple high-pass spatial filter which simulates an AO system as having a perfect spatial response, but with a time delay. That is specifically a set of DM commands is calculated by applying a high pass filter to the atmospheric phase at time t. This set of DM commands is that applied some time later, $t + \Delta t$, in order to simulate the time delay of an AO system. This rather simple algorithm assumes perfect measurement noise, so it is important to have a realistic estimate of the appropriate time delay to use. The appropriate time delay was calculated using analytical equations to balance the measurement and bandwidth errors to minimize the RMS wavefront.

9.2.5 DSS

This section will describe the DSS component of the end-to-end simulation. A detailed description of the architecture and performance of the nuller, which is the architecture employed in the end-to-end simulation, is provided in Section 4.2.3.

The model implements a linear, two-stage nuller with the Lyot stop for this configuration displayed in Figure 55. The single stage shear is 3.12 meters, which corresponds to 50% transmittance at 3.04 λ/D for $\lambda = 1.65\mu m$. The inputs for each time step are the atmosphere amplitude and phase screens, which are converted to a complex amplitude field.

Three copies of the input field are made. One is the nominal field, centered on the origin of the pupil. The other two are identical to the nominal, except for the fact that one is shifted along the +x-axis by the value of the baseline and the other is shifted along the –x-axis by the value of the baseline. The null field is generated by combining the pupils with the correct set of weighting factors to generate destructive interference in the pupil overlap region. The bright field is generated by combining the pupils with a set of weighting factors to generate constructive interference in the pupil overlap region. Finally both of the fields are masked by the Lyot stop.

9.2.6 Post-DSS AO

This section will describe the implementation of the Post-DSS AO algorithm in the end-to-end simulation. A detailed explanation of the algorithm is provided in Section 6 of this report.

The Post-DSS AO algorithm simulates the homodyne interference of the first mode of the wavefront from the parent star with the nulled wavefront. The nulled wavefront is spatially filtered with a pinhole sized to the Nyquist frequency of the front-end AO actuator spacing. This interference of the spatially filtered wavefronts takes place in an image of the pupil. The phase error at each location in the pupil that corresponds to the location of an actuator on the back-end DM is calculated using the standard four-bin {A,B,C,D} algorithm. The map to apply to the DM is calculated from Y = D – B, which is a measure of the phase error. A user-specified influence function is used to interpolate the phase corrections to locations in the pupil that lie between DM actuator locations. We

used a pyramid influence function in these simulations. The phase correction is applied to the unfiltered nulled wavefront, which is then saved to disk.

The implementation has the ability to tune the gain and the ability to run the back-end DM update at a slower rate than the front-end AO system.

9.2.7 Science imager

Once the final wavefront corrections have been applied, the nulled output pupils are sent to the "science imager" routine to generate the final images and to compute the associated contrast metric. All the pupils are zero-padded to ensure the Nyquist sampling criterion is met at a wavelength of 1.65 μm – this results in zero-padded arrays of size somewhat less than 2048×2048 pixels. We then compute the Point-Spread Function (PSF) by taking the absolute square of the Fourier transform of the zero-padded complex pupil.

We choose as our contrast metric a measure of the statistical noise in the resultant PSF. The algorithm for determining this contrast is as follows. First, we median-filter the PSF with a square median filter of width ~4 λ/D. The filter width is chosen so that the PSF speckles are not completely filtered out, but some overall structure still is visible. This balances filtering away the speckles (which we don't want to do) and filtering away smooth trends due to the atmosphere (which we do). Next, the filtered PSF is subtracted from the original PSF to obtain a detrended, highpass-filtered PSF. At increasing radii, we then calculate the azimuthal standard deviation within annuli of width ½ λ/D. Studies suggest that five times this azimuthal standard deviation is a detection threshold. Note that this metric (and in fact any that plots contrast only as a function of radius) assumes azimuthal symmetry in the PSF; while this is not applicable far out in the field (e.g., near the borders of a rectangular "dark hole"), it is a reasonable approximation near the inner region of interest (around 3 λ/D).

9.2.8 Post processing

Using the time averaged complex pupil measured by the post-DSS AO WFS .a PSF estimate can be generated. Using post processing this image can be used to improve the contrast of the actual science image. In pervious simulations for the Gemini Planet Imager it was shown that a factor of 30 improvement in contrast could be achieved. This result took into account detector read noise as well as photon noise in the WFS.. As part of future work this should be re-studied for the PFI specific parameters.

9.3 End-to-end simulation results

Using the algorithm chain described above we have analyzed several representative cases (Table 19). Cases one and two correspond to target stars in the solar neighborhood – PSF science missions 1 and 3 (2.2.4 and 2.2.5). Cases 3 and 4 correspond to imaging T Tauri stars in the Taurus star-forming region, representing PFI science mission 2 (2.2.6). The baseline instrument parameters are described in section 9.1.2.

Table 19 Representative science cases. Flux is in units of photons/sec/subaperture at the WFS. The I/Z band flux is used for the front end WFS and the H band for the post-DSS WFS.

	I/Z band	H Band	Notes
Magnitude	4.2	3.6	Case 1
Flux	4059	794	G5 star @ 10pc
Magnitude	6.6	5.9	Case 2
Flux	451	88	G5 star @ 30pc
Magnitude	10.7	7.6	Case 3
Flux	10.3	19.2	“Bright” T Tauri star
Magnitude	14.2	9.3	Case 4
Flux	0.42	4.0	“Typical” T Tauri star

Figure 116 and Figure 117show the performance for case 1 and 2 in after 1.5 and 0.5 seconds of simulation respectively. In both cases the pre-DSS AO system was simulated using a Pyramid WFS running at 2000Hz and the post-DSS AO running at 1000Hz. It is clear from these simulations that final contrast is limited by servo lag error at 2000 Hz – the optimal frame rate would be in the 4000 Hz range, similar to that calculated analytically. As a result the contrast achieved case 1 and 2 is very similar. The 1 hr contrast was obtained by scaling the simulation results by the square-root of 1hr divided by the greater of the simulation duration and the atmospheric speckle lifetime (~1 second for TMT, Macintosh et al. 2005.). The requirements of contrast 10^{-8} is nearly met even without speckle suppression and easily met for methane-dominated objects that achieve significant speckle suppression.

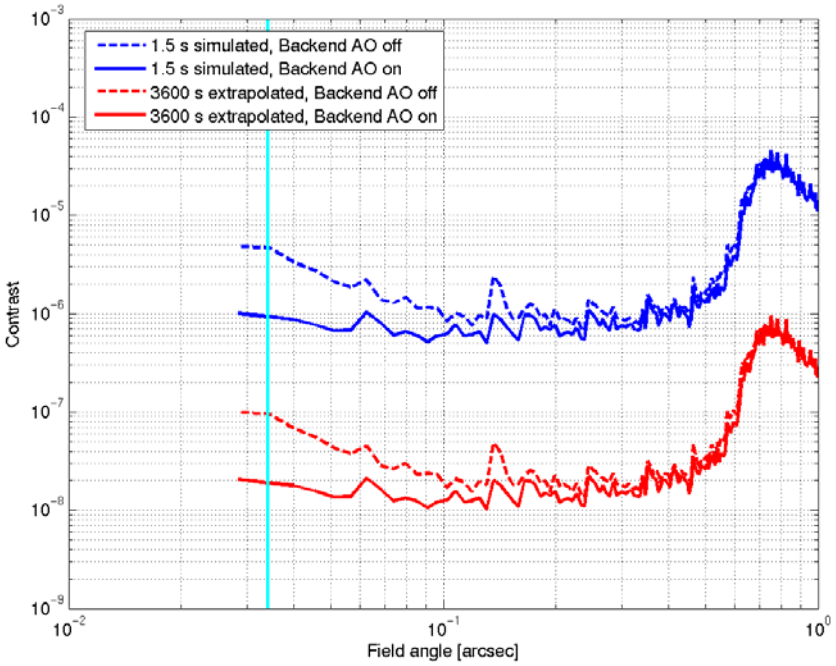

Figure 116: The contrast is shown as a function of field angle for case 1 in Table 19. The simulation was run for 1.5 seconds and the results scaled as the square root of time divided by 1.5 to obtain the contrast in 1 hr.

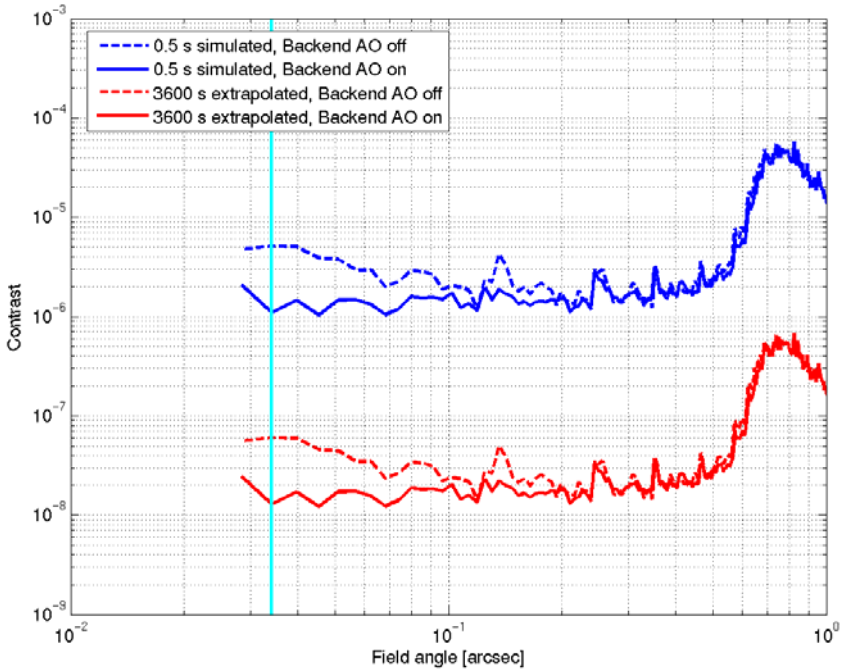

Figure 117: The contrast is shown as a function of field angle for case 2 in Table 19. The simulation was run for 0.5 seconds and the results scaled as the square root of time divided by 0.5 to obtain the contrast in 1 hr.

Figure 118 and Figure 119 show the performance for case 3 and 4 in Table 19 after four seconds of simulation. Due to computation and time limits, it was decided to approximate the pre-DSS AO system as a spatial filter with a time delay (section 9.2.4.4). The contrast in these case should be dominated by the performance of the post-AO WFS, as the post-AO WFS has two to ten times as many photons. The optimal frame rate of the post-DSS AO WFS was calculated via two methods, first analytically and secondly by running several short (0.1 sec) simulations at various frame rates. Both methods showed that the optimal frame rate was 500Hz for case3 and 125hz for case 4. In case 4 the AO system was run with 64 (instead of 128) actuators across the pupil, this results in better SNR and better contrast close into the star at the cost of a decrease in performance from ~0.3 to 0.6 arcseconds. In this particular science case this is a good trade as the search space is in close (< 0.3 arcseconds) to the central star. The 1 hr contrast was obtained by scaling the simulation results by the square-root of 1hr divided by simulation time (4.0 seconds). The requirements here are 10^{-6} and are again met, though the goal of 10^{-7} requires significant postprocessing speckle suppression, which may prove more challenging for young, hot planets. In both of these cases (3 and 4) the final contrast is limited by the atmosphere not telescope.

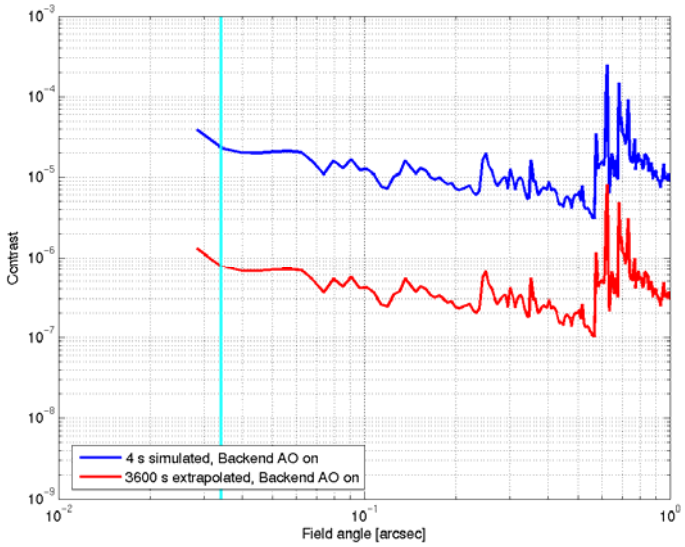

Figure 118: The contrast is shown as a function of field angle for case 3 (Bright T Tauri star) in Table 19. The simulation was run for 4.0 seconds and the results scaled as the square root of time divided by 4.0 to obtain the contrast in 1 hr.

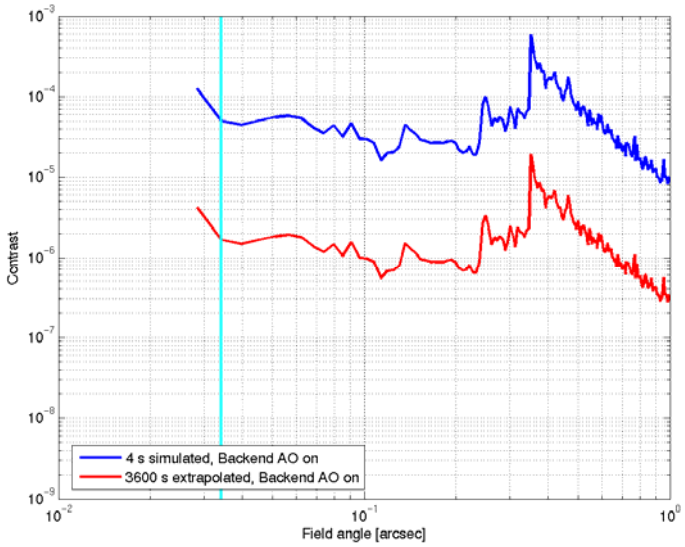

Figure 119: The contrast is shown as a function of field angle for case 4 (Dim T Tauri star) in Table 19. The simulation was run for 4.0 seconds and the results scaled as the square root of time divided by 4.0 to obtain the contrast in 1 hr.

9.4 Scintillation and Chromaticity simulations

9.4.1 Scintillation

The effects of scintillation were studied by Fresnel propagating an electromagnetic wave at the science wavelength through the standard seven-layer TMT atmospheric model. This model is shown in Table 17. The effective r_o at 1.65 μm is 0.63 m. As the electromagnetic wave propagates through the turbulence, the phase variations turn into amplitude variations. This occurs over a distance know as the Talbot length, which is dependent on the square of the spatial frequency. The Talbot length is therefore shorter for higher spatial frequencies; however, the energy contained in the higher spatial frequencies is less due to the frequency dependent power spectrum for atmospheric turbulence which leads to a fairly flat contrast ratio.

After propagating the electromagnetic wave through the seven-layer atmospheric model, the resulting pupil is scintillated. By placing an apodization mask in that plane and performing a Fourier transform, the field in the Fourier plane is determined. The instantaneous contrast ratio can be determined by squaring the field to get the intensity and looking at the resulting lineout. For these tests, a prolate spheroid mask was chosen for the apodizer as shown in Figure 120. This apodizer allows for contrast ratios of 10^{-11} to be achieved along the horizontal axis, as shown by the solid black line in Figure 123.

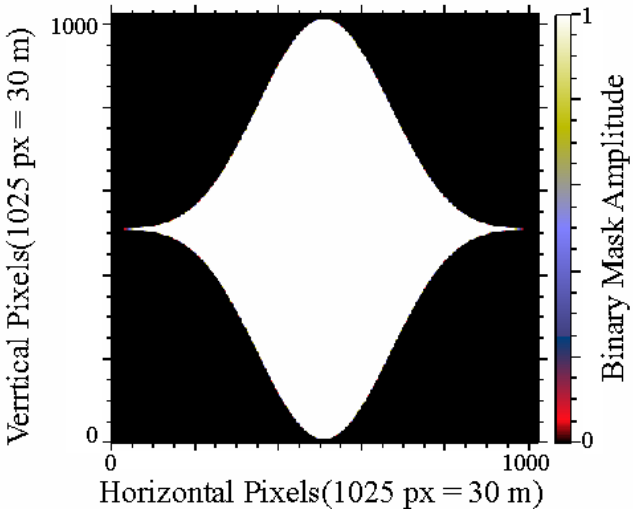

Figure 120: Binary mask used for the apodizer.

Two separate codes were used to generate the atmospheres. An IDL program was used to generate a Von Karman turbulence spectrum with an outer scale of $L=2\pi/\kappa_o= 25$ m and ARROYO was used to generate a Kolmogorov turbulence spectrum. Both codes used an effective r_o of 0.63 m at 1.65 μm. The 1.65 μm light, after Fresnel propagation through the screens, had an intensity in the pupil as shown in Figure 121 for the ARROYO code.

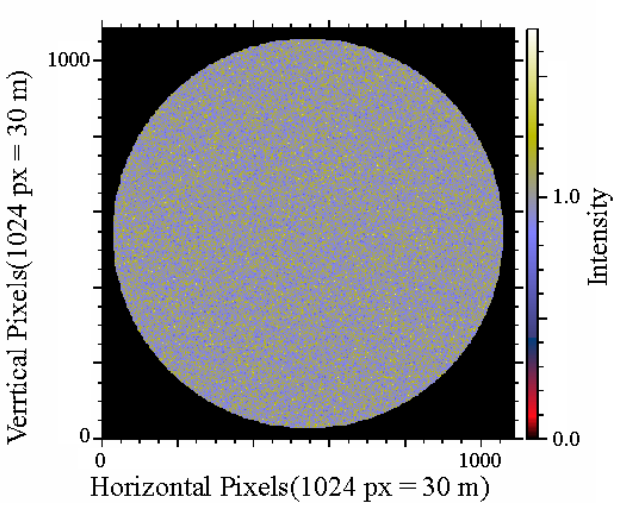

Figure 121: Intensity scintillation at 1650 nm from the ARROYO simulation.

A complex wave consisting of the field corresponding to the intensity pattern in Figure 121, a zero phase component and the apodization mask shown in Figure 120 was then used to determine the point-spread-function and from that the instantaneous contrast ratio. The resultant point-spread-function is shown below in Figure 122.

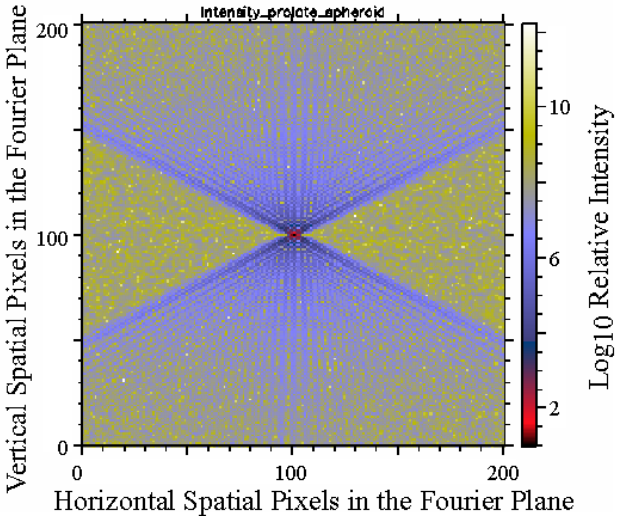

Figure 122: Far-field image of scintillated pupil aperture with the apodizer above.

A horizontal lineout through the center of the Far-field image above shows that the instantaneous contrast ratio due to scintillation alone is ~10^{-7}, dashed line in Figure 123. This contrast ratio was obtained for both the ARROYO code using a Kolmogorov spectrum and the IDL code using a Von Karman spectrum. The contrast ratio is fairly flat over the 100 λ/D plotted (2 pixels ~ λ/D).

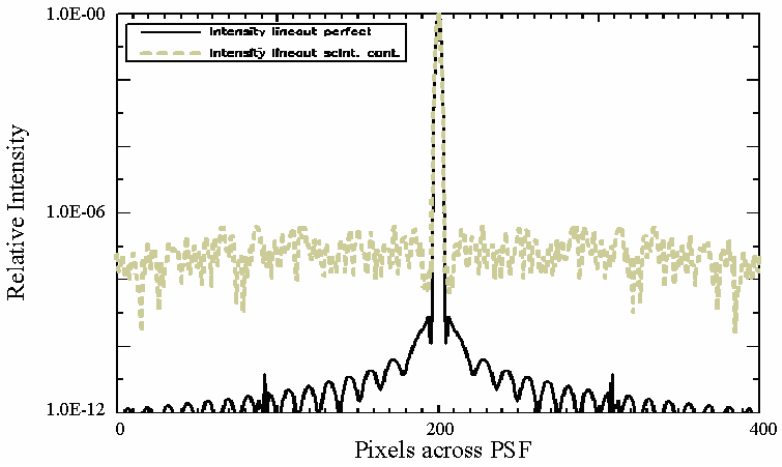

Figure 123: Relative intensity lineout of Arroyo generated intensity at the pupil.

9.4.2 Chromaticity

The chromatic effects associated with Fresnel propagating two wavelengths through the atmosphere were also studied. This is of interest in the case where the science wavelength and the wave-front sensor operate at different wavelengths. Chromatic effects will limit the contrast ratio achievable. This effect was again studied by Fresnel propagating two electromagnetic waves, one at the science wavelength of 1.65 μm and another at the wave-front sensor wavelength of 0.85 μm, through the standard seven layer TMT atmospheric model. Due to the wave-length dependence of the Fried parameter, the shorter wavelength electromagnetic wave passes through stronger turbulence and is hence more scintillated than the longer wavelength electromagnetic wave. This difference in the scintillation between the two wavelengths will result in differences in the phase at the pupil for the different electromagnetic waves.

The IDL code was used to generate pupil fields at two wavelengths and therefore allowed an evaluation of the effects of chromaticity and also allowed a comparison between the relative importance in chromatic phase and scintillation. The two fields generated had their phases wrapped modulo 2π and so before the difference in phase could be determined the phases were unwrapped using a path-following algorithm. To study the effects of chromatically induced contrast ratio alone, a uniform field was generated with the apodizer mask shown in Figure 120 applied and the difference between the phases of the two wavelengths was used for the phase. Fourier transforming the field and finding the intensity allowed this effect to be examined. The instantaneous contrast ratio due to chromatic Fresnel propagation is shown in Figure 124 as the solid black line. In this case the instantaneous contrast ratio due to chromatic Fresnel propagation is below 10^{-9} as far out as 30 λ/D.

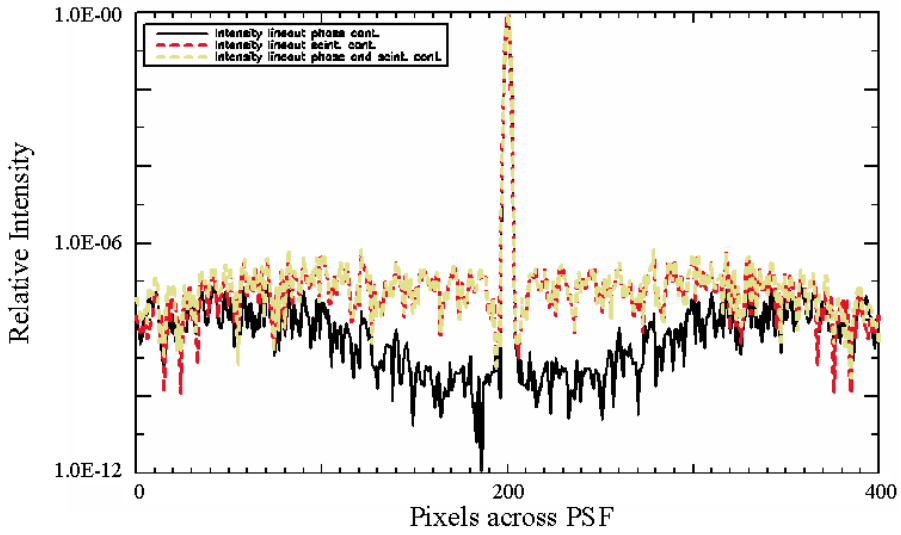

Figure 124: Contrast limitations due to scintillation and chromatic phase differences.

Figure 124 also has the total instantaneous contrast ratio due to both chromatic Fresnel propagation, dashed yellow line, and due to scintillation, dashed red line. This plot shows that scintillation is more important than chromatic Fresnel propagation for this particular case out to~60 λ/D. The scintillation contrast limit of 10^{-7} agrees with the contrast above from the Arroyo simulations which had a similar variance in the intensity at the pupil.

By running the codes at different wavelengths, an expression for the RMS error between the wave-front sensor wavelength and the science camera wave-length can be derived as a function of the wavelength separation. The results (see **Error! Reference source not found.**) indicate that there is a linear change in the RMS error with wavelength separation, but that it saturates at large wavelength separation, ~ 400 nm.

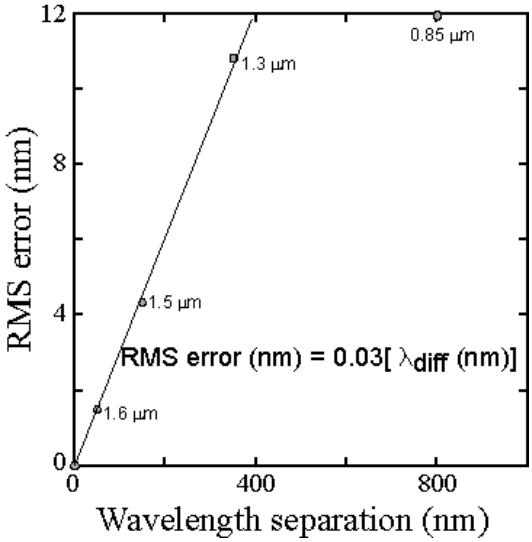

Figure 125: RMS error (nm) between the science camera and the wave-front sensor as the relative wavelength separation is varied.

10 Optical interactions with telescope

The goal of this section is to determine how key TMT design decisions impact the contrast achievable with the TMT. The current Science Requirements Document (TMT TMT.PSC.DRD.05.001.REL15) and Planet Formation Instrument Feasibility Study RFP (TMT TMT.INS.CON.05.005.REL01) specify that the telescope should ultimately be able to achieve contrast levels of 10^{-8} (requirement) or 10^{-9} (goal) from 3λ/D (34 mas) to 88λ/D (1 arcsecond). Similarly, the PFI OCDD gives a science-motivated goal of contrast (on bright stars) of 10^{-8} from 3λ//D outward, and 2×10^{-9} at 10 λ//D. As discussed in Section 9.1 static error sources can set a contrast floor no matter how long an exposure is taken. This is a key point in this section; our goal it to make sure we design and build a telescope that is ultimately able to reach these contrast levels independent of the capabilities of the first generation of PFI.

If the telescope aberrations are stable then the speckles from the telescope can be subtracted out (to the photon noise level). In practice, this type of subtraction has achieved a factor of ~5-10 improvement in contrast (Marois 2004, Metchev 2005). The most successful technique to date uses the parallactic rotation to distinguish planets from artifact speckles. Multi-wavelength imaging can also be used as discussed in Section 7. Both techniques are most effective at large field angles – performance at 3–5 λ/D will be almost completely determined by the telescope.

In this analysis we assume that in the future we will be able to achieve a factor of 10 using multiwavelength speckle suppression and subtraction techniques. The telescope error budget for high-contrast imaging must be smaller than the required final imaging performance. As a conservative value the telescope error budget should be no greater then the sum of all other AO error terms (See Section 9.1). This increases the specification on contrast by a factor of 2. If we include the factor of 10 improvement from speckle suppression then the total contrast error budget for the telescope is 5×10^{-8} (requirement) or 5×10^{-9} (goal).

In the following sections we investigate static telescope effects. These effects include M1 and M3 phase errors, M1 segment reflectivity variations as well as the segment gaps, secondary obscuration and support. To start with we investigate each error term individually. We then combine all error terms together to estimate a final contrast level. The general approach we take is to model all errors using wave-optics based simulations as described in Section 9.2 .

10.1 M1 static errors

In this section we consider the effects of static errors on M1 and how these limit achievable contrast levels. In each of the following sub-sections each error term is considered independent of the rest. In Section 10.1.4 all error terms are combined to provide a limiting contrast level from the primary mirror.

We start by modeling the pupil as described in Section 9.2.2. After that the pupils are run through a wave-optics simulation of an AO system as described in Section 9.2.4.2. The result is the AO corrected complex pupil from which we can calculate the residual RMS wavefront error and apply a diffraction suppression system to estimate the achievable contrast. All M1 analysis uses the baseline diffraction suppression system (see Section 9.2.5 for a description) so that a fair comparison can be made for all error terms.

10.1.1 Segment gaps and secondary obscuration

This has been analyzed in section 4.2.3. Figure 42 shows an image of the TMT pupil with the segment gaps and secondary obscuration and support we used. The resulting contrast level from these effects, shown in Figure 134, is 2.3×10^{-10}. This is not a significant limitation in the achievable contrast.

10.1.2 Phase Errors

10.1.2.1 Segment aberrations

In this section we model the expected segment aberrations to understand how they impact the performance on the PFI.

10.1.2.1.1 Wavefront Errors

The input phase errors are those specified by the TMT Specification for Finished Primary Mirror Segments Document (TMT.OPT.SPE.05.002). Gary Chanan has taken the specifications and translated them into equivalent Zernike aberrations; the Zernike mode coefficients are listed in column 5 of Table 20. The wavefront errors were generated by taking the specified coefficient values for each segment, then applying a random sign for each coefficient on each segment. As a result, by design each segment will just meet the specifications in the RFP.

The current plan is for each segment to have a 27 point warping harness (WH) with 18 actuators per segment. The warping harnesses should help to improve the lower-order segment aberrations. According to analysis performed by Hytec (TMT.OPT.PRE.05.077.REL01) the WH should reduce the segment aberrations by the factors given in column 2 of Table 20. These predicted reductions in wavefront error are 5 to 15 times greater then those currently achieved at Keck (column 3 of Table 20). It is currently unclear what is limiting the Keck WH performance. TMT has funded an effort to in collaboration with Keck investigate this critical issue.

It will be necessary to measure the segment Zernikes and set the warping harness to correct for the phase errors. There will be some error in this process. The Alignment and Phasing System (APS) will be used to measure the segment aberrations and set the warping harness. Analysis of data from the Keck Phasing Camera System shows that the 2nd order segment Zernikes are measured to an RMS of 7.8 nm (wavefront) after 6 minutes of integration. If we assume these measurement errors are limited by measurement of subimage motion through the atmosphere then they can be scaled to higher order Zernike modes. The calculated measurement noise at Keck is shown in column 4 of Table 20. The measurement noise should scale as the segment diameter to the power of 5/6, so the Keck measurement values can be scaled to TMT.

Given an assumption about the WH reduction factors (none, Keck or TMT (Ideal)) and the measurement noise (none, Keck or TMT) the segment aberrations can be calculated by combining the wavefront errors from each term in quadature. Columns 5 through 9 of Table 20 show the segment aberrations for several combinations of assumptions. As a baseline for analysis we have chosen the TMT WH reduction factors combined with the Keck measurement errors (column 7 Table 20). We have selected the Keck measurement errors in part because while the APS measurement errors might be smaller at Keck there are likely other error terms that will contribute to errors in setting the WH.

Table 20: Segment warping harness parameters used to simulate the TMT segment aberrations. The details of each column are described in the text.

Zernike index (n,m)	TMT WH Reduction Factors (Theory)	Keck WH Reduction Factors (Experiment)	Keck Meas. noise (nm WFE)	No Warping Harness (nm WFE)	TMT WH (nm WFE)	TMT WH with Keck meas. error (nm WFE)	Keck WH with Keck meas. error (nm WFE)	TMT WH with TMT meas. error (nm WFE)
2,0	17.2	4	7.8	19	1.1	7.9	9.1	4.9
2,-2	50.8	4	7.8	20	0.4	7.8	9.3	4.8
2,2	50.8	4	7.8	20	0.4	7.8	9.3	4.8
3,-1	5.1	1	3.9	8	1.6	4.2	8	2.9
3,1	5.1	1	3.9	8	1.6	4.2	8	2.9
3,-3	12.2	1	3.9	16	1.3	4.1	16	2.7
3,3	15.2	1	3.9	0	0	3.9	0	2.4
4,0	1	1	2.6	2	2	3.3	2	2.6
4,-2	3.1	1	2.6	2	0.6	2.7	2	1.7
4,2	3.1	1	2.6	2	0.6	2.7	2	1.7
4,-4	6.4	1	2.6	2	0.3	2.6	2	1.6
4,4	6.4	1	2.6	2	0.3	2.6	2	1.6

Figure 126 shows the phase map used in the simulations that corresponds to the wavefront errors from column 7 of Table 20. The left hand side is the full pupil and the right hand is a zoomed in view. The RMS wavefront error is 17.3 nm and the peak-to-valley wavefront error is 242 nm. The large peak-to-valley errors occur primarily at the segment boundaries. These large discontinuities are a direct result of the relatively large low spatially frequencies. In this analysis there are approximately 1024 pixels across the aperture or 3 cm/pixel. A few cases were run with 1.5 cm/pixel and the results were similar, so we opted for the smaller pupils and faster run times.

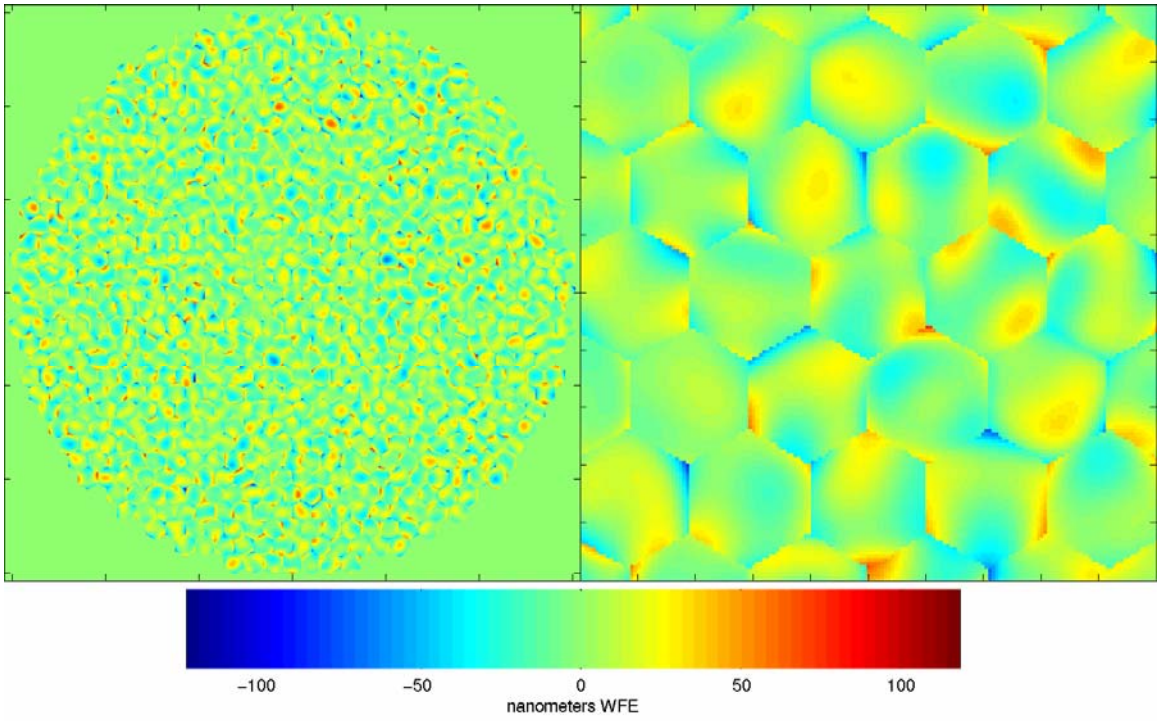

Figure 126: TMT pupil using the expected segment aberrations from column 7 of Table 20. The units are nanometers of wavefront error. The RMS wavefront error is 17.3 nm and the peak-to-valley is 242 nm. The right hand side shows a 5 m by 5 m zoomed view of the pupil. There are clearly large edge discontinues that cause the 242 nm of peak-to-valley of error.

Table 21: RMS and Peak-Valley wavefront errors in nanometers for segment aberrations both before and after AO correction with 127 actuators across the pupil. The AO system is able to reduce the RMS wavefront error, but has little effect on the P-V errors which are dominated by edge discontinuities from the segment aberrations.

		RMS	P-V
Without warping harnesses	Pre-AO	39.6	352
	Post-AO	21.1	354
WH: Keck-like Zernike reduction	Pre-AO	25.5	272
	Post-AO	14.2	257
Ideal WH: Keck measurement errors	Pre-AO	17.3	242
	Post-AO	9.1	199
Ideal WH: TMT measurement errors	Pre-AO	11	153
	Post-AO	5.7	126
Perfect WH Zernike reduction	Pre-AO	3.7	38
	Post-AO	1.7	36

The AO pupils corresponding to the five cases in Table 21 were run though a wave-optics simulation of the AO system. Figure 127 shows the AO corrected phase map that corresponds to the wavefront errors from column 7 of Table 20. The left hand side is the full pupil and the right hand is a 5 m by 5m zoomed-in view. The AO-corrected RMS wavefront error is 9.1 nm and the peak-to-valley wavefront error is 199 nm. The AO system has reduced the RMS wavefront error by a factor of 1.9, but the segment to segment edge discontinuities are not significantly changed from the input wavefront values. Table 21 shows a summary of the before and after AO wavefront for all five cases.

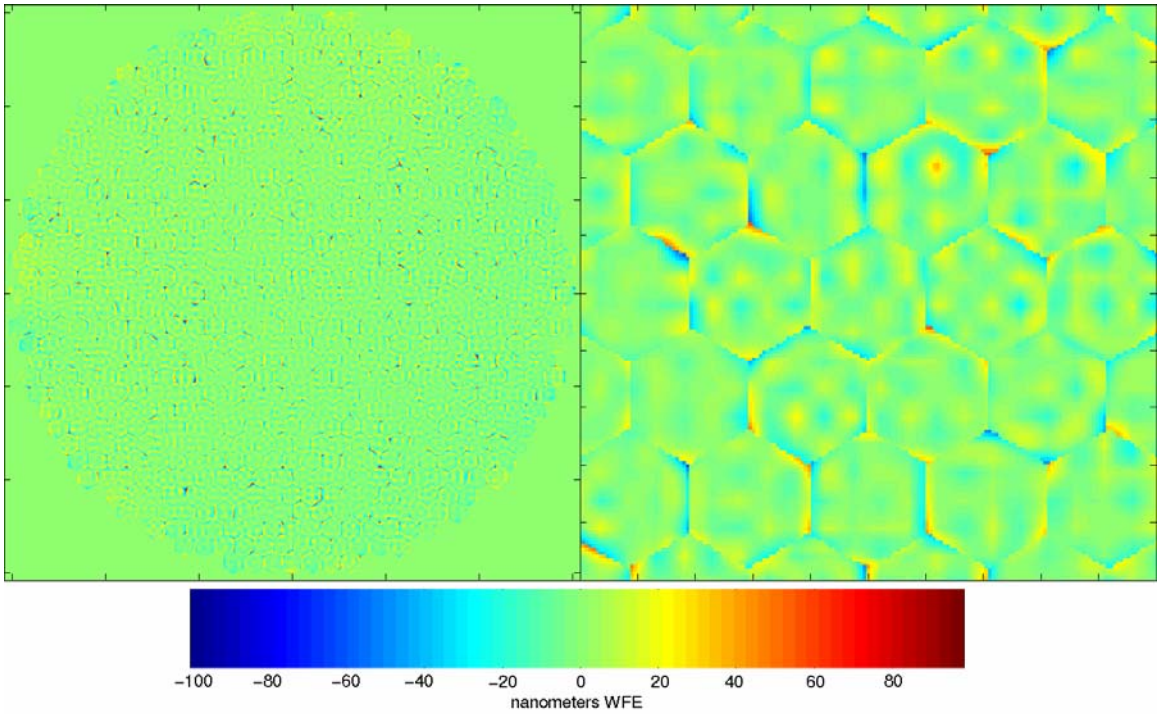

Figure 127: The TMT pupil with segment aberrations from column 7 of Table 20, corrected with 127 actuators across the pupil. The units are nanometers of wavefront error. The RMS wavefront error is 9.1 nm and the peak-to-valley is 199 nm. The right hand side shows a zoomed-in view of the pupil. The AO system has significantly reduced the wavefront error interior to a segment, but the large edge discontinues are not changed significantly from the non-AO corrected wavefront (Figure 126).

10.1.2.1.2 Achievable contrast levels

The AO corrected pupils were then run through the baseline diffraction suppression system. Figure 128 shows the contrast as a function of angular distance from the star for the five cases described in Table 20. The different cases in Figure 128 vary in contrast by over a factor of 100. It is clear that different performance levels of the WH will have a significant impact on the achievable contrast. In the baseline case using the TMT predicted WH reduction factors and the Keck measurement errors, a contrast value of 2.0×10^{-7} is reached at 3λ/D with an average value of 4.5×10^{-8} from 3 to 10 λ/D. We show in Section 10.1.4 that this is the limiting factor in achievable contrast. As a result it is critical to understand how well the WH will perform and why they don't perform well at Keck. Specific recommendations for following work are specified in Section 10.3 .

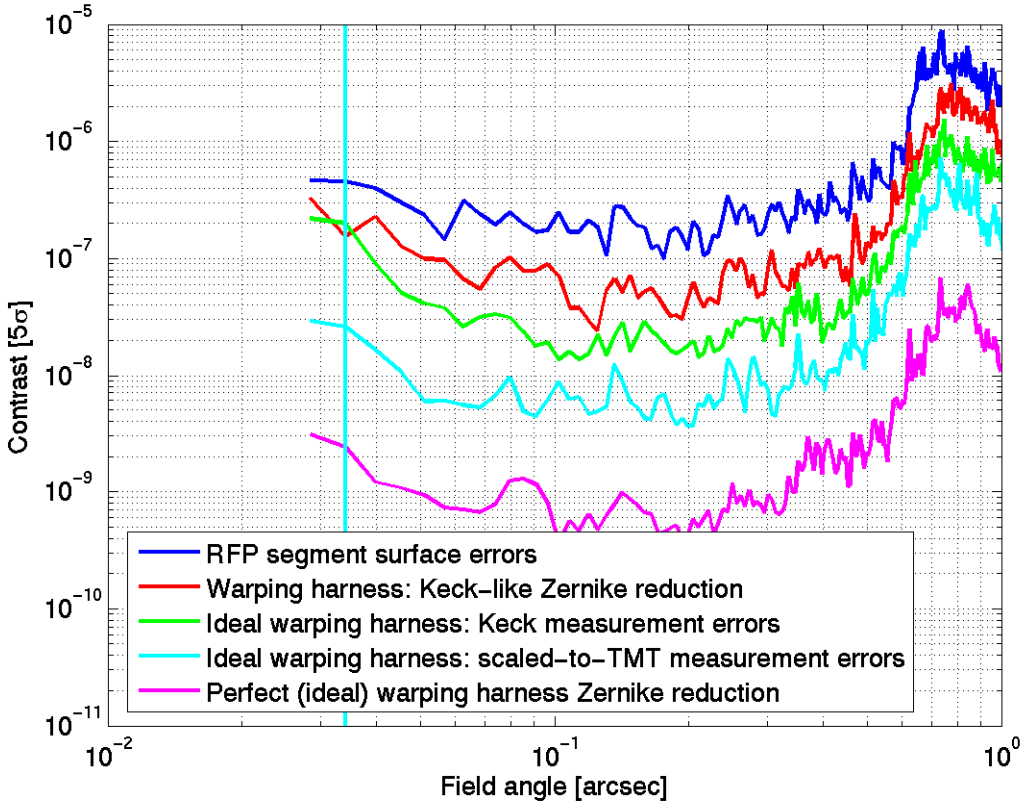

Figure 128: The contrast is shown as a function of field angle. From top to bottom: The dark blue curve is the contrast using the RFP specification for the mirror segments. The red curve shows the result of WH that work as well as those on Keck. The green curve is the contrast from TMT WH working at their theoretical efficacy, with Keck PCS measurement noise. The cyan curve is also from TMT WH at their theoretical efficacy, but with the Keck PCS measurement noise scaled to match the size of TMT's segments. The magenta curve is the contrast resulting from WH performing at their theoretical efficacy, without measurement noise.

10.1.2.2 Whiffletree print through

The primary mirror segments are supported by a whiffle tree structure. Gravity sag and other forces result in a residual wavefront error from this support system. A finite element analysis of the segment print though was performed by Hytec (TMT.OPT.PRE.05.025.REL01) and the optical implications analyzed by Myung Cho (TMT.OPT.PRE.05.034.DRF03). The input data was provided on a uniform square sampling of 2.34 mm for three cases: Axial, Lateral-y and Lateral-x forces.

10.1.2.2.1 Wavefront Errors

Following discussions with the project scientist we combined the data as follows to make up a wavefront error as a function of telescope zenith angle. We assume the segments will be polished so that the axial error will be zero at Zenith and increase to a maximum at the horizon. We further assume the wavefront errors will scale as the cosine of the Zenith angle. The Lateral-y error is also assumed to be zero at the Zenith and scale as cosine of the Zenith angle reaching a maximum at the Horizon. We assume the lateral-x error is zero

as there is no component of gravity in this direction when the segments are installed in the telescope.

Given the previous assumptions we can now calculate the whiffletree print through as a function of Zenith angle. The RMS error is zero at the Zenith and increases to a maximum of 21 nm at the Horizon. In this analysis we will assume a Zenith angle of 45 degrees which has an RMS wavefront error of 15nm. This wavefront error is replicated for all 738 TMT segments to form a TMT pupil. The left panel of Figure 129 shows the wavefront error for a portion of the TMT pupil.

The right panel of Figure 129 shows the AO corrected wavefront, for which the RMS wavefront error is 13 nm. As expected, the RMS wavefront error did not significantly decrease because most of the phase errors are outside of the controllable bandpass of the 127 actuator (across the pupil) AO system.

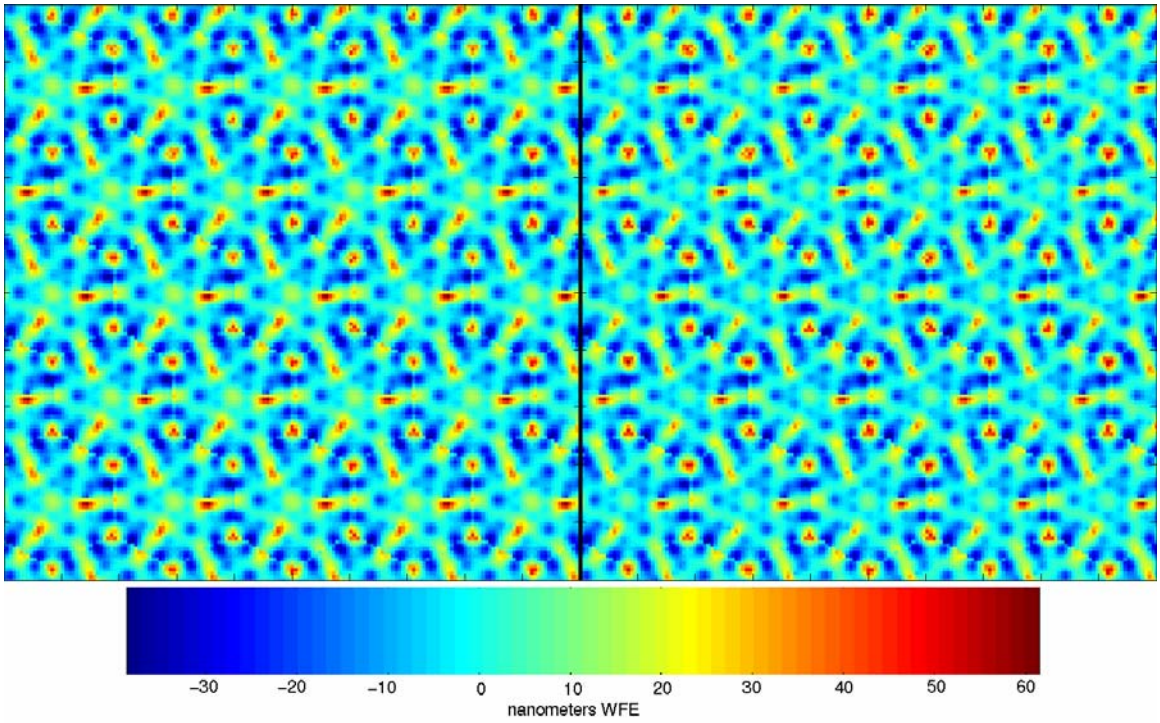

Figure 129: A zoomed in image of the TMT pupil with whiffletree print through. The left side is before AO correction (14 nm RMS) and the right side after AO correction (13 nm RMS). As expected the AO system has done little to reduce these high-spatial frequency errors.

10.1.2.2.2 Achievable contrast levels

The print through error is at a very high spatial frequency (significantly higher then 64 λ/D) and has a relatively small amplitude so as a result one would not expect it to contribute significantly to contrast inside the AO dark hole. The AO corrected pupils were run through the diffraction suppression, and Figure 130 shows the contrast as a function of angular distance from the star. A contrast value of 1.0×10^{-8} is reached at 3λ/D, and an

average value of 2.9×10^{-9} is reached from 3 to 10 λ/D. As expected, these are relatively small values and do not drive the error budget. The same analysis was also performed using a sampling of 1.5 cm/pixel (or 2048 pixels across the pupil) and similar results were obtained; for consistency we present the 3 cm/pixel results.

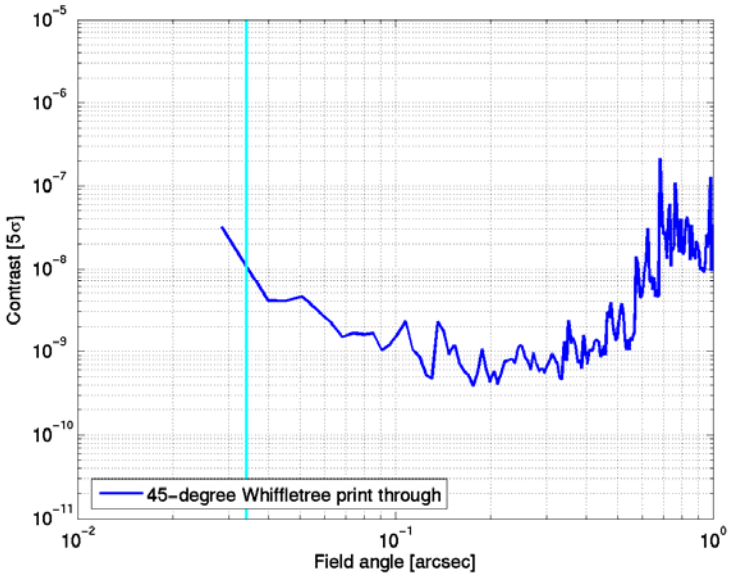

Figure 130: The contrast is shown as a function of field angle for whiffletree print through errors for the telescope at a Zenith angle of 45 Degrees.

10.1.2.3 Alignment errors

In this section we consider the effects from errors in the piston, tip, and tilt degrees of freedom of each segment. It is assumed that these errors are dominated by the ability of the APS system to measure and align the primary mirror segments.

10.1.2.3.1 Wavefront errors

We assume that the segment piston errors will be 12 nm RMS wavefront. This value is consistent with the performance of the Keck PCS. On Keck the random tip/tilt alignment error between segments is dominated by low spatial frequencies with the AO system will easily correct. However, random tip/tilt errors between segments lead to edge discontinuities which the AO system can not correct. At Keck the RMS edge discontinuity from tip/tilt is 48nm wavefront which corresponds to 17 nm RMS wavefront error of random tip/tilt errors. If we assume the Keck values are limited by measurement of subimage motion through the atmosphere then they should scale as the segment diameter to the power of 5/6 scaling the Keck segment results (1.8 m) to the TMT segment size (1.2 m) results in a predicted value of 11 nm. We assume some improvement over the current Keck performance will be achieved. The level of potential improvement is being modeled as part of the CoDR for APS. We have used a value if 6 nm RMS wavefront for each of the tip and tilt terms, this is modeled as normally distributed tip/tilts between the segments Table 22 shows the RMS surface errors for the three alignment cases: piston only, tip/tilt

only and piston and tip/tilt combined for both before and after AO correction. As above the generated pupils were run through a wave-optics simulation of an AO system with 127 actuators across the pupil. In the combined case the RMS wavefront error was reduced by a factor of 2.6 from 15.2 nm to 5.8 nm.

Table 22: RMS and Peak-Valley wavefront errors in nanometers for segment alignment errors both before and after AO correction with 127 actuators across the pupil. The AO system is able to reduce the RMS wavefront error by a factor of 2.7, but has little effect on the P-V errors which are dominated by edge discontinuities from the misalignment errors.

		RMS (nm Wavefront)	P-V (nm Wavefront)
Segment piston	Pre-AO	12.7	80
	Post-AO	4.5	77
Segment tip/tilt	Pre-AO	8.3	97
	Post-AO	3.8	75
Combined segment alignment error	Pre-AO	15.2	132
	Post-AO	5.8	107

10.1.2.3.2 Achievable contrast levels

The AO corrected pupils were then run through the baseline diffraction suppression system. Figure 131 shows the contrast as a function of angular distance from the star for the three cases described in Table 22. In the combined case of piston and tip/tilt errors a contrast value of 2.3×10^{-8} is reached at $3\lambda/D$ and an average value of 1.3×10^{-8} from 3 to 10 λ/D. The piston and tip/tilt contrast limits are approximately equal. If the tip/tilt errors are not improved from the current Keck values the contrast will increase by about a factor of 4 as contrast is proportional to the phase error squared. Further study is needed to insure that APS can meet the required tip/tilt accuracy. If APS can not meet the required accuracy then an alternative would be to study the benefit from feeding the segment tip/tilt measured from PFI back to the primary mirror control system; PFI's interferometric wavefront sensors with ~4 samples per segment should be highly capable of reconstructing segment piston/tip/tilt.

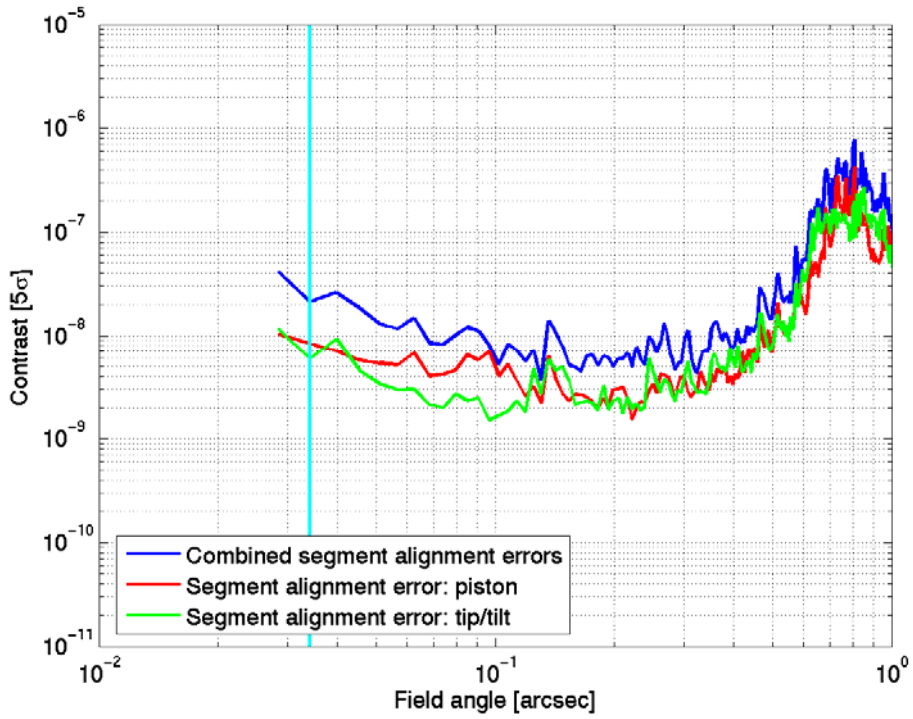

Figure 131: The contrast is shown as a function of field angle. The green curve (bottom) is the contrast from segment piston errors. The red curve (middle) is the contrast from tip/tilt alignment errors. The blue curve (top) is the contrast from both piston and tip/tilt alignment errors.

10.1.2.4 Combination of all phase error terms

In this section we investigate the contrast limits placed by all of the phase errors described in this section. We generated a TMT pupil that includes segment aberrations, whiffletree print through and segment alignment errors. The initial RMS wavefront error is 27 nm and is reduced to 17 nm after AO correction.

The AO-corrected pupils were then run through the diffraction suppression system. Figure 132 shows the contrast as a function of angular distance from the star for all phase errors and each of the individual error terms. Phase errors limit the contrast value to 1.4×10^{-7} at $3\lambda/D$ and an average value of 5.6×10^{-8} from 3 to 10 λ/D. The contrast is dominated by the segment aberrations, highlighting the importance of gaining an increased understanding of how well the warping harnesses will perform.

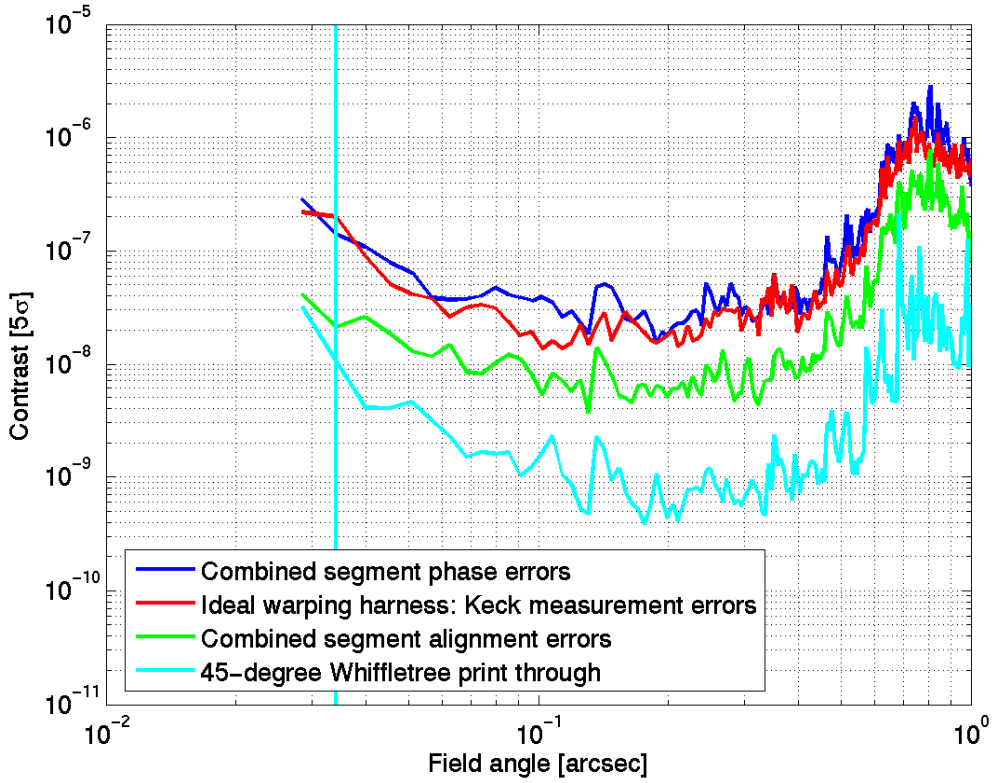

Figure 132: The contrast is shown as a function of field angle. From bottom to top: the light blue curve is the contrast from whiffletree print through. The green curve is the contrast from segment alignment errors. The red curve is the contrast from segment aberrations. The blue curve shows the contrast from combining all of the error terms. The contrast in the combined case is dominated by affects from the segment aberrations.

10.1.3 Amplitude errors

In this section we consider the effects of segment reflectivity variations on M1. At Keck the dominant effect is segment-to-segment variations in reflectivity, so we have concentrated our analysis in this area. We have analyzed two cases: the first has a mean segment reflectivity of 99% with a 1% variation and the second has a mean of 95% with a 5% variation. The 1% variation case is consistent with the specification in the SRD for reflectivities better then 99% at wavelengths longer then 1.5 microns.

Complex pupils where generated with the specified segment reflectivity variations and zero phase errors. After running through an AO system the AO system added a very small amount of wavefront error, 0.02 nm. The AO corrected pupils were then run through the diffraction suppression system. Figure 133 shows the contrast as a function of radial distance from the star for both the 1% and 5% reflectivity variation cases. The 1% variation case produced a contrast value to 1.4×10^{-7} at 3λ/D and an average value of 1.2×10^{-7} from 3 to 10 λ/D. When the amplitude variations are small the contrast should be proportional to the amplitude error squared. This is indeed true: the 5% reflectivity variation case has a contrast value that is, on average, 26 times worse then the 1% variation case.

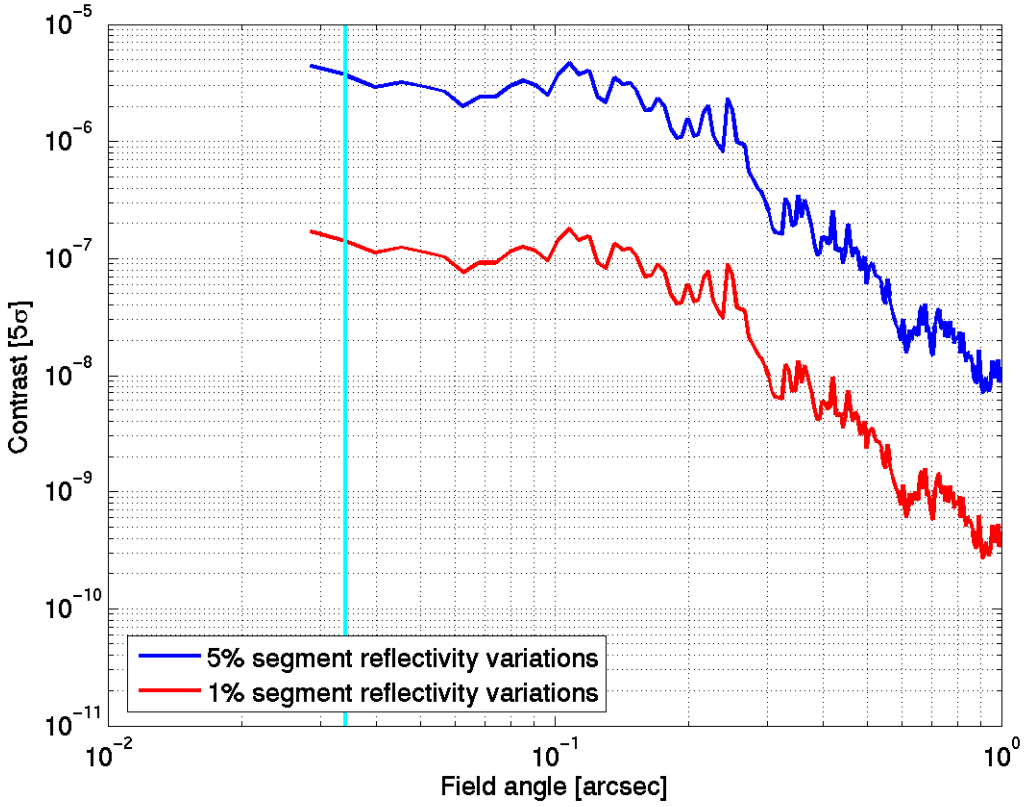

Figure 133: The contrast is shown as a function of field angle. The red curve (bottom) is the contrast from 1% segment reflectivity variations and the blue curve (top) is for 5% segment reflectivity variations. The contrast scales as the amplitude variation squared. Even with 1% segment reflectivity variations (the SRD specification) the contrast is significantly impacted.

If the segment-to-segment reflectivity variations are 1% then this contrast error term is approximately equal to that from the phase errors. If the segment reflectivity variations increase this error term will quickly dominate the achievable contrast. There are several techniques that can be used to mitigate amplitude variations. The ideal solution is a device that can modulate the amplitude of the pupil; however, there are currently no devices that can do this well in broadband unpolarized light. A second solution is to use a single DM to carve out a half-dark hole in the image plane. It has been shown (Trauger et al. 2004) that, in monochromatic light, by use of the correct control algorithm a DM can be used to correct for intensity errors if the contrast is allowed to increase in half of the image plane. A third solution is to install another DM in the system, but not at a pupil, to allow control of phase and amplitude errors as is described in Section 11.3 .

10.1.4 Combination of all M1 error terms

In this section we analyze the contrast limits placed by all of the M1 error terms. We have generated a TMT pupil that includes all of the effects discussed in this Section: phase errors (Section 10.1.2.4), reflectivity (Section 10.1.3), and the segment gaps and secondary obscurations (Section 10.1.1). Figure 134 shows the contrast as a function of radial

distance from the star for the combined cases as well as the individual error terms. Primary mirror related errors limit the contrast value to 3.4×10^{-7} at 3λ/D and an average value of 2.1×10^{-7} from 3 to 10 λ/D. **Error! Reference source not found.** shows the PSF including the affects from all of the primary mirror related errors, the PSF shows the characteristic dark hole at a radius of 0.6 arcseconds and the vertical "stripes" associated with the nuller pattern.

The limiting factor in contrast is segment to segment reflectivity, however, as mentioned this can be corrected for with the addition of a 2nd DM, which we will assume can be done. Thus, the fundamental limiting factor in telescope contrast is set by the residual segment aberrations. The resulting contrast is 1.4×10^{-7} at 3λ/D and an average value of 5.6×10^{-8} from 3 to 10 λ/D. We have as a requirement (section 10) 5×10^{-8} and meet this requirement at field angles greater then ~55mas (5λ/D). The contrast is dominated by the segment aberrations, the impacts of this and future work needed is discussed in Section 10.3 .

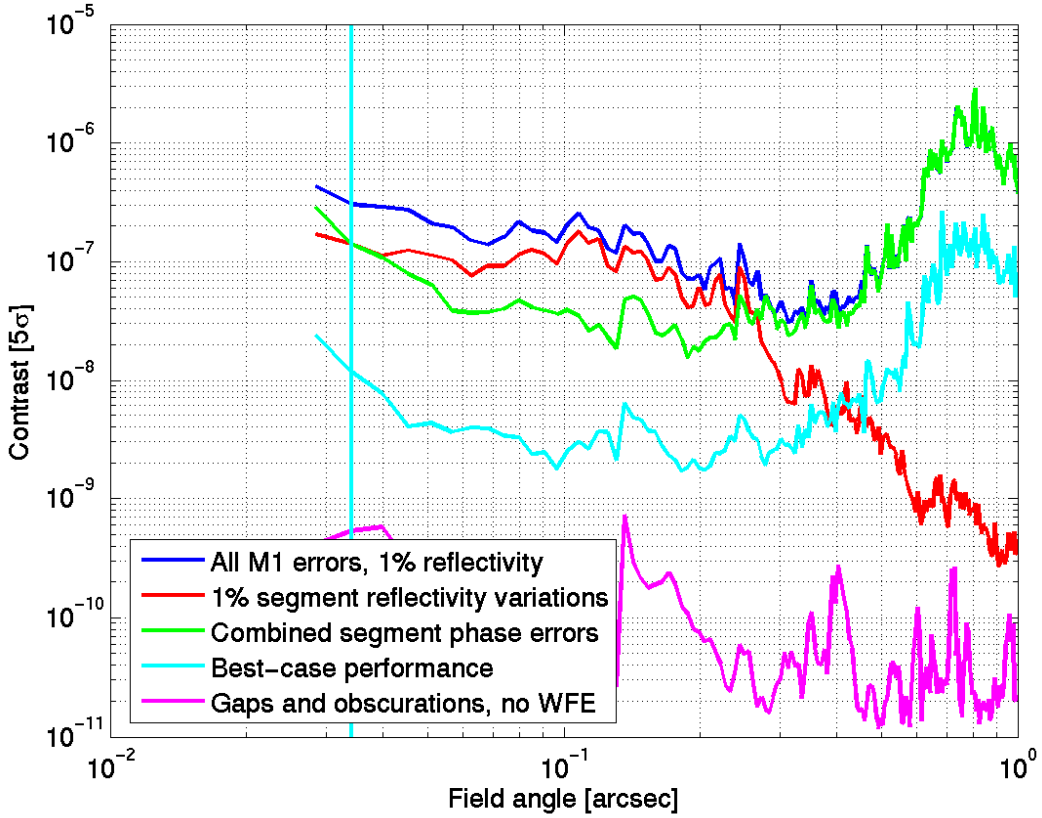

Figure 134: The contrast is shown as a function of field angle. From bottom to top: the magenta curve is the contrast from the telescope obscuration and segment gaps. The light blue curve is the contrast for the "optimal" M1 described below. The green curve is the contrast from all segment phase errors. The red curve is the contrast from 1% segment to segment reflectivity variations. The blue curve shows the contrast from combining reflectivity with all of the M1 error terms. The contrast in the combined case is dominated by segment reflectivity for angles less then ~0.25 arcseconds.

It is instructive to ask what the contrast performance is in a best-case scenario. To do this we have assumed segment WH errors with ideal reduction factors and no measurement errors, along with PFI feedback providing piston and tip/tilt errors a factor of two better then previously assumed. The resulting contrast is the light blue line in Figure 134.We achieve $1x10^{-8}$ at $3\lambda/D$ and an average value of 6.6x10-9 from 3 to 10 L/D. In this case we meet the requirement for telescope performance at all field angles.

10.2 M3 static errors

In this section the impact of the phase and phase-induced amplitude errors from the tertiary mirror (M3) are analyzed. A finite element analysis was provided by Myung Cho over the full 2.95 m by 4.12 m of M3. We have analyzed the "ring" support design; the finite element model had 2500 nodes and was re-sampled onto a uniform grid of 128×128 points across the larger diameter of M3. This analysis includes only the axial support print-through and the telescope was modeled at the Zenith (M3 at 45 degrees).

10.2.1 Wavefront and phase-induced amplitude errors

The finite element data was re-sampled onto a square 1 mm grid for use in our simulations. The RMS wavefront error is 13.4 nm with a P-V of 80 nm over the entire M3. Using Fresnel propagation the wavefront error at M3 was then propagated to a pupil. Figure 135 shows on-axis portion of the phase and amplitude errors at the pupil. The RMS wavefront error over the on-axis beam is 20 nm with a P-V of 135 nm. The RMS of the phase-induced amplitude errors is 3.4×10^{-4} with a P-V of 0.017.

Three pupils where run through the AO system, amplitude errors only, phase errors only, and both phase and amplitude errors. In the presence of no amplitude errors the AO corrected wavefront had an RMS value of 0.03 nm (essentially zero). In both of the cases with phase errors the initial 20 nm of wavefront error was reduced to less then 1 nm.

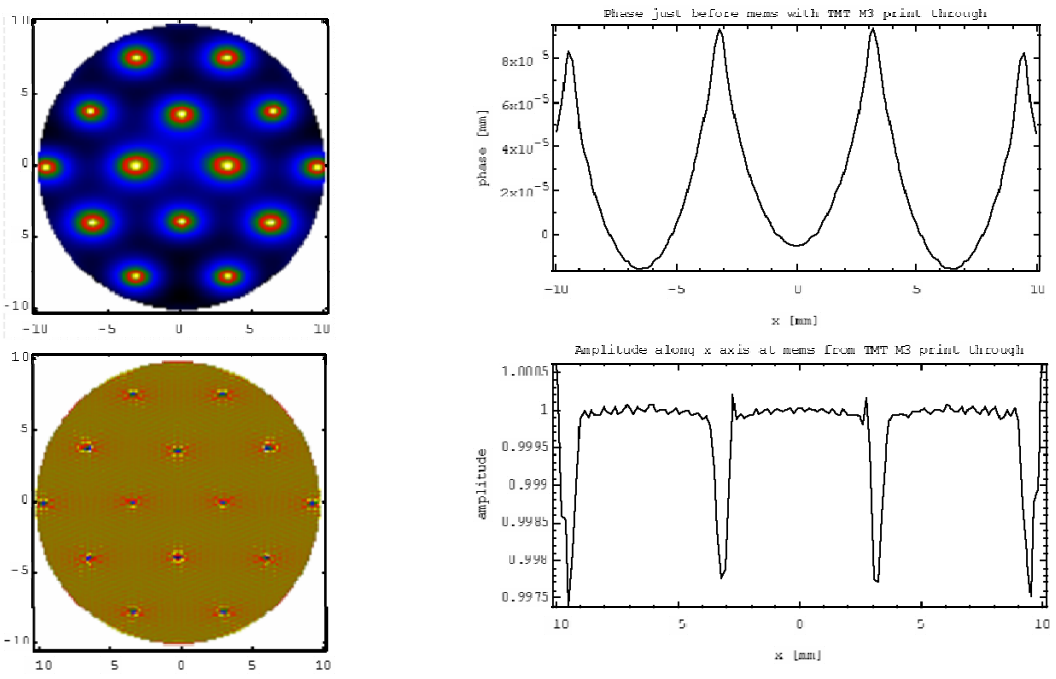

Figure 135: The upper left image shows the phase error along with a cut across the diameter (on the right). The lower left image shows the phase-induced amplitude errors along with a cut across the diameter (on the right).

10.2.2 Achievable Contrast

The AO corrected pupils were then run through a diffraction suppression system. In this case an earlier version of the DSS was used, namely a simple Lyot Coronagraph using a Sinc squared occulting function. This has the same aberration and amplitude sensitivity as the nuller, so the results presented are still valid. Figure 136 shows the contrast as a function of angular distance from the star. The phase-induced amplitude errors dominate the contrast. A contrast value of 1.0×10^{-10} is reached at 3λ/D and an average value of 1.1×10^{-10} from 3 to 10 λ/D. This contrast is significantly better then the requirement. In the case of small amplitude and phase errors the contrast is proportional to the phase errors squared. As a result even if the wavefront errors on M3 were 4 times worse we would still achieve a contrast of 2×10^{-9}.

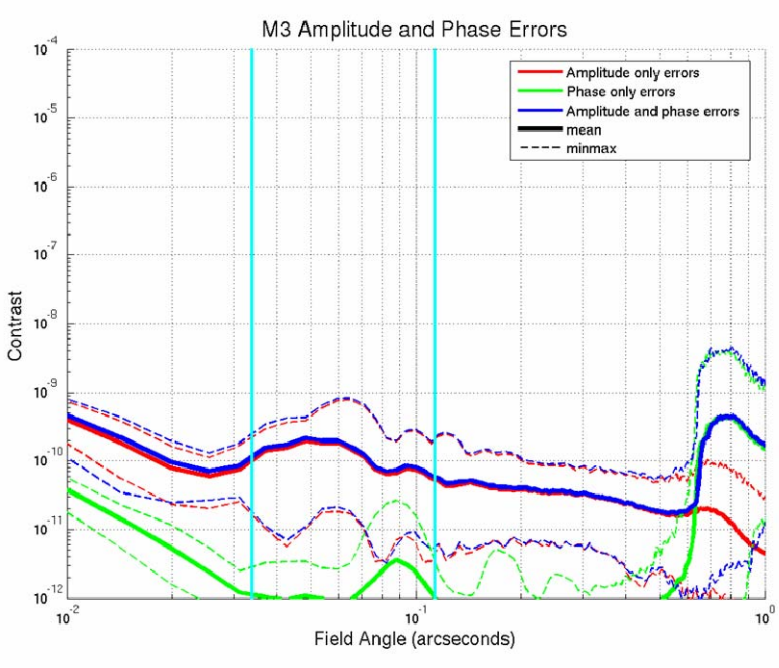

Figure 136: The contrast is shown as a function of field angle. The green curve (bottom) is the contrast from the phase errors only on M3. The red curve (middle) shows the contrast from the phase-induced amplitude errors. The blue curve (top) shows the contrast from the combined phase and phase-induced amplitude errors after AO correction. It is clear that the phase-induced amplitude errors dominate the contrast. This contrast is not a significant driver in the error budget.

10.2.3 Notes on segmented M3

Late in this feasibility study an idea was proposed to have a segmented M3 (TMT.OPT.TEC.06.001.DRF01). This idea is not yet mature enough to have concrete specifications for which simulations could be performed. However, a segmented M3 could clearly have a large impact on the performance and/or design of PFI. In particular, the 4 mm segment gaps map to non-square large gaps in the pupil plane (63 by 88 mm). In principle such gaps could be masked out in the DSS, but in the current nuller design this would result in three such copies of the masks and could significantly impact throughput and contrast. In addition the impact of segment-to-segment reflectivity's not at a pupil plane need to be studied as well as methods to mitigate them. Clearly from a PFI perspective a monolithic M3 or one with a central segment as large as the on-axis beam is beneficial. If the project decides to study the segment M3 design further PFI simulations and analysis should be performed to help define the requirements.

10.3 Conclusions

The key question is to what extent this will limit PFI's scientific performance. Figure 137 shows a plot of case 1 with the contrast achieved in 1.5 seconds (dark blue curve) and 1 hour (red curve) along with the contrast limit from the telescope (light blue curve). In a 1.5 second exposure the telescope does not limit the contrast, however in a ~2.5 minute integration the telescope and atmospheric contrast limits are the same. In a 1 hour exposure

the telescope contrast floor ranges from 2 to 10 times larger then that set by the atmosphere with the largest difference occurring for small working angles.

The above assumes no speckle suppression. If we assume that atmospheric speckles are suppressed by a factor of 10 (Metchev 2005) and that the telescope speckles are suppressed by a factor of 20, then we get the green and magenta curves in Figure 137. In this case the telescope is limiting the contrast to ~1×10^{-8} out to ~0.1 arcseconds. The telescope limits the achievable contrast out to ~0.3 arcseconds. Figure 138shows the scientific impact of this baseline primary mirror.

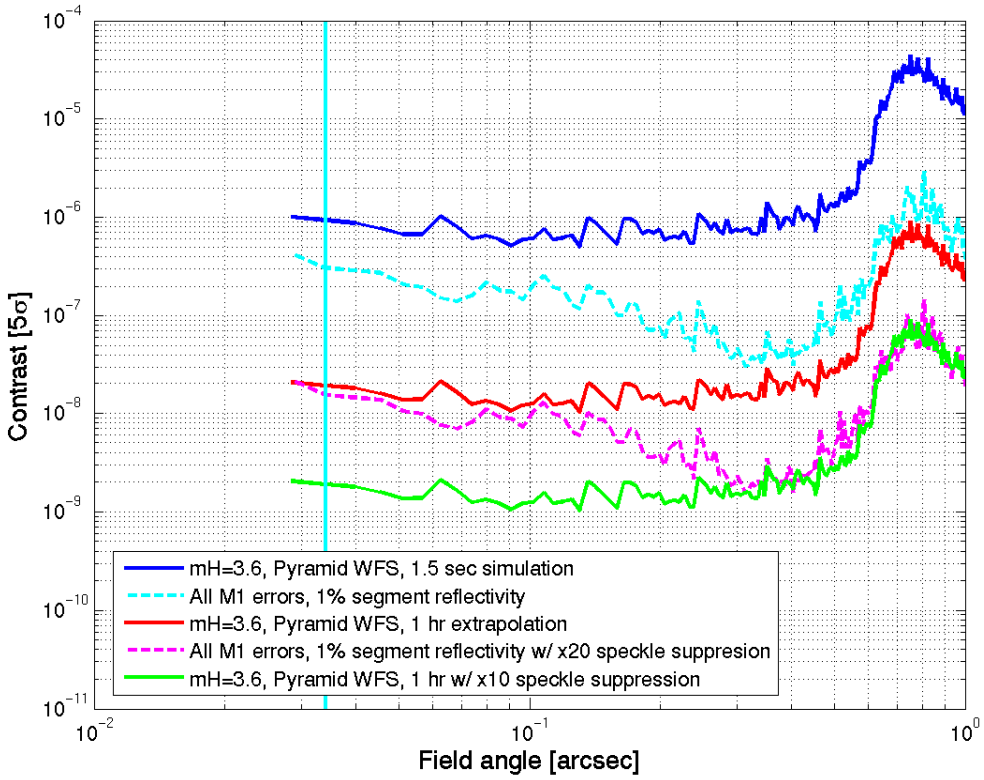

Figure 137: The contrast is shown as a function of field angle for case 1 in Table 19 along with the effect of telescope errors. In a one hour integration the telescope limits the achievable contrast to ~1x10^{-8}, where as the atmospheric limit is almost an order of magnitude better, ~2x10^{-9}.

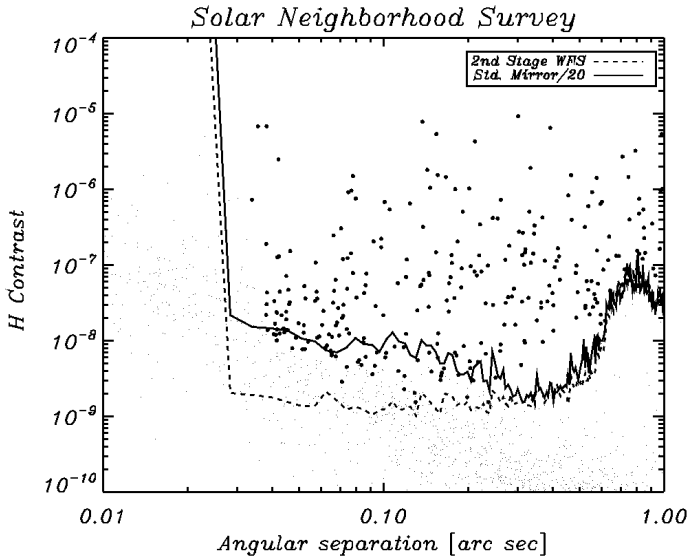

Figure 138: Performance of PFI with and without the baseline mirror errors in planet detection. As in Figure 11, each dot represents a planet; large dots are found by the baseline PFI system. Lines show contrast for a representative star with and without the primary mirror errors (note that contrast, especially photon noise, varies from target to target and hence not all targets above the line are detected.) Primary mirror errors cut the overall detection rate by 33% and prevent almost all detections of reflected-light planets.

We have shown that the current telescope design should set an instantaneous contrast floor of ~6×10^{-8}, (assuming the addition of a 2nd DM to correct the affects from segment reflectivity variations) which after speckle suppression techniques should produce a contrast floor of ~6×10^{-9}. This level of performance is better then the requirement of 10^{-8}, but (as shown in Section 9.1) uses up a substantial fraction of the available error budget, and will likely limit performance on many science targets.

We have looked at M1 segment aberrations, segment tip/tilt and piston alignment errors, whiffle tree print through, the effects of gaps and the secondary obscuration (and supports), reflectivity variations and M3 phase and phase-induced amplitude errors. The dominant error term, segment-to-segment reflectivity variations will require the addition of an amplitude control architecture such as an additional DM located not at a pupil. This leaves segment aberrations as the dominant error term. We have shown (Section 10.1.2.1.2) that depending on the assumptions on the performance of the warping harness (Figure 128) the resulting contrast can change by over two orders of magnitude.

Once the telescope is built many of these error terms will be difficult (if not impossible) to correct as a result we suggested that the following be investigated in a future study:

1) Investigation of the effects of dynamic disturbances. This should include wind shake (global tip/tilt) as well as segment vibration. This is believed to be one of the limiting factors in the performance of the Keck AO system on bright guide stars.
2) Further investigation and a better understanding of the expected segment aberrations after correction by warping harness. This should include:
    a. Understanding the poor performance of the Keck warping harness

b. A better understanding of the predicted measurement errors for setting the warping harnesses.

c. Investigate measurement and control techniques using the APS to minimize the edge discontinuities and there impact on contrast.

d. Investigate feeding segment tip/tilt information from the PFI WFS to the primary mirror control system.

3) Investigate the expected spatial distribution of tip/tilt errors and associated edge discontinuities.

4) Developing a metric to express edge discontinuities that can be quickly evaluated for a given M1

5) If a segmented M3 is selected further studies should be done to understand it's wavefront requirements and impact on the design of the DSS

11 Alternative Architectures and PFI Options

11.1 NFIRAOS

At the request of the TMT Project Office, we have carried out a basic evaluation of NFIRAOS for high-contrast imaging purposes. NFIRAOS, even in its first-light configuration, will be a powerful AO system, designed for higher Strehl ratios than currently achieved on 8-10m telescopes. Combined with the aperture advantage of TMT, if it were limited only by atmospheric wavefront errors it would be capable of contrasts an order of magnitude higher than current AO systems, approaching the dim-star performance of the Gemini Planet Imager, especially at wider angles (0.5-2"). To achieve this performance, some class of simple coronagraph would be needed to suppress diffraction from the telescope outer and inner edge and possibly secondary supports; diffraction off the segment gaps would not be significant at this level.

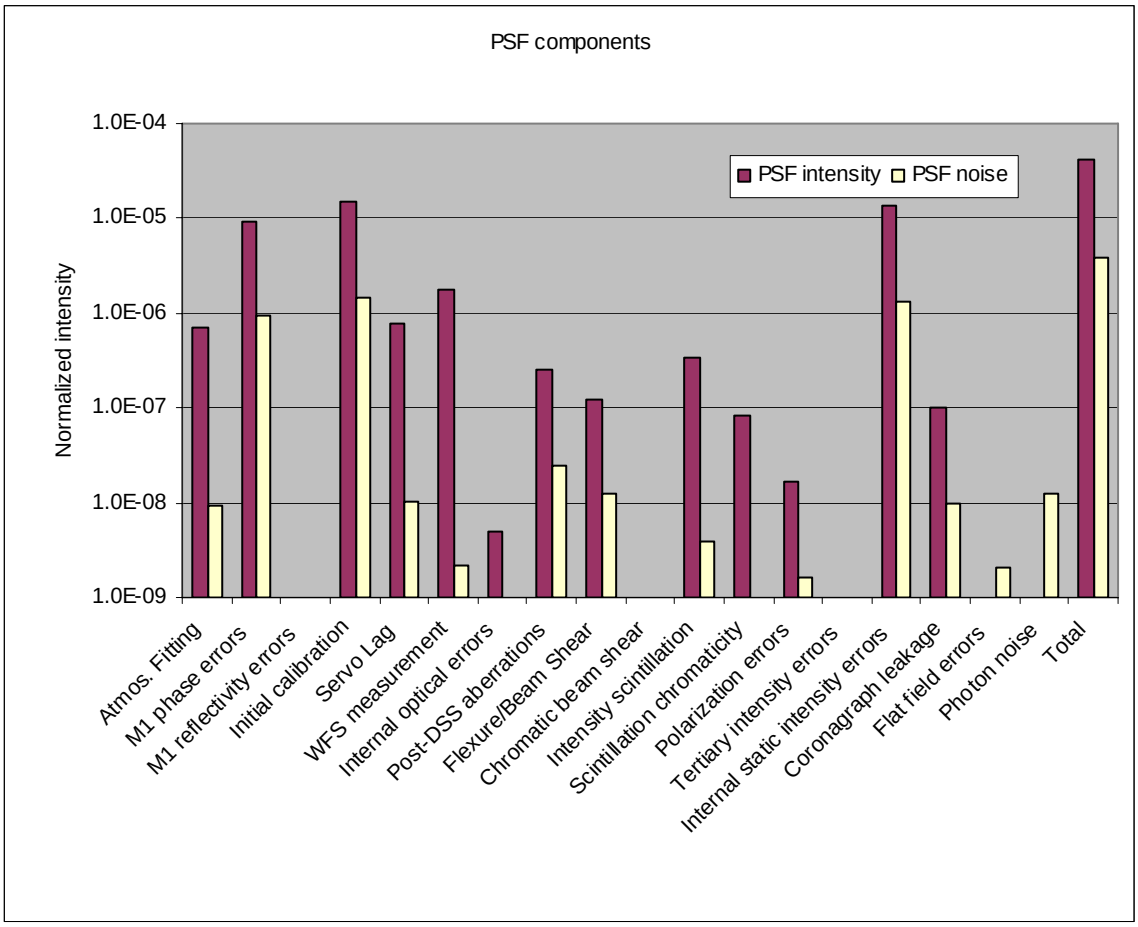

Figure 139: Contrast error budget for NFIRAOS on a bright NGS, 0.4" separation in a 60-second exposure, with a simple Lyot coronagraph.

However, as discussed in Section 9, the ultimate metric for ExAO systems is not Strehl ratio, but final contrast, and that contrast is dominated by static wavefront errors. NFIRAOS' error budget allocated 35 nm for residual non-common path errors, and

NFIRAOS' large optics are specified to λ/5 quality (though this may be changed.) In even a one-minute exposure this would complete swamp contrast due to dynamic wavefront error sources, and NFIRAOS would achieve a 5σ contrast of ~10^{-5} at 0.4" – an order of magnitude better than current Keck AO imaging at these separations, but not competitive with GPI, let alone PFI.

Figure 139 shows a sample contrast error budget (similar to section 9.1) for NFIRAOS.

In longer exposure sequences, siderial rotation could be used to remove some of these static artifacts as is done on Keck (Marois et al 2006.) This technique would be somewhat limited in that it appears to be impossible to keep the telescope pupil fixed with respect to NFIRAOS, so the PSF will evolve as telescope-related artifacts rotate with respect to NFIRAOS artifacts.

11.1.1 NFIRAOS with IRIS

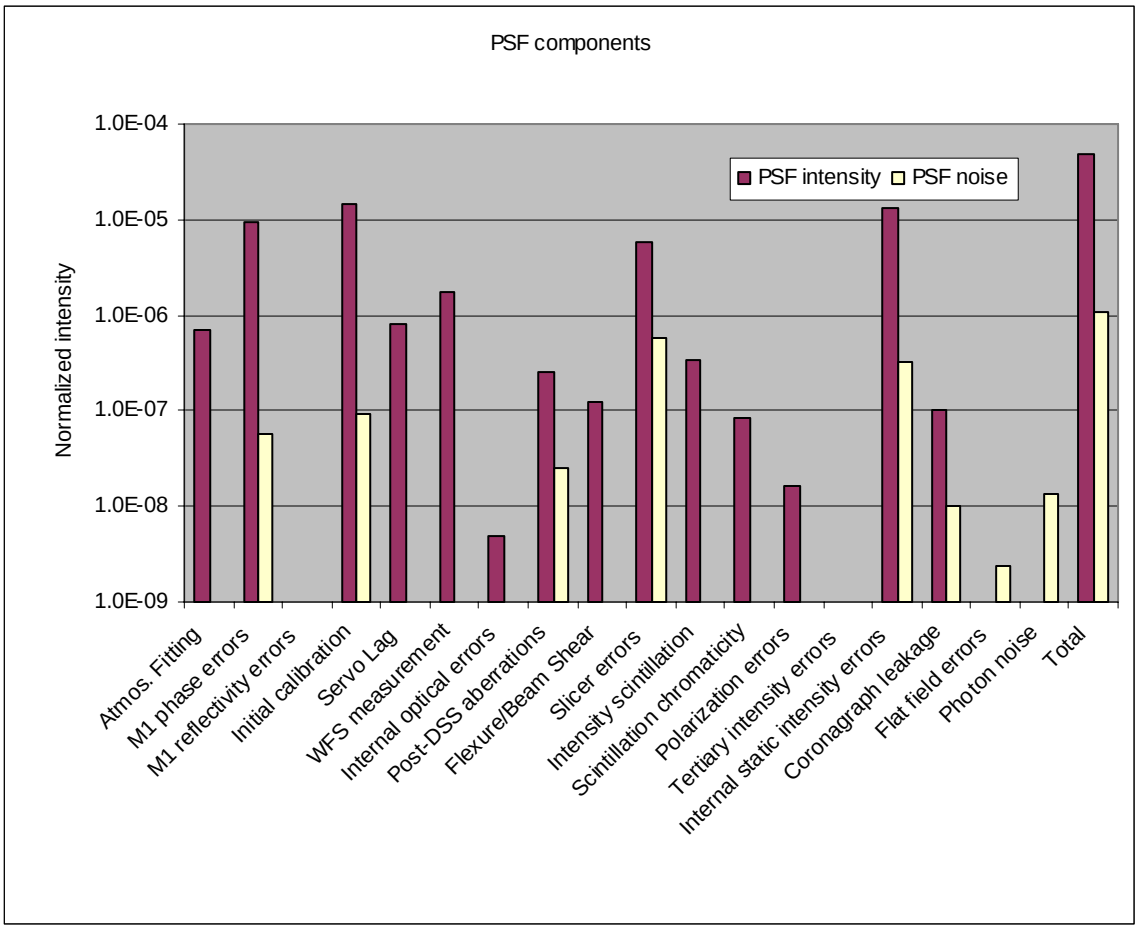

Figure 140: NFIRAOS error budget with a slicer-based IRIS used for speckle suppression.

Many of these errors will be removable through multi-wavelength speckle suppression, especially at large field angles. However, IRIS may not be the ideal instrument for such observations. First, its narrow field of view when Nyquist-sampled would limit companion

searches to <0.35" separation. Second, if a slicer-based architecture is selected for IRIS, differential aberrations between each slice and along each spectral direction will scatter light from speckle to speckle. The lenslet-based IRIS has a restricted bandpass per exposure (currently 5%), which again limits speckle suppression since it is impossible to fully span the methane band, and the "length" of diagonal speckles in the data cube will be reduced. We expect that IRIS would therefore only achieve contrast in the 10^{-6} range. This could potentially be used for high-SNR spectroscopy of bright planets discovered by GPI or similar instruments.

11.1.2 NFIRAOS with the U. Montreal speckle-suppression IFS

By contrast to IRIS, The IFS described in Section 7 is optimized for extrasolar planet detection and observations of circumstellar dust. The baseline spectral resolution (R~70) is much lower than IRIS, but sufficient for detecting and characterizing extrasolar planets, and the low resolution allows a large Nyquist-sampled field of view. The lenslet-based architecture is designed to minimize chromatic errors and maximize speckle rejection. The instrument includes a rotating selectable pupil stop suitable for Lyot coronagraphy. A dual-channel polarimetry for characterization of circumstellar dust is available.

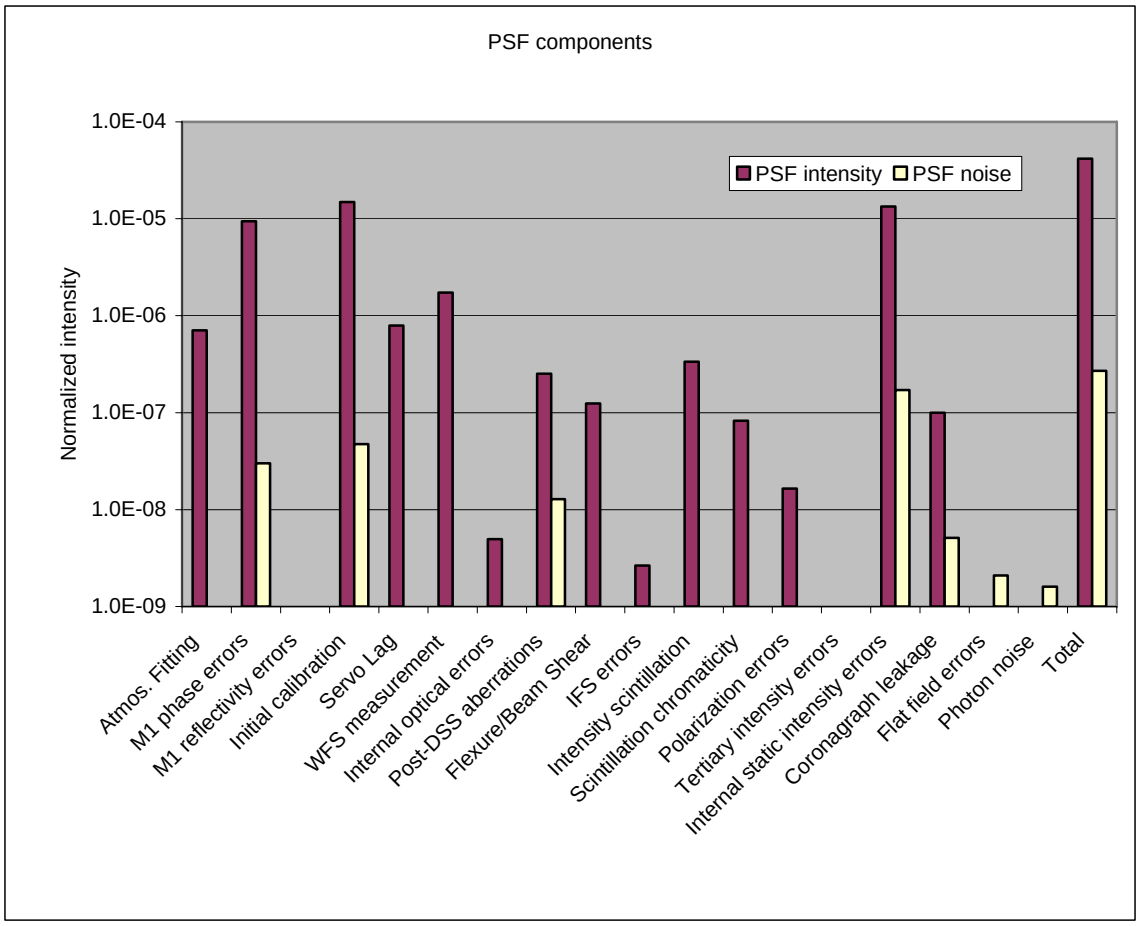

Figure 141: Error budget for the Montreal IFS operated behind NFIRAOS

Placed behind NFIRAOS, this instrument would enable first-light planet detection for TMT, and could also be used for other science missions that do not require high spectral resolution or OH suppression, such as 3–5 micron spectroscopy of solar system objects, circumstellar dust, or brown dwarfs.

Figure 141 shows the error budget for this mode in a 1-hour exposure; performance is limited by errors on out-of-plane optics to the 10^{-7} to 10^{-6} level. Since telescope errors are much less significant than errors internal to NFIRAOS, siderial rotation could be used to further improve performance. Using NFIRAOS laser guide star mode, planet searches could be carried out around dim targets (M, L and T dwarfs, reddened T Tauri stars, and the more distant young stars) inaccessible to GPI. With a simple Lyot coronagraph, planet-like contrast levels would only be achieved at >0.1-0.2 arcseconds, corresponding to 5-10 AU a target at 50 pc, probing the outer parts of target solar systems. The 30-m aperture would allow high SNR spectroscopy of planets already discovered by GPI.

To allow for this (and use of IRIS with PFI, discussed below), PFI has been designed to duplicate the NFIRAOS optical output (and hence match the TMT Naysmith focus.) The PFI instrument mount and rotator would also be duplicates of the NFIRAOS design.

A lower-cost version of the IFS could omit the 3–5 micron channel. Even a well-designed dual-channel imager would substantially improve NFIRAOS' planet-detection capability.

Residual tip/tilt, causing leakage through the DSS nuller, would be a concern for this mode; it may be necessary to include fast tip/tilt capability in the DSS itself.

11.1.3 NFIRAOS options summary

Although NFIRAOS (particularly in its second incarnation) is a very powerful AO system, a general-purpose AO system with large optics can never match the performance of a compact dedicated precision ExAO system. Nonetheless, NFIRAOS may be capable of significant early planet-detection science. A combination of NFIRAOS, a Lyot coronagraph, and the Montreal IFS could have sensitivity comparable to GPI, for high-SNR follow-up of GPI planet detections and planet searches around targets too dim for 8–10-m ExAO.

Adding the full PFI DSS and its wavefront sensing and control capabilities would remove many static errors and could allow NFIRAOS to carry out a portion of the core PFI science on young or high-contrast planets, though not the mature-planet survey discussed in section 2.

To fully evaluate these options, collaborative modeling between the NFIRAOS and PFI teams is needed. The key issues would be

(1) Contrast evaluation of simulated NFIRAOS atmospheric PSFs

(2) NFIRAOS ability to correct M1 errors and the resulting PSF contrast

(3) Simulation of likely NFIRAOS non-common-path and calibration errors and studies of algorithms to reduce them

(4) Wave-optics modeling of out-of-plane optical errors in NFIRAOS

(5) Evaluation of different simple coronagraph architectures

(6) Speckle-suppression modeling of NFIRAOS PSFs

(7) Refining the design of the Montreal IFS for compatibility with NFIRAOS

In addition, several design requirements should be considered for NFIRAOS and PFI to allow these options and improve performance

(8) Improving the quality of NFIRAOS internal optics, particularly those optics closest to focus.

(9) Designing and leaving room for a non-cryogenic simple coronagraph that could be introduced between NFIRAOS and science instruments

(10) Incorporating a spatial filter into the NGS WFS of the NFIRAOS upgrade (N=60 NFIRAOS is too low-order to benefit from a spatial filter.)

(11) Compatible instrument interfaces to allow IRIS and the PFI IFS to be moved between NFIRAOS and PFI

11.2 IRIS for PFI follow-on science

The baseline PFI IFS has a R=700 spectral mode, but only over limited spectral bands, for characterization of already-known exoplanets. Alternatively, IRIS could be used for such characterization. The narrow field of view would not be a limitation since the planet location would already be known. Scattered light from IRIS slicers would probably prevent speckle suppression, but this mode would be used for characterization of bright exoplanets. In the current architecture, the DSS Lyot stop is located inside the science instrument, but this could be changed, or a selectable pupil inside IRIS itself could fill this role. NIRES could fill a similar role at even higher spectral resolution, allowing for spatially-resolved studies of gas lines in circumstellar disks. As with the options above, this requires attention to the PFI and NFIRAOS science interfaces.

11.3 PFI intensity control upgrade and advanced wavefront control

As discussed in Section 10, reflectivity variations from segment to segment in M1 could limit PFI contrast at the 10^{-8} level, even for 1% RMS reflectivity variations. Internal intensity errors, particularly those induced by out-of-plane wavefront errors, are a similar source of scattered light that cannot be corrected by a single DM.

Pairs of DMs operated in a coordinated fashion, however, can be used to control both phase and intensity. One such architecture is a Michelson interferometer with a DM in each arm (Littman et al 2003); another is a pair of sequential deformable mirrors at different conjugates. Including an additional DM in PFI would therefore allow it to remove intensity errors and reduce the specifications on both M1 and internal optics. (Note that all of PFI's baseline DMs are necessarily at the same conjugate and hence cannot provide intensity correction.)

Although the Michelson configuration would be easiest to implement in PFI, analytic calculations by Shaklan and Green (2006) show that it provides a strongly chromatic correction and would only correct a ~4% bandpass for our conditions. A DM located in the main optical train ~0.5–1 m from the "main" DM, however, would provide enough control authority over our entire bandpass. This DM could be driven from signals from the post-DSS wavefront sensor, though the reconstruction and control problem is nontrivial. Nonetheless, based the modeling in Section 10, we expect that some form of intensity control will have to be adopted by PFI if M1 reflectivity variations are any worse than 1%.

11.4 Fiber mode filter array to reject M1 errors

The nulling coronagraph can be combined with a single mode fiber array that rejects high spatial frequency aberrations (Figure 142). In a pupil plane after the nuller the beam is fed into an array of fibers, typically with one fiber per DM actuator. Wavefront errors at high spatial frequency, on a spatial scale smaller than any deformable mirror (DM) actuator, will be filtered by the single-mode fibers instead of propagating to the science focal plane. If the optical fibers all have the same length (to λ/20) the planet light from each fiber combines to form a coherent off-axis image. The phase of the residual star light exiting the fiber array is random, hence it is scattered evenly across the whole field of view. In summary, wavefront errors on a spatial scale larger than the DM actuator spacing are corrected by active control of the DM and wavefront errors smaller than the actuator spacing are filtered out by the fiber array.

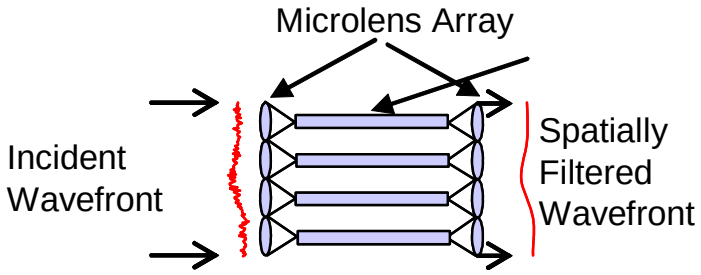

Figure 142: A single mode spatial filter array consists of an input lens array that divides incident wavefront into multiple segments and couple light in each segment into a single mode fiber, which removes higher-order spatial modes from each light segment; a second lens array recollimates the light from each fiber.

The role of the fiber array is well suited to a segmented mirror primary mirror. This architecture mitigates segment gaps and discontinuities, which induce high frequency phase and amplitude errors. Alternatively, each segment can be matched one or more fibers (essentially one fiber per airy spot). In a nulling coronagraph being studied for TPF, it is planned to have a fiber array of up to 10,000 fibers. At the fiber input, if the starlight is suppressed by say x, then at the output a total of x of the starlight will appear as scattered light. Because in general the nulling process will produce residual leaked light with random phase, this starlight will be scattered uniformly over the airy spot output field of

view, with an average scattered light level of x/airy spot, hence greatly enhancing the contrast of the final image.

A single mode spatial filter array module consists of a coherent fixed-length fiber array sandwiched by two lenslet arrays as shown in Figure 1. The current technology path calls for development of a 1000 fiber array for TPF –C within 3 years. Assembly methods can be scaled to the sizes needed for TMT. There are two array fabrication approaches under investigation. For an IR instrument, these fibers would necessarily be cryogenic, and hence this would be complicated to implement.

11.5 Phased and descoped PFI options

The Planet Formation Instrument as described here is an extremely powerful instrument, capable of a broad range of science missions. To re-iterate, the core science missions are:

1. Broad and systematic studies of the extrasolar planet population in the solar neighborhood, spanning semimajor axis separations $0.5 < a/\mathrm{AU} < 50$ and exoplanet mass $0.5 < M/M_J < 12$.
2. Imaging very young planets (0–15 Myr) in the process of forming or migrating at > 100 pc distances
3. High-SNR studies of planetary atmospheres and their astrophysics
4. The studies of circumstellar disks ranging from young protoplanetary disks through debris disks to high-density extrasolar zodiacal debris in inner solar systems.

Descoped versions of PFI could still carry out 50% of these missions for a moderate reduction in initial cost. There are two primary descopes to PFI that could be considered, in addition to phased instrument capabilties. Any descope will reduce the uniqueness of PFI science compared to the highly capable Gemini and VLT instruments, and the relative merits of the PFI science roles should be carefully weighed and analyzed.

11.5.1 Conventional coronagraph

The first is to relax the inner working angle requirement from 0.03 arcseonds at *H* band to ~0.1 arcseconds. It is this requirement, more than anything else, that drives the choice of the nuller-based DSS. Replacing the nuller with a band-limited or apodized Lyot coronagraph would sharply reduce contrast at <0.1 arcsecond, but also reduce the complexity of the instrument. We estimate the total cost savings would be on the order of $1M.

Such a descope fails to take full advantage of the angular resolution provided by the TMT aperture. Science mission 2 would be almost completely precluded; a scale of 10 AU in star-forming regions at 100 pc corresponds to 0.1 arcseconds. Science mission 1 would be moderately impacted. The mature reflected-light Jovian planets that are inaccessible to the Gemini Planet Imager would remain inaccessible to TMT. However, TMT would still extend the space searched by GPI down to much lower masses and a larger sample of stars, perhaps detecting a factor of two more planets than GPI. The conventional coronagraph would have some operational advantages, including near-complete coverage of a given search space. This could be combined with a descope of the second-stage IR AO system to a lower frame rate and removing its dedicated DM, resulting in a GPI-like architecture

with a slight performance penalty on bright stars but removing the capability to observe red TTauri stars. With careful design, room could be reserved for the nuller and high-speed post-DSS AO system for future upgrades.

11.5.2 Low-order initial AO

PFI's powerful first-stage AO system is driven by science requirement 1 – the need to achieve very high contrast (10^{-8}) on bright solar-neighborhood stars. Reducing the capabilities of this system, for example leaving only the woofer DM, would likely reduce achievable contrast to 10^{-7} for most targets. This would leave TMT with sensitivity comparable to GPI, though still with a much smaller inner working angle. That would still be sufficient to detect young planets (mission 2) and to carry out spectroscopy of planets previously discovered by GPI (mission 3), but would preclude science mission 1 and to some extent mission 4. The cost savings would be on the order of $3M (no advanced control computers or wavefront sensors, though the back AO system would still require a MEMS). It would be straightforward to leave room for a future MEMS/high-order AO upgrade on the optical bench. Intermediate descopes would include replacing the pyramid sensor with a Shack-Hartmann wavefront sensor; this would reduce limiting magnitude by 1-2, reducing the sample size for all science missions but not precluding most observations.

11.5.3 Science instrument changes

The 3-5 micron capability of the science instrument is primarily driven by spectral characterization (see the appendix of the OCDD) rather than initial searches. A 1-2.5 micron IFS would have a somewhat reduced cost, with simpler optics and fewer moving parts, for a cost savings of $0.5M. Removing the narrowband / high spectral resolution capability might result in similar savings. Simpler science instruments such as single-channel or even dual-channel imagers with limited ability to reject speckles would substantially increase the risk of PFI, increase the sensitivity to primary-mirror aberrations, and would not be competitive with GPI, except for science mission 2.

12 Proposed future activities

Although a full PFI Conceptual Design is not likely until at least 2009, there are several areas in which near-term investment by TMT is important, as well as ongoing research funded by other institutions that bears on TMT. The single most important area is continued modeling of M1 interactions with PFI and other AO systems

12.1 Telescope interactions

As detailed in this report (sections 9 and 10) the ultimate performance of the PFI (or any high-Strehl AO corrected instrument) on TMT will likely be limited by telescope effects. This is a particularly important area for comparisons between different 20-30m telescope designs. We believe it is critical to continue to fund an effort during the TMT design phase to insure that the TMT will support high-contrast imaging, even if PFI is not selected as a first light instrument. To re-iterate specific areas of concern and need future work:

1) Investigation of the effects of dynamic disturbances. This should include wind shake (global tip/tilt) as well as segment vibration.
2) The uncertainty in warping harness performance is currently the single largest uncertainty in predicted the performance of the static telescope. The contrast from this term can change by a factor of 100 depending on the assumptions made. Further investigation and a better understanding of the expected segment aberrations after correction by warping harness is needed. In addition advanced wavefront control techniques to minimize edge discontinuities should be investigated.
3) Investigate the expected spatial distribution of tip/tilt errors and associated edge discontinuities.
4) If a segmented M3 is selected further studies should be done to understand its wavefront requirements and impact on the design of the DSS.
5) If a segmented M2 is selected its impact on contrast and on the design of the DSS should be studied.
6) Integrated modeling including observing scenarios, post-processing of images, and dynamic variations in M1

12.2 Early planet search options

As discussed in Section 11.1.4, it is possible that NFIRAOS together with PFI's integral field spectrograph could be a capable planet characterization tool. However, to fully evaluate this requires more detailed modeling in cooperation with the NFIRAOS team. It is possible that some low-cost improvements to NFIRAOS (e.g. better calibration) could significantly enhance its capability, and members of the PFI team are willing to collaborate on identifying such improvements.

To fully evaluate these options, collaborative modeling between the NFIRAOS and PFI teams is needed. The key issues would be

- Contrast evaluation of simulated NFIRAOS atmospheric PSFs
- NFIRAOS ability to correct M1 errors and the resulting PSF contrast
- Simulation of likely NFIRAOS non-common-path and calibration errors and studies of algorithms to reduce them
- Wave-optics modeling of out-of-plane optical errors in NFIRAOS
- Evaluation of different simple coronagraph architectures
- Speckle-suppression modeling of NFIRAOS PSFs
- Refining the design of the Montreal IFS for compatibility with NFIRAOS

Building a dedicated Planet-finding Integral Field Spectrograph (PIFS) (that would later serve as PFI's IFS) behind NIFRAOS is an attractive option which would require the following work:

- Further development of the science case for NIFRAOS+PIFS.
- Contrast performance estimation with a simple coronagraph and conservative IFS speckle-suppression attenuation.
- Detailed cost estimate of the IFS

12.3 Technology development

PFI is dependent on technology development in the areas of deformable mirrors and infrared detectors.

12.3.1 Deformable Mirrors

As discussed in Section 3.4.2 there does not currently exist an off the shelf "tweeter" DM that meets the PFI requirements. With further development work both the Boston MEMS and Xinetics "photonix" devices have the potential to meet PFI requirements. The laboratory for AO in collaboration with the GPI project is funding Boston to develop a 64 x 64 actuator device. Xinetics is currently under contract to provide a 64 x 64 actuator device to JPL for use in the Palomar AO system. However, this device will still be too large (~1.7 mm pitch) to meet the needs of PFI. In order to not be dependent on a single source of DMs it would be advantageous to discus with Xinetics the potential for development of a 0.5 mm pitch device with sufficient stroke for PFI.

12.3.2 Infrared detectors

The post-DSS AO WFS needs a 128 x 128 pixel IR device that has 3 e^- of read noise at a frame rate of 1000Hz. These requirements have been adjusted to be similar to those of TMT tip/tilt sensors. Rockwell currently has a development effort (see Section 5.4) to build such a device. TMT should actively pursue this with Rockwell and consider co-funding development if appropriate.

12.4 Experiments

Several active projects will contribute indirectly to the PFI concept:

12.4.1 Gemini Planet Imager

A team led by the Planet Formation Instrument PI, with significant JPL and Montreal participation, is currently negotiating contracts with Gemini for the construction of the Gemini Planet Imager, one of two 8-m telescope ExAO systems now beginning construction. GPI has an architecture similar to (but less ambitious than) PFI and represents a 2 order of magnitude improvement in contrast over conventional AO. The construction of GPI is a necessary pre-requisite to the construction of PFI, and the lessons learned will certainly strongly influence the PFI process. GPI is scheduled to have first light in 2010.

12.4.2 JPL missions

JPL is currently developing nulling interferometer technology to support both TPF and future Astrobiology themed missions. In addition to conceptual studies and sensitivity studies, JPL is working on an end-to-end- laboratory demonstration of the nulling interferometer based concept. This includes work in demonstrating deep and wide interfereometric nulls, the development of arrays of optical fibers for spatial filtering, and MEMS deformable mirror technology in both continuous facesheet and segmented architectures. JPL is also working to develop inexpensive and reliable electronics to drive these DM's and has recently fielded a prototype based on newly available commercial integrated amplifier devices.

JPL is building a calibration test-bed to develop technology in post-coronagraphic wavefront sensing. This work is supporting the Gemini Planet Imager selected as an 'Aspen Round' Instrument. The test bed will also serve as a pathfinder for advanced development project devoted to planet imaging and spectroscopy. This includes a near-term sounding rocket experiment (PICTURE 2007), for future Terrestrial Planet Finder Concepts, and for the development of advanced spectrometers for detection of molecular species indicative of life for future astrobiology missions.

JPL also plays the lead role in the Palomar AO system. Using a 1.5 meter unobscured subaperture (of the full 5 meter Hale telescope) residual wavefront errors of less then 100 nm have been achieved. There are currently plans for testing several coronagraphic techniques on the sky using this mode of AO operation. In addition JPL and Caltech are working on plans for PALM-3000, an upgrade of the current AO system to 64 x 64 actuators. The upgraded system will achieve diffraction limited visible performance over much of the sky. In addition on bright natural guide stars it will achieve 70 nm RMS wavefront error.

12.5 Science case development

TMT should continue to support development of the planet detection science case, both to refine the instrument requirements and build community knowledge of and support for this instrument. Areas for research include:

1) Computation of dense grids of atmospheres for cooling tracks to evaluate accurate narrow band indices and exploration of the effects of clouds and condensation
2) Improved treatment of insolation and cooling of planets in close orbits
3) Developing techniques for establishing Teff, log(g), & composition measurements from spectra
4) Integrated modeling of young planets in disks, including self- consistent radiative transfer modeling to refine our expectations of what young systems look like
5) Further modeling of operational modes and methods to maximize the science efficiency with the nuller
6) Hosting a workshop, perhaps through the Center for Adaptive Optics or in collaboration with HIA in Canada, on exoplanet science with 20-50m telescopes

13 Cost Estimate: See Volume 2.

14 References

Angel, J. R. P., "Ground-based imaging of extra solar planets using adaptive optics," *Nature* **368**, 203–207, 1994.

Bauman, B., Optical Design for Extremely Large Telescope Adaptive Optics Systems, Ph.D. dissertation, University of Arizona, 2003.

Bloemhof, E. E. and J. K. Wallace, "Phase contrast wavefront sensing for adaptive optics," *Proc. SPIE*, Vol. 5553, 159–169, 2004.

Bloemhof, E. E. and J. K. Wallace, "Simple broadband implementation of a phase contrast wavefront sensor for adaptive optics," *Optics Express* **12**, 6240, 2004.

Chanan, G. and M. Troy, *Appl. Opt.* **38**, 6642, 1999.

Christophe, V., "On the nature of the measurements provided by a pyramid wave-front sensor," *Opt. Comm.* **233**, 27–38, 2004.

François, R., J. P. Véran, and O. Lai, "An analytical model for Shack-Hartmann-based adaptive optics systems," *Proc. SPIE*, Vol. 3353, 1038–1048, 1998.

Ghedina, A., M. Cecconi, R. Ragazzoni, J. Farinato, A. Baruffolo, G. Crimi, E. Diolaiti, S. Esposito, L. Fini, M. Ghigo, E. Marchetti, T. Niero, A. Puglisi, *Proc. SPIE 4839*, 869–877, 2003.

Giveon, A, et al. 2004 Proc. SPIE 5490, 1438

Giveon, A, et al. 2006 Applied Optics, in press

Goodman, J., *Introduction to Fourier Optics*, McGraw-Hill, New York, 1996.

Gratadour, D., Rouan, D., Boccaletti, A., Riaud, P., & Clénet, Y. 2005, A&A, 429, 433

Green, J. ???###

Guyon, O., *ApJ* **615**, 562, 2004.

Haguenauer, P., Serabyn, E., Bloemhof, E. E., Troy, M., Wallace, J. K., 2005, Proc SPIE 5905

Hardy, J. W., *Adaptive Optics for Astronomical Telescopes*, 1998.

Kuchner, M. J. and W. A. Traub, *ApJ* **570**, 900, 2002.

Lyot, B., *MNRAS* **99**, 580, 1939.

Macintosh, B., et al., Gemini Planet Imager Design Study, 2004.

Marois, C. 2004, Ph.D. thesis, Univ. Montréal

Marois, C., et al. 2006 Ap.J. in press, http://www.arxiv.org/abs/astro-ph/0512335

Mennesson, B., M. Shao, B. M. Levine, J. K. Wallace, D. T. Liu, E. Serabyn, S. C. Unwin, and C. A. Beichman, *SPIE Conf.* **4860**, 32, 2004.

Metchev, S., Ph.D. thesis, Caltech, 2005.

Metchev, Stanimir A.; Eisner, Joshua A.; Hillenbrand, Lynne A.; Wolf, Sebastian

Metchev S.A., Eisner J.A., Hillenbrand L.A., Wolf S., 2005, ApJ, 622, 451

Millerd, J., J. Hayes, M. North-Morris, M. Novak, and J. Wyant, "Pixellated Phase-Mask Dynamic Interferometer," *Proc. SPIE* **5531**, 304, 2004.

Nicolle, M., T. Fusco, G. Rousset, and V. Michau, "Improvement of Shack-Hartmann wave-front sensor measurement for extreme adaptive optics," *Optics Letters*, Vol. 29, 2743, 2004.

Palacios, D. M., *Proc. SPIE*, Vol. 5905, Article 5905-25, 2005.

Perrin, M., et al.

Phillion, D. W., "General methods for generating phase-shifting interferometry algorithms," *Applied Optics* **36**, 8098–8115, 1997.

Poyneer, L. A. and B. Macintosh, "Spatially filtered wave-front sensor for high-order adaptive optics," *J. Opt. Soc. Am. A*, Vol. 21, 810, 2004.

Poyneer, L, and Veran, J., 2005 JOSA A 22 1515

Racine et. al., "Speckle Noise and the Detection of Faint Companions," 1999 *PASP* 111, 587

Ragazzoni, R., *J. Mod. Opt.* **43**, 289, 1996.

Riaud, P., A. Boccaletti, D. Rouan, F. LeMarquis, and A. Labeyrie, "The Four-Quadrant Phase-Mask Coronagraph. II: Simulations," *Pub. Astronom. Soc. Pac.*, Vol. 113, 1145–1154, 2001.

Rigaut, F., J. P. Véran, and O. Lai, "An analytical model for Shack-Hartmann-based adaptive optics systems," *Proc. SPIE*, Vol. **3353**, 1038–1048, 1998.

Rouan, D., Riaud, P., Boccaletti, A., Clenet, Y., & Labeyrie, A. 2000, PASP, 112, 1479 (Paper I).

Swartzlander, G. A., *Opt. Lett.* **30**, 2876, 2005.

Trauger, J., et al. "Coronagraph contrast demonstrations with the high-contrast imaging testbed," *Proc. SPIE 5487*, 1330–1336, 2004.

Troy, M., et al., 2004.

Troy, M., G. Chanan, E. Sirko, and E. Leffert. "Residual misalignments of the Keck Telescope primary mirror segments: classification of modes and implications for adaptive optics," *Proc. SPIE 3352*, 307–317, 1998.

Verinaud, C., "On the nature of the measurements provided by a pyramid wave-front sensor," *Opt. Comm.* **233,** 27–38, 2004.

Verinaud, C., M. Le Louarn, V. Korkiakoski, and M. Carbillet, "Adaptive Optics for high-contrast imaging:pyramid sensor versus spatially filtered Shack-Hartmann sensor," *Mon. Not. R. Astron. Soc.* **357**, L26–L30, 2005.

AUTHORS:

Lawrence Livermore National Laboratory:
Bruce Macintosh (PI)
Kevin Baker
Brian Bauman
Victor Karpenko
Christian Marois
David Palmer
Donald Phillion
Lisa Poyneer

COM DEV:
Neil Rowlands
Ken Tam

Immervision Inc.:
Simon Thibault

University of California:
Travis Barman
James Graham (Project Scientist)

Jet Propulsion Laboratory:
Mitchell Troy (Co-PI)
Ian Crossfield
Philip Dumont
Joseph Green
Marty Levine
Bertrand Mennesson
David Palacios
Michael Shao
Gene Serabyn
Chris Shelton
Gautam Vasisht
James Wallace

Universite de Montreal:
Rene Doyon (Co-PI)
Jean-Francois Lavigne
Philippe Vallee

Portions of this work were carried out under the auspices of the United States Department of Energy by the University of California, Lawrence Livermore National Laboratory under contract 7405-Eng-48.